

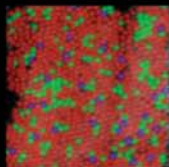
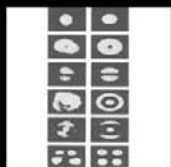
Third Edition

Sponsored by the Optical Society of America

# HANDBOOK OF OPTICS

## **Volume III**

*Vision and Vision Optics*



*Editor-in-Chief:*  
**Michael Bass**

*Associate Editors:*  
Casimer M. DeCusatis  
Jay M. Enoch  
Vasudevan Lakshminarayanan  
Guifang Li  
Carolyn MacDonald  
Virendra N. Mahajan  
Eric Van Stryland

OSA<sup>®</sup>

# HANDBOOK OF OPTICS

---

## ABOUT THE EDITORS

---

**Editor-in-Chief:** Dr. Michael Bass is professor emeritus at CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida.

**Associate Editors:**

Dr. Casimer M. DeCusatis is a distinguished engineer and technical executive with IBM Corporation.

Dr. Jay M. Enoch is dean emeritus and professor at the School of Optometry at the University of California, Berkeley.

Dr. Vasudevan Lakshminarayanan is professor of Optometry, Physics, and Electrical Engineering at the University of Waterloo, Ontario, Canada.

Dr. Guifang Li is a professor at CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida.

Dr. Carolyn MacDonald is a professor at the University at Albany, and director of the Center for X-Ray Optics.

Dr. Virendra N. Mahajan is a distinguished scientist at The Aerospace Corporation.

Dr. Eric Van Stryland is a professor at CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, Florida.

# HANDBOOK OF OPTICS

---

Volume III  
Vision and Vision Optics

---

THIRD EDITION

---

Sponsored by the  
OPTICAL SOCIETY OF AMERICA

**Michael Bass** Editor-in-Chief  
*CREOL, The College of Optics and Photonics  
University of Central Florida,  
Orlando, Florida*

**Jay M. Enoch** Associate Editor  
*School of Optometry, University of California at Berkeley  
Berkeley, California  
and  
Department of Ophthalmology  
University of California at San Francisco  
San Francisco, California*

**Vasudevan Lakshminarayanan** Associate Editor  
*School of Optometry and Departments of Physics and Electrical Engineering  
University of Waterloo  
Waterloo, Ontario, Canada*



New York Chicago San Francisco Lisbon London Madrid  
Mexico City Milan New Delhi San Juan Seoul  
Singapore Sydney Toronto



Copyright © 2010 by The McGraw-Hill Companies, Inc. All rights reserved. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

ISBN: 978-0-07-162928-7

MHID: 0-07-162928-9

The material in this eBook also appears in the print version of this title: ISBN: 978-0-07-149891-3, MHID: 0-07-149891-5.

All trademarks are trademarks of their respective owners. Rather than put a trademark symbol after every occurrence of a trademarked name, we use names in an editorial fashion only, and to the benefit of the trademark owner, with no intention of infringement of the trademark. Where such designations appear in this book, they have been printed with initial caps.

McGraw-Hill eBooks are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. To contact a representative please e-mail us at [bulksales@mcgraw-hill.com](mailto:bulksales@mcgraw-hill.com).

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. (“McGraw-Hill”) from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

#### TERMS OF USE

This is a copyrighted work and The McGraw-Hill Companies, Inc. (“McGraw-Hill”) and its licensors reserve all rights in and to the work. Use of this work is subject to these terms. Except as permitted under the Copyright Act of 1976 and the right to store and retrieve one copy of the work, you may not decompile, disassemble, reverse engineer, reproduce, modify, create derivative works based upon, transmit, distribute, disseminate, sell, publish or sublicense the work or any part of it without McGraw-Hill’s prior consent. You may use the work for your own noncommercial and personal use; any other use of the work is strictly prohibited. Your right to use the work may be terminated if you fail to comply with these terms.

THE WORK IS PROVIDED “AS IS.” MCGRAW-HILL AND ITS LICENSORS MAKE NO GUARANTEES OR WARRANTIES AS TO THE ACCURACY, ADEQUACY OR COMPLETENESS OF OR RESULTS TO BE OBTAINED FROM USING THE WORK, INCLUDING ANY INFORMATION THAT CAN BE ACCESSED THROUGH THE WORK VIA HYPERLINK OR OTHERWISE, AND EXPRESSLY DISCLAIM ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. McGraw-Hill and its licensors do not warrant or guarantee that the functions contained in the work will meet your requirements or that its operation will be uninterrupted or error free. Neither McGraw-Hill nor its licensors shall be liable to you or anyone else for any inaccuracy, error or omission, regardless of cause, in the work or for any damages resulting therefrom. McGraw-Hill has no responsibility for the content of any information accessed through the work. Under no circumstances shall McGraw-Hill and/or its licensors be liable for any indirect, incidental, special, punitive, consequential or similar damages that result from the use of or inability to use the work, even if any of them has been advised of the possibility of such damages. This limitation of liability shall apply to any claim or cause whatsoever whether such claim or cause arises in contract, tort or otherwise.

## COVER ILLUSTRATIONS

---

**A Photograph Taken of a Lady Viewing Her Face Using One of the World's Oldest Ground and Polished Mirrors.** The oldest known manufactured mirrors (ground and polished), made of obsidian (volcanic glass) have been found in ancient Anatolia in the ruins of the City of Çatal Hüyük = “mound at a road-fork.” The locations where the mirrors were discovered were dated 6000 to 5900 B.C.E. by Mellaart and his coworkers. That city is located in the South Konya Plane of Modern Turkey. Thus, these mirrors are about 8000 years old (B.P.). The obsidian was transported over a distance of more than one hundred miles to the city for processing. These mirrors can be found at the Museum of Anatolian Civilizations in Ankara. One cannot fail to be impressed by the quality of this image seen by reflectance from this ancient mirror! These mirrors had been buried twice. There is an extended history of processing of obsidian at that site for scrapers, spear, and arrow points and other tools. This very early city contained an estimated 10,000 individuals at that time(!); it was a center for development of modern agriculture, Indo-European languages, various crafts, etc., and had established road connections and trade relations [Enoch, J., *Optom. Vision Sci.* **83**(10):775–781, 2006]. (*This figure is published with permission of Prof. Mellaart, the Director of the Museum of Anatolian Civilizations, the author, and the editor of the Journal.*)

**Waveguide Modal Patterns in Vertebrate Eyes (Including Human).** This illustration demonstrates the variety of waveguide modal patterns observed in freshly removed retinas obtained from normal human, monkey, and rat retinas [Enoch, J., *J. Opt. Soc. Am.* **53**(1):71–85, 1963]. These modal patterns have been recorded in paracentral retinal receptors. Reverse path illumination was employed. These modes were photographed in near monochromatic light. This figure provides representative modal patterns observed and recorded near terminations of these photoreceptor waveguides. With variation of wavelength, at cutoff (please refer to the “V” parameter), it is possible to witness sharp modal pattern alterations. In this figure, the intent was to show the classes of modal patterns observed in these retinal receptors. (*This figure is reproduced with permission of JOSA and the author.*)

**Photoreceptors in the Human Eye.** This figure shows the first map ever made of the spatial arrangement of the three cone classes in the human retina. The three colors (red, green, and blue) indicate cones that are sensitive to the long, middle, and short wavelength ranges of the visible spectrum and are classified as L, M, and S cones. The image was recorded from a living human eye using the adaptive optics ophthalmoscope, which was developed by David Williams' lab at the University of Rochester [Liang, J., Williams, D. R., and Miller, D. (1997). Supernormal vision and high-resolution retinal imaging through adaptive optics, *J. Opt. Soc. Am. A* **14**:2884–2892]. This image was first published in the journal *Nature* [Roorda, A., and Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye, *Nature* **397**:520–522]. (*Courtesy of Austin Roorda and David Williams.*)

*This page intentionally left blank*

---

# CONTENTS

---

<b>Contributors</b>	<b>xiii</b>
<b>Brief Contents of All Volumes</b>	<b>xv</b>
<b>Editors' Preface</b>	<b>xxi</b>
<b>Preface to Volume III</b>	<b>xxiii</b>
<b>Glossary and Fundamental Constants</b>	<b>xxvii</b>

---

## Chapter 1. Optics of the Eye *Neil Charman* 1.1

1.1	Glossary / 1.1
1.2	Introduction / 1.3
1.3	Ocular Parameters and Ametropia / 1.4
1.4	Ocular Transmittance and Retinal Illuminance / 1.8
1.5	Factors Affecting In-Focus Retinal Image Quality / 1.12
1.6	Final Retinal Image Quality / 1.21
1.7	Depth-of-Focus and Accommodation / 1.28
1.8	Eye Models / 1.36
1.9	Two Eyes and Stereopsis / 1.38
1.10	Movements of the Eyes / 1.42
1.11	Conclusion / 1.45
1.12	References / 1.45

---

## Chapter 2. Visual Performance *Wilson S. Geisler and Martin S. Banks* 2.1

2.1	Glossary / 2.1
2.2	Introduction / 2.2
2.3	Optics, Anatomy, Physiology of the Visual System / 2.2
2.4	Visual Performance / 2.14
2.5	Acknowledgments / 2.41
2.6	References / 2.42

---

## Chapter 3. Psychophysical Methods *Denis G. Pelli and Bart Farell* 3.1

3.1	Introduction / 3.1
3.2	Definitions / 3.2
3.3	Visual Stimuli / 3.3
3.4	Adjustments / 3.4
3.5	Judgments / 3.6
	Magnitude Estimation / 3.8
3.6	Stimulus Sequencing / 3.9
3.7	Conclusion / 3.9
3.8	Tips from the Pros / 3.10
3.9	Acknowledgments / 3.10
3.10	References / 3.10

---

## Chapter 4. Visual Acuity and Hyperacuity *Gerald Westheimer* 4.1

4.1	Glossary / 4.1
4.2	Introduction / 4.2
4.3	Stimulus Specifications / 4.2

- 4.4 Optics of the Eye's Resolving Capacity / 4.4
- 4.5 Retinal Limitations—Receptor Mosaic and Tiling of Neuronal Receptive Fields / 4.5
- 4.6 Determination of Visual Resolution Thresholds / 4.6
- 4.7 Kinds of Visual Acuity Tests / 4.7
- 4.8 Factors Affecting Visual Acuity / 4.9
- 4.9 Hyperacuity / 4.14
- 4.10 Resolution, Superresolution, and Information Theory / 4.15
- 4.11 Summary / 4.16
- 4.12 References / 4.16

---

**Chapter 5. Optical Generation of the Visual Stimulus** *Stephen A. Burns and Robert H. Webb* **5.1**

---

- 5.1 Glossary / 5.1
- 5.2 Introduction / 5.1
- 5.3 The Size of the Visual Stimulus / 5.2
- 5.4 Free or Newtonian Viewing / 5.2
- 5.5 Maxwellian Viewing / 5.4
- 5.6 Building an Optical System / 5.8
- 5.7 Light Exposure and Ocular Safety / 5.18
- 5.8 Light Sources / 5.19
- 5.9 Coherent Radiation / 5.19
- 5.10 Detectors / 5.21
- 5.11 Putting It Together / 5.21
- 5.12 Conclusions / 5.24
- 5.13 Acknowledgments / 5.24
- 5.14 General References / 5.25
- 5.15 References / 5.26

---

**Chapter 6. The Maxwellian View: with an Addendum on Apodization** *Gerald Westheimer* **6.1**

---

- 6.1 Glossary / 6.1
- 6.2 Introduction / 6.2
- 6.3 Postscript (2008) / 6.13

---

**Chapter 7. Ocular Radiation Hazards** *David H. Sliney* **7.1**

---

- 7.1 Glossary / 7.1
- 7.2 Introduction / 7.2
- 7.3 Injury Mechanisms / 7.2
- 7.4 Types of Injury / 7.3
- 7.5 Retinal Irradiance Calculations / 7.7
- 7.6 Examples / 7.8
- 7.7 Exposure Limits / 7.9
- 7.8 Discussion / 7.11
- 7.9 References / 7.15

---

**Chapter 8. Biological Waveguides** *Vasudevan Lakshminarayanan and Jay M. Enoch* **8.1**

---

- 8.1 Glossary / 8.1
- 8.2 Introduction / 8.2
- 8.3 Waveguiding in Retinal Photoreceptors and the Stiles-Crawford Effect / 8.3
- 8.4 Waveguides and Photoreceptors / 8.3
- 8.5 Photoreceptor Orientation and Alignment / 8.5
- 8.6 Introduction to the Models and Theoretical Implications / 8.8
- 8.7 Quantitative Observations of Single Receptors / 8.15
- 8.8 Waveguide Modal Patterns Found in Monkey/Human Retinal Receptors / 8.19
- 8.9 Light Guide Effect in Cochlear Hair Cells and Human Hair / 8.24

- 8.10 Fiber-Optic Plant Tissues / 8.26
- 8.11 Sponges / 8.28
- 8.12 Summary / 8.29
- 8.13 References / 8.29

---

**Chapter 9.** The Problem of Correction for the Stiles-Crawford Effect of the First Kind in Radiometry and Photometry, a Solution *Jay M. Enoch and Vasudevan Lakshminarayanan* 9.1

---

- 9.1 Glossary / 9.1
- 9.2 Introduction / 9.2
- 9.3 The Problem and an Approach to Its Solution / 9.3
- 9.4 Sample Point-by-Point Estimates of SCE-1 and Integrated SCE-1 Data / 9.6
- 9.5 Discussion / 9.13
- 9.6 Teleological and Developmental Factors / 9.14
- 9.7 Conclusions / 9.14
- 9.8 References / 9.15

---

**Chapter 10.** Colorimetry *David H. Brainard and Andrew Stockman* 10.1

---

- 10.1 Glossary / 10.1
- 10.2 Introduction / 10.2
- 10.3 Fundamentals of Colorimetry / 10.3
- 10.4 Color Coordinate Systems / 10.11
- 10.5 Matrix Representations and Calculations / 10.24
- 10.6 Topics / 10.32
- 10.7 Appendix—Matrix Algebra / 10.45
- 10.8 References / 10.49

---

**Chapter 11.** Color Vision Mechanisms *Andrew Stockman and David H. Brainard* 11.1

---

- 11.1 Glossary / 11.1
- 11.2 Introduction / 11.3
- 11.3 Basics of Color-Discrimination Mechanisms / 11.9
- 11.4 Basics of Color-Appearance Mechanisms / 11.26
- 11.5 Details and Limits of the Basic Model / 11.31
- 11.6 Conclusions / 11.79
- 11.7 Acknowledgments / 11.85
- 11.8 References / 11.86

---

**Chapter 12.** Assessment of Refraction and Refractive Errors and Their Influence on Optical Design *B. Ralph Chou* 12.1

---

- 12.1 Glossary / 12.1
- 12.2 Introduction / 12.3
- 12.3 Refractive Errors / 12.3
- 12.4 Assessment of Refractive Error / 12.5
- 12.5 Correction of Refractive Error / 12.8
- 12.6 Binocular Factors / 12.15
- 12.7 Consequences for Optical Design / 12.17
- 12.8 References / 12.17

---

**Chapter 13.** Binocular Vision Factors That Influence Optical Design *Clifton Schor* 13.1

---

- 13.1 Glossary / 13.1
- 13.2 Combining the Images in the Two Eyes into One Perception of the Visual Field / 13.3
- 13.3 Distortion of Space by Monocular Magnification / 13.13

- 13.4 Distortion of Space Perception from Interocular Aniso-Magnification (Unequal Binocular Magnification) / 13.16
- 13.5 Distortions of Space from Convergence Responses to Prism / 13.19
- 13.6 Eye Movements / 13.19
- 13.7 Coordination and Alignment of the Two Eyes / 13.20
- 13.8 Effects of Lenses and Prism on Vergence and Phoria / 13.25
- 13.9 Prism-Induced Errors of Eye Alignment / 13.27
- 13.10 Head and Eye Responses to Direction (Gaze Control) / 13.29
- 13.11 Focus and Responses to Distance / 13.30
- 13.12 Video Head Sets, Head's Up Displays and Virtual Reality: Impact on Binocular Vision / 13.31
- 13.13 References / 13.35

**Chapter 14. Optics and Vision of the Aging Eye** *John S. Werner, Brooke E. Scheffrin, and Arthur Bradley* 14.1

---

- 14.1 Glossary / 14.1
- 14.2 Introduction / 14.2
- 14.3 The Graying of the Planet / 14.2
- 14.4 Senescence of the Eye's Optics / 14.4
- 14.5 Senescent Changes in Vision / 14.14
- 14.6 Age-Related Ocular Diseases Affecting Visual Function / 14.22
- 14.7 The Aging World from the Optical Point of View: Presbyopic Corrections / 14.27
- 14.8 Conclusions / 14.30
- 14.9 Acknowledgments / 14.30
- 14.10 References / 14.30

**Chapter 15. Adaptive Optics in Retinal Microscopy and Vision** *Donald T. Miller and Austin Roorda* 15.1

---

- 15.1 Glossary / 15.1
- 15.2 Introduction / 15.2
- 15.3 Properties of Ocular Aberrations / 15.4
- 15.4 Implementation of AO / 15.7
- 15.5 Application of AO to the Eye / 15.15
- 15.6 Acknowledgments / 15.24
- 15.7 References / 15.24

**Chapter 16. Refractive Surgery, Correction of Vision, PRK and LASIK** *L. Diaz-Santana and Harilaos Ginis* 16.1

---

- 16.1 Glossary / 16.1
- 16.2 Introduction / 16.2
- 16.3 Refractive Surgery Modalities / 16.9
- 16.4 Laser Ablation / 16.15
- 16.5 Acknowledgments / 16.19
- 16.6 References / 16.19

**Chapter 17. Three-Dimensional Confocal Microscopy of the Living Human Cornea** *Barry R. Masters* 17.1

---

- 17.1 Glossary / 17.1
- 17.2 Introduction / 17.3
- 17.3 Theory of Confocal Microscopy / 17.3
- 17.4 The Development of Confocal Instruments / 17.3
- 17.5 The Scanning Slit and Laser Scanning Clinical Confocal Microscopes / 17.6
- 17.6 Clinical Applications of Confocal Microscopy / 17.8
- 17.7 Perspectives / 17.9
- 17.8 Summary / 17.10
- 17.9 Acknowledgments / 17.10
- 17.10 References / 17.10

---

## Chapter 18. Diagnostic Use of Optical Coherence Tomography in the Eye *Johannes F. de Boer* 18.1

---

- 18.1 Glossary / 18.1
- 18.2 Introduction / 18.2
- 18.3 Principle of OCT: Time Domain OCT / 18.3
- 18.4 Principle of OCT: Spectral Domain OCT / 18.5
- 18.5 Principle of OCT: Optical Frequency Domain Imaging / 18.7
- 18.6 SD-OCT Versus OFDI / 18.9
- 18.7 Sensitivity Advantage of SD-OCT Over TD-OCT / 18.9
- 18.8 Noise Analysis of SD-OCT Using Charge Coupled Devices (CCDs) / 18.9
- 18.9 Signal to Noise Ratio and Autocorrelation Noise / 18.11
- 18.10 Shot-Noise-Limited Detection / 18.12
- 18.11 Depth Dependent Sensitivity / 18.13
- 18.12 Motion Artifacts and Fringe Washout / 18.15
- 18.13 OFDI at 1050 NM / 18.15
- 18.14 Functional Extensions: Doppler OCT and Polarization Sensitive OCT / 18.18
- 18.15 Doppler OCT and Phase Stability / 18.18
- 18.16 Polarization Sensitive OCT (PS-OCT) / 18.20
- 18.17 PS-OCT in Ophthalmology / 18.24
- 18.18 Retinal Imaging with SD-OCT / 18.27
- 18.19 Conclusion / 18.29
- 18.20 Acknowledgment / 18.30
- 18.21 References / 18.30

## Chapter 19. Gradient Index Optics in the Eye *Barbara K. Pierscionek* 19.1

---

- 19.1 Glossary / 19.1
- 19.2 Introduction / 19.2
- 19.3 The Nature of an Index Gradient / 19.2
- 19.4 Spherical Gradients / 19.2
- 19.5 Radial Gradients / 19.3
- 19.6 Axial Gradients / 19.5
- 19.7 The Eye Lens / 19.5
- 19.8 Fish / 19.6
- 19.9 Octopus / 19.7
- 19.10 Rat / 19.7
- 19.11 Guinea Pig / 19.8
- 19.12 Rabbit / 19.8
- 19.13 Cat / 19.9
- 19.14 Bovine / 19.9
- 19.15 Pig / 19.11
- 19.16 Human/primate / 19.12
- 19.17 Functional Considerations / 19.14
- 19.18 Summary / 19.15
- 19.19 References / 19.15

## Chapter 20. Optics of Contact Lenses *Edward S. Bennett* 20.1

---

- 20.1 Glossary / 20.1
- 20.2 Introduction / 20.2
- 20.3 Contact Lens Material, Composition, and Design Parameters / 20.3
- 20.4 Contact Lens Power / 20.6
- 20.5 Other Design Considerations / 20.20
- 20.6 Convergence and Accommodation Effects / 20.25
- 20.7 Prismatic Effects / 20.30
- 20.8 Magnification / 20.31
- 20.9 Summary / 20.34
- 20.10 Acknowledgments / 20.34
- 20.11 References / 20.34



<b>Chapter 21.</b>	<b>Intraocular Lenses</b>	<i>Jim Schwiegerling</i>	<b>21.1</b>
21.1	Glossary	/ 21.1	
21.2	Introduction	/ 21.2	
21.3	Cataract Surgery	/ 21.4	
21.4	Intraocular Lens Design	/ 21.5	
21.5	Intraocular Lens Side Effects	/ 21.20	
21.6	Summary	/ 21.22	
21.7	References	/ 21.22	
<b>Chapter 22.</b>	<b>Displays for Vision Research</b>	<i>William Cowan</i>	<b>22.1</b>
22.1	Glossary	/ 22.1	
22.2	Introduction	/ 22.2	
22.3	Operational Characteristics of Color Monitors	/ 22.3	
22.4	Colorimetric Calibration of Video Monitors	/ 22.20	
22.5	An Introduction to Liquid Crystal Displays	/ 22.34	
22.6	Acknowledgments	/ 22.40	
22.7	References	/ 22.40	
<b>Chapter 23.</b>	<b>Vision Problems at Computers</b>	<i>Jeffrey Anshel and James E. Sheedy</i>	<b>23.1</b>
23.1	Glossary	/ 23.1	
23.2	Introduction	/ 23.4	
23.3	Work Environment	/ 23.4	
23.4	Vision and Eye Conditions	/ 23.9	
23.5	References	/ 23.12	
<b>Chapter 24.</b>	<b>Human Vision and Electronic Imaging</b>	<i>Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Jan P. Allebach</i>	<b>24.1</b>
24.1	Introduction	/ 24.1	
24.2	Early Vision Approaches: The Perception of Imaging Artifacts	/ 24.2	
24.3	Higher-Level Approaches: The Analysis of Image Features	/ 24.6	
24.4	Very High-Level Approaches: The Representation of Aesthetic and Emotional Characteristics	/ 24.9	
24.5	Conclusions	/ 24.10	
24.6	Additional Information on Human Vision and Electronic Imaging	/ 24.11	
24.7	References	/ 24.11	
<b>Chapter 25.</b>	<b>Visual Factors Associated with Head-mounted Displays</b>	<i>Brian H. Tsou and Martin Shenker</i>	<b>25.1</b>
25.1	Glossary	/ 25.1	
25.2	Introduction	/ 25.1	
25.3	Common Design Considerations among All HMDs	/ 25.2	
25.4	Characterizing HMD	/ 25.7	
25.5	Summary	/ 25.10	
25.6	Appendix	/ 25.10	
25.7	Acknowledgments	/ 25.12	
25.8	References	/ 25.12	

---

# CONTRIBUTORS

---

- Jan P. Allebach** *Electronic Imaging Systems Laboratory, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana (CHAP. 24)*
- Jeffrey Anshel** *Corporate Vision Consulting, Encinitas, California (CHAP. 23)*
- Martin S. Banks** *School of Optometry, University of California, Berkeley, California (CHAP. 2)*
- Edward S. Bennett** *College of Optometry, University of Missouri, St. Louis, Missouri (CHAP. 20)*
- Arthur Bradley** *School of Optometry, Indiana University, Bloomington, Indiana (CHAP. 14)*
- David H. Brainard** *Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania (CHAPS. 10, 11)*
- Stephen A. Burns** *School of Optometry, Indiana University, Bloomington, Indiana (CHAP. 5)*
- Neil Charman** *Department of Optometry and Vision Sciences, University of Manchester, Manchester, United Kingdom (CHAP. 1)*
- B. Ralph Chou** *School of Optometry, University of Waterloo, Waterloo, Ontario, Canada (CHAP. 12)*
- William Cowan** *Department of Computer Science, University of Waterloo, Waterloo, Ontario, Canada (CHAP. 22)*
- Johannes F. de Boer** *Department of Physics, VU University, Amsterdam, and Rotterdam Ophthalmic Institute, Rotterdam, The Netherlands (CHAP. 18)*
- Jay M. Enoch** *School of Optometry, University of California at Berkeley, Berkeley, California (CHAPS. 8, 9)*
- Bart Farell** *Institute for Sensory Research, Syracuse University, Syracuse, New York (CHAP. 3)*
- Wilson S. Geisler** *Department of Psychology, University of Texas, Austin, Texas (CHAP. 2)*
- Harilaos Ginis** *Institute of Vision and Optics, University of Crete, Greece (CHAP. 16)*
- Vasudevan Lakshminarayanan** *School of Optometry and Departments of Physics and Electrical Engineering, University of Waterloo, Waterloo, Ontario, Canada (CHAPS. 8, 9)*
- Barry R. Masters** *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts (CHAP. 17)*
- Donald T. Miller** *School of Optometry, Indiana University, Bloomington, Indiana (CHAP. 15)*
- Thrasylvoulos N. Pappas** *Department of Electrical and Computer Engineering, Northwestern University Evanston, Illinois (CHAP. 24)*
- Denis G. Pelli** *Psychology Department and Center for Neural Science, New York University, New York (CHAP. 3)*
- Barbara K. Pierscionek** *Department of Biomedical Sciences, University of Ulster, Coleraine, United Kingdom (CHAP. 19)*
- Bernice E. Rogowitz** *IBM T. J. Watson Research Center, Hawthorne, New York (CHAP. 24)*
- Austin Roorda** *School of Optometry, University of California, Berkeley, California (CHAP. 15)*
- L. Diaz-Santana** *Department of Optometry and Visual Science, City University, London, United Kingdom (CHAP. 16)*
- Brooke E. Schefrin** *Department of Psychology, University of Colorado, Boulder, Colorado (CHAP. 14)*
- Clifton Schor** *School of Optometry, University of California, Berkeley, California (CHAP. 13)*
- Jim Schwiegerling** *Department of Ophthalmology, University of Arizona, Tucson, Arizona (CHAP. 21)*

- James E. Sheedy** *College of Optometry, Pacific University, Forest Grove, Oregon (CHAP. 23)*
- Martin Shenker** *Martin Shenker Optical Design, Inc., White Plains, New York (CHAP. 25)*
- David H. Sliney** *Consulting Medical Physicist, Fallston, Maryland, and Retired, U.S. Army Center for Health Promotion and Preventive Medicine, Laser/Optical Radiation Program, Aberdeen Proving Ground, Maryland (CHAP. 7)*
- Andrew Stockman** *Department of Visual Neuroscience, UCL Institute of Ophthalmology, London, United Kingdom (CHAPS. 10, 11)*
- Brian H. Tsou** *Air Force Research Laboratory, Wright Patterson AFB, Ohio (CHAP. 25)*
- Robert H. Webb** *The Schepens Eye Research Institute, Boston, Massachusetts (CHAP. 5)*
- John S. Werner** *Department of Ophthalmology & Vision Science, University of California, Davis, Sacramento, California (CHAP. 14)*
- Gerald Westheimer** *Division of Neurobiology, University of California, Berkeley, California (CHAPS. 4, 6)*

---

# BRIEF CONTENTS OF ALL VOLUMES

---

## VOLUME I. GEOMETRICAL AND PHYSICAL OPTICS, POLARIZED LIGHT, COMPONENTS AND INSTRUMENTS

---

### PART 1. GEOMETRICAL OPTICS

---

Chapter 1. General Principles of Geometrical Optics *Douglas S. Goodman*

### PART 2. PHYSICAL OPTICS

---

Chapter 2. Interference *John E. Greivenkamp*

Chapter 3. Diffraction *Arvind S. Marathay and John F. McCalmont*

Chapter 4. Transfer Function Techniques *Glenn D. Boreman*

Chapter 5. Coherence Theory *William H. Carter*

Chapter 6. Coherence Theory: Tools and Applications *Gisele Bennett, William T. Rhodes, and J. Christopher James*

Chapter 7. Scattering by Particles *Craig F. Bohren*

Chapter 8. Surface Scattering *Eugene L. Church and Peter Z. Takacs*

Chapter 9. Volume Scattering in Random Media *Aristide Dogariu and Jeremy Ellis*

Chapter 10. Optical Spectroscopy and Spectroscopic Lineshapes *Brian Henderson*

Chapter 11. Analog Optical Signal and Image Processing *Joseph W. Goodman*

### PART 3. POLARIZED LIGHT

---

Chapter 12. Polarization *Jean M. Bennett*

Chapter 13. Polarizers *Jean M. Bennett*

Chapter 14. Mueller Matrices *Russell A. Chipman*

Chapter 15. Polarimetry *Russell A. Chipman*

Chapter 16. Ellipsometry *Rasheed M. A. Azzam*

### PART 4. COMPONENTS

---

Chapter 17. Lenses *R. Barry Johnson*

Chapter 18. Afocal Systems *William B. Wetherell*

Chapter 19. Nondispersive Prisms *William L. Wolfe*

Chapter 20. Dispersive Prisms and Gratings *George J. Zisis*

Chapter 21. Integrated Optics *Thomas L. Koch, Frederick J. Leonberger, and Paul G. Suchoski*

Chapter 22. Miniature and Micro-Optics *Tom D. Milster and Tomasz S. Tkaczyk*

Chapter 23. Binary Optics *Michael W. Farn and Wilfrid B. Veldkamp*

Chapter 24. Gradient Index Optics *Duncan T. Moore*

### PART 5. INSTRUMENTS

---

Chapter 25. Cameras *Norman Goldberg*

Chapter 26. Solid-State Cameras *Gerald C. Holst*

Chapter 27. Camera Lenses *Ellis Betensky, Melvin H. Kreitzer, and Jacob Moskovich*

Chapter 28. Microscopes *Rudolf Oldenbourg and Michael Shribak*

Chapter 29. Reflective and Catadioptric Objectives *Lloyd Jones*

- Chapter 30. Scanners *Leo Beiser and R. Barry Johnson*
- Chapter 31. Optical Spectrometers *Brian Henderson*
- Chapter 32. Interferometers *Parameswaran Hariharan*
- Chapter 33. Holography and Holographic Instruments *Lloyd Huff*
- Chapter 34. Xerographic Systems *Howard Stark*
- Chapter 35. Principles of Optical Disk Data Storage *Masud Mansuripur*

## VOLUME II. DESIGN, FABRICATION, AND TESTING; SOURCES AND DETECTORS; RADIOMETRY AND PHOTOMETRY

---

### PART 1. DESIGN

---

- Chapter 1. Techniques of First-Order Layout *Warren J. Smith*
- Chapter 2. Aberration Curves in Lens Design *Donald C. O'Shea and Michael E. Harrigan*
- Chapter 3. Optical Design Software *Douglas C. Sinclair*
- Chapter 4. Optical Specifications *Robert R. Shannon*
- Chapter 5. Tolerancing Techniques *Robert R. Shannon*
- Chapter 6. Mounting Optical Components *Paul R. Yoder, Jr.*
- Chapter 7. Control of Stray Light *Robert P. Breault*
- Chapter 8. Thermal Compensation Techniques *Philip J. Rogers and Michael Roberts*

### PART 2. FABRICATION

---

- Chapter 9. Optical Fabrication *Michael P. Mandina*
- Chapter 10. Fabrication of Optics by Diamond Turning *Richard L. Rhorer and Chris J. Evans*

### PART 3. TESTING

---

- Chapter 11. Orthonormal Polynomials in Wavefront Analysis *Virendra N. Mahajan*
- Chapter 12. Optical Metrology *Zacarias Malacara and Daniel Malacara-Hernández*
- Chapter 13. Optical Testing *Daniel Malacara-Hernández*
- Chapter 14. Use of Computer-Generated Holograms in Optical Testing *Katherine Creath and James C. Wyant*

### PART 4. SOURCES

---

- Chapter 15. Artificial Sources *Anthony LaRocca*
- Chapter 16. Lasers *William T. Silfvast*
- Chapter 17. Light-Emitting Diodes *Roland H. Haitz, M. George Craford, and Robert H. Weissman*
- Chapter 18. High-Brightness Visible LEDs *Winston V. Schoenfeld*
- Chapter 19. Semiconductor Lasers *Pamela L. Derry, Luis Figueroa, and Chi-shain Hong*
- Chapter 20. Ultrashort Optical Sources and Applications *Jean-Claude Diels and Ladan Arissian*
- Chapter 21. Attosecond Optics *Zenghu Chang*
- Chapter 22. Laser Stabilization *John L. Hall, Matthew S. Taubman, and Jun Ye*
- Chapter 23. Quantum Theory of the Laser *János A. Bergou, Berthold-Georg Englert, Melvin Lax, Marian O. Scully, Herbert Walther, and M. Suhail Zubairy*

### PART 5. DETECTORS

---

- Chapter 24. Photodetectors *Paul R. Norton*
- Chapter 25. Photodetection *Abhay M. Joshi and Gregory H. Olsen*
- Chapter 26. High-Speed Photodetectors *John E. Bowers and Yih G. Wey*
- Chapter 27. Signal Detection and Analysis *John R. Willison*
- Chapter 28. Thermal Detectors *William L. Wolfe and Paul W. Kruse*

### PART 6. IMAGING DETECTORS

---

- Chapter 29. Photographic Films *Joseph H. Altman*
- Chapter 30. Photographic Materials *John D. Baloga*

- Chapter 31. Image Tube Intensified Electronic Imaging *C. Bruce Johnson and Larry D. Owen*  
 Chapter 32. Visible Array Detectors *Timothy J. Tredwell*  
 Chapter 33. Infrared Detector Arrays *Lester J. Kozlowski and Walter F. Kosonocky*

## PART 7. RADIOMETRY AND PHOTOMETRY

- Chapter 34. Radiometry and Photometry *Edward F. Zalewski*  
 Chapter 35. Measurement of Transmission, Absorption, Emission, and Reflection *James M. Palmer*  
 Chapter 36. Radiometry and Photometry: Units and Conversions *James M. Palmer*  
 Chapter 37. Radiometry and Photometry for Vision Optics *Yoshi Ohno*  
 Chapter 38. Spectroradiometry *Carolyn J. Sher DeCusatis*  
 Chapter 39. Nonimaging Optics: Concentration and Illumination *William Cassarly*  
 Chapter 40. Lighting and Applications *Anurag Gupta and R. John Koschel*

## VOLUME III. VISION AND VISION OPTICS

- Chapter 1. Optics of the Eye *Neil Charman*  
 Chapter 2. Visual Performance *Wilson S. Geisler and Martin S. Banks*  
 Chapter 3. Psychophysical Methods *Denis G. Pelli and Bart Farrell*  
 Chapter 4. Visual Acuity and Hyperacuity *Gerald Westheimer*  
 Chapter 5. Optical Generation of the Visual Stimulus *Stephen A. Burns and Robert H. Webb*  
 Chapter 6. The Maxwellian View with an Addendum on Apodization *Gerald Westheimer*  
 Chapter 7. Ocular Radiation Hazards *David H. Sliney*  
 Chapter 8. Biological Waveguides *Vasudevan Lakshminarayanan and Jay M. Enoch*  
 Chapter 9. The Problem of Correction for the Stiles-Crawford Effect of the First Kind in Radiometry and Photometry, a Solution *Jay M. Enoch and Vasudevan Lakshminarayanan*  
 Chapter 10. Colorimetry *David H. Brainard and Andrew Stockman*  
 Chapter 11. Color Vision Mechanisms *Andrew Stockman and David H. Brainard*  
 Chapter 12. Assessment of Refraction and Refractive Errors and Their Influence on Optical Design  
*B. Ralph Chou*  
 Chapter 13. Binocular Vision Factors That Influence Optical Design *Clifton Schor*  
 Chapter 14. Optics and Vision of the Aging Eye *John S. Werner, Brooke E. Scheffrin, and Arthur Bradley*  
 Chapter 15. Adaptive Optics in Retinal Microscopy and Vision *Donald T. Miller and Austin Roorda*  
 Chapter 16. Refractive Surgery, Correction of Vision, PRK, and LASIK *L. Diaz-Santana and Harilaos Giniis*  
 Chapter 17. Three-Dimensional Confocal Microscopy of the Living Human Cornea *Barry R. Masters*  
 Chapter 18. Diagnostic Use of Optical Coherence Tomography in the Eye *Johannes F. de Boer*  
 Chapter 19. Gradient Index Optics in the Eye *Barbara K. Pierscionek*  
 Chapter 20. Optics of Contact Lenses *Edward S. Bennett*  
 Chapter 21. Intraocular Lenses *Jim Schwiegerling*  
 Chapter 22. Displays for Vision Research *William Cowan*  
 Chapter 23. Vision Problems at Computers *Jeffrey Anshel and James E. Sheedy*  
 Chapter 24. Human Vision and Electronic Imaging *Bernice E. Rogowitz, Thrasylvoulos N. Pappas, and Jan P. Allebach*  
 Chapter 25. Visual Factors Associated with Head-Mounted Displays *Brian H. Tsou and Martin Shenker*

## VOLUME IV. OPTICAL PROPERTIES OF MATERIALS, NONLINEAR OPTICS, QUANTUM OPTICS

### PART 1. PROPERTIES

- Chapter 1. Optical Properties of Water *Curtis D. Mobley*  
 Chapter 2. Properties of Crystals and Glasses *William J. Tropf, Michael E. Thomas, and Eric W. Rogala*  
 Chapter 3. Polymeric Optics *John D. Lytle*

- Chapter 4. Properties of Metals *Roger A. Paquin*  
Chapter 5. Optical Properties of Semiconductors *David G. Seiler, Stefan Zollner, Alain C. Diebold, and Paul M. Amirtharaj*  
Chapter 6. Characterization and Use of Black Surfaces for Optical Systems *Stephen M. Pompea and Robert P. Breault*  
Chapter 7. Optical Properties of Films and Coatings *Jerzy A. Dobrowolski*  
Chapter 8. Fundamental Optical Properties of Solids *Alan Miller*  
Chapter 9. Photonic Bandgap Materials *Pierre R. Villeneuve*

---

## PART 2. NONLINEAR OPTICS

---

- Chapter 10. Nonlinear Optics *Chung L. Tang*  
Chapter 11. Coherent Optical Transients *Paul R. Berman and D. G. Steel*  
Chapter 12. Photorefractive Materials and Devices *Mark Cronin-Golomb and Marvin Klein*  
Chapter 13. Optical Limiting *David J. Hagan*  
Chapter 14. Electromagnetically Induced Transparency *Jonathan P. Marangos and Thomas Halfmann*  
Chapter 15. Stimulated Raman and Brillouin Scattering *John Reintjes and M. Bashkansky*  
Chapter 16. Third-Order Optical Nonlinearities *Mansoor Sheik-Bahae and Michael P. Hasselbeck*  
Chapter 17. Continuous-Wave Optical Parametric Oscillators *M. Ebrahim-Zadeh*  
Chapter 18. Nonlinear Optical Processes for Ultrashort Pulse Generation *Uwe Siegner and Ursula Keller*  
Chapter 19. Laser-Induced Damage to Optical Materials *Marion J. Soileau*

---

## PART 3. QUANTUM AND MOLECULAR OPTICS

---

- Chapter 20. Laser Cooling and Trapping of Atoms *Harold J. Metcalf and Peter van der Straten*  
Chapter 21. Strong Field Physics *Todd Ditmire*  
Chapter 22. Slow Light Propagation in Atomic and Photonic Media *Jacob B. Khurgin*  
Chapter 23. Quantum Entanglement in Optical Interferometry *Hwang Lee, Christoph F. Wildfeuer, Sean D. Huver, and Jonathan P. Dowling*

---

# VOLUME V. ATMOSPHERIC OPTICS, MODULATORS, FIBER OPTICS, X-RAY AND NEUTRON OPTICS

---

---

## PART 1. MEASUREMENTS

---

- Chapter 1. Scatterometers *John C. Stover*  
Chapter 2. Spectroscopic Measurements *Brian Henderson*

---

## PART 2. ATMOSPHERIC OPTICS

---

- Chapter 3. Atmospheric Optics *Dennis K. Killinger, James H. Churnside, and Laurence S. Rothman*  
Chapter 4. Imaging through Atmospheric Turbulence *Virendra N. Mahajan and Guang-ming Dai*  
Chapter 5. Adaptive Optics *Robert Q. Fugate*

---

## PART 3. MODULATORS

---

- Chapter 6. Acousto-Optic Devices *I-Cheng Chang*  
Chapter 7. Electro-Optic Modulators *Georgianne M. Purvinis and Theresa A. Maldonado*  
Chapter 8. Liquid Crystals *Sebastian Gauza and Shin-Tson Wu*

---

## PART 4. FIBER OPTICS

---

- Chapter 9. Optical Fiber Communication Technology and System Overview *Ira Jacobs*  
Chapter 10. Nonlinear Effects in Optical Fibers *John A. Buck*  
Chapter 11. Photonic Crystal Fibers *Philip St. J. Russell and G. J. Pearce*  
Chapter 12. Infrared Fibers *James A. Harrington*

- Chapter 13. Sources, Modulators, and Detectors for Fiber Optic Communication Systems *Elsa Garmire*  
 Chapter 14. Optical Fiber Amplifiers *John A. Buck*  
 Chapter 15. Fiber Optic Communication Links (Telecom, Datacom, and Analog) *Casimer DeCusatis and Guifang Li*  
 Chapter 16. Fiber-Based Couplers *Daniel Nolan*  
 Chapter 17. Fiber Bragg Gratings *Kenneth O. Hill*  
 Chapter 18. Micro-Optics-Based Components for Networking *Joseph C. Palais*  
 Chapter 19. Semiconductor Optical Amplifiers *Jay M. Wiesenfeld and Leo H. Spiekman*  
 Chapter 20. Optical Time-Division Multiplexed Communication Networks *Peter J. Delfyett*  
 Chapter 21. WDM Fiber-Optic Communication Networks *Alan E. Willner, Changyuan Yu, Zhongqi Pan, and Yong Xie*  
 Chapter 22. Solitons in Optical Fiber Communication Systems *Pavel V. Mamyshev*  
 Chapter 23. Fiber-Optic Communication Standards *Casimer DeCusatis*  
 Chapter 24. Optical Fiber Sensors *Richard O. Claus, Ignacio Matias, and Francisco Arregui*  
 Chapter 25. High-Power Fiber Lasers and Amplifiers *Timothy S. McComb, Martin C. Richardson, and Michael Bass*

## PART 5. X-RAY AND NEUTRON OPTICS

### Subpart 5.1. Introduction and Applications

- Chapter 26. An Introduction to X-Ray and Neutron Optics *Carolyn A. MacDonald*  
 Chapter 27. Coherent X-Ray Optics and Microscopy *Qun Shen*  
 Chapter 28. Requirements for X-Ray diffraction *Scott T. Misture*  
 Chapter 29. Requirements for X-Ray Fluorescence *George J. Havrilla*  
 Chapter 30. Requirements for X-Ray Spectroscopy *Dirk Lützenkirchen-Hecht and Ronald Frahm*  
 Chapter 31. Requirements for Medical Imaging and X-Ray Inspection *Douglas Pfeiffer*  
 Chapter 32. Requirements for Nuclear Medicine *Lars R. Furenlid*  
 Chapter 33. Requirements for X-Ray Astronomy *Scott O. Rohrbach*  
 Chapter 34. Extreme Ultraviolet Lithography *Franco Cerrina and Fan Jiang*  
 Chapter 35. Ray Tracing of X-Ray Optical Systems *Franco Cerrina and M. Sanchez del Rio*  
 Chapter 36. X-Ray Properties of Materials *Eric M. Gullikson*

### Subpart 5.2. Refractive and Interference Optics

- Chapter 37. Refractive X-Ray Lenses *Bruno Lengeler and Christian G. Schroer*  
 Chapter 38. Gratings and Monochromators in the VUV and Soft X-Ray Spectral Region *Malcolm R. Howells*  
 Chapter 39. Crystal Monochromators and Bent Crystals *Peter Siddons*  
 Chapter 40. Zone Plates *Alan Michette*  
 Chapter 41. Multilayers *Eberhard Spiller*  
 Chapter 42. Nanofocusing of Hard X-Rays with Multilayer Laue Lenses *Albert T. Macrander, Hanfei Yan, Hyon Chol Kang, Jörg Maser, Chian Liu, Ray Conley, and G. Brian Stephenson*  
 Chapter 43. Polarizing Crystal Optics *Qun Shen*

### Subpart 5.3. Reflective Optics

- Chapter 44. Reflective Optics *James Harvey*  
 Chapter 45. Aberrations for Grazing Incidence Optics *Timo T. Saha*  
 Chapter 46. X-Ray Mirror Metrology *Peter Z. Takacs*  
 Chapter 47. Astronomical X-Ray Optics *Marshall K. Joy and Brian D. Ramsey*  
 Chapter 48. Multifoil X-Ray Optics *Ladislav Pina*  
 Chapter 49. Pore Optics *Marco Beijersbergen*  
 Chapter 50. Adaptive X-Ray Optics *Ali Khounsary*  
 Chapter 51. The Schwarzschild Objective *Franco Cerrina*  
 Chapter 52. Single Capillaries *Donald H. Bilderback and Sterling W. Cornaby*  
 Chapter 53. Polycapillary X-Ray Optics *Carolyn MacDonald and Walter M. Gibson*



**Subpart 5.4. X-Ray Sources**

- Chapter 54. X-Ray Tube Sources *Susanne M. Lee and Carolyn MacDonald*
- Chapter 55. Synchrotron Sources *Steven L. Hulbert and Gwyn P. Williams*
- Chapter 56. Laser Generated Plasmas *Alan Michette*
- Chapter 57. Pinch Plasma Sources *Victor Kantsyrev*
- Chapter 58. X-Ray Lasers *Greg Tallents*
- Chapter 59. Inverse Compton X-Ray Sources *Frank Carroll*

**Subpart 5.5. X-Ray Detectors**

- Chapter 60. Introduction to X-Ray Detectors *Walter M. Gibson and Peter Siddons*
- Chapter 61. Advances in Imaging Detectors *Aaron Couture*
- Chapter 62. X-Ray Spectral Detection and Imaging *Eric Lifshin*

**Subpart 5.6. Neutron Optics and Applications**

- Chapter 63. Neutron Optics *David Mildner*
- Chapter 64. Grazing-Incidence Neutron Optics *Mikhail Gubarev and Brian Ramsey*

---

# EDITORS' PREFACE

---

The third edition of the *Handbook of Optics* is designed to pull together the dramatic developments in both the basic and applied aspects of the field while retaining the archival, reference book value of a handbook. This means that it is much more extensive than either the first edition, published in 1978, or the second edition, with Volumes I and II appearing in 1995 and Volumes III and IV in 2001. To cover the greatly expanded field of optics, the *Handbook* now appears in five volumes. Over 100 authors or author teams have contributed to this work.

Volume I is devoted to the fundamentals, components, and instruments that make optics possible. Volume II contains chapters on design, fabrication, testing, sources of light, detection, and a new section devoted to radiometry and photometry. Volume III concerns vision optics only and is printed entirely in color. In Volume IV there are chapters on the optical properties of materials, nonlinear, quantum and molecular optics. Volume V has extensive sections on fiber optics and x ray and neutron optics, along with shorter sections on measurements, modulators, and atmospheric optical properties and turbulence. Several pages of color inserts are provided where appropriate to aid the reader. A purchaser of the print version of any volume of the *Handbook* will be able to download a digital version containing all of the material in that volume in PDF format to one computer (see download instructions on bound-in card). The combined index for all five volumes can be downloaded from [www.HandbookofOpticsOnline.com](http://www.HandbookofOpticsOnline.com).

It is possible by careful selection of what and how to present that the third edition of the *Handbook* could serve as a text for a comprehensive course in optics. In addition, students who take such a course would have the *Handbook* as a career-long reference.

Topics were selected by the editors so that the *Handbook* could be a desktop (bookshelf) general reference for the parts of optics that had matured enough to warrant archival presentation. New chapters were included on topics that had reached this stage since the second edition, and existing chapters from the second edition were updated where necessary to provide this compendium. In selecting subjects to include, we also had to select which subjects to leave out. The criteria we applied were: (1) was it a specific application of optics rather than a core science or technology and (2) was it a subject in which the role of optics was peripheral to the central issue addressed. Thus, such topics as medical optics, laser surgery, and laser materials processing were not included. While applications of optics are mentioned in the chapters there is no space in the *Handbook* to include separate chapters devoted to all of the myriad uses of optics in today's world. If we had, the third edition would be much longer than it is and much of it would soon be outdated. We designed the third edition of the *Handbook of Optics* so that it concentrates on the principles of optics that make applications possible.

Authors were asked to try to achieve the dual purpose of preparing a chapter that was a worthwhile reference for someone working in the field and that could be used as a starting point to become acquainted with that aspect of optics. They did that and we thank them for the outstanding results seen throughout the *Handbook*. We also thank Mr. Taisuke Soda of McGraw-Hill for his help in putting this complex project together and Mr. Alan Tourtlotte and Ms. Susannah Lehman of the Optical Society of America for logistical help that made this effort possible.

We dedicate the third edition of the *Handbook of Optics* to all of the OSA volunteers who, since OSA's founding in 1916, give their time and energy to promoting the generation, application, archiving, and worldwide dissemination of knowledge in optics and photonics.

*Michael Bass, Editor-in-Chief*

*Associate Editors:*

*Casimer M. DeCusatis*

*Jay M. Enoch*

*Vasudevan Lakshminarayanan*

*Guifang Li*

*Carolyn MacDonald*

*Virendra N. Mahajan*

*Eric Van Stryland*

*This page intentionally left blank*

---

# PREFACE TO VOLUME III

---

Volume III of the *Handbook of Optics*, Third Edition, addresses topics relating to vision and the eye which are applicable to, or relate to the study of optics. For reasons we do not understand fully, in recent years, there seems to have been a tendency for the optics and the vision science communities (the latter group was known in earlier times as “physiological optics”) to drift somewhat apart. Physiological optics had become a meaningful component within optics during the latter part of the nineteenth century. As but one example, we urge interested readers to read H. von Helmholtz’s masterful three-volume *Handbook of Physiological Optics* (third edition) which was translated by J. P. C. Southall, of Columbia University, into English by the Optical Society of America in the 1920s.<sup>1</sup> It should also be noted that Allvar Gullstrand received the Nobel Prize in Physiology/Medicine in 1911 for his work on model eyes, which was a direct application of thick lens theory. Gullstrand was not only a professor of ophthalmology at the University of Uppsala, but was also a professor of physiological and physical optics at that institution. He also added five new chapters to the first volume of Helmholtz’s treatise published in 1909. Not only is this a remarkable scientific work, but much of it remains applicable today! The simple fact is that the two groups, optical science and vision science need each other, or, alternatively, are effectively “joined at the hip.” Thus, here, we seek to provide a broad view of vision, vision processes, and discussions of areas where vision science interacts with the ever broadening field of optics.

Obviously, no treatment such as this one can be complete, but we have tried here to present applicable topics in an orderly manner. In the current edition, we have taken a wide-ranging view of vision and its relationship with optics. In particular, in recent years, we have seen a rapid increase of interest in new technologies and applications in the areas of adaptive optics (AO), scanning laser ophthalmoscopy (SLO), and optical coherence tomography (OCT), amongst others. Separately, there has been rapid growth of refractive surgery (LASIK, etc.), use of intraocular lenses (IOLs), and other forms of visual corrections. And, we do not overlook the incredible expansion of information technology, the broad utilization of computer, video, and other forms of displays which have been employed in myriad applications (with associated implications in vision).

We want to call the reader’s attention to the three cover illustrations. Here, of course, our choices were many! We have chosen one for its historical value, it is a photograph taken of a modern young lady viewing herself in an obsidian mirror in bright sunlight. That obsidian mirror, buried twice for extended time periods in its history, is ca. 8000 years old (!), and it is one of a number of the oldest known mirrors. These items are displayed in the Museum of Anatolian Civilizations, located in Ankara, Turkey<sup>2</sup> and/or at the Konya Museum in Konya which is in the south-central valley of Turkey and is located near to the dig site. This photograph falls into the evolving field of archaeological optics (not treated in this edition).<sup>3,4</sup> Please consider the quality of the image in that “stone-age” mirror which was manufactured during the mesolithic or epipaleolithic period! A second figure displays a wave-guide modal pattern obtained radiating from a single human or primate photoreceptor in the early 1960s. There is further discussion of this topic in Chap. 8 on biological waveguides. The third figure is of human parafoveal cone photoreceptors taken from the living human eye by Austin Roorda (see Chap. 15). It was obtained using adaptive optics technology. The long (seen in red), middle (seen in green), and short (seen in blue) wavelength absorbing pigments contained in these individual cone photoreceptors are readily defined.

Please note, with the formation of the section on radiometry and photometry in this edition, the chapter addressing such measurements (as they pertain to visual optics), written by Dr. Yoshi Ohno, was relocated to Volume II, Chap. 37. A new, relatively brief, chapter on radiometry and photometry associated with the Stiles-Crawford effect (of the first kind) (Chap. 9) has been added. It was added after the chapter on biological waveguides (Chap. 8). This chapter fitted more logically there (where

the Stiles-Crawford effects are discussed) than in the new section on radiometry and photometry. The new chapter raises issues suggesting that revision is needed for specification of the visual stimulus (retinal illuminance) in certain test situations, particularly if the entrance pupil of the eye is larger than about 3 mm in diameter.

The outline of the “Vision” section of the this edition offers the reader a logical progression of topics addressed. Volume III leads off with an extensive chapter on optics of the eye written by Neil Charman (Chap. 1). This material has been considerably expanded from the earlier version in the second edition. William S. Geisler and Martin S. Banks reproduced their earlier chapter on visual performance (Chap. 2); and Denis G. Pelli and Bart Farrell similarly repeated their chapter on psychophysical methods used to test vision (Chap. 3). We are pleased that Prof. Gerald Westheimer wrote a new chapter on visual acuity and hyperacuity for this edition (Chap. 4). Professors Stephen A. Burns and Robert H. Webb repeated their chapter on optical generation of the visual stimulus (Chap. 5). Gerald Westheimer also kindly allowed the editors to reproduce a “classic” article he wrote some years ago in *Vision Research* on the topic of the Maxwellian view for inclusion in this edition. He also added a valuable addendum updating material in that discussion (Chap. 6). These chapters as a group provide a valuable introduction to this volume. As in other sections of the *Handbook of Optics*, all material written is intended to be at the first-year graduate student level. So saying, it is intended to be readable and readily appreciated by all parties.

The next set of topics is intended to broaden the discussion in several specific areas of interest. Chap. 7 by David H. Sliney addresses radiation hazards associated with vision and vision testing. Vasudevan Lakshminarayanan and Jay M. Enoch address biological waveguides in Chap. 8. This rapidly broadening subject grew out of work on the Stiles-Crawford effects, that is, “the directional sensitivity of the retina,” first reported in 1933. The 75th anniversary of this discovery was celebrated recently at the meeting of the Optical Society of America held in Rochester, New York, in 2008. In Chap. 9, Enoch and Lakshminarayanan speak of issues associated with the specification of the visual stimulus and the integration of the Stiles-Crawford effect of the first kind (SCE-1). These are meaningful matters associated with photometric and radiometric characterization of visual stimuli.

In Chaps. 10 and 11 David H. Brainard and Andrew Stockman address issues associated with color vision and colorimetry. The associate editors felt strongly that it was necessary to expand coverage of these topics in the third edition of the handbook. The chapter on refraction and refractive techniques (Chap. 12) was prepared by a different author, B. Ralph Chou, in this edition. He also broadened the topics covered from those treated in the second edition. Clifton Schor updated his chapter on binocular vision (Chap. 13), and John S. Werner, Brooke E. Scheffrin, and Arthur Bradley updated the discussion on the optics of vision and the aging eye (Chap. 14), a very important topic due to the rapid increase in aging of populations occurring worldwide!

The next portion of this volume addresses new/emerging technology. Donald T. Miller and Austin Roorda teamed-up to write about the rapidly evolving field of adaptive optics (AO) (Chap. 15). The reader will find that AO techniques are being combined with other emerging techniques in order to enhance utility of instruments, those both in development and also now appearing on the market. Included are scanning laser ophthalmoscopy (SLO), optical coherence tomography (OCT), and flood illumination. That is, these emerging technologies offer additional unique advantages. Interestingly, SLO technology originated some years ago in the laboratory of Prof. Robert H. Webb (Chap. 5). Optical coherence tomography (OCT), a powerful new tool useful in ophthalmic examinations (both for study of the anterior and posterior segments of the eye), is discussed by the researcher, Prof. Dr. Johannes F. deBoer (Chap. 18). New techniques for refractive surgery have been addressed by Harilaos Ginis and L. Diaz-Santana in Chap. 16. Dr. Barry R. Masters has considered confocal imaging of the cornea in Chap. 17; and Prof. Barbara K. Pierscionek has addressed the current state of graded index of refraction in the eye lens (GRIN profiles) in Chap. 19. Perhaps the Pierscionek chapter might have been better placed earlier in the order of things. Edward S. Bennett addresses the always very lively field of contact lens optics in Chap. 20 and Dr. Jim Schweigerling considers the optics of intraocular lenses in Chap. 21.

Clearly, we cannot overlook imaging and display problems associated with modern optical science. Thus, from an information processing point of view as applied to optical problems, William Cowan considers displays for vision research in Chap. 22. And Jeffery Anshel has added to, modified

and updated, the chapter which James E. Sheedy had written in the second edition of the *Handbook of Optics* (Chap. 23). That chapter addressed visual problems and needs of the observer when using computers—often for long periods of time. Bernice Rogowitz, Thrasyvoulos N. Pappas, and Jan P. Allerbach discussed human vision and electronic imaging, which is a major issue in the modern environment (Chap. 24). Finally, Brian H. Tsou and Martin Shenker address visual problems associated with heads-up displays (Chap. 25). The latter problems have been a particular challenge in the aviation industry.

Thus, we have tried to cover reasonably “the waterfront” of interactions between man/eye and instruments (human engineering/ergonomics, if one prefers). And we have sought to address directly issues related to optics per se. These changes have resulted in a longer volume than in the past. So saying, we wish to emphasize that the material is not encyclopedic; that is, we wish we had more material on eye movements, as well as subjects such as aniseikonia or problems encountered by individuals with unequal image sizes in their two eyes. Relating man to the optical instruments or images presented is no small thing from a number of points of view. So saying, we sought to achieve a reasonable balance in topics presented, and in the lengths of these discussions. Obviously, we were constrained by time and availability of authors. We thank all of our diligent group of authors spread out around the world and their helpers, as well as the editorial team at McGraw-Hill and those representing the Optical Society of America, particularly Editor-in-Chief, Michael Bass!

## REFERENCES

1. H. von Helmholtz, *Handbuch der Physiologischen Optik*, 3d ed., 3 volumes, 1911; with commentary by A. Gullstrand, J. von Kries, W. Nagel., and a chapter by Christine Ladd-Franklin. *Handbook/Treatise on Physiological Optics*, translated to English by H. P. C. Southall. Published by the Optical Society of America, 1924. Reprinted by Dover Press, New York City, 1962. (The three original volumes were combined into two print volumes in the Dover Edition.)
2. Jay M. Enoch, “History of Mirrors Dating Back 8000 Years,” *Optometry and Vision Science* **83**:775–781, 2006.
3. Jay M. Enoch, “Archaeological Optics,” chap. 27 in Arthur H. Guenther (ed.): *International Trends in Applied Optics*, International Commission on Optics, vol. 5, Bellingham, Wash., SPIE Press, Monograph: PM 119. 2002, pp. 629–666. (ISBN: 0-8194-4510-X).
4. Jay M. Enoch, “Archeological Optics,” *Journal of Modern Optics* **54**:1221–1239, 2007.

*Jay M. Enoch and Vasudevan Lakshminarayanan*  
Associate Editors

*This page intentionally left blank*

---

# GLOSSARY AND FUNDAMENTAL CONSTANTS

---

## Introduction

This glossary of the terms used in the Handbook represents to a large extent the language of optics. The symbols are representations of numbers, variables, and concepts. Although the basic list was compiled by the author of this section, all the editors have contributed and agreed to this set of symbols and definitions. Every attempt has been made to use the same symbols for the same concepts throughout the entire Handbook, although there are exceptions. Some symbols seem to be used for many concepts. The symbol  $\alpha$  is a prime example, as it is used for absorptivity, absorption coefficient, coefficient of linear thermal expansion, and more. Although we have tried to limit this kind of redundancy, we have also bowed deeply to custom.

## Units

The abbreviations for the most common units are given first. They are consistent with most of the established lists of symbols, such as given by the International Standards Organization ISO<sup>1</sup> and the International Union of Pure and Applied Physics, IUPAP.<sup>2</sup>

## Prefixes

Similarly, a list of the numerical prefixes<sup>1</sup> that are most frequently used is given, along with both the common names (where they exist) and the multiples of ten that they represent.

## Fundamental Constants

The values of the fundamental constants<sup>3</sup> are listed following the sections on SI units.

## Symbols

The most commonly used symbols are then given. Most chapters of the Handbook also have a glossary of the terms and symbols specific to them for the convenience of the reader. In the following list, the symbol is given, its meaning is next, and the most customary unit of measure for the quantity is presented in brackets. A bracket with a dash in it indicates that the quantity is unitless. Note that there is a difference between units and dimensions. An angle has units of degrees or radians and a solid angle square degrees or steradians, but both are pure ratios and are dimensionless. The unit symbols as recommended in the SI system are used, but decimal multiples of some of the dimensions are sometimes given. The symbols chosen, with some cited exceptions, are also those of the first two references.

---

## RATIONALE FOR SOME DISPUTED SYMBOLS

---

The choice of symbols is a personal decision, but commonality improves communication. This section explains why the editors have chosen the preferred symbols for the *Handbook*. We hope that this will encourage more agreement.



## Fundamental Constants

It is encouraging that there is almost universal agreement for the symbols for the fundamental constants. We have taken one small exception by adding a subscript  $B$  to the  $k$  for Boltzmann's constant.

## Mathematics

We have chosen  $i$  as the imaginary almost arbitrarily. IUPAP lists both  $i$  and  $j$ , while ISO does not report on these.

## Spectral Variables

These include expressions for the wavelength  $\lambda$ , frequency  $\nu$ , wave number  $\sigma$ ,  $\omega$  for circular or radian frequency,  $k$  for circular or radian wave number and dimensionless frequency  $x$ . Although some use  $f$  for frequency, it can be easily confused with electronic or spatial frequency. Some use  $\tilde{\nu}$  for wave number, but, because of typography problems and agreement with ISO and IUPAP, we have chosen  $\sigma$ ; it should not be confused with the Stefan-Boltzmann constant. For spatial frequencies we have chosen  $\xi$  and  $\eta$ , although  $f_x$  and  $f_y$  are sometimes used. ISO and IUPAP do not report on these.

## Radiometry

Radiometric terms are contentious. The most recent set of recommendations by ISO and IUPAP are  $L$  for radiance [ $\text{Wcm}^{-2}\text{sr}^{-1}$ ],  $M$  for radiant emittance or exitance [ $\text{Wcm}^{-2}$ ],  $E$  for irradiance or incidence [ $\text{Wcm}^{-2}$ ], and  $I$  for intensity [ $\text{Wsr}^{-2}$ ]. The previous terms,  $W$ ,  $H$ ,  $N$ , and  $J$ , respectively, are still in many texts, notably Smith<sup>4</sup> and Lloyd<sup>5</sup> but we have used the revised set, although there are still shortcomings. We have tried to deal with the vexatious term *intensity* by using *specific intensity* when the units are  $\text{Wcm}^{-2}\text{sr}^{-1}$ , *field intensity* when they are  $\text{Wcm}^{-2}$ , and *radiometric intensity* when they are  $\text{Wsr}^{-1}$ .

There are two sets of terms for these radiometric quantities, which arise in part from the terms for different types of reflection, transmission, absorption, and emission. It has been proposed that the *ion* ending indicate a process, that the *ance* ending indicate a value associated with a particular sample, and that the *ivity* ending indicate a generic value for a "pure" substance. Then one also has reflectance, transmittance, absorptance, and emittance as well as reflectivity, transmissivity, absorptivity, and emissivity. There are now two different uses of the word emissivity. Thus the words *exitance*, *incidence*, and *sterance* were coined to be used in place of emittance, irradiance, and radiance. It is interesting that ISO uses radiance, exitance, and irradiance whereas IUPAP uses radiance, exitance [*sic*], and irradiance. We have chosen to use them both, i.e., emittance, irradiance, and radiance will be followed in square brackets by exitance, incidence, and sterance (or vice versa). Individual authors will use the different endings for transmission, reflection, absorption, and emission as they see fit.

We are still troubled by the use of the symbol  $E$  for irradiance, as it is so close in meaning to electric field, but we have maintained that accepted use. The spectral concentrations of these quantities, indicated by a wavelength, wave number, or frequency subscript (e.g.,  $L_\lambda$ ) represent partial differentiations; a subscript  $q$  represents a photon quantity; and a subscript  $\nu$  indicates a quantity normalized to the response of the eye. Thereby,  $L_\nu$  is luminance,  $E_\nu$  illuminance, and  $M_\nu$  and  $I_\nu$  luminous emittance and luminous intensity. The symbols we have chosen are consistent with ISO and IUPAP.

The refractive index may be considered a radiometric quantity. It is generally complex and is indicated by  $\tilde{n} = n - ik$ . The real part is the relative refractive index and  $k$  is the extinction coefficient. These are consistent with ISO and IUPAP, but they do not address the complex index or extinction coefficient.

## Optical Design

For the most part ISO and IUPAP do not address the symbols that are important in this area.

There were at least 20 different ways to indicate focal ratio; we have chosen FN as symmetrical with NA; we chose  $f$  and  $e\text{fl}$  to indicate the effective focal length. Object and image distance, although given many different symbols, were finally called  $s_o$  and  $s_i$  since  $s$  is an almost universal symbol for distance. Field angles are  $\theta$  and  $\phi$ ; angles that measure the slope of a ray to the optical axis are  $u$ ;  $u$  can also be  $\sin u$ . Wave aberrations are indicated by  $W_{ijk}$ , while third-order ray aberrations are indicated by  $\sigma_i$  and more mnemonic symbols.

## Electromagnetic Fields

There is no argument about  $\mathbf{E}$  and  $\mathbf{H}$  for the electric and magnetic field strengths,  $Q$  for quantity of charge,  $\rho$  for volume charge density,  $\sigma$  for surface charge density, etc. There is no guidance from Refs. 1 and 2 on polarization indication. We chose  $\perp$  and  $\parallel$  rather than  $p$  and  $s$ , partly because  $s$  is sometimes also used to indicate scattered light.

There are several sets of symbols used for reflection transmission, and (sometimes) absorption, each with good logic. The versions of these quantities dealing with field amplitudes are usually specified with lower case symbols:  $r$ ,  $t$ , and  $a$ . The versions dealing with power are alternately given by the uppercase symbols or the corresponding Greek symbols:  $R$  and  $T$  versus  $\rho$  and  $\tau$ . We have chosen to use the Greek, mainly because these quantities are also closely associated with Kirchhoff's law that is usually stated symbolically as  $\alpha = \epsilon$ . The law of conservation of energy for light on a surface is also usually written as  $\alpha + \rho + \tau = 1$ .

## Base SI Quantities

length	m	meter
time	s	second
mass	kg	kilogram
electric current	A	ampere
temperature	K	kelvin
amount of substance	mol	mole
luminous intensity	cd	candela

## Derived SI Quantities

energy	J	joule
electric charge	C	coulomb
electric potential	V	volt
electric capacitance	F	farad
electric resistance	$\Omega$	ohm
electric conductance	S	siemens
magnetic flux	Wb	weber
inductance	H	henry
pressure	Pa	pascal
magnetic flux density	T	tesla
frequency	Hz	hertz
power	W	watt
force	N	newton
angle	rad	radian
angle	sr	steradian

## Prefixes

Symbol	Name	Common name	Exponent of ten
F	exa		18
P	peta		15
T	tera	trillion	12
G	giga	billion	9
M	mega	million	6
k	kilo	thousand	3
h	hecto	hundred	2
da	deca	ten	1
d	deci	tenth	-1
c	centi	hundredth	-2
m	milli	thousandth	-3
$\mu$	micro	millionth	-6
n	nano	billionth	-9
p	pico	trillionth	-12
f	femto		-15
a	atto		-18

## Constants

$c$	speed of light vacuo [ $299792458 \text{ ms}^{-1}$ ]
$c_1$	first radiation constant = $2\pi^5 c^2 h = 3.7417749 \times 10^{-16} [\text{Wm}^2]$
$c_2$	second radiation constant = $hc/k = 0.0144838769 [\text{mK}]$
$e$	elementary charge [ $1.60217733 \times 10^{-19} \text{ C}$ ]
$g_n$	free fall constant [ $9.80665 \text{ ms}^{-2}$ ]
$h$	Planck's constant [ $6.6260755 \times 10^{-34} \text{ Js}$ ]
$k_B$	Boltzmann constant [ $1.380658 \times 10^{-23} \text{ JK}^{-1}$ ]
$m_e$	mass of the electron [ $9.1093897 \times 10^{-31} \text{ kg}$ ]
$N_A$	Avogadro constant [ $6.0221367 \times 10^{23} \text{ mol}^{-1}$ ]
$R_\infty$	Rydberg constant [ $10973731.534 \text{ m}^{-1}$ ]
$\epsilon_0$	vacuum permittivity [ $\mu_0^{-1} \text{ c}^{-2}$ ]
$\sigma$	Stefan-Boltzmann constant [ $5.67051 \times 10^{-8} \text{ Wm}^{-2} \text{ K}^{-4}$ ]
$\mu_0$	vacuum permeability [ $4\pi \times 10^{-7} \text{ NA}^{-2}$ ]
$\mu_B$	Bohr magneton [ $9.2740154 \times 10^{-24} \text{ JT}^{-1}$ ]

## General

<b>B</b>	magnetic induction [ $\text{Wbm}^{-2}, \text{ kgs}^{-1} \text{ C}^{-1}$ ]
<b>C</b>	capacitance [ $\text{f}, \text{ C}^2 \text{ s}^2 \text{ m}^{-2} \text{ kg}^{-1}$ ]
<b>C</b>	curvature [ $\text{m}^{-1}$ ]
$c$	speed of light in vacuo [ $\text{ms}^{-1}$ ]
$c_1$	first radiation constant [ $\text{Wm}^2$ ]
$c_2$	second radiation constant [ $\text{mK}$ ]
<b>D</b>	electric displacement [ $\text{Cm}^{-2}$ ]
<b>E</b>	incidence [irradiance] [ $\text{Wm}^{-2}$ ]
$e$	electronic charge [coulomb]
$E_v$	illuminance [ $\text{lux}, \text{ lmm}^{-2}$ ]
<b>E</b>	electrical field strength [ $\text{Vm}^{-1}$ ]
$E$	transition energy [ $\text{J}$ ]
$E_g$	band-gap energy [ $\text{eV}$ ]
$f^g$	focal length [ $\text{m}$ ]
$f_c$	Fermi occupation function, conduction band
$f_v$	Fermi occupation function, valence band

FN	focal ratio ( $f$ /number) [—]
$g$	gain per unit length [ $\text{m}^{-1}$ ]
$g_{\text{th}}$	gain threshold per unit length [ $\text{m}^{-1}$ ]
$\mathbf{H}$	magnetic field strength [ $\text{Am}^{-1}$ , $\text{Cs}^{-1} \text{m}^{-1}$ ]
$h$	height [m]
$I$	irradiance (see also $E$ ) [ $\text{Wm}^{-2}$ ]
$I$	radiant intensity [ $\text{Wsr}^{-1}$ ]
$I$	nuclear spin quantum number [—]
$I$	current [A]
$i$	$\sqrt{-1}$
$\text{Im}()$	imaginary part of
$J$	current density [ $\text{Am}^{-2}$ ]
$\mathbf{j}$	total angular momentum [ $\text{kg m}^2 \text{sec}^{-1}$ ]
$J_1()$	Bessel function of the first kind [—]
$k$	radian wave number $=2\pi/\lambda$ [ $\text{rad cm}^{-1}$ ]
$\mathbf{k}$	wave vector [ $\text{rad cm}^{-1}$ ]
$k$	extinction coefficient [—]
$L$	sterance [radiance] [ $\text{Wm}^{-2} \text{sr}^{-1}$ ]
$L_v$	luminance [ $\text{cdm}^{-2}$ ]
$L$	inductance [ $\text{h}$ , $\text{m}^2 \text{kg C}^2$ ]
$L$	laser cavity length
$L, M, N$	direction cosines [—]
$M$	angular magnification [—]
$M$	radiant exitance [radiant emittance] [ $\text{Wm}^{-2}$ ]
$m$	linear magnification [—]
$m$	effective mass [kg]
MTF	modulation transfer function [—]
$N$	photon flux [ $\text{s}^{-1}$ ]
$N$	carrier (number) density [ $\text{m}^{-3}$ ]
$n$	real part of the relative refractive index [—]
$\tilde{n}$	complex index of refraction [—]
NA	numerical aperture [—]
OPD	optical path difference [m]
$P$	macroscopic polarization [ $\text{C m}^{-2}$ ]
$\text{Re}()$	real part of [—]
$R$	resistance [ $\Omega$ ]
$\mathbf{r}$	position vector [m]
$S$	Seebeck coefficient [ $\text{VK}^{-1}$ ]
$s$	spin quantum number [—]
$s$	path length [m]
$S_o$	object distance [m]
$S_i$	image distance [m]
$T$	temperature [K, C]
$t$	time [s]
$t$	thickness [m]
$u$	slope of ray with the optical axis [rad]
$V$	Abbe reciprocal dispersion [—]
$V$	voltage [V, $\text{m}^2 \text{kgs}^{-2} \text{C}^{-1}$ ]
$x, y, z$	rectangular coordinates [m]
$Z$	atomic number [—]

## Greek Symbols

$\alpha$	absorption coefficient [ $\text{cm}^{-1}$ ]
$\alpha$	(power) absorptance (absorptivity)

$\epsilon$	dielectric coefficient (constant) [—]
$\epsilon$	emittance (emissivity) [—]
$\epsilon$	eccentricity [—]
$\epsilon_1$	Re ( $\epsilon$ )
$\epsilon_2$	Im ( $\epsilon$ )
$\tau$	(power) transmittance (transmissivity) [—]
$\nu$	radiation frequency [Hz]
$\omega$	circular frequency = $2\pi\nu$ [rads <sup>-1</sup> ]
$\omega$	plasma frequency [Hz]
$\lambda$	wavelength [ $\mu\text{m}$ , nm]
$\sigma$	wave number = $1/\lambda$ [cm <sup>-1</sup> ]
$\sigma$	Stefan Boltzmann constant [Wm <sup>-2</sup> K <sup>-1</sup> ]
$\rho$	reflectance (reflectivity) [—]
$\theta, \phi$	angular coordinates [rad, °]
$\xi, \eta$	rectangular spatial frequencies [m <sup>-1</sup> , r <sup>-1</sup> ]
$\phi$	phase [rad, °]
$\phi$	lens power [m <sup>-2</sup> ]
$\Phi$	flux [W]
$\chi$	electric susceptibility tensor [—]
$\Omega$	solid angle [sr]

## Other

$\mathfrak{R}$	responsivity
$\exp(x)$	$e^x$
$\log_a(x)$	log to the base $a$ of $x$
$\ln(x)$	natural log of $x$
$\log(x)$	standard log of $x$ : $\log_{10}(x)$
$\Sigma$	summation
$\Pi$	product
$\Delta$	finite difference
$\delta x$	variation in $x$
$dx$	total differential
$\partial x$	partial derivative of $x$
$\delta(x)$	Dirac delta function of $x$
$\delta_{ij}$	Kronecker delta

## REFERENCES

---

1. Anonymous, *ISO Standards Handbook 2: Units of Measurement*, 2nd ed., International Organization for Standardization, 1982.
2. Anonymous, *Symbols, Units and Nomenclature in Physics*, Document U.I.P. 20, International Union of Pure and Applied Physics, 1978.
3. E. Cohen and B. Taylor, "The Fundamental Physical Constants," *Physics Today*, 9 August 1990.
4. W. J. Smith, *Modern Optical Engineering*, 2nd ed., McGraw-Hill, 1990.
5. J. M. Lloyd, *Thermal Imaging Systems*, Plenum Press, 1972.

**William L. Wolfe**  
*College of Optical Sciences*  
*University of Arizona*  
*Tucson, Arizona*

# OPTICS OF THE EYE

Neil Charman

*Department of Optometry and Vision Sciences  
University of Manchester  
Manchester, United Kingdom*

## 1.1 GLOSSARY

$F, F'$	focal points
$N, N'$	nodal points
$P, P'$	principal points

### Equation (1)

$r$	distance from axis
$R_0$	radius of curvature at corneal pole
$p$	corneal asphericity parameter

### Equation (2)

$s$	distance from Stiles-Crawford peak
$\eta/\eta_{\max}$	relative luminous efficiency
$\rho$	coefficient in S-C equation

### Equation (3)

$d$	pupil diameter
$P_A$	ratio of effective to true pupil area

### Transmittance and reflectance

$T_E(\lambda)$	total transmittance of the eye media
$R_R(\lambda)$	reflectance of the retina
$\lambda$	wavelength

### Equation (4)

$(\theta, \phi)$	angular direction coordinates in visual field
$\delta(\lambda)$	wavelength interval
$L_{e\lambda}(\theta, \phi)$	spectral radiance per unit wavelength interval per unit solid angle in direction $(\theta, \phi)$
$p(\theta, \phi)$	area of pupil as seen from direction $(\theta, \phi)$

$t(\theta, \phi, \lambda)$  fraction of incident radiation flux which is transmitted by the eye  
 $m(\theta, \phi, \lambda)$  areal magnification factor

Equation (5)

$I$  normalized illuminance  
 $z$  dimensionless diffraction unit

Equation (6)

$\gamma$  angular distance from center of Airy diffraction pattern  
 $d$  pupil diameter

Equation (7)

$\theta_{\min}$  angular resolution by Rayleigh criterion

Equation (8)

$R$  spatial frequency  
 $R_R$  reduced spatial frequency

Equation (9)

$\Delta F$  dioptric error of focus  
 $g$  number of Rayleigh units of defocus

Equation (10)

$\beta$  angular diameter of retinal blur circle

Equation (11)

$T(R)$  modulation transfer function

Equation (13)

$R_x(\lambda)$  chromatic difference in refraction with respect to 590 nm

Equation (14)

$L_{\text{eq}}$  equivalent veiling luminance  
 $E$  illuminance produced by glare source at eye  
 $\omega$  angle between direction of glare source and visual axis

Equations (15) and (16)

$M_{IT}(R)$  threshold modulation on the retina  
 $M_{OT}(R)$  external threshold modulation

Equation (17)

$\text{DOF}_{\text{go}}$  total depth-of-focus for an aberration-free eye according to geometrical optics  
 $\Delta F_{\text{tol}}$  tolerable error of focus  
 $\beta_{\text{tol}}$  tolerable angular diameter of retinal blur circle

Equation (18)

$\text{DOF}_{\text{po}}$  total depth-of-focus for an aberration-free eye according to physical optics

Equation (19)

OA objective amplitude of accommodation

Equation (20)

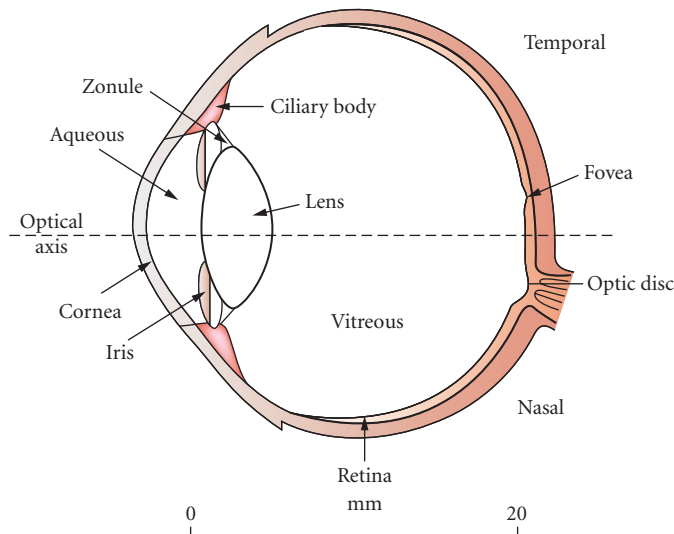
$l$  object distance  
 $p$  interpupillary distance  
 $\delta l$  minimum detectable difference in distance  
 $\delta\theta$  stereo acuity

## Equation (21)

- $M$  transverse magnification  
 $N$  factor by which effective interpupillary distance is increased

## 1.2 INTRODUCTION

The human eye (Fig. 1) contains only a few optical components. However, in good lighting conditions when the pupil is small (2 to 3 mm), it is capable of near diffraction-limited performance close to its axis. Each individual eye also has a very wide field of view (about 65, 75, 60, and 95 deg in the superior, inferior, nasal, and temporal semimeridians, respectively, for a fixed frontal direction of gaze, the exact values being dependent upon the individual's facial geometry). The binocular field, where the two monocular fields overlap, has a lateral extent of about 120 deg. Optical image quality, while somewhat degraded in the peripheral field is, in general, adequate to meet the needs of the neural network which it serves, since the spatial resolution of the neural retina falls rapidly away from the visual axis (the latter joins the point of regard, the nodal points and the fovea). The orientation of the visual axis typically differs by a few degrees from that of the optical axis, as the fovea, where neural resolution is optimal, is usually slightly displaced from the intersection of the optical axis with the retina.<sup>1,2</sup> Control of ocular aberrations is helped by aspheric optical surfaces and by the gradients of refractive index in the lens, the lens index progressively reducing from the lens center toward its outer layers. Off-axis aberrations are further reduced by the eye's approximation to a homocentric system, in which the optical and detector surfaces are concentric with a common center of curvature at the aperture stop.<sup>3</sup> Although aberration levels increase and optical image quality falls as the pupil dilates at lower light levels (to reach a maximum diameter of about 8 mm, corresponding to a numerical aperture of about 0.25 mm), neural performance also declines, so that optical and neural performances remain reasonably well matched. When the eye is in its basic "relaxed" state it is nominally in focus for distant objects. In the younger eye (<50 years) the power of the crystalline lens can be increased to allow near objects to be clearly focused, a process known as *accommodation*. These general characteristics will now be discussed in more detail.



**FIGURE 1** Schematic horizontal section of the eye. The bar gives the approximate scale.



### 1.3 OCULAR PARAMETERS AND AMETROPIA

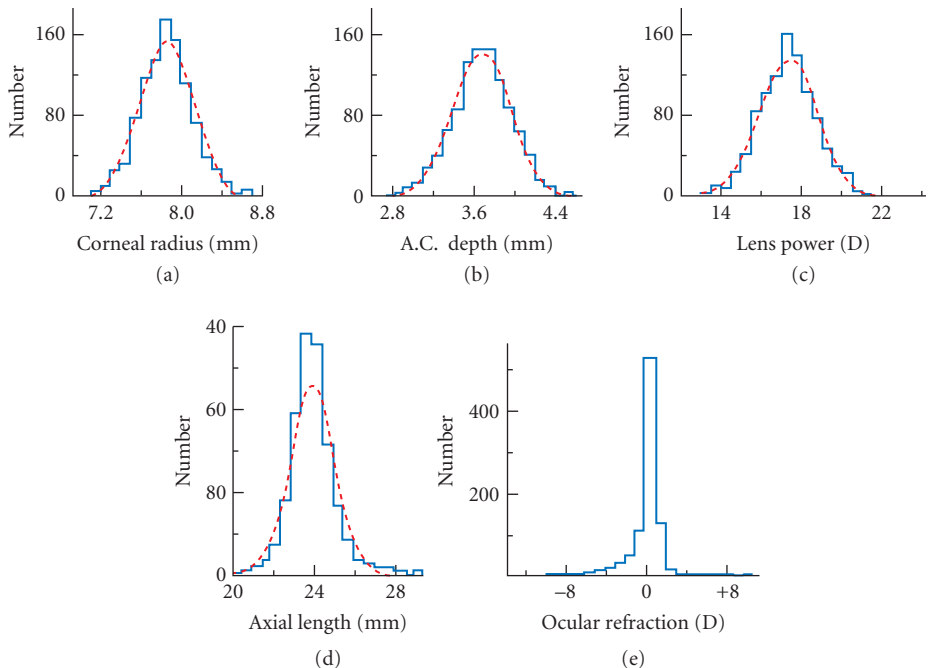
#### Variation in Ocular Parameters across the Population

In the first few years of life there is substantial overall ocular growth, with axial length increasing from about 17 mm at birth to about 23 mm by the age of 3 years.<sup>4-6</sup> Some growth continues thereafter, the axial length in adults being about 24 mm.<sup>7-12</sup> Although most dimensions remain almost stable during adulthood, the lens continues to grow in thickness and volume, but not diameter, throughout life. The thickness increases approximately linearly by about 50 percent between birth and the age of 70,<sup>10,13-15</sup> there being accompanying changes in its index distribution.<sup>16-19</sup> Surface curvatures, component separations, and axial lengths show considerable variation (~10 percent) between individuals of the same age, although the refractive indices of the cornea, vitreous, and aqueous humours are essentially constant.<sup>7-9,20,21</sup> Figure 2 shows some typical measured adult distributions of the values of several parameters: most of the distributions shown are approximately normal (dashed curves), but this is not true for the axial length.

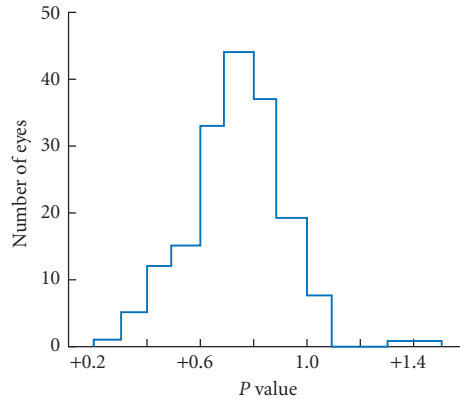
As noted earlier, the optical surfaces may be aspheric. The form of the anterior cornea is particularly significant. Owing to the large refractive index change at its anterior surface it contributes about three-quarters of the total refractive power of the eye. Its topography also has obvious relevance to contact lens design and to the monochromatic aberrations of the eye. It is often modeled as a conicoid, of the form

$$r^2 + pz^2 - 2R_0z = 0 \quad (1)$$

where the axis of symmetry lies in the  $z$  direction,  $r$  is the distance perpendicular to the axis, and  $R_0$  is the radius of curvature at the pole of the surface. Figure 3 shows experimental measurements of the distribution of the parameter  $p$ .<sup>22,23</sup> The distribution is fairly wide and peaks at a value of



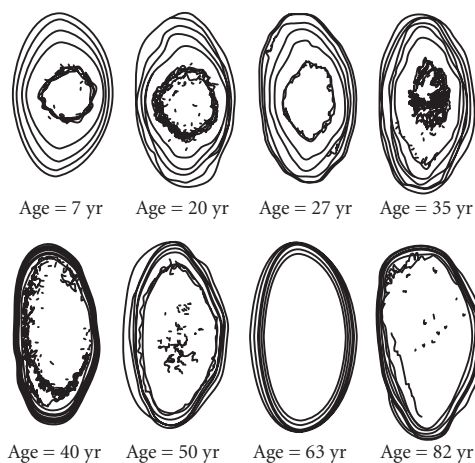
**FIGURE 2** Distributions of some dimensional parameters for the adult human eye. (After Stenström.<sup>21</sup>)



**FIGURE 3** Histogram showing the distribution of the corneal asphericity parameter  $p$ . A value of  $p = 1$  corresponds to a spherical surface, values between 0 and 1.0 represent prolate (flattening) ellipsoids and  $p > 1.0$  oblate (steepening) ellipsoids. (Based on Refs. 22 and 23.)

about  $+0.8$ , corresponding to a flattening ellipsoid in which the radius of curvature is smallest at the center of the cornea and increases toward the corneal periphery. At the corneal vertex the radius of curvature is about  $7.8 \pm 0.25$  mm.<sup>22,23</sup>

The least understood optical feature of the eye is the distribution of refractive index within the lens. As noted earlier, the lens grows throughout life,<sup>10,13,14</sup> with new material being added to the surface layers (the cortex). The oldest part of the lens is its central region (the nucleus). While there is general agreement that the refractive index is highest at the lens center and falls toward its outer layers, the exact form of the gradients involved has proved difficult to measure.<sup>16–19</sup> Description is complicated by the fact that the shape of the lens and its gradients change when the eye accommodates to view near objects and with age. To illustrate the general trend of the changes with age, Fig. 4 shows some



**FIGURE 4** Contours of refractive index in lenses of different ages (7 to 82 years). The contour interval is 0.01. (After Ref. 18.)

recent *in vitro* iso index contours for isolated lenses, obtained using magnetic resonance imaging: when measured *in vitro* the lenses take up their fully accommodated form. It can be seen that, with age, the region of relatively constant index at the lens center increases in volume and that the gradient becomes almost entirely confined to the surface layers of the lens. The central index remains constant at  $1.420 \pm 0.075$  and the surface index at  $1.371 \pm 0.004$ .<sup>18</sup> These index distributions have been modeled as a function of age and accommodation by several authors. (see, e.g., Refs. 24–30)

In addition to these general variations, each eye may have its own idiosyncratic peculiarities, such as small tilts or lateral displacements of surfaces, lack of rotational symmetry about the axis, or irregularities in the shape or centration of the pupil.<sup>1,2,31</sup> These affect both the refractive error (*ametropia*) and the higher-order aberrations.

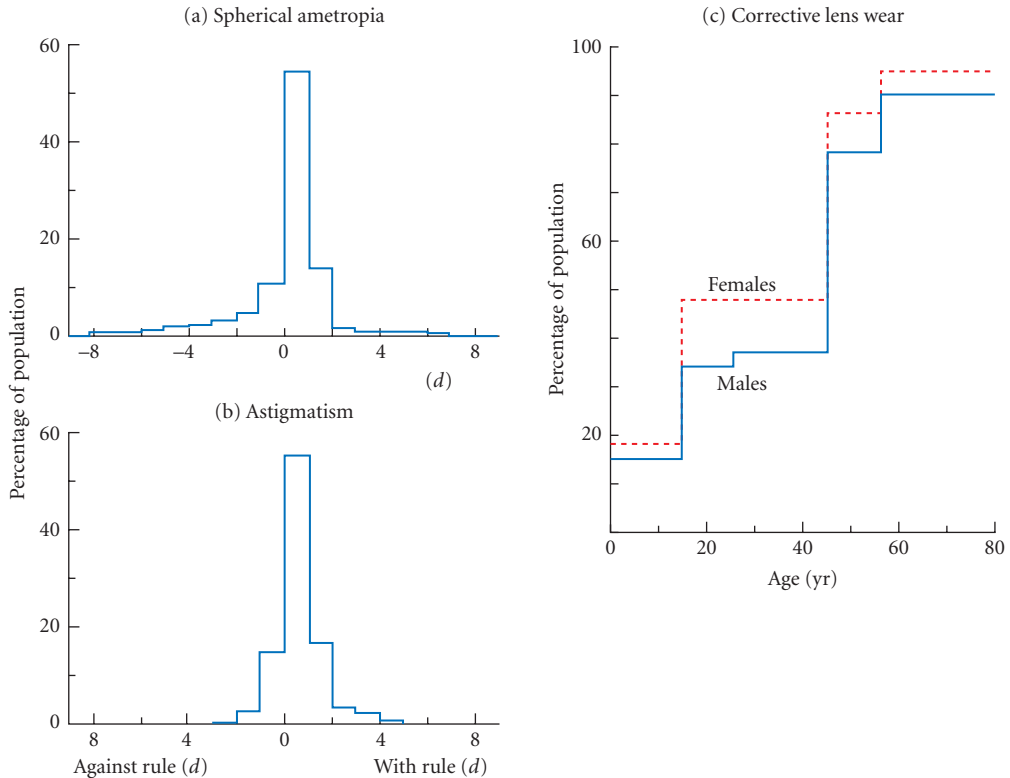
## Ocular Ametropia

If the combination of ocular parameters is such that, with accommodation relaxed, a distant object of regard is focused on the retinal fovea, the region where the density of the cone receptors is highest and photopic neural performance is optimal (Fig. 1), the eye is *emmetropic*. This condition is often not achieved, in which case the eye is *ametropic*. If the power of the optical elements is too great for the axial length, so that the image of the distant object lies anterior to the retina, the eye is *myopic*. If, however, the power is insufficient, the eye is *hypermetropic* (or *hyperopic*). These defects can be corrected by the use of appropriately powered diverging (myopia) or converging (hypermetropia) spectacle or contact lenses to respectively reduce or increase the power of the lens-eye combination. (see Chap. 20 by Edward S. Bennett and William J. Benjamin for reviews.) Spherical ametropia tends to be associated with axial length differences, that is, myopic eyes tend to be longer than emmetropic eyes while hyperopic eyes are shorter.<sup>32</sup>

In some individuals the ocular dioptrics lack rotational symmetry, one or more optical surfaces being toroidal, tilted, or displaced from the axis. This leads to the condition of ocular astigmatism, in which on the visual axis two longitudinally separated, mutually perpendicular, line images of a point object are formed. In the vast majority of cases (*regular astigmatism*) the meridians of maximal and minimal power are perpendicular: there is a strong tendency for the principal meridians to be approximately horizontal and vertical but this is not always the case. Eyes in which the more powerful meridian is vertical are often described as having *with-the-rule astigmatism* and those in which it is horizontal as having *against-the-rule astigmatism*. The former is more common. Correction of astigmatism can be achieved by including an appropriately oriented cylindrical component in any correcting lens. It is sometimes convenient to talk of the *best-, mean-, or equivalent-sphere* correction. This is the power of spherical lens which brings the circle of least confusion onto the retina: its value is  $S + C/2$ , where  $S$  and  $C$  are respectively, the spherical and cylindrical dioptric components of the correction.

In addition to spectacle and contact lens corrections, surgical methods of correction for both spherical and astigmatic errors are now widely used. Most common are those using excimer lasers which essentially reshape the anterior surface of the cornea by selectively ablating its tissue across the chosen area to appropriately modify its spherocylindrical power. A popular current method is laser-assisted keratomileusis (LASIK) which involves cutting a thin, uniform “flap” of material from the anterior cornea and then ablating the underlying corneal stroma to change its curvature. The flap is then replaced. (see Chap. 16 by L. Diaz-Santana and Harilaos Ginis for details.) Intraocular lenses can be used to replace the crystalline lens when the latter has lost transparency due to cataract: single-vision, bifocal and multifocal designs are available and efforts are being made to develop lenses of dynamically varying power to simulate the accommodative abilities of the younger eye. (see Chap. 21 by Jim Schwiegerling.)

Figure 5 shows representative data for the frequency of occurrence of different spherical<sup>21</sup> and astigmatic<sup>34</sup> errors in western adults. Myopia is more common in many eastern populations. Note particularly that the spherical errors are not normally distributed and that a state of near-emmetropia is most common. It is believed that the correlation of component values required to achieve near-emmetropia is achieved partly as a result of genetic factors and partly as a result of



**FIGURE 5** Typical adult data for the frequency of occurrence of spherical and cylindrical (astigmatic) refractive errors, together with the fraction of the population wearing corrective spectacle or contact lenses. (a) Spherical errors (diopters) in young adult males. (After Stenstrom.<sup>21</sup>) (b) Cylindrical errors. (Based on Lyle.<sup>34</sup>) Cases in which the meridian of greatest power is within 30 deg of the horizontal (against-the-rule) and within 30 deg of the vertical (with-the-rule) are shown: the remaining 4 percent of the population have axes in oblique meridians. (c) Percentage of the population wearing lenses, as a function of age. (After Farrell and Booth.<sup>35</sup>)

environmentally influenced growth processes which drive the development of the young eye toward emmetropia (*emmetropization*).<sup>33</sup>

Not all individuals with ametropia actually wear a correction; the fraction of the population that typically does so is shown in Fig. 5. The increase in lens wear beyond the age of 40 is due to the need for a near correction for close work, a condition known as *presbyopia*. This arises as a result of the natural, progressive failure with age of the eye's own accommodation system (see "Age-dependent Changes in Accommodation").

The widespread existence of ametropia among users of visual instruments such as telescopes and microscopes, means that it is desirable to make provision for focusing the eyepiece to compensate for any spherical refractive error of the observer. This is particularly the case where the eyepiece contains a graticule. Since the refractive errors of the two eyes of an individual may not be identical (*anisometropia*), differential focusing should be provided for the eyepieces of binocular instruments. As correction for cylindrical errors is inconvenient to incorporate into eyepieces, astigmatic users of instruments must usually wear their normal refractive correction. For spectacle wearers, where the distance of the lenses in front of the eyes is usually 10 to 18 mm, this implies that the exit pupil of the instrument must have an adequate eye clearance or eye relief (at least 20 mm and preferably 25 mm) to avoid contact between the spectacle lens and the eyepiece and allow the instrument's full field to be seen.

## 1.4 OCULAR TRANSMITTANCE AND RETINAL ILLUMINANCE

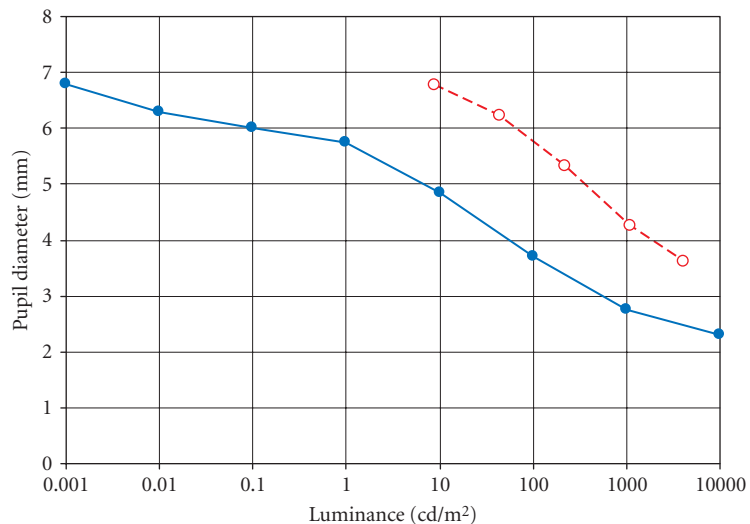
The amount, spectral distribution, and polarization properties of the light reaching the retina are modified with respect to the original stimulus in a way that depends upon the pupil diameter and the transmittance characteristics of the eye.

### Pupil Diameter

The circular opening in the iris, located approximately tangential to the anterior surface of the lens, plays the important role of aperture stop of the eye. It therefore controls the amount of light flux reaching the retina, as well as influencing retinal image quality through its effects on diffraction, aberration, and depth-of-focus (see Sec. 1.7). It may also affect the amount of scattered light reaching the retina, particularly in older eyes where cataract is present.

What is normally measured and observed is the image of the true pupil as viewed through the cornea, that is, the entrance pupil of the eye. This is some 13 percent larger in diameter than the true pupil. Although ambient lighting and its spatial distribution have the most important influence on entrance pupil diameter (Fig. 6) the latter is also affected by many other factors including age, accommodation, emotion, and drugs.<sup>36,37</sup> For any scene luminance, the pupils are slightly smaller under binocular conditions of observation.<sup>38</sup> The gradual constriction with age<sup>36</sup> helps to account for the poorer visual performance of older individuals under dim lighting conditions in comparison with younger individuals.<sup>39</sup>

The pupil can respond to changes in light level at frequencies up to about 4 Hz.<sup>37</sup> Shifts in pupil center of up to 0.6 mm may occur when the pupil dilates<sup>40,41</sup> and these may be of some significance in relation to the pupil-dependence of ocular aberration and retinal image quality.



**FIGURE 6** Entrance pupil diameter as a function of scene luminance for an extended visual field and young observers: the filled symbols and full curve show the weighted average of 6 studies. (After Farrell and Booth.<sup>35</sup>) Larger pupils are observed when the illuminated field is of smaller area: the dashed curve and open symbols show data for a 10 deg illuminated field. (After Winn et al.<sup>36</sup>)

It has been suggested<sup>42,43</sup> that the major value of light-induced pupillary constriction is that it reduces retinal illuminance and hence prepares the eye for a return to darkness: following a change to a dark environment the dilation of the mobile pupil allows substantially better stimulus detection during the first few minutes of dark adaptation than would be found with a fixed pupil.

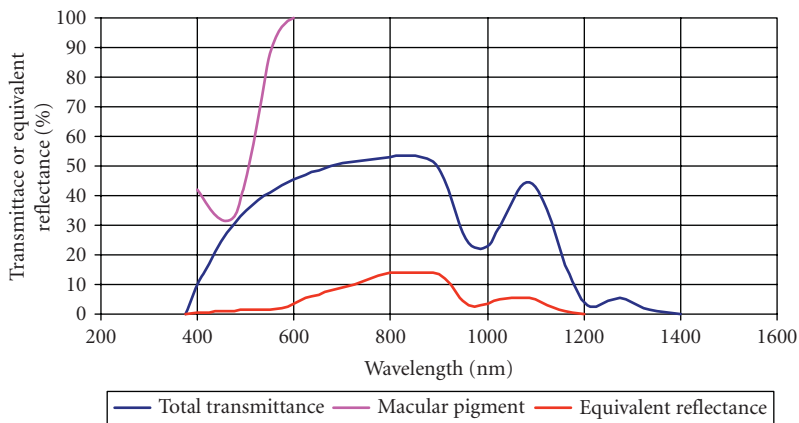
## Transmittance

Light may be lost by spectrally varying reflection, scattering, and absorption in any of the media anterior to the retina.<sup>44</sup> Fresnel reflection losses are in general small, the maximum being 3 to 4 percent at the anterior cornea. Wavelength-dependent absorption and scattering are much more important: both tend to increase with age.

The measured transmittance<sup>45–49</sup> depends to some extent on the measuring technique, in particular on the extent to which scattered light is included, but representative data are shown in Fig. 7. The transmittance rises rapidly above about 400 nm, to remain high in the longer wavelength visible and near infrared. It then falls through several absorption bands, due mainly to water, to reach zero at about 1400 nm.

Although most of the absorption at ultraviolet wavelengths below 300 nm occurs at the cornea (where it may disrupt the surface cells, leading to photokeratitis, e.g., snow blindness or welder's flash, the lowest damage thresholds of about  $0.4 \text{ J} \cdot \text{cm}^{-2}$  being at  $270 \text{ nm}$ <sup>50</sup>), there is also substantial absorption in the lens at the short wavelength (roughly 300–400 nm) end of the visible spectrum. This lenticular absorption increases markedly with age,<sup>51–54</sup> the lens becoming progressively yellower in appearance,<sup>55</sup> and can adversely affect color vision.<sup>56,57</sup> Most of the absorption occurs in the lens nucleus.<sup>58</sup> There is evidence that UV absorption in the lens may be a causative factor in some types of cataract.<sup>59</sup> Excessive visible light at the violet-blue end of the spectrum is thought to cause accelerated aging and resultant visual loss at the retinal level<sup>60</sup> so that lenticular absorption may have a protective function. (See also Chap. 7 by David H. Sliney.)

In the foveal region a thin layer of macular pigment, extending over the central few degrees of the retina<sup>61,62</sup> and lying anterior to the receptor outer segments<sup>63,64</sup> absorbs heavily at shorter wavelengths (Fig. 7). It has been argued that this absorption is helpful in reducing the blurring effects of longitudinal chromatic aberration<sup>65</sup> and in protecting the foveal receptors, which are responsible for detailed pattern vision, against blue-light damage.<sup>66</sup> It is, however, notable that the amount of macular pigment varies widely between individuals.<sup>64</sup>



**FIGURE 7** Spectral dependence of the overall transmittance of the ocular media<sup>46</sup> and the equivalent reflectance of the retina.<sup>75–77</sup> Also shown is the transmittance at the fovea of the macular pigment.<sup>65</sup>

Since the cornea, lens, and retinal nerve fibre layer all show birefringence, the polarization characteristics of the light entering the eye are modified before it reaches the outer segments of the retinal receptors. In general these effects are of little practical significance, although they can be demonstrated and measured by suitable methods.<sup>67</sup>

## The Stiles-Crawford Effect

One complicating factor when considering the effectiveness of the light flux which enters the eye as a stimulus to vision is the Stiles-Crawford effect of the first kind SCEI.<sup>68</sup> (See also Chaps. 8 and 9 by Jay M. Enoch and Vasudevan Lakshminarayanan in this volume.) This results in light which enters the periphery of the entrance pupil to reach a given retinal location being less effective at stimulating the retina than light which passes through the pupil center (Fig. 8). The effect varies slightly with the individual and is not always symmetric about the center of the pupil. Under photopic conditions, giving cone vision, there is typically a factor of about 8 between the central effectiveness and that at the edge of a fully dilated 8-mm pupil; the effect is much weaker under rod-dominated, scotopic conditions (see Fig. 8).<sup>69-71</sup> In practice, unless pupil-dilating drugs are used, the natural pupil will only be large under scotopic conditions and will normally be constricted at photopic levels (see Fig. 6): the influence of SCE I is still significant, however.

Many equations have been proposed to fit photopic data of the type illustrated in Fig. 8. The simplest, due to Stiles,<sup>72</sup> can be written:

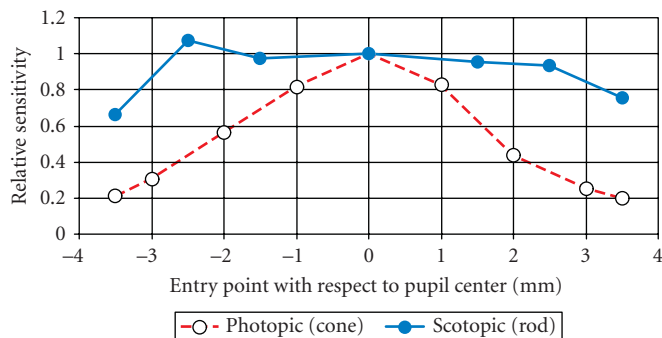
$$\log_{10}(\eta/\eta_{\max}) = -\rho s^2 \quad (2)$$

where  $\eta/\eta_{\max}$  is the relative luminous efficiency and  $s$  is the distance within the entrance pupil from the function peak (in mm). Values of the constant  $\rho$  equal to about 0.07 are typical, the value varying somewhat with wavelength.<sup>72</sup>

It is evident that the Stiles-Crawford effect of the first kind results in the photopic retinal stimulus being somewhat weaker than that predicted on the basis of visible pupil area. This can be accounted for by using an effective pupil area instead of the actual entrance pupil area. Moon and Spencer<sup>73</sup> suggested that the ratio  $P_A$  of the effective to the true pupil areas could be approximated by:

$$P_A = 1 - 0.0106d^2 + 0.0000417d^4 \quad (3)$$

where  $d$  is the pupil diameter in mm.



**FIGURE 8** The Stiles-Crawford effect (SCE I) under photopic (open circles) and scotopic (filled circles) conditions for the same observer, measured at a position 6 deg from the fovea. The relative luminous efficiency is plotted as a function of the horizontal pupillary position of the beam. (After van Loo and Enoch.<sup>69</sup>)

The Stiles-Crawford effect of the second kind (SCE II) involves a shift in the hue of monochromatic light as it enters different areas of the pupil.<sup>70,72,74</sup>

Although Fresnel reflection and lenticular absorption variations with pupil position may play minor roles, there seems little doubt that the Stiles-Crawford effects mainly involve the waveguide properties of the outer segments of the small-diameter receptors. (See Refs. 63 and 64, and Chap. 8 by Vasudevan Lakshminarayanan and Jay M. Enoch for reviews.) Effectively, the receptor can only trap light efficiently if the latter is incident in directions within a few degrees of the receptor axis. For this reason, the receptor outer segments must always be aligned toward the exit pupil of the eye, rather than perpendicular to the local surface of the eyeball. An obvious advantage of such directional sensitivity is that it helps to suppress the degrading effects of intraocular stray light. Such light will be ineffective at stimulating the receptors if it is incident at oblique angles on the retina. As will be discussed in Chap. 8, it appears that SCE I acts as an amplitude apodizing filter in its effects upon retinal image quality.

## Retinal Reflectance

The function of the optical system of the eye is to deliver an image to the retina. Nevertheless, a significant amount of this light is reflected when it reaches the retina and underlying structures. Although such light does not contribute usefully to the process of vision, its existence allows the development of a variety of clinical instruments for examination of the retina, measurement of refraction, and other purposes.

Due to its double passage through the eye media, the emergent flux at any wavelength is proportional to the equivalent reflectance  $T_E(\lambda)^2 R_R(\lambda)$ , where  $T_E(\lambda)$  is the total transmittance of the eye media and  $R_R(\lambda)$  is the true retinal reflectance. Equivalent reflectance rises with wavelength across the visible spectrum, to become quite high in the near infrared.<sup>62,75–78</sup> Representative values are given Fig. 7. Absolute levels of equivalent reflectance are affected by the pigmentation of an individual eye. At the violet end of the visible spectrum the equivalent reflectance falls markedly with age,<sup>79</sup> due to the decreased transmittance of the lens. In the same spectral region, equivalent reflectance is usually lower within the immediate area of the fovea, due to the low transmittance of the macular pigment.

The high equivalent reflectance in the infrared is particularly useful in allowing measurements to be made of, for example, refraction or aberration, at wavelengths which are essentially invisible to the patient or subject. In practice, depending upon the wavelength, light may be reflected from structures anywhere between the anterior surface of the retina and the choroid/sclera interface. Although details of the nature of the reflections are still imperfectly understood, shorter visible wavelengths appear to penetrate less deeply before reflection occurs, while the infrared is reflected from the anterior sclera. At the shorter wavelengths the reflection is almost specular but becomes more diffuse at longer visible and infrared wavelengths.<sup>79–81</sup> Waveguiding effects within the receptors may play some role in determining the nature of the reflection and the angular distribution of the reflected light.<sup>82–85</sup>

## Ocular Radiometry and Retinal Illuminance

If we confine ourselves to uniform object fields subtending at least 1 deg at the eye, so that blurring due to diffraction, aberration or defocus has little effect, the retinal image at moderate field angles would also be expected to be uniform across its area, except at the edges. Wyszecki and Stiles<sup>37</sup> show that the retinal irradiance in a wavelength interval  $\delta\lambda$  corresponding to an external stimulus of spectral radiance  $L_{e\lambda}(\theta, \phi)$  per unit wavelength interval per  $\text{cm}^2$  per unit solid angle of emission, in a direction with respect to the eye given by the angular coordinates  $(\theta, \phi)$  is:

$$\frac{L_{e\lambda}(\theta, \phi) \cdot \delta\lambda \cdot p(\theta, \phi) \cdot t(\theta, \phi, \lambda)}{m(\theta, \phi, \lambda)} \quad (4)$$

where  $p(\theta, \phi, \lambda)$   $\text{cm}^2$  is the apparent area of the pupil as seen from the direction  $(\theta, \phi)$ ;  $t(\theta, \phi, \lambda)$  is the fraction of the incident radiant flux transmitted through the eye; and  $m(\theta, \phi, \lambda)$  is an areal



magnification factor ( $\text{cm}^2$ ) relating the area of the retinal image to the angular subtense of the stimulus at the eye, which will vary somewhat with the parameters of the individual eye. If required, the pupil area  $p(\theta, \phi, \lambda)$  can be modified using Eq. (3) to take account of the Stiles-Crawford effect.

Use of Eq. (4) near to the visual axis is straightforward. In the peripheral field, however, complications arise. With increasing field angle the entrance pupil appears as an ellipse of increasing eccentricity and reduced area: the ratio of the minor diameter to the major diameter falls off somewhat more slowly than the cosine of the field angle.<sup>86,87</sup> Also, due to the retina lying on the curved surface of the quasi-spherical eyeball, both the distance between the exit pupil of the eye and the retina, and the retinal area corresponding to the image of an object of constant angular subtense diminish with field angle. Remarkably, theoretical calculations<sup>88-91</sup> show that these pupil and retinal effects tend to compensate one another, so that an extended field of constant luminance (i.e. a *Ganzfeld*) results in a retinal illuminance which is almost constant with peripheral field angle. This theoretical result is broadly confirmed by practical measurements,<sup>91</sup> showing that, from the photometric point of view, the design of the eye as a wide-angle system is remarkably effective.

A useful discussion of the photometric aspects of point and extended sources in relation to the eye is given by Wright.<sup>92</sup>

## 1.5 FACTORS AFFECTING IN-FOCUS RETINAL IMAGE QUALITY

The optical quality of the retinal image is degraded by the effects of diffraction, monochromatic and chromatic aberration, and scattering. The image is often further blurred by defocus, due to errors in refraction and accommodation. Under many conditions the latter may be the dominant cause of image degradation. It will, however, be convenient to consider the question of focus in a separate section.

### The Aberration-Free (Diffraction-Limited) Eye

In the absence of aberration or variation in transmittance across the pupil, the only factors influencing the retinal image quality in monochromatic light at optimal focus would be the diffraction effects associated with the finite wavelength,  $\lambda$ , of the light and the pupil diameter,  $d$ . For such an eye, the point-spread function (PSF) is the well-known Airy diffraction pattern<sup>93</sup> whose normalized illuminance distribution,  $I(z)$ , takes the form:

$$I(z) = \left[ \frac{2J_1(z)}{z} \right]^2 \quad (5)$$

where  $J_1(z)$  is the Bessel function of the first kind of order 1 of the variable  $z$ . In the case of the eye, the dimensionless distance  $z$  has the value:

$$z = \frac{d\pi \cdot \sin\gamma}{\lambda} \quad (6)$$

$\gamma$  is the angular distance from the center of the pattern, measured at the second nodal point, this being equal to the corresponding angular distance in the object space, measured at the first nodal point.<sup>94,95</sup> The angular resolution  $\theta_{\min}$  for two neighboring equally luminous incoherent object points, as given by the Rayleigh criterion is then:

$$\theta_{\min} = \frac{1.22\lambda}{d} \text{ rad} \quad (7)$$

$\theta_{\min}$  is about 1 minute of arc when  $d$  is 2.3 mm and  $\lambda$  is 555 nm. Evidently the Rayleigh criterion is somewhat arbitrary, since it assumes that the visual system can just detect the 26 percent drop in irradiance between the adjacent image peaks; for small pupils actual visual performance is usually somewhat better than this limit.<sup>96</sup> The size of the PSF increases on either side of optimal focus.<sup>97-99</sup>

Images of more complex objects can be obtained by regarding the image as the summation of an appropriate array of PSFs, that is, by convolving the PSF with the object radiance distribution.<sup>94</sup> The in-focus line-spread function (LSF), the edge image,<sup>100</sup> and the modulation transfer function (MTF) also take standard forms. The phase transfer function (PTF) is always zero because the diffractive blur is rotationally symmetrical: this also makes the LSF, edge image, and MTF independent of orientation. Figure 9 shows the form of the MTF as a function of focus.<sup>101</sup> Extensive tables of numerical values are given by Levi.<sup>102</sup> Relative spatial frequencies,  $R_R$ , in Fig. 9 have been normalized in terms of the cutoff value beyond which the modulation transfer is always zero. Errors of focus have been expressed in what Levi calls “Rayleigh units,” i.e., the number of quarter wavelengths of wavefront aberration at the edge of the pupil. Note from Fig. 9 that the modulation transfer is most sensitive to defocus at intermediate normalized spatial frequencies ( $R_R \approx 0.5$ ).<sup>103,104</sup>

To convert the units of relative spatial frequency and Rayleighs to true spatial frequencies  $R$  c/deg and dioptric errors of focus  $\Delta F$  respectively, the following relations may be used:

$$R = \frac{10^6 \times d \times R_R}{\lambda} \text{ c/rad}$$

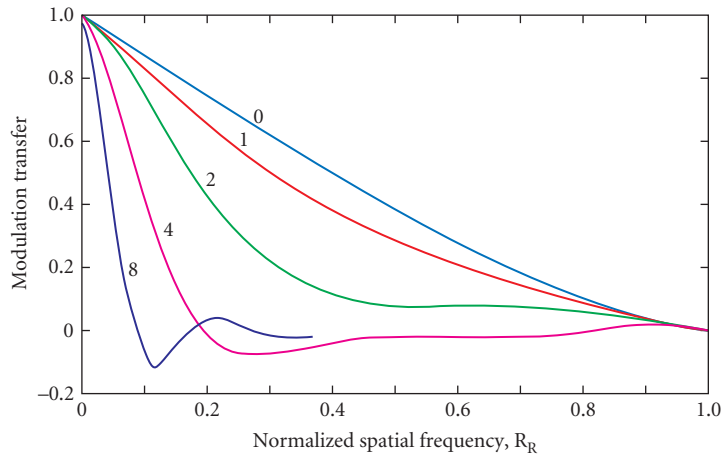
$$= \frac{1.746 \times 10^4 \times d \times R_R}{\lambda} \text{ c/deg} \quad (8)$$

$$\Delta F = \frac{2 \times 10^{-3} \lambda g}{d^2} \text{ diopters} \quad (9)$$

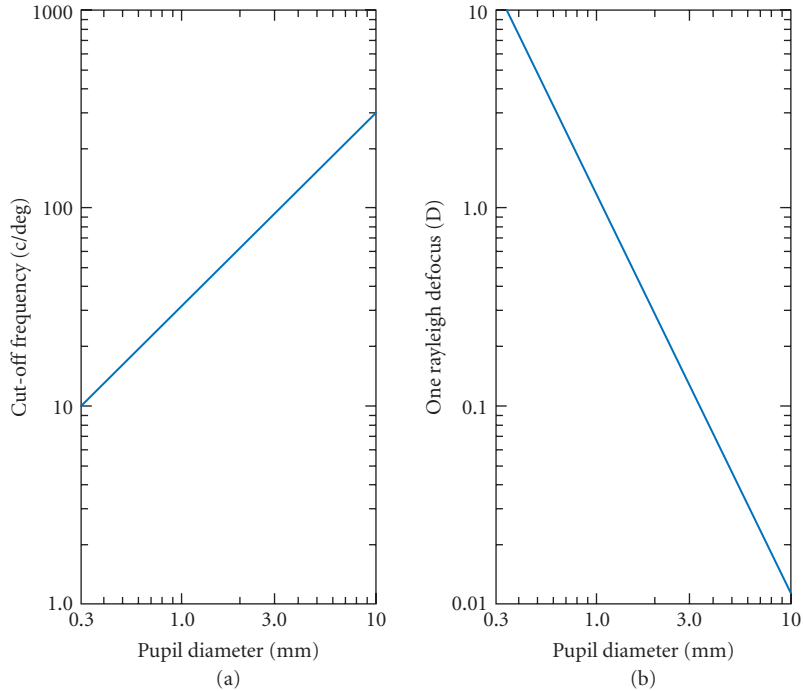
where the entrance pupil diameter  $d$  is in mm, the wavelength  $\lambda$  is in nm, and  $g$  is the number of Rayleighs of defocus.

Figure 10 illustrates the variation in these parameters as a function of the ocular pupil diameter,  $d$ , for the case where  $\lambda = 555$  nm. In such green light and with the typical photopic pupil diameter of 3 mm, the cutoff frequency imposed by diffraction is about 100 c/deg and one Rayleigh of defocus corresponds to about 0.12 D.

Sets of diffraction-limited ocular MTF curves for various specific combinations of pupil diameter, wavelength, and defocus have been illustrated by several authors (e.g., Refs. 105–111).



**FIGURE 9** Modulation transfer functions for a diffraction-limited optical system with a circular pupil working in monochromatic light. It suffers from the errors of focus indicated. Defocus is expressed in Rayleighs, that is, the number of quarter-wavelengths of defocus wavelength aberration. (Based on Levi.<sup>102</sup>)



**FIGURE 10** Values of (a) cutoff frequency  $R_R = 1.0$  and (b) dioptric defocus equivalent to one Rayleigh, for a diffraction-limited eye working in monochromatic light of wavelength 555 nm.

When errors of focus become large, the geometric approximation in which the defocus PSF is a uniform blur circle becomes increasingly valid.<sup>101,110,112–115</sup> The angular diameter,  $\beta$  of the retinal blur circle for a pupil diameter  $d$  mm and error of focus  $\Delta F$  diopters is

$$\beta = \frac{0.18 \cdot d \cdot \Delta F}{\pi} \text{ deg} \quad (10)$$

The corresponding geometrical optical MTF is

$$T(R) = \frac{2J_1(\pi\beta R)}{\pi\beta R} \quad (11)$$

where  $R$  is the spatial frequency (c/deg) and  $J_1(\pi\beta R)$  is the Bessel function of the first kind of order 1 of  $(\pi\beta R)$ . Smith<sup>114</sup> gives detailed consideration to the range of ocular parameters under which the geometrical optical approximation may reasonably be applied and Chan et al.<sup>115</sup> have demonstrated experimentally that Eq. (10) usefully predicts blur circle diameters for pupils between 2 and 6 mm in diameter and defocus between 1 and 12 D.

## Monochromatic Ocular Aberrations

In many early studies of ocular aberration, it was common to assume that the eye was symmetrical about a unique optical axis which differed only slightly from the visual axis. Thus it was expected that spherical aberration would dominate on axis, with oblique astigmatism, field curvature, and

coma becoming progressively more important as the field angle increased, although the curved surface of the retina would tend to reduce the impact of field curvature. The assumption was that the brain would adapt to any distortion, so that this aberration was not important. This simple picture has been progressively modified with the realisation that exact symmetry about a unique optical axis rarely occurs and that the fovea does not normally lie where the nominal optical axis intercepts the retina (the difference is usually a few degrees—see e.g., Refs. 116–118 for discussion of possible axes for eye). As a result, the patterns of both on- and off-axis aberrations are more complex than was expected on the basis of early, simple models of the eye.

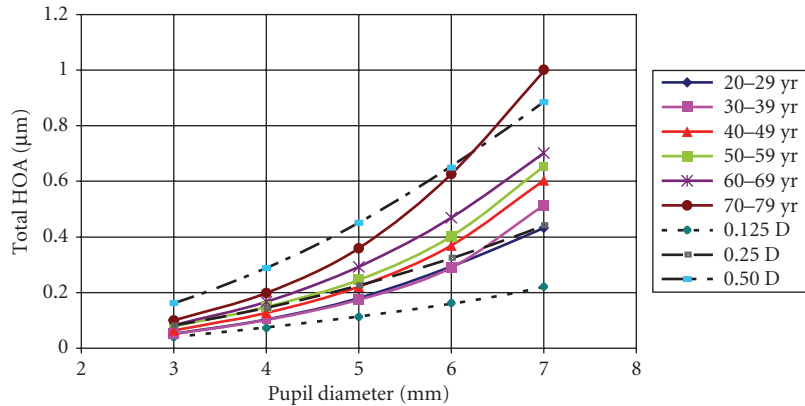
Recent years have, in fact, seen an enormous expansion in the literature of the aberrations of the eye, fueled by the development of commercial aberrometers capable of measuring the characteristics of the individual eye within a few seconds<sup>119,120</sup> and the demands of refractive surgery, where it was recognized that although earlier techniques nominally corrected refractive error, they often resulted in poor visual outcomes due to higher than normal levels of residual aberration.<sup>119</sup> Currently, the optical defects of the eye are usually described in terms of its wavefront aberration under specified conditions, the overall level of aberration being expressed as the root-mean-square (RMS) wavefront error. Different types of aberration are quantified in terms of the coefficients of the corresponding Zernike polynomials of polar coordinates in the pupil (e.g., Refs. 119, 121–124, see also Chap. 11 by Virendra N. Mahajan in Vol. II as well as Chap. 4 by Virendra N. Mahajan and Chap. 5 by Robert Q. Fugate in Vol. V), defined according to OSA recommendations,<sup>123,124</sup> although this approach has been criticized as being inappropriate for eyes in which the wavefront aberration shows locally abrupt variation, as in, for example, some postsurgical cases.<sup>125,126</sup> In the recommended formulation, each Zernike coefficient gives the root-mean-square (RMS) wavefront error (in microns) contributed by the particular Zernike mode: the overall RMS error is given by the square root of the sum of the squares of the individual coefficients. The set of Zernike coefficients thus gives detailed information on the relative and absolute importance of the different aberrational defects of any particular eye for the specified conditions of measurement. In the Zernike description, first-order polynomials simply describe wavefront tilt (i.e. prismatic effects) and have no effect on image quality. Second-order polynomials describe the spherocylindrical errors of focus which can normally be negated by optical corrections, such as spectacles or contact lenses. It is the higher-order (third and greater) polynomials which represent the aberrations. The third-order modes include vertical and horizontal primary coma, and the fourth-order primary spherical aberration. Iskander et al.<sup>127</sup> have illustrated the effects of some of the individual Zernike aberrations on the retinal images of a selection of objects. The values of the Zernike coefficients for any particular eye will, of course, vary with pupil diameter, accommodation, and field angle.

**Aberrations on the Visual Axis** Several large-scale studies have addressed the question of the variation of aberrations between individuals.<sup>128–133</sup> Others have considered changes of aberration with such specific factors as pupil diameter,<sup>134</sup> age,<sup>134–138</sup> accommodation,<sup>139–144</sup> refractive error,<sup>146,147</sup> and time.<sup>148–152</sup>

Figure 11 shows recent mean data for the variation in total higher-order, axial, RMS wavefront error with pupil diameter for different age groups.<sup>134</sup> As would be expected, aberration levels tend to increase with pupil diameter: they also increase with age. The Maréchal criterion<sup>153</sup> suggests that near diffraction-limited performance will be given if the RMS error is less than  $\lambda/14$ , corresponding to about 0.04 microns in the green region of the spectrum. It can be seen that this level of aberration is typically present when the pupil diameter is about 3 mm in younger eyes. As illustrated in Fig. 9, such a pupil diameter is generally found under luminance conditions of a few hundred cd/m<sup>2</sup>, corresponding to that occurring on cloudy days. Thus, in most eyes, wavefront aberration is likely to have only a minor impact on vision under daylight conditions.

To give some insight into the image degradation caused by any level of RMS wavefront aberration, we can roughly evaluate its blurring effects by equating them with those of an “equivalent defocus,” that is the spherical error in focus which produces the same magnitude of RMS aberration for the same pupil size. The equivalent defocus is given by:

$$\text{Equivalent defocus (diopters)} = \frac{16.3^{1/2} \cdot \text{RMS error}}{d^2} \quad (12)$$



**FIGURE 11** Plots of total mean RMS wave aberration as a function of pupil size for different age groups. (After Applegate et al.<sup>134</sup>) The dashed curves show levels of equivalent defocus of 0.125, 0.25, and 0.50 D.

where the RMS aberration is measured in microns and the pupil diameter,  $d$ , in mm. As examples, the dashed curves in Fig. 11 indicate equivalent defocus levels of 0.125, 0.25, and 0.50 D. For younger subjects (20–39 years), the mean HOA is always lower than an equivalent defocus level of 0.25 D, except at the largest, 7 mm, pupil diameter. For comparison, the reliability of clinical refractive techniques is around  $\pm 0.3$  D.<sup>154,155</sup> Although the assumption that equal RMS errors produce equal degradation of vision is not completely justified,<sup>156,157</sup> it is evident that, in most younger, normal eyes, the impact on vision of optical blur due to axial monochromatic aberrations is likely to be modest under most conditions, although this may not be true for a minority of individuals.

When the coefficients of the individual Zernike polynomials are considered for eyes in which the accommodation is relaxed for distance vision, several large-scale studies involving many hundred normal individuals give very similar results.<sup>128–134</sup> As an example, the study by Applegate and his colleagues<sup>134</sup> generated mean values for the magnitudes of different types of third- and fourth-order aberration for different pupil sizes and ages (coefficients for still higher-order Zernike modes are usually much smaller). Table 1 gives examples of their values for different age groups. Note that,

**TABLE 1** Mean Values of the Coefficients ( $\mu\text{m}$ ) and Their Standard Deviations for Individual Third- and Fourth-Order Zernike Modes for 3- and 6-mm Pupils and Different Subject Age Groups\*

Age (yrs)	Pupil Diameter (mm)	RMS WFE ( $\mu\text{m}$ ) Trefoil	RMS WFE ( $\mu\text{m}$ ) Coma	RMS WFE ( $\mu\text{m}$ ) Tetrafoil	RMS WFE ( $\mu\text{m}$ ) 2nd Astig.	RMS WFE ( $\mu\text{m}$ ) Sph. ab.
20–29	3	0.029 ± 0.018	0.028 ± 0.019	0.011 ± 0.010	0.011 ± 0.007	0.013 ± 0.013
30–39	3	0.027 ± 0.017	0.031 ± 0.022	0.010 ± 0.004	0.015 ± 0.008	0.014 ± 0.010
40–49	3	0.038 ± 0.023	0.036 ± 0.020	0.014 ± 0.008	0.014 ± 0.009	0.016 ± 0.011
50–59	3	0.043 ± 0.027	0.048 ± 0.028	0.019 ± 0.016	0.018 ± 0.011	0.014 ± 0.011
60–69	3	0.041 ± 0.021	0.047 ± 0.026	0.023 ± 0.019	0.017 ± 0.011	0.027 ± 0.013
70–79	3	0.059 ± 0.031	0.055 ± 0.026	0.024 ± 0.014	0.020 ± 0.010	0.030 ± 0.022
20–29	6	0.141 ± 0.089	0.137 ± 0.076	0.051 ± 0.025	0.063 ± 0.035	0.132 ± 0.108
30–39	6	0.139 ± 0.089	0.136 ± 0.087	0.056 ± 0.030	0.055 ± 0.027	0.130 ± 0.090
40–49	6	0.187 ± 0.083	0.169 ± 0.089	0.073 ± 0.048	0.071 ± 0.037	0.193 ± 0.110
50–59	6	0.189 ± 0.097	0.198 ± 0.145	0.072 ± 0.051	0.073 ± 0.039	0.197 ± 0.115
60–69	6	0.196 ± 0.115	0.238 ± 0.134	0.088 ± 0.068	0.097 ± 0.070	0.235 ± 0.141
70–79	6	0.292 ± 0.175	0.339 ± 0.170	0.113 ± 0.064	0.093 ± 0.060	0.311 ± 0.153

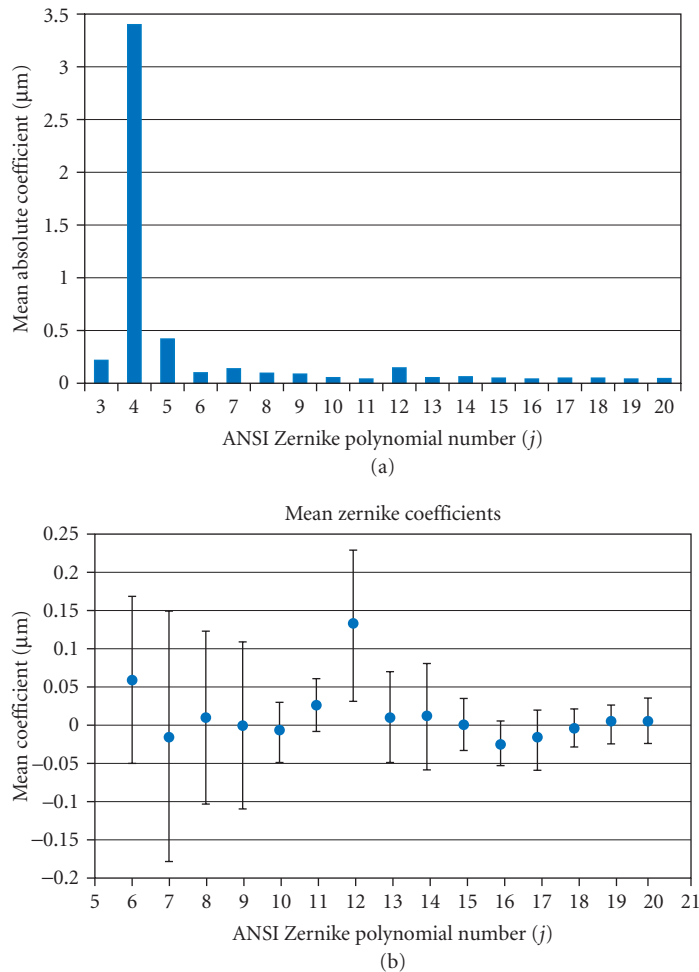
\*The third-order modes are third-order trefoil and coma, the fourth-order are tetrafoil, secondary astigmatism (2nd astig.) and spherical aberration (sph. ab.). The eyes are accommodated for distance vision.

Source: Applegate et al.<sup>134</sup>

where appropriate, the coefficients for similar, but differently oriented, polynomials have been combined. Evidently for smaller, 3-mm pupils, third-order coma and trefoil aberrations tend to dominate over fourth-order aberrations but spherical aberration becomes comparable to coma for the larger 6-mm pupil.

The results of another study<sup>128</sup> are shown in Fig. 12a, where in this case the second-order coefficients are included. Note that the second-order coefficients are much larger than those of the higher orders implying, not surprisingly, that the optical defects of many eyes are dominated by simple spherocylindrical refractive errors.

A somewhat different picture emerges if we average the signed coefficients of the higher-order Zernike modes, rather than their absolute values (Fig. 12b). It is striking that the coefficients of most modes now have means close to zero, although individual eyes may have substantial aberration, as is shown by the relatively large standard deviations. A notable exception is the  $j = 12$ ,  $Z_4^0$  spherical aberration



**FIGURE 12** Typical data for the wavefront aberration of normal eyes with relaxed accommodation: 109 subjects, 5.7-mm pupil diameter. (a) Means of the absolute values of the coefficients of the Zernike modes from the second to the fifth orders. (b) Means of the signed values of each of the coefficients: among the higher-order coefficients, only that for  $j = 12$  ( $C_4^0$ , spherical aberration) has a value which differs significantly from zero. (Based on Porter et al.<sup>128</sup>)

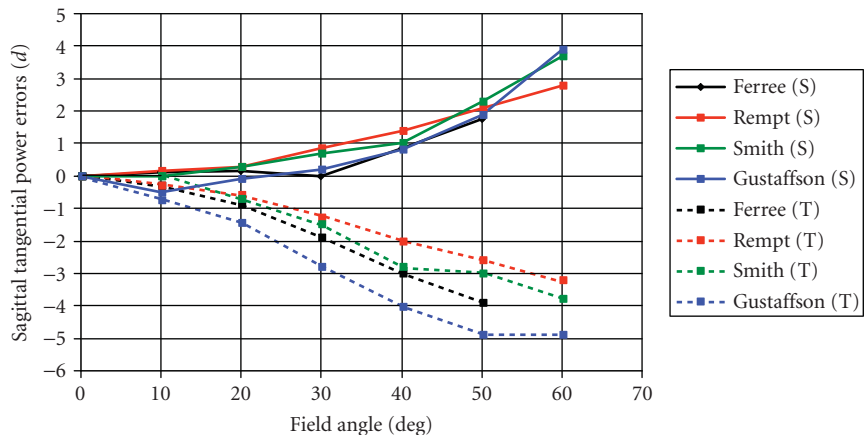
mode, where the mean is positive and differs significantly from zero. Thus the picture that emerges is that most eyes have a central tendency to be free of all higher-order aberration, except for spherical aberration, which shows a significant bias toward slightly positive (under corrected) values. The Zernike coefficients of individual eyes vary randomly about these mean values in a way that presumably depends upon the idiosyncratic surface tilts, decentrations, and other asymmetries of the eye.

When the contributions made by the different optical components of the eye are considered, it appears that, with accommodation relaxed, the overall level of ocular aberration in the young adult is reduced by there being a balance between the contributions of the cornea and the lens. This is particularly the case for spherical aberration and horizontal coma,<sup>158–162</sup> although the compensation may not be found in young eyes with high levels of total aberration<sup>163</sup> or in older eyes.<sup>164</sup> The mechanism by which compensation might be achieved has been discussed by Artal and his colleagues.<sup>2,165</sup> It is interesting to note that there is at least some evidence that there may be additional neural compensation for the aberrations, this being specific to the individual.<sup>166</sup>

Since the shape and gradient index characteristics of the lens change with both accommodation and age, this affects the balance between the corneal and internal aberrations. As accommodation increases, spherical aberration tends to change from positive to negative.<sup>139–145</sup> With age, aberrations with relaxed accommodation at fixed pupil diameter also increase.<sup>134–138</sup> However, under normal conditions, the pupil diameter at constant light level decreases with age,<sup>37</sup> reducing the ocular aberration: image quality therefore remains almost constant with age, although retinal illuminance is lower.<sup>135</sup> Higher-order aberrations generally show, at most, only a very weak dependence on refractive error,<sup>147,167</sup> although the balance between the horizontal coma of the cornea and internal optics may be affected.<sup>168</sup>

Finally we note that measured higher-order aberrations of any individual eye show small fluctuations over time<sup>148–152</sup> with frequencies up to at least 20 Hz. Although the causes of these fluctuations remain to be fully elucidated, the lower-frequency components undoubtedly involve such factors as tear film changes and the cardiopulmonary system.<sup>152,169,170</sup> Lid pressures during such activities as reading may also produce longer-term changes.<sup>171–173</sup>

**Off-Axis Aberrations** Off-axis, on average increasing amounts of second-order aberration (defocus and astigmatism) are encountered (Fig. 13). These may show substantial variations with the individual and with the meridian under study, as may also the relationship between the tangential and sagittal image shells and the retinal surface.<sup>174–187</sup> While there is little systematic change in the mean oblique astigmatism with the axial refraction of the eye, myopes tend to have a relatively hyperopic peripheral mean-sphere refraction, while that in hyperopes tends to be relatively myopic.<sup>177,178,183,185</sup> There may be small changes in peripheral refraction with accommodation<sup>181</sup> and age.<sup>186,187</sup>



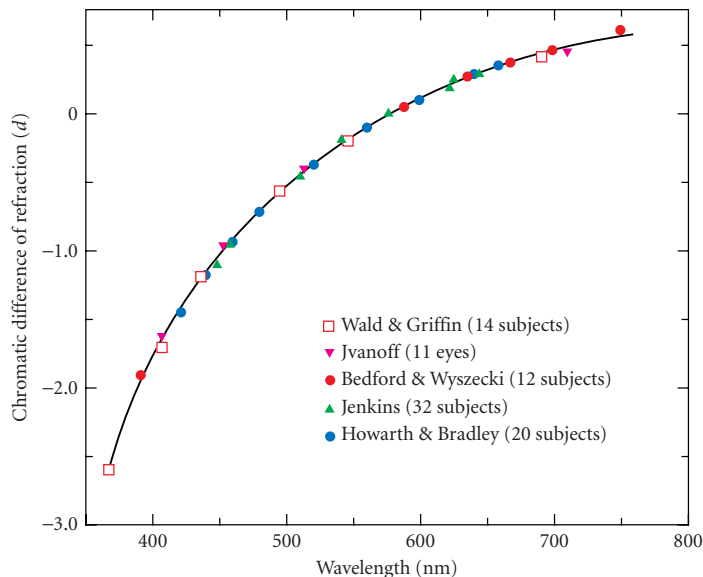
**FIGURE 13** Oblique astigmatism in human eyes. T and S refer to the tangential and sagittal image shells, respectively. (After Ferree et al.,<sup>124</sup> Jenkins,<sup>175</sup> Rempt et al.,<sup>176</sup> Smith et al.,<sup>178</sup> Gustafson et al.<sup>184</sup>)

Higher-order wave aberrations have been studied as a function of field angle by several groups.<sup>188–192</sup> They are generally much less important than the second-order defocus terms but third-order, coma-like terms rise with field angle to be considerably higher than those in axial vision.<sup>188–192</sup> As on axis, there appears to be a degree of balancing between the aberrations associated with the anterior cornea and those of the lens.<sup>192</sup> One problem in using Zernike polynomials to describe off-axis aberrations is that the associated entrance pupils are elliptical rather than circular as required if the Zernike approach is to be used: scaling methods to overcome this difficulty have been devised.<sup>193–195</sup>

## Chromatic Aberration

Chromatic aberration arises from the dispersive nature of the ocular media, the refractive index, and hence the ocular power, being higher at shorter wavelengths. Constringence values for the ocular media are generally quoted<sup>196</sup> as being around 50, although there is evidence that this may need modification.<sup>197,198</sup> Atchison and Smith<sup>199</sup> have recently discussed the available data for the different media and recommend the use of Cauchy's equation to fit experimental values in the visible and allow extrapolation into the near infrared. Both longitudinal or axial chromatic aberration (LCA) and transverse or lateral chromatic aberration (TCA) occur (see, e.g., Refs. 200 and 201 for reviews). For LCA, what is normally measured experimentally is not the change in power of the eye across the spectrum but rather the change in its refractive error, or the chromatic difference of refraction. There are only minor differences in the results of different studies of LCA (e.g., Refs. 175, 202–205) and the basic variation in ocular refraction with wavelength, equivalent to about 2 D of LCA across the visible spectrum is well established (Fig. 14). Atchison and Smith<sup>199</sup> suggest that when the chromatic difference data are set to be zero at 578 nm they can be well fitted by the Cauchy equation

$$R_x(\lambda) = 1.60911 - 6.70941 \times \frac{10^5}{\lambda^2} + 5.55334 \times \frac{10^{10}}{\lambda^4} - 5.5998 \times \frac{10^{15}}{\lambda^6} \text{ diopters} \quad (13)$$



**FIGURE 14** Representative sets of average data for the longitudinal chromatic aberration of the eye. The data for all subjects and studies have been displaced in power so that the chromatic aberration is always zero at 578 nm. (Based on Jenkins,<sup>175</sup> Wald and Griffin,<sup>202</sup> Bedford and Wyszecki,<sup>204</sup> Howarth and Bradley.<sup>205</sup>)



where the wavelength is in nanometers and the difference in refraction,  $R_x(\lambda)$ , is in diopters. LCA has little effect on visual acuity for high-contrast objects in white light.<sup>206</sup> This appears to be because the spectral weighting introduced by the photopic luminosity curve, which is heavily biased toward the central region of the spectrum, results in the primary effect of the LCA being to degrade modulation transfer at intermediate spatial frequencies rather than at the high spatial frequencies involved in the high-contrast acuity limit.<sup>107</sup> Alternatively, McLellan et al.<sup>207</sup> have argued that it is the interaction between the monochromatic and chromatic aberrations of the eye that helps to minimize any change in image quality with wavelength across the spectrum.

TCA results in a wavelength-dependent change in magnification in the retinal image, the red image being larger than the blue. Thus in white light the image of an off-axis point is drawn out into a short radial spectrum whose length in a paraxial model increases with the field angle.<sup>208</sup> TCA therefore affects modulation transfer for tangentially oriented grating components of images.<sup>201,209,210</sup>

For a centered system, TCA would be zero on the optical axis. Since in the eye there is a  $\sim 5$  deg difference (called angle  $\alpha$ ) in orientation between the visual axis joining the fixation point, nodal points and fovea of the eye, and the approximate optical axis, some foveal TCA might be expected. Remarkably, however, the center of the pupil in most eyes lies almost exactly on the visual axis<sup>1,200,211</sup> so that actual values of foveal TCA are typically only of the order of 0.7 min arc.<sup>1,211</sup> Although this value is less than predicted by simple eye models, it is still large enough to cause some orientation-dependent image degradation. This may be substantially increased if an artificial pupil is used which, for any reason, becomes decentered: such a situation may arise when using visual instrumentation having exit pupils which are much smaller than the entrance pupil of the eye.

With binocular viewing, the TCA associated with small decentrations of the natural pupils and of the foveas from the optical axis leads to the phenomenon of *chromostereopsis*, whereby objects of different colors placed at the same physical distance may appear to the observer to be at different distances.<sup>212–218</sup> The exact effect varies with the individual and, when artificial pupils are used, with the separation of the pupils. It is thus of some practical significance in relation to the design of instruments, such as binocular microscopes, in which interpupillary distance settings may not always be optimal for the observer.<sup>216</sup>

In the periphery, TCA may play a significant role in limiting the detection of tangential as opposed to radial gratings.<sup>210,219</sup> There is as yet no consensus as to its magnitude, although some measurements have been made.<sup>220</sup>

## Intraocular Scattered Light and Lenticular Fluorescence

A variety of regular and irregular small-scale inhomogeneities exist within the optical media of the eye. These may serve to scatter light during its passage between the anterior cornea and the retinal receptors. Further stray light may arise for reflections at the various optical surfaces and the retina itself, and some light may also penetrate the nominally opaque iris and sclera to reach the internal eye (*diaphany*), particularly in eyes with low pigmentation, as in albinos. The main effect of such light is to reduce the contrast in the retinal image.

Quantitative studies of the effects of stray light on vision were pioneered by Holladay<sup>221</sup> and Stiles,<sup>222</sup> who expressed its impact in terms of an equivalent veiling luminance,  $L_{eq}$   $\text{cd} \cdot \text{m}^{-2}$ , that would produce the same masking effect as a glare source giving an illuminance  $E$  lux at the eye, as a function of the angular distance,  $\omega$  deg, between the glare source and the fixation point. Vos et al.<sup>223</sup> have summarized more recent work by the approximate relationship:

$$L_{eq} = \frac{29E}{(\omega + 0.13)^{2.8}} \quad \text{where} \quad 0.15 \text{ deg} < \omega < 8 \text{ deg} \quad (14)$$

This expression relates to young adult eyes. Scattering increases throughout life by a factor of at least 2 to 3 times<sup>224–227</sup> and glare formulas can be modified to take account of this (e.g., Ref. 228). Roughly a quarter of the stray light comes from the cornea<sup>229,230</sup> and a further quarter from the retinal reflections.<sup>231,232</sup> The rest comes almost entirely from the lens,<sup>233</sup> there being little contribution from the aqueous or vitreous humors in normal healthy eyes.

Ohzu and Enoch<sup>234</sup> attempted to measure a retinal MTF which included the effects of forward scatter, by focusing grating images onto the anterior surface of an excised retina *in vitro* and measuring the image transfer to the far side. They argued that the receptor outer segments effectively act as a fiber optics bundle and that transmission through this bundle produces image degradation which supplements to the main optical elements of the eye. More recent psychophysical measurements<sup>235,236</sup> suggest, however, that forward scatter in the inner retina is negligible, implying that postmortem changes in the retina may have degraded Ohzu and Enoch's MTFs.

In addition to the general effects described above, a variety of more regularly organized, wavelength-dependent, scattering effects in the form of annular haloes or "star" patterns may occur (see e.g., Refs. 237 and 238 for reviews). These are most easily observed with point sources in an otherwise dark field, particularly if the pupil is large. Some of these result from diffraction due to ocular structures with quasi-regular spacing, for example, the corneal epithelial cells or lens fibers,<sup>239,240</sup> others relate to small-scale refractive irregularities or to higher-order aberrations.<sup>241</sup>

Forward scattering of light affects the precision of aberrometry,<sup>137,241</sup> while back scatter is of importance in allowing anterior ocular structures to be viewed by clinical examination techniques such as slit-lamp biomicroscopy and Scheimpflug photography.<sup>242</sup>

Stray light may also arise as a result of fluorescence in the crystalline lens. This increases with age and with the occurrence of cataract,<sup>243</sup> largely through the progressive accumulation of fluorogens. Under some circumstances, the emitted fluorescent light can cause a slight reduction in low-contrast acuity in older individuals<sup>244</sup> but in younger adults the effects are probably of little practical significance,<sup>245</sup> since the cornea absorbs most of the potentially activating short wavelength light.

## 1.6 FINAL RETINAL IMAGE QUALITY

Experimental estimates of the final quality of the retinal image can be made in three main ways: by calculation from wavefront aberration or similar data, by psychophysical methods, and by direct measurement of the light distribution on the retina using a double-pass ophthalmoscopic technique. Although each method has its limitations, the various methods yield compatible results in the same subjects and collectively produce a reasonably consistent picture of the changes in retinal image quality with pupil diameter, retinal location, and age.

### Image Quality on the Visual Axis

**Calculation from Aberration Data** The optical transfer function (OTF) can be calculated by autocorrelation of the complex pupil function with its complex conjugate, using methods originally devised by Hopkins.<sup>246</sup> The pupil function gives the variation in amplitude and phase across the exit pupil of the system. The phase at each point can be deduced from the corresponding value of the wavefront aberration (each wavelength of aberration corresponds to  $2\pi$  radians of phase). It is often assumed that the amplitude across the pupil is uniform but if imagery under photopic conditions is being considered, it may be more correct to take account of the Stiles-Crawford effect (SCE I) by including appropriate amplitude apodization, ideally on an individual basis;<sup>209,247,248</sup> this suggestion is supported by some experimental evidence.<sup>249,250</sup> The point- and line-spread functions can also be directly calculated from the wavefront aberration (see Chap. 4 by Glenn D. Boreman in Vol. I and Chap. 4 by Virendra N. Mahajan and Chap. 5 by Robert Q. Fugate in Vol. V) and appropriate software is often included with current commercial aberrometers.

The attractive feature of this approach is that it can allow the OTF (i.e., both the modulation and phase transfer functions) to be calculated for any orientation. On the other hand, it fails to include the effects of any scattered light and hence may give too optimistic a view of the final retinal image quality, particularly in older eyes in which scattering is high: high levels of intraocular scatter may have the additional effect of reducing the reliability and validity of the aberrometer estimates of wavefront aberration, the exact effects depending upon the design of the particular aberrometer.<sup>251</sup>

Van Meeteren<sup>209</sup> argued that it was appropriate to multiply the aberration-derived MTFs by the MTF derived by Ohzu and Enoch<sup>234</sup> for image transfer through the retina. When this is done, the results agree quite well with those found by the double-pass ophthalmoscope technique;<sup>252</sup> although, as noted earlier, the Ohzu and Enoch MTF may overestimate the effects of retinal degradation. A further problem with MTFs derived from wavefront measurements is that most aberrosopes estimate the form of the wavefront from measurements made at a limited number of points across the pupil.<sup>253,254</sup> Variations in aberration on a small spatial scale may therefore remain undetected, with consequent uncertainties in the derived OTFs; this problem was probably greatest with the early designs of aberroscope, such as the crossed cylinder device.<sup>252</sup>

**Psychophysical Comparison Method** This method depends upon the comparison of modulation (contrast) thresholds for a normally viewed series of sinusoidal gratings of differing spatial frequencies with those for similar gratings which are produced directly on the retina by interference techniques.<sup>256,257</sup>

Suppose an observer directly views a sinusoidal grating of spatial frequency  $R$ . Then if the grating has modulation  $M_0(R)$ , the modulation of the retinal image will be  $M_0(R) \cdot T(R)$ , where  $T(R)$  is the modulation transfer of the eye at this spatial frequency, under the wavelength and pupil diameter conditions in use. If now the modulation of the grating is steadily reduced until it appears to be just at threshold, the threshold modulation  $M_{IT}(R)$  on the retina will be given by

$$M_{IT}(R) = M_{OT}(R) \cdot T(R) \quad (15)$$

where  $M_{OT}(R)$  is the measured modulation of the external grating at threshold. The reciprocal of  $M_{OT}(R)$  is the corresponding conventional contrast sensitivity and its measurement as a function of  $R$  corresponds to the procedure used to establish the contrast sensitivity function.

It is clear that  $M_{IT}(R)$  corresponds to the threshold for the retina/brain portion of the visual system. If its value can be independently established, it will be possible to determine  $T(R)$ .  $M_{IT}(R)$ , in fact, can be measured by bypassing the dioptics of the eye and their aberrations and forming a system of interference fringes directly on the retina. This procedure was originally suggested by Le Grand<sup>256,257</sup> and has since been progressively improved<sup>258–264</sup> (see Ref. 257 for review). Two mutually coherent point sources are produced close to the nodal points of the eye and the two resultant divergent beams overlap on the retina to generate a system of Young's fringes, whose angular separation,  $\gamma$  rads, is given by  $\gamma = \lambda/a$ , where  $\lambda$  is the wavelength and  $a$  is the source separation, both measured in air. If the sources have equal intensity, the fringes will nominally be of unit modulation. Fringe modulation can be controlled by varying the relative intensities of the two sources, by adding a uniform background, or by modulating the two sources with a temporal square-wave and introducing a phase difference between the two modulations. The contrast threshold  $M_{IT}(R)$  for the retina brain can then be measured as a function of  $R$ , allowing the modulation transfer for the ocular dioptics,  $T(R)$ , to be deduced from the relationship:

$$T(R) = \frac{M_{IT}(R)}{M_{OT}(R)} \quad (16)$$

There are some problems with this approach. Both sets of thresholds are affected by stray light, but in different ways. The determination of the external modulation threshold involves light entering the full pupil whereas for the interferometric measurements only two small regions of the pupil are used. There may also be problems in maintaining the same threshold criterion for the two types of grating, particularly when they may differ in color, field size, and possibly speckle characteristics. Lastly, although the method can in principle give ocular MTFs for any orientation, it yields no phase information and hence the PTF cannot be determined.

Some other psychophysical methods have been suggested<sup>265</sup> but as yet they have not been widely employed.

**Ophthalmoscopic (Double-Pass) Methods** When the image of an object is thrown on the retina, some of the light will be reflected back out of the eye and can be collected by an appropriate observing system

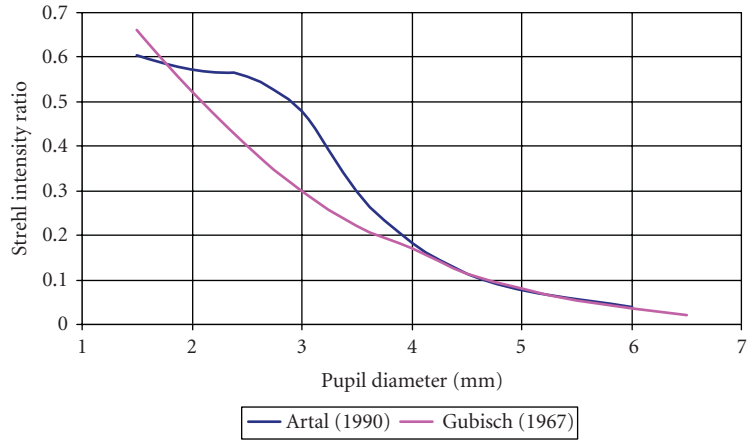
to form an external image. If, for example, the object is a narrow line, the external image will be the LSF for the double passage of the eye. It is usual to make the assumption that the retina acts as a diffuse reflector,<sup>266</sup> coherence of the image light being lost in the reflection. In this case, earlier workers assumed that the image suffered two identical stages of image degradation, so that the MTF deduced from the Fourier transform of the external LSF was the square of the single-pass MTF. Flamant's pioneering study<sup>267</sup> with this method used photographic recording, but later workers have all used electronic imaging methods, initially slit-scanning arrangements with photomultipliers to record LSFs, and latterly low-noise CCD cameras which allow PSFs to be recorded.<sup>266,268–276</sup> An important advance was the realization that in the simple form of the method as employed in earlier studies, when the same pupil acted as aperture stop for both the entering and exiting light paths, information on odd-order aberrations and on transverse chromatic aberration was lost.<sup>274</sup> While the estimates of MTF were potentially correct, the PTF could not be measured. This problem can be overcome for monochromatic aberrations by arranging the entering and exiting beams so that the entrance pupil is smaller than the exit pupil.<sup>275,276</sup> If the entrance pupil is small enough for the initial image to be effectively diffraction limited, the true single-pass OTF (i.e., both the MTF and the PTF) can be deduced from the double-pass PTF, at least up to the cutoff frequency imposed by the small entrance pupil. Some theoretical aspects of this problem have been discussed by Diaz-Santana and Dainty.<sup>277</sup>

The double-pass method has been used to explore the extent to which poorer retinal image quality contributes to the deterioration in visual performance that is observed in older eyes,<sup>278</sup> and to demonstrate the changes in retinal image quality with accommodation that are caused by aberrational change in the crystalline lens.<sup>279</sup> An adaptation allows the basic method to be used to determine an "index of diffusion" designed to characterize the optical deficit in eyes with age- and disease-related abnormalities of the anterior segment, using encircled energy measurements of the double-pass PSF.<sup>280</sup>

In all variations of the double-pass method, one problem is that light levels in the outer parts of any spread function are low, leading to possible truncation errors and to overestimation of the MTF.<sup>281</sup> Vos et al.<sup>223</sup> attempted to overcome the truncation problem by combining the ophthalmoscopic estimates of the PSF with measurements of wider-angle entoptic stray light, to produce a realistic estimate of the full light profiles in the foveal white-light PSF. A vexing question which has yet to be fully answered is the identity of the layer or layers at which the retinal reflection occurs: it seems likely that this is wavelength dependent. If more than one layer is involved, the estimated MTF will be somewhat too low. However there is evidence that any effect of retinal thickness on the estimated MTF is small<sup>282</sup> and that scattered light from the choroid and deeper retina is guided through the receptors on its return through the pupil.<sup>283</sup>

**Comparison between Methods** Only a few direct comparisons of MTF measurements have been made by different techniques on the same eyes. Campbell and Gubisch<sup>266</sup> found that their double-pass MTFs were lower than those determined by the interferometric psychophysical method.<sup>260</sup> A similar result was found by Williams et al.,<sup>284</sup> who noted that agreement between the two techniques was better if green rather than red light was used for the double-pass measurements, presumably as a result of reduced retinal and choroidal scatter.<sup>285</sup> Although MTFs derived from early aberrometers, which only sampled the pupil at a small number of points, tended to be markedly higher than those derived by other methods, if green light is used with young eyes the three basic methods appear to yield very similar MTFs: increased entoptic scatter in older eyes may cause larger differences. Liang and Williams<sup>286</sup> give comparative MTF results obtained by the three techniques with three subjects with 3-mm pupils. The greatest discrepancies appear to be at intermediate spatial frequencies, where values of modulation transfer increase in the order double-pass, interferometric, and wave aberration derived.

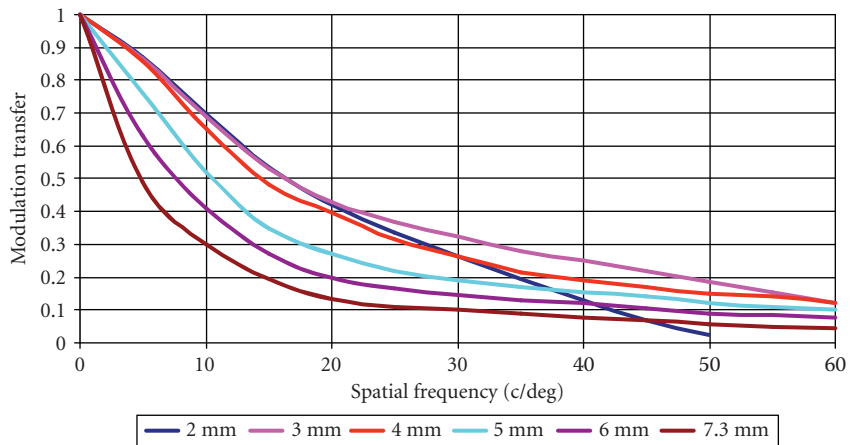
**Summary of Observed Optical Performance** When the eye is corrected for any spherocylindrical refractive error, all investigators agree that, near the visual axis, the eye's performance in monochromatic light is reasonably close to the limit set by diffraction for pupil diameters up to 2 mm. As pupil size is increased further, aberration starts to play a more important role. The increasing impact of aberration is illustrated by a consideration of the changes in the Strehl intensity ratio,<sup>287</sup> the ratio of the maximum irradiance in the PSF to that which would be found in a truly diffraction-limited



**FIGURE 15** Changes in Strehl ration with pupil diameter. (Based on data from Artal<sup>288</sup> and Gubisch.<sup>288</sup>)

system. Typical published values<sup>89,288</sup> are shown in Fig. 15. Although differences between individuals may be expected, the ratio falls steadily as the pupil diameter is increased, indicating the increasing impact of aberrations. For the smallest natural pupil diameters, the value approaches, but does not quite reach, the figure of 0.8 which is usually accepted as the minimum required for an optical system to closely approximate to being diffraction limited:<sup>287</sup> direct measures of the MTF for a 1.5-mm pupil in comparison with the diffraction-limited case support this finding.<sup>276</sup>

The changing balance between diffractive and aberrational effects results in optimal overall performance usually being achieved with pupil diameters of about 2.5 to 3 mm,<sup>266,286,288,289</sup> corresponding to the diameters of natural pupils under bright, photopic conditions. For still larger pupils, the degrading effects of aberration dominate and modulation transfer falls. Examples<sup>286</sup> of typical estimates of MTF, in this case based on wavefront measurements for different pupil diameters in young eyes, are shown in Fig. 16. The MTF at constant pupil diameter tends to deteriorate with



**FIGURE 16** Examples of typical foveal MTFs in young, adult eyes for the pupil diameters indicated. (After Liang and Williams.<sup>286</sup>)

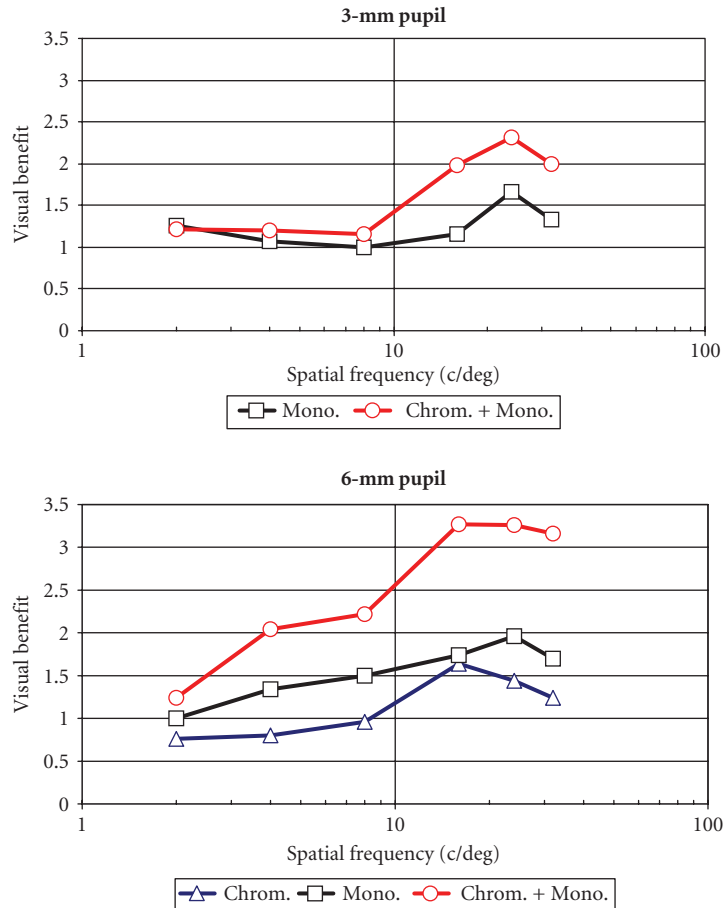
age, due to both increased aberration and scatter.<sup>278</sup> It should be noted that, although transmission through the eye undoubtedly changes the polarization state of the incident light, the retinal image quality is apparently nearly independent of the initial state of polarization. However, quality as recorded by double-pass measurements may vary if polarizing elements are included in both the entering and exiting light paths.<sup>290</sup>

**Effects of Aberration Correction on Visual Performance** Early attempts<sup>291</sup> to improve axial visual performance by correction of monochromatic aberrations were failures, largely because it was assumed that spherical aberration was always dominant and the same in all eyes. Correction of longitudinal chromatic aberration also brought no improvement in acuity<sup>292</sup> although modest improvements at intermediate spatial frequencies could be demonstrated.<sup>99</sup> The advent of aberrometers capable of rapidly measuring the aberrations of an individual eye has refocused attention on the possibility that customized aberration correction might yield significantly enhanced vision (e.g., Refs. 293–298).

Several potential methods for correcting the monochromatic aberrations have been suggested. The most effective of these is adaptive optics.<sup>299,300</sup> This is impractical for everyday use but is being vigorously explored as a way of obtaining improved images of the retina for clinical diagnostic or research purposes (e.g., Ref. 301), since near diffraction-limited performance can be achieved through the full, 7 to 8 mm, drug-dilated eye pupil. For normal purposes, three methods have been suggested. In the first, spatially modulated excimer laser ablation of the cornea is used to correct the measured aberrations of the individual eye, as well as any second-order refractive error.<sup>293–298</sup> While such “wavefront-guided” ablation has been effective in reducing ocular aberrations consequent upon laser refractive surgery, it has yet to succeed in producing eyes which are completely aberration-free, largely because of factors such as subject-dependent variations in healing which influence the final level of aberration. The second possible method is the wearing of customized contact lenses, where the local thicknesses of the lens are manipulated to yield variations in path length which compensate for the ocular wavefront aberration.<sup>302,303</sup> The problem with this approach is that the lens must be stable against both rotation and decentration on the eye if aberration correction is to be maintained.<sup>304–308</sup> The limited control of lens movement occurring in practice means that adequate correction is difficult to achieve in normal eyes. Improved performance may, however, be given in clinically abnormal eyes with high levels of aberration, such as those of keratoconics.<sup>307</sup> The last suggested method, which has yet to be successfully demonstrated, is to incorporate the aberration correction in an intraocular lens: this has the theoretical advantage that the position of such a lens and its correction should be stable in the optical path.

Some of the potential benefits of aberration correction have been demonstrated by Yoon and Williams.<sup>300</sup> Figure 17 shows some of their results for eyes under cycloplegia. The contrast sensitivity function for either a 3- or a 6-mm pupil was measured under four conditions: white light with only spherocylindrical (second-order) refractive error corrected; with chromatic (but not monochromatic) aberration additionally removed by viewing through a narrow-band green filter; with monochromatic (but not chromatic) aberrations corrected using adaptive optics; with both monochromatic and chromatic aberrations corrected by using adaptive optics with a green filter. The retinal illuminances under the various conditions were kept constant with neutral density filters at 14.3 trolands for the 3-mm pupil and 57 trolands for the 6-mm pupil: these correspond to the lower end of the photopic range (a natural 6-mm pupil diameter is reasonably typical for this level of retinal illuminance). Yoon and Williams<sup>300</sup> express their results in terms of the “visual benefit” at each spatial frequency, that is, the ratio of the contrast sensitivity under a particular condition to that achieved with white-light gratings and just the spherocylindrical refractive error corrected.

It can be seen that useful performance gains are given if both monochromatic and chromatic aberration can be corrected, particularly for the larger pupil. The benefits are less impressive (<2 at all spatial frequencies) if only monochromatic aberration is corrected, as would in practice be the case with corneal ablation, contact lens, or other corrections. It has been argued that these gains may be still smaller under real-life conditions, due to such factors as inaccurate accommodation, aberrational changes, and, at the larger natural pupil diameters occurring under mesopic and scotopic conditions, the limits to performance set by the neural parts of the visual system.<sup>309,310</sup> Although in principle



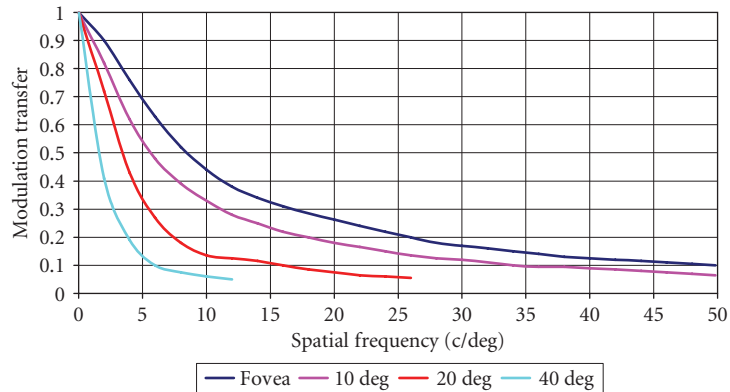
**FIGURE 17** Visual benefit in contrast sensitivity for a subject with (a) a 3-mm pupil and (b) a 6-mm pupil. In all cases the eye has an optimal spherocylindrical refractive correction. The benefit is shown for (i) correction of chromatic aberration only (green light), (ii) correction of monochromatic aberration only (adaptive optics, white light), and (iii) correction of both chromatic and monochromatic aberration (adaptive optics, green light). (Based on Yoon and Williams.<sup>300</sup>)

chromatic aberration of the eye can be corrected,<sup>204,291,311–315</sup> the required multielement lens systems are relatively bulky and decentration effects may mean that new problems arise in controlling transverse chromatic aberration.<sup>316</sup>

## Off-Axis Image Quality

Fewer measurements have been made of off-axis image quality. Psychophysical measurements are difficult to carry out and aberration data, although available, do not yet appear to have been used to calculate MTFs. Most of the available measurements have therefore been made by the double-pass technique.<sup>272, 317–319</sup> As noted earlier, astigmatism usually increases with field angle, so that the PSFs





**FIGURE 18** MTFs at different field angles for a 3-mm pupil. The position of focus is that corresponding to the circle of least confusion, with no correction for oblique astigmatism. (Based on Williams et al.<sup>319</sup>)

normally lack rotational symmetry and different results for MTF and PTF are obtained at different orientations and levels of focus. Some representative data are shown in Fig. 18. The results shown are for the case when the circle of least confusion lies on the retina. Note that the MTF falls with field angle and that, for objects with strongly oriented structure, image quality will be improved if the appropriate focal line can be brought onto the retina, or if the peripheral astigmatism is corrected. Williams et al.<sup>319</sup> have suggested that the falloff in optical performance with field angle is advantageous in helping to prevent aliasing in peripheral vision.

Off-axis MTFs for different pupil diameters and field angles have been approximated as the sum of two exponential functions<sup>318,319</sup> or as a single exponential.<sup>320</sup>

## Retinal Image Quality with Visual Instruments

When an object (e.g., a display screen) is observed with the naked eye, with no intervening optics, the spatial frequency spectrum of the retinal image is simply the spectrum of the object multiplied by the OTF of the eye under the observing conditions in use. The situation is, however, more complex when instruments such as microscopes, telescopes, or binoculars are used. In general, it is incorrect to assume that the OTF of the instrument-eye combination is simply the product of the individual OTFs. Instead, for any field position the aberrated wavefront from the instrument is modified by the wavefront aberrations of the eye, the two wavefronts being added together point by point across the common pupil. Under favorable circumstances where the component aberrations are of opposite sign, this *coherent coupling*<sup>321–324</sup> can obviously result in the summed aberration being smaller than the individual wavefront aberrations, and hence in performance being superior to that which would be expected on the basis of the product of the individual OTFs.

Two factors are of particular importance in this coupling: first, the size and positional relationship between the instrument's exit pupil and the ocular entrance pupil and, second, the state of focus of the eye. It is clear that if the entrance pupil of the eye is smaller than the exit pupil of the instrument (the two pupils coinciding) the eye pupil will constitute the aperture stop of the combined system. Movement of the smaller eye pupil within the instrument's exit pupil will sample different areas of the instrumental wavefront aberration and may therefore influence retinal image quality. Conversely, if the exit pupil is smaller than the eye pupil, the instrument's pupil provides the aperture stop and controls the aberrations. For example, with an exit pupil of 2-mm diameter or less centered on the eye pupil, the combined performance would be governed almost entirely by the instrumental aberrations,



since the eye aberrations would normally be small enough to make a negligible contribution to the combined wavefront aberration. Any focus error in the instrument or accommodation error in the observer will add an additional defocus term to the wavefront aberration. This may again be beneficial under some circumstances: for example, instrumental field curvature may be compensated for by changes in ocular accommodation (Refs. 325 and 326, and see later under “Vergence Input”).

Considerations of this type mean that the MTF of any instrument alone may not be a very good guide to the visual performance that can be achieved with it.<sup>327,328</sup> One possible solution is to test the instrument in combination with a model eye which simulates as accurately as possible the aberrations and focus of the real eye.<sup>329</sup> Burton and Haig<sup>330</sup> explored the tolerance of the visual system to different levels and types of wavefront aberration (see also Ref. 331) and concluded that instrumental criteria based on the Strehl ratio may ultimately be more useful than MTF tests.<sup>328</sup> This suggestion has, however, been disputed by Mouroulis and Zhang<sup>332</sup> who feel that other criteria are more informative.

Chromatic aberrations are also of importance. The role of longitudinal chromatic aberration has been considered by Mouroulis and Woo,<sup>332,323</sup> who find that under some circumstances quite substantial amounts of instrumental LCA can be tolerated. Further studies support a tolerance of the order of 2.5 min arc for transverse chromatic aberration, with TCA having greater effects on contrast sensitivity than high-contrast resolution.<sup>334</sup> The general question of the design and testing of visual instruments is well reviewed in Mouroulis.<sup>335</sup>

## 1.7 DEPTH-OF-FOCUS AND ACCOMMODATION

In the foregoing it has been tacitly assumed that the eye is always optimally focused. In practice, since clear vision is required over a range of distances, exact focus may not always be achieved. It is of interest that, even if objects are at a fixed distance, small changes in focus may be required to optimize modulation transfer at different spatial frequencies in the presence of some types of aberration (e.g., spherical aberration<sup>334–339</sup>). With the increasing availability of aberrometers, considerable efforts are being devoted to establishing robust criteria for “best focus” (i.e., ocular refraction) from wavefront data.<sup>340–342</sup>

### Ocular Depth-of-Focus

As in any optical system, very small changes in ocular focus have little effect on the retinal image but image quality deteriorates progressively as the error of focus increases (e.g., Fig. 9). For many practical purposes we would like to know how large the dioptric error of focus can be before image quality becomes unsatisfactory. However, this immediately raises the question of what we mean by “unsatisfactory.” Atchison et al. have, for example, defined the concepts of “noticeable,” “troublesome,” and “objectionable” blur.<sup>343</sup> Noticeable blur is the defocus level at which blur of a set of letters first becomes detectable, troublesome blur is where the lack of clarity in the letters starts to be irritating, although the letters may still be readable, and objectionable blur is the level of blur which is unacceptable: the three dioptric limits of blur were found to be in the ratio of about 1.0:1.7:2.3, respectively. Ciuffreda et al. find very similar results when using minor variants of the same criteria.<sup>344</sup> In fact, as will be discussed below, values of depth-of-focus are strongly dependent upon the methodology and conditions used (see Ref. 345 for a recent review).

In the geometrical optical approximation, for an aberration-free eye the angular diameter of the retinal blur circle  $\beta$  degrees increases linearly with the error of focus  $\Delta F$  diopters and pupil diameter  $d$  mm [Eq. (10)]. Thus the limits of the depth-of-focus correspond to the blur circle diameter reaching some assumed tolerable limit  $\beta_{\text{tol}}$  degrees: the corresponding value of  $\Delta F_{\text{tol}}$  is then determined from [Eq. (10)] to yield for the total geometrical depth-of-focus  $\text{DOF}_{\text{go}}$

$$\text{DOF}_{\text{go}} = 2\Delta F_{\text{tol}} = \frac{(34.9\beta_{\text{tol}})}{d} \text{ diopters} \quad (17)$$

The exact value obtained is, then, dependent on the assumed value of  $\beta_{\text{tol}}$ . If, say, a value of 2 min arc is taken, and the pupil diameter is 3 mm, the geometric DOF is about 0.4 D. Note that the geometric DOF is inversely dependent on the pupil diameter.

When the effects of physical optics are considered, for a diffraction-limited system it is conventional to use the Rayleigh criterion and to say that the limits of depth-of-focus are set by the requirement that the optical path difference between light from the center and edge of the pupil should not exceed a quarter-wavelength [one Rayleigh unit of defocus, see Eq. (9)]. For the eye, this implies that the total physical optical depth-of-focus ( $\text{DOF}_{\text{po}}$ ) should correspond to 2 Rayleigh units, that is,

$$\text{DOF}_{\text{po}} = \frac{4 \cdot 10^{-3} \lambda}{d^2} \text{ diopters} \quad (18)$$

where  $\lambda$  is in nm and  $d$  is in mm. Optimal focus will lie midway through this total depth-of-focus. Note that, unlike the geometrical approximation, [Eq. (18)] predicts that depth-of-focus will be inversely proportional to the square of the pupil diameter.

In reality, the ocular depth-of-focus depends on a variety of additional factors. From the purely optical point of view, even if the eye is diffraction-limited the rate of loss in modulation transfer with defocus is spatial-frequency-dependent (Fig. 9), so that the detectable error of focus is likely to depend upon the spatial-frequency content of the object under observation, as well as the pupil diameter. Low spatial frequencies are relatively insensitive to focus change.<sup>94–96,346</sup> Both monochromatic and chromatic aberrations will further modify the through-focus characteristics of the retinal image focus and, in general, will tend to increase the depth-of-focus.

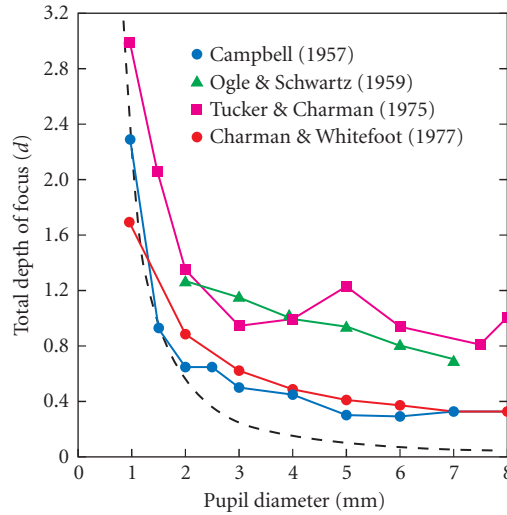
Equally importantly, the perceptible focus changes will depend upon the neural characteristics of the visual system.<sup>347</sup> Under many conditions, the limited capabilities of the retina/brain system will mean that defocus tolerance may become larger than that expected on purely optical grounds. With larger pupils and photopic conditions, the Stiles-Crawford effect may play a role in reducing the effective pupil diameter and increasing the depth-of-focus (see “The Stiles-Crawford effect”). At low luminances, since only low spatial frequencies can be perceived<sup>348</sup> there is an increased tolerance to defocus blur:<sup>349,350</sup> a similarly increased tolerance is found in low-vision patients at photopic levels.<sup>96</sup>

Figure 19 shows a selection of experimental depth-of-focus data obtained by various techniques.<sup>349,351–354</sup> Although the exact results depend upon the methods and criteria used, it is clear that for larger pupils the visual depth-of-focus substantially exceeds the purely optical Rayleigh-limit predictions. Of more importance from the practical viewpoint, errors of focus in excess of about 0.2 to 0.5 D are likely to lead to perceptible image blur under photopic conditions with pupils of diameter 3 to 4 mm. Thus the eye must either change its focus (accommodate) for the range of distances that are of interest in everyday life (e.g., from infinity down to 0.2 m or less, corresponding to vergences from 0 to 5 D), or use some form of optical aid such as reading glasses.

It is of interest that Goss and Grosvenor<sup>355</sup> concluded from their review of the available clinical literature that conventional refraction is repeatable to within 0.25 D in approximately 75 percent of cases and to within 0.50 D in 95 percent of cases: errors in prescription of 0.25 D have been shown to produce dissatisfaction in many patients and a significant loss in both acuity and contrast sensitivity.<sup>356</sup>

## The Accommodation Response

As with any focusing system, several aspects of accommodation are of interest: its range or amplitude, its speed, its stability, and its time-averaged steady-state characteristics. All of these are age-dependent and in most of what follows the behavior of young, adult (age around 15 to 35 years) eyes is described. Discussion of the response is complicated by the fact that, under normal binocular conditions of observation, accommodation (i.e., the focusing system) is intimately linked with the vergence system which ensures that the eyes converge appropriately to bring the images of any object of regard onto the foveas of the two eyes (see “Vergence Input” and “Movements of the Eyes”). Due to this linkage, accommodation can drive convergence and vice versa. The pupil usually contracts



**FIGURE 19** Examples of experimental measurements of photopic, total monocular depth-of-focus as a function of pupil diameter (optimal focus lies midway through the total depth-of-focus). The criteria for determining the depths-of-focus are Campbell<sup>349</sup>—just-perceptible blur for a small disc, one subject, white light; Ogle and Schwartz<sup>351</sup>—50-percent probability of resolving a 20/25 checkerboard, mean of three subjects, white light; Tucker and Charman<sup>352</sup>—80-percent probability of achieving 90 percent of the optimal Snellen acuity, mean of two subjects, white light; Charman and Whitefoot<sup>353</sup>—detectable movement of laser speckles, mean of six subjects, 633 nm. The dashed line gives the depth-of-focus based on Rayleigh's quarter-wavelength criterion for an aberration-free eye in 555-nm light.

during near vision (often called accommodative miosis), although it appears that such contraction is not directly driven by accommodation but is a comovement.<sup>357–362</sup> The three functions (accommodation, vergence, pupil) are sometimes known as the *near triad*. One important result of this linkage is that both the dynamic and static accommodation responses in the two eyes are always essentially the same, even when one eye is occluded.<sup>363,364</sup> Accommodation is driven by the responses of the cones of the retina and becomes progressively less effective as lighting levels are lowered from photopic levels through the mesopic region: at scotopic levels it is inoperative (see “Accuracy of Response”).<sup>365,366</sup>

Although details of the neural and physiological mechanisms responsible for accommodation are beyond the scope of this section (see, e.g., Refs. 367–370 for reviews), it will be helpful to remember that, in simple terms, the crystalline lens is supported around its equator by a system (the zonule) of approximately radially oriented zonular fibres. The far ends of these fibres are anchored in the ciliary body or muscle and neighboring structures. The ciliary body effectively forms a ring surrounding the lens (Fig. 1). The lens and its thin enclosing capsule are elastic and, if the lens is free of the forces applied to it by the zonular fibres, its structure is such that its surfaces will naturally assume the relatively steep curvatures required for near vision.

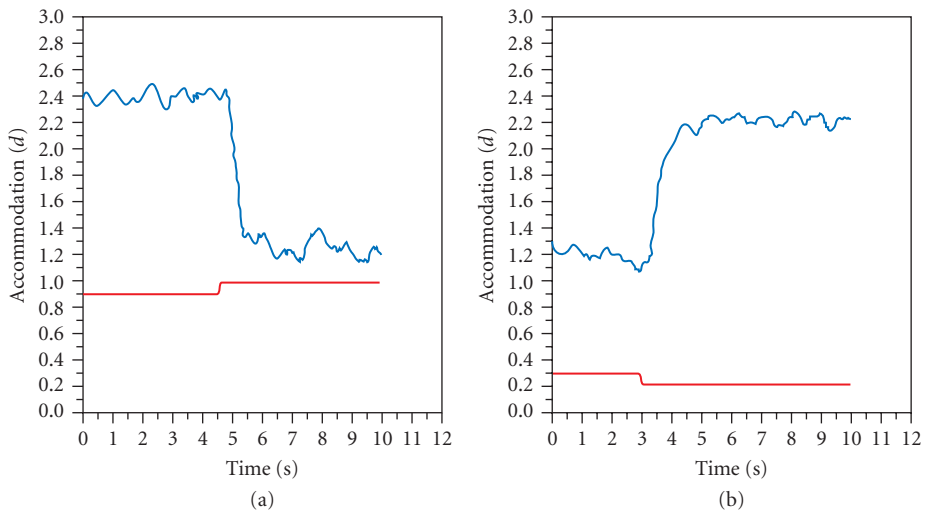
Under conditions of distant viewing, the inner diameter of the ciliary ring is relatively large. This leads to a correspondingly high tension in the zonular fibers. These tensional forces when applied to

the periphery of the elastic lens and capsule cause the lens surfaces to flatten slightly, reducing the optical power to the value required for a clear distance focus. During active accommodation, the ciliary ring reduces in diameter. This relaxes the tension in the zonular fibres and allows the surfaces of the elastic lens to assume a steeper curvature and the power of the lens to increase. Attempts have been made to model this process (e.g., Refs. 371–374).

**Dynamics of Response** Typical records of a young subject's responses to abrupt changes in the position of an accommodation target are shown in Fig. 20. It can be seen that the change in position is followed by a short reaction time, or latency (about 0.4 s) during which the response remains unchanged. The response then progresses to its new level, the minimum response time typically being around 0.6 s.<sup>360,375–380</sup> Under conditions where there are plentiful cues to target distance this response usually occurs in a single sweep, but if binocular and other cues to target position apart from defocus blur are eliminated, the response may become uncertain and may initially be in the wrong direction, leading to longer response times.<sup>381–383</sup> The response times become larger for larger dioptric changes.<sup>377–380</sup>

Another way of characterizing the response dynamics is in terms of their frequency response characteristics.<sup>384</sup> These may be assessed by determining the gain and phase of the response obtained when a target is observed whose vergence is changing sinusoidally with time, as a function of the temporal frequency. It appears that when gain and phase are plotted as a function of the frequency both vary in an essentially linear fashion, with the cutoff frequency at which response no longer tracks the stimulus being about 2 Hz.<sup>385</sup> It should be stressed that these characteristics are not the output of a simple reflex system but depend upon higher-order involvement. They are strongly influenced by training and motivation,<sup>386–389</sup> and, with repetitive stimuli, by the knowledge that the required response has a predictable periodic form.<sup>390</sup> When the response to an abrupt unexpected step change in target vergence is analyzed in terms of the corresponding frequency response, much larger phase lags are given.<sup>385,390</sup> Thus the system does not behave linearly.<sup>385</sup>

The importance of perceptual factors in relation to accommodation is exemplified by studies in which the distance of the target (a Maltese cross) is kept constant but its lateral scale is varied sinusoidally.



**FIGURE 20** Examples of monocular accommodation responses to abrupt step changes in the vergence of an accommodation stimulus. Note the time interval (latency or reaction time) before each response starts to change and the time taken to complete the response (response time). Fluctuations in accommodation can be seen. The lower traces merely show the times at which the stimulus changes (between vergence levels of 2.38 and 1.33 D) occur.

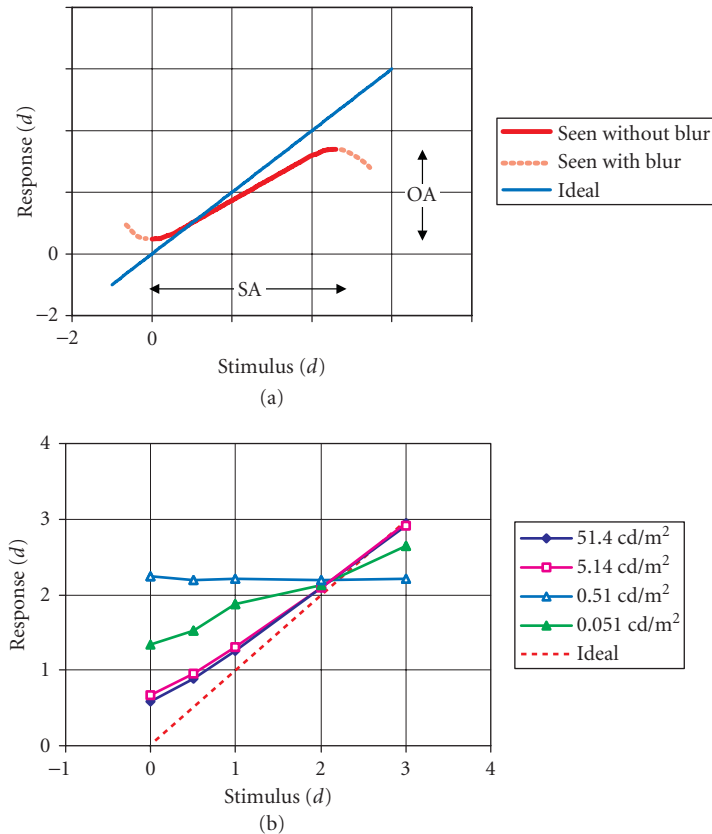
Observers usually interpret this size variation at constant distance as a variation in distance of a target of constant linear size, and change their accommodation accordingly, even though this blurs the retinal image.<sup>381</sup> A similar effect is observed when binocular accommodation is stimulated by apparent depth elicited by pairs of stereograms at fixed distance.<sup>392</sup>

**Stability of Response** When a target at fixed distance is viewed with steady fixation, the accommodation response typically shows small fluctuations ( $\sim 0.25$  D) occurring at frequencies up to about 5 Hz (see Fig. 20): the fluctuations are correlated in the two eyes (see Refs. 393 and 394 for reviews). It appears that the lower frequency (0–0.6 Hz) components of the fluctuations may be under neural control but that higher frequencies (1.0–2.3 Hz) are related to physiological rhythms and other sources of noise.<sup>395</sup> The power in the low-frequency components increases as the target luminance is reduced, while that in the high-frequency components remains stable.<sup>395,396</sup> Similarly, the low-frequency components increase when the pupil is small and the depth-of-focus is large.<sup>397</sup> A peak in the frequency spectrum at frequencies of 1 to 2 Hz is often observed<sup>398,399</sup> and is correlated with the arterial pulse.<sup>400,401</sup> The fluctuations appear to have their origin in the crystalline lens and tend to increase in magnitude as the mean accommodation level increases,<sup>397,402–404</sup> although there is some suggestion that the higher-frequency components diminish at very high stimulus levels.<sup>395</sup> There is some disagreement on the exact nature of the changes with age.<sup>395,405</sup>

The possible role of these fluctuations in relation to accommodation control remains contentious. Some have argued that they simply represent plant noise<sup>384,406</sup> and are of no utility. Others suggest that they could both guide the initial direction of the response and help to maintain accurate focus,<sup>101,399,407–411</sup> the basic hypothesis being that if a fluctuation in one direction improves the clarity of the image the control system responds by moving the mean level of accommodation in that direction. Current opinion generally favors the concept that the high-frequency fluctuations represent plant noise and that any role in accommodation control is played by the lower-frequency components (below about 0.6 Hz), and that these would be primarily involved in the maintenance of the response at a steady level, rather than in rapid response changes.<sup>410</sup> In any case, under most circumstances the fluctuations appear to produce little variation in visual acuity,<sup>412</sup> although their effects can just be detected.<sup>413,414</sup>

**Accuracy of Response** Following pioneering work by Morgan,<sup>415</sup> numerous studies have shown that, rather than there being one-to-one matching between the dioptric stimulus level and the corresponding accommodation response, steady-state errors are an intrinsic part of the accommodation control system. This is true under both monocular and binocular conditions, although the magnitude of the errors differs in the two states.<sup>415,416</sup> These focus errors are often the major cause of foveal retinal image degradation, rather than the higher-order aberrations discussed earlier (see “Factors Affecting Retinal Image Quality”). Figure 21a illustrates a schematic response/stimulus curve. This emphasizes that the steady-state is usually characterized by overaccommodation (accommodative “lead”) for distant objects and underaccommodation (“lag”) for near objects. The range of stimulus vergence over which there is no noticeable image blur is termed the *subjective amplitude of accommodation*; it obviously includes depth-of-focus effects. The corresponding, somewhat smaller, range of actual accommodation response is the *objective amplitude of accommodation*. Note that a clinically “emmetropic” or “corrected” subject is usually left with a small myopic refractive error for objects at infinity, ocular depth-of-focus being relied upon to give clear vision under these circumstances.

The slope of the quasi-linear region of the curve depends upon the observing conditions, including target form and contrast,<sup>100,101,417–426</sup> ocular pupil diameter,<sup>427–429</sup> luminance level,<sup>350,365,366,417</sup> and the acuity of the observer.<sup>418,430–433</sup> The common feature is that, as the quality of the stimulus available to the visual system degrades, the slope of the central region of the response/stimulus curve diminishes (see, e.g., Fig. 21b, where the slope falls as the luminance is reduced and cone vision and visual acuity are gradually lost). To a close approximation, for any individual, as the stimulus degrades the curve pivots about the point for which stimulus and response are equal, that is, where the curve crosses the ideal one-to-one stimulus/response line. It is of interest that in many studies it appears that there is

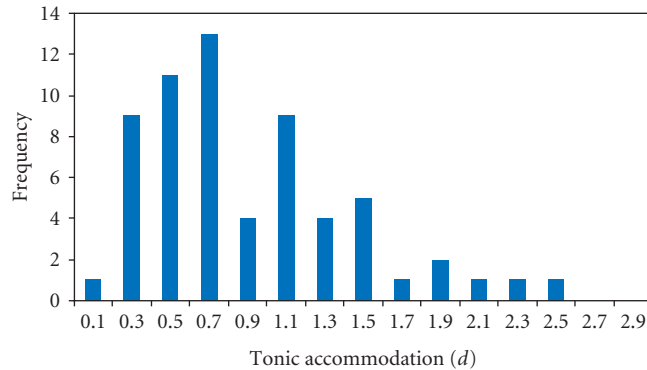


**FIGURE 21** (a) Schematic appearance of a typical accommodation response/stimulus curve under photopic conditions. The black portion of the curve represents the range over which the target is seen without noticeable blur, the grey portions where the lags or leads exceed the ocular depth-of-focus and the target appears blurred. OA is the objective amplitude of accommodation and SA the subjective amplitude. The “ideal” line corresponds to equal response and stimulus values. (b) Typical changes in the lower part of the response/stimulus curve with target luminance. (Based on Johnson<sup>366</sup>) Target luminances are as indicated and the independently measured dark focus for the same subject is 2.0 D, corresponding closely to the cross-over point of the two curves.

a linear relationship between the slope of the central region of the response/stimulus curve and the eye’s minimum angle of resolution under the target and observing conditions in use.<sup>434</sup>

The extreme case in this process is where the stimulus is either a uniform photopic field completely lacking in detail (a Ganzfeld), or the field is completely dark. In both cases no spatial information is available to the accommodation control system and the response/stimulus curve becomes completely flat. The response remains constant at the somewhat myopic value at which, under normal photopic conditions with a structured target, response equals stimulus (see, e.g., Fig. 21b). The refractive states under these stimulus-free, light and dark conditions are known as *empty field myopia* and *dark focus* respectively, and for any individual these states are essentially the same.<sup>435–438</sup>

These observations have led to the concept that this intermediate myopic level of accommodation is, in fact, the *tonic* level of accommodation (sometimes called the resting state or equilibrium level)



**FIGURE 22** Frequency distribution of tonic accommodation as estimated by the dark focus. The values shown represent the difference in autorefractor measurements of refraction in photopic conditions and in darkness. (After McBrien and Millodot.<sup>439</sup>) The mean value of tonic accommodation and its standard deviation are  $0.91 \pm 0.53$  D (62 young adult subjects).

to which the system returns in the absence of an adequate stimulus. This implies that accommodation must actively change from this level to view both distant and near targets. The tonic level varies somewhat with the individual (Fig. 22) but appears to have a mean value of around 1 D.<sup>437,439</sup>

Current theories suggest that these characteristics are dictated by the balance between the sympathetic and parasympathetic innervations to the system.<sup>368,440–442</sup>

**Vergence Input** As noted earlier, the convergence required to maintain single vision of a near object during binocular observation provides an input to the control system which supplements that due to pure defocus blur. This may be particularly useful in stimulating accommodation under conditions when the accommodation stimulus alone is relatively ineffective, for example, under mesopic night-driving conditions where the onset of *night (or twilight) myopia* is delayed when viewing is binocular.<sup>443</sup>

Control system models for the combined accommodation-convergence system, involving both fast feedback loops and slow adaptive elements, have been proposed by various authors (see, e.g., Refs. 368, 440, 444–446).

**Application to Instrumentation** From the practical point of view, it is evident that errors in accommodation and the resultant degradation in retinal image quality will be minimized if targets such as displays which are viewed by the naked eye are placed at vergences which approximately match the typical tonic levels of observers, that is, distances of about 1 m. Acuity will then be optimal<sup>366</sup> and the accommodative effort, and hence potential fatigue, will be minimized. With color displays, the longitudinal chromatic aberration of the eye means that more accommodative effort is required for red/black symbol combinations than for blue/black combinations. Only minor variations in response have, however, been reported when two-color isoluminant symbol/background combinations are used:<sup>447–450</sup> there is no evidence that viewing multichromatic targets results in a less stable response than that for black-and-white targets.<sup>451</sup>

It is of interest that, when focusing visual instruments, most observers set the focus so that the light leaving the eyepiece is slightly divergent, that is, they prefer to accommodate slightly when viewing the instrumental image.<sup>452</sup> This *instrument myopia*<sup>453</sup> correlates closely with the empty field myopia and dark focus of the individual observer,<sup>435–438</sup> suggesting that these myopic states (the anomalous myopias) have a common origin and that the instrument focus is selected to minimize

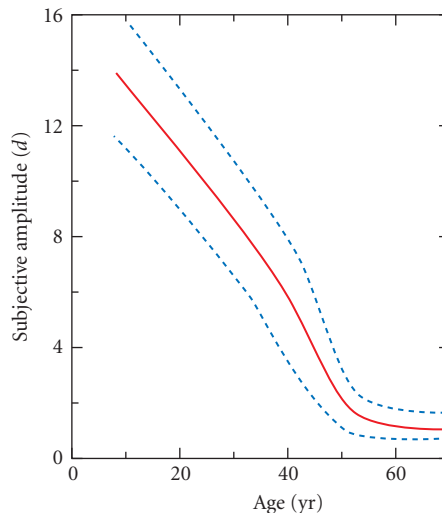
accommodative effort. This implies that fixed-focus visual instruments should not be designed to have parallel light leaving their eyepieces but that focus should be set so that the image appears at a vergence of about  $-1$  D.<sup>452</sup> In any binocular instrument where the imagery is at infinity and the optical axes are arranged to be parallel, proximal (psychic) convergence and accommodation often occur, that is most young users tend to converge and accommodate slightly, leading to a loss in visual performance.<sup>454,455</sup> This is presumed to be due to the perception that the targets lie within the “black box” constituted by the instrument. Smith et al.<sup>456</sup> have discussed the various problems that arise when the focus of binocular instruments is not compatible with the angle between the eyepiece tubes.

Many visual instruments display field curvature and, provided that the vergences involved are negative, this can be at least partly compensated for by accommodation by the observer as the field is scanned.<sup>325,326</sup>

**Age-Dependent Changes in Accommodation** As the lens ages and thickens, its elastic constants change. This, in combination with other ocular changes, causes the efficiency of the accommodation system to diminish (see, e.g., Refs. 367–370 for reviews).

The most obvious change is in the amplitude of accommodation (Fig. 23). The mean subjective amplitude declines steadily from the age of about 10, to reach a small, constant level of about 1 D by the age of 50–55.<sup>457</sup> This residual subjective amplitude represents depth-of-focus rather than true accommodation and the corresponding objective amplitude is zero.<sup>458</sup> Longitudinal measurements of objective amplitudes suggest that the decline for any individual is linear with age.<sup>459–461</sup> Although the exact values of the constants depend upon the individual subject and the method of measurement, the amplitude changes might typically be described by an equation of the form

$$OA = 12.7 - 0.27 (\text{age}) \quad (19)$$



**FIGURE 23** Transverse data for the changes in subjective amplitude of accommodation with age. (After Duane.<sup>456</sup>) The full curve represents the mean result, the dashed curves the limit of the range found at each age. The data refer to 4200 eyes, amplitudes being measured at the spectacle point 14 mm anterior to the cornea.



where OA is the objective amplitude in diopters and the age is in years.<sup>461</sup> Note that the conventional use of a value of 250 mm (i.e., an amplitude of accommodation of 4 D) for the “least distance of distinct vision” in the calculation of the nominal magnification of eyepieces and other devices is unlikely to reflect the effective magnification obtained by the individual user.

Apart from the change in amplitude, the general characteristics of the response/stimulus curve remain robust against age change up to about 35 years, when the slope starts to decline.<sup>362,462,463</sup> The frequency response of the accommodation system maintains its cutoff value at around 2 Hz up to the age of about 40 but at any lower frequency the gain tends to decrease and the phase lag to increase with age. There are also changes in the time constants of the responses to step changes in stimulus.<sup>360,465–467</sup>

The loss in amplitude becomes of practical significance when it declines sufficiently to begin to make it difficult to read at normal reading distances (around 1/3 m). It is tiring to exercise the full available amplitude of accommodation, so that problems usually start to arise at around the age of 40 when the subjective amplitude has fallen to about 5 D. These difficulties steadily increase as the amplitude declines further. Initially these problems may be eased by increases in reading distance, or the use of high light levels to increase depth-of-focus by constricting the pupil, but by the age of about 45 all emmetropes or corrected ametropes inevitably require some form of optical correction for near work. (See Chap. 14 by John S. Werner, Brooke E. Scheffrin, and Arthur Bradley.)

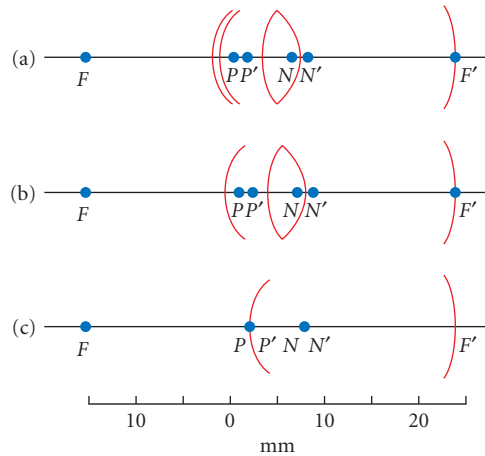
## 1.8 EYE MODELS

Optical models of the eye have long been used in attempts to better understand its imaging characteristics, the development of refraction, and the optical effects of spectacle and contact lenses (see, e.g., Refs. 468–470 for reviews). More recently such models have assumed additional importance in relation to the effects of a variety of surgical procedures for modifying the optics of the eye, such as corneal refractive surgery and intraocular lens implantation.<sup>471</sup> They have also been used in the evaluation of retinal radiation hazards.<sup>472,473</sup> The earlier models were usually limited to describing behavior in the paraxial region, where rays make small angles with the axis: surfaces were assumed to be spherical. However recent years have seen the development of a succession of increasingly sophisticated paraxial and wide-angle models, incorporating such features as aspheric surfaces, gradient index optics, and accommodation.

### Paraxial Models

Figure 24 shows examples of three typical types of paraxial model with spherical refracting surfaces and media characterized by single values of refractive index, the relevant parameters being given in Table 2. These paraxial models are useful for predicting the approximate dimensions of paraxial images (1 deg in the visual field corresponds to about 0.29 mm on the retina) and their changes with accommodation and correcting lenses. However, they are of more doubtful value in indicating retinal image quality on axis or in describing images in the peripheral field, and cannot predict the aberrations of real eyes.

In constructing such models, the values of the parameters are normally selected to be reasonably representative of those found in real eyes but are then adjusted slightly to make the eye emmetropic. In *schematic eyes*, the cornea and lens are each represented by a pair of surfaces (although sometimes the lens is divided into central nuclear and outer cortical regions, these being assigned different refractive indices). In the *simplified schematic eye* a single surface is used for the cornea. It can be seen in Fig. 24 that in both these eyes the two nodal points are very close together, as also are the two principal points. This has encouraged the development of *reduced eye* models, containing a single refractive surface, where each pair of nodal and principal points collapses to a single point. The cardinal points of reduced eyes are very close to those of the more complex paraxial models. Accommodation has been incorporated into several schematic eyes,<sup>469,476–478</sup> some of which incorporate aspheric surfaces and lenticular index gradients.



**FIGURE 24** Examples of paraxial eye models of the human eye. In each case  $F, F'$ ;  $P, P'$ ;  $N, N'$  represent the first and second focal, principal, and nodal points respectively. (a) Unaccommodated schematic eye with four refracting surfaces. (*Le Grand and El Hage*.<sup>474</sup>) (b) Simplified schematic eye with three refracting surfaces. (*Emsley*.<sup>475</sup>) (c) Reduced eye with a single refracting surface. (*Emsley*.<sup>475</sup>) Note that progressive reduction in the number of surfaces used in the model produces only minor changes in the positions of the cardinal points.

**TABLE 2** Parameters of Some Paraxial Models of the Human Eye

		Schematic Eye <sup>474</sup>	Simplified Schematic Eye <sup>475</sup>	Reduced Eye <sup>475</sup>
Radii of surfaces (mm)	Anterior cornea	7.80	7.80	5.55
	Posterior cornea	6.50	-	-
	Anterior lens	10.20	10.00	-
	Posterior lens	-6.00	-6.00	-
Distances from anterior cornea (mm)	Posterior cornea	0.55	-	-
	Anterior lens	3.60	3.60	-
	Posterior lens	7.60	7.20	-
	Retina	24.20	23.90	-
	1st principal point P	1.59	1.55	0
	2nd principal point P'	1.91	1.85	0
	1st nodal point N	7.20	7.06	5.55
	2nd nodal point N'	7.51	7.36	5.55
	1st focal point F	-15.09	-14.99	-16.67
	2nd focal point F'	24.20	23.90	22.22
	Refractive indices	Cornea	1.3771	-
Aqueous humour		1.3374	1.333	4/3
Lens		1.4200	1.416	4/3
Vitreous humour		1.3360	1.333	4/3

## Wide-Angle Models

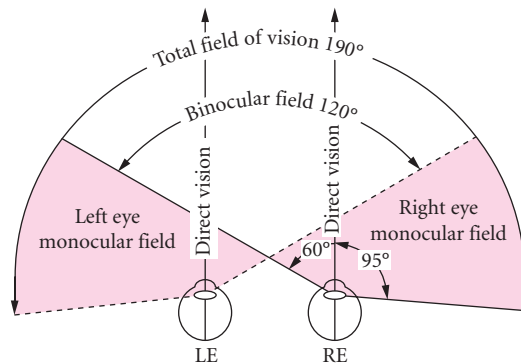
Following gradual advances in our knowledge of the form of the aspheric surfaces of the cornea and lens, together with the lenticular gradients of refractive index, several authors have produced sophisticated wide-angle eye models and have attempted to validate them by comparing the model predictions with the off-axis aberrations of real eyes.<sup>478–487</sup> It is of interest that a very simple reduced eye model with a single aspheric refracting surface and a suitably placed pupil (the “Indiana Eye”) can simulate the chromatic, spherical, and oblique astigmatic aberrations typically found in real eyes.<sup>488–491</sup>

As yet, none of these models is completely successful, but progressive refinement in the experimental data should lead to corresponding improvements in the modeling of the eye’s overall optical performance. Since, as discussed earlier, there are wide variations in both the on- and off-axis performance of individual eyes, it may be that “personalized” eye models, incorporating parameters as measured for the particular eye, will eventually prove to be of greatest value for use in predicting the outcomes of refractive surgery.

## 1.9 TWO EYES AND STEREOPSIS

Although binocular vision confers a variety of advantages, ranging from an extension of the field of view to a lowering of contrast thresholds,<sup>492</sup> attention here will be largely confined to its relevance for stereopsis and stereoscopic instruments. The relationship between the typical monocular and the binocular fields within which stereopsis can occur is shown in Fig. 25.

Just as the second-order wavefront errors (ocular refractions) of the two eyes are usually broadly similar, recent evidence suggests that the higher-order aberrations tend to show mirror symmetry between the two eyes:<sup>128–132,286,493</sup> cone directionality (SCE I) also appears to be mirror symmetric.<sup>494</sup> In cases where marked interocular differences in image quality occur, it is possible that the brain can make selective use of the better of the two retinal images under any particular set of observing conditions due to some form of probability summation. Certainly the apparently drastic technique of monovision contact or intraocular lens correction, in which one eye of a presbyopic observer receives a distance correction and the other a near correction, appears to work well for many patients and to yield acceptable vision over a range of distances.<sup>495–497</sup> It may be the studies of monocular performance that can sometimes give an unduly pessimistic view of the optical information available under binocular conditions.



**FIGURE 25** Approximate horizontal angular extents of the monocular and binocular fields of vision.

## Basics of Stereoscopic Acuity

Due to the lateral separation of the eyes in the head, the apparent angular separations of objects at differing distances are slightly different for the two eyes. The resultant disparities between the two retinal images can be used by the visual system to estimate the relative distances of the objects, although absolute judgment is usually much more dependent upon such monocular cues as perspective and size constancy.

In Fig. 26*a*, suppose that the observer can just detect that the two object points A and B lie at different distances,  $l$  and  $l + \delta l$ , respectively. Then the corresponding stereo acuity  $\delta\theta$  is given by:

$$\delta\theta = \theta_R - \theta_L = \alpha_1 - \alpha_2$$

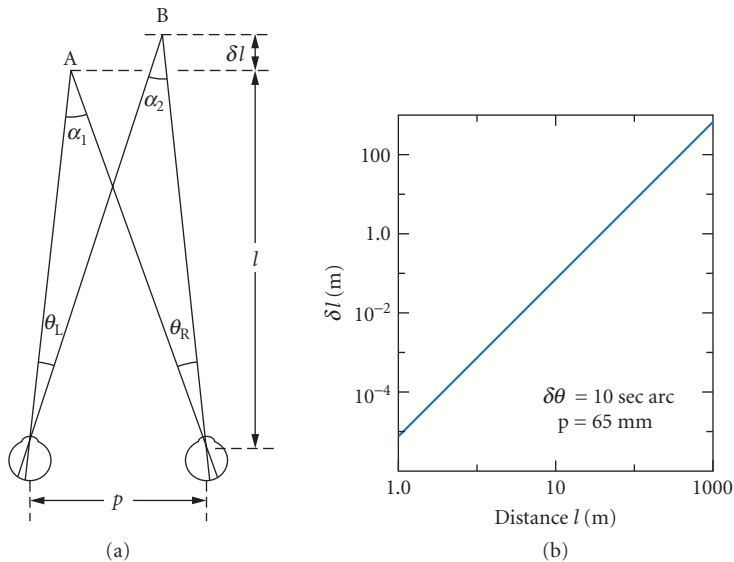
Approximating all angles as being small [i.e.,  $l \gg \delta l, p$ , where  $p$  is the lateral separation of the nodal points of the two eyes, or interpupillary distance (IPD)], and using the binomial expansion with omission of higher-order terms in  $\delta l$  yields:

$$\delta\theta = \frac{p \cdot \delta l}{l^2} \quad \text{or} \quad \delta l = \frac{l^2 \cdot \delta\theta}{p} \quad (20)$$

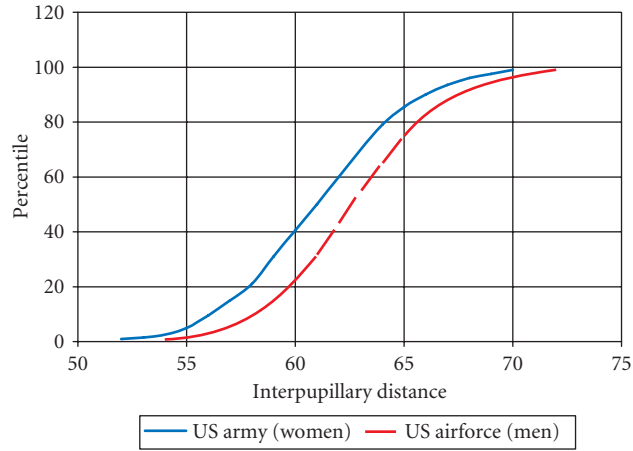
where  $\delta\theta$  is in radians. Thus the minimum detectable difference in object distance is directly proportional to the square of the viewing distance and inversely proportional to the separation between the eyes (see, e.g., Schor and Flom<sup>498</sup> for a more detailed analysis).

Figure 26*b* plots this approximate predicted value of the just-detectable difference in distance as a function of the distance  $l$ , on the assumption that  $p = 65$  mm and  $\delta\theta = 10$  sec arc. Note that discrimination of depth becomes very poor at distances in excess of about 500 m.

The interpupillary distance (IPD) varies somewhat between individuals and population groups.<sup>499</sup> Typical distributions for men and women are illustrated in Fig. 27. Values range between about 50 and 76 mm. In the real world, of course, binocular cues to distance are supplemented by



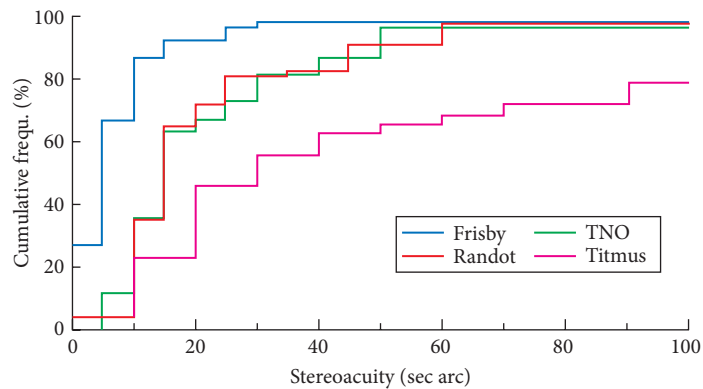
**FIGURE 26** (a) Geometry of stereopsis. It is assumed that points A and B can just be discriminated in depth. (b) Theoretical just discriminable distance  $\delta l$  as a function of the object distance  $l$  for the assumed values of  $p$  and  $\delta\theta$  indicated.



**FIGURE 27** Typical cumulative frequency distributions of interpupillary distance for men and women.<sup>499</sup>

a variety of monocular cues such as perspective, overlay (interposition), size constancy, texture, motion parallax, etc.<sup>500</sup>

For any observer,  $\delta\theta$  varies with such parameters as the target luminance and angular distance from fixation, and the observation time allowed (e.g., Refs. 501–504), being optimal close to fixation, with high luminance and extended observation times.<sup>505</sup> Values also vary with the nature of the task involved. Clinical tests of stereoacuity (e.g., Refs. 506 and 507) which are usually carried out at a distance of 40 cm and are calibrated for an assumed IPD,  $p$ , of 65 mm, typically yield normal stereoacuities of about 20 to 40 sec arc (see Fig. 28).<sup>508</sup> Two- or three-needle<sup>509</sup> or similar tests carried out at longer distances usually give rather smaller values, of around 5 to 10 sec arc.



**FIGURE 28** Cumulative frequency distribution for stereoscopic acuity, as measured by various clinical tests at a viewing distance of 400 mm, based on a sample of 51 adult subjects with normal binocular vision.<sup>508</sup>

## Stereoscopic and Related Instruments

The stereoscopic acuity of natural viewing can be enhanced both by extending the effective IPD from  $p$  to  $Np$  with the aid of mirrors or prisms and by introducing transverse magnification  $M$  into the optical paths before each eye. This nominally has the effect of changing the just-detectable distance to

$$\delta l = \frac{l^2 \cdot \delta\theta}{MNp} \quad (21)$$

although in practice this improvement in performance is not always fully realized. Such changes will in general vary the spatial relationships in the perceived image, so that the object appears as having either enhanced or reduced depth in proportion to its lateral scale. If, for example, the magnification  $M$  is greater than one but  $N$  is unity, the object appears nearer but foreshortened. Simple geometrical predictions of such effects are, however, complicated by a variety of factors such as the reduced depth-of-field of magnifying system (see, e.g., Refs. 501, 510–512).

As a result of the range of IPD values encountered among different potential users (Fig. 27), it is important that adequate adjustment of IPD be provided (preferably covering 46 to 78 mm), with a scale so that users can set their own IPD. Convergence should be appropriate to the distance at which the image is viewed, that is, the angle between the visual axes should be approximately 3.7 deg for each diopter of accommodation exercised.<sup>513</sup>

## Tolerances in Binocular Instrumentation and the Problem of Aniseikonia

In the foregoing, it has been assumed that the images available to the two eyes can be successfully fused to yield full binocular vision. Although this does not demand exact alignment of the two images, since, for example, horizontal misalignment can be compensated for by appropriate convergence or divergence between the visual axes of the user's eyes, such compensation is only possible over a limited range. Several authors have suggested instrumental tolerances appropriate to different circumstances (particularly duration of use) for vertical misalignment, convergence error, divergence error, magnification difference, and rotation in the images presented to the eyes by binocular instruments (see, e.g., Refs. 513–515 for reviews). Recommended tolerances on divergence and vertical misalignment (usually around 3 to 10 min arc) are smaller than those on convergence (around 10 to 20 min arc). These tolerances relate to the eyes: evidently if, for example, an afocal binocular instrument provides lateral magnification, the tolerances on the alignment of the tubes of the right- and left-eye optical systems will be reduced accordingly. Cyclofusion to compensate for relative image rotation is difficult to maintain on a sustained basis: any tolerance is independent of magnification.

Particular interest attaches to the possibility of a magnification difference between the two images (*aniseikonia*).<sup>516</sup> In practice, such magnification differences are probably most likely to occur as a result of refractive spectacle corrections in patients with markedly different refractions in the two eyes (*anisometropia*).<sup>517</sup> Strictly speaking we are not concerned with the retinal images as such but with the corresponding perceived images, since it is possible that these may differ in size as a result of neural as well as purely optical factors. It is conventional to express the relevant magnification differences in percentage terms. As a rough guide, magnification differences less than 1 percent usually present no problems and those greater than 5 percent often make it difficult to maintain fusion, leading to diplopia or suppression. It is the range between about 1 and 5 percent of perceived size difference where disturbing problems in spatial perception may arise but binocular vision is still possible. These may in turn lead to symptoms such as eyestrain, nausea, and headaches.

In eyes demanding a correction for both spherical and astigmatic errors, size differences may vary in different meridians, depending upon the axes of the correcting cylinders. The spatial distortions resulting from horizontal and vertical size differences are different in nature but, in general, objects may appear tilted and unpleasant instabilities in the perception of space may arise as the head or eyes turn. To give a rough estimate of the conditions under which significant aniseikonia may occur, we make use of the crude approximation that the spectacle magnification, expressed in percentage terms

is about  $100aF$ , where  $a$  mm is the distance of the spectacle lens from the eye and  $F$  diopter is its power. Thus the spectacle magnification is around 1.5 percent per diopter (positive for converging lenses, negative for diverging lenses). Thus, with a spectacle correction, problems may begin to arise even with quite modest degrees of anisometropia: these are reduced with contact lenses,<sup>518</sup> which give much lower spectacle magnification since  $a \approx 0$ . Even for ametropic patients without significant anisometropia, size differences will arise if one eye is corrected with, for example, corneal refractive surgery and the other with a spectacle lens. Similarly, correction of one eye with an intraocular lens following cataract surgery while the other eye remains spectacle corrected may also cause difficulties.<sup>516</sup> Methods for measuring aniseikonia and for its control are discussed in Refs. 516 and 517; fortunately many patients adapt to modest image size differences.

## 1.10 MOVEMENTS OF THE EYES

Movements of the eyes are of significance in relation to visual optics from two major points of view. First, since under photopic conditions both optical and neural performance are optimal on the visual axis, the eye movement system must be capable of rapidly directing the eyes so that the images of the detail of interest fall on the central foveas of both eyes, where visual acuity is highest (*gaze shifting* leading to *fixation*). A scene is explored through a series of such fixational movements (*saccades*) between different points within the field.<sup>519,520</sup> Second, the system must be capable of maintaining the images on the two foveas both when the object is fixed in space (*gaze holding*) and, ideally, when it is moving. Any lateral movement of the images with respect to the retina is likely to result in degraded visual performance, due to the limited temporal resolution of the visual system and the falloff in acuity with distance from the central fovea.

These challenges to the eye movement control system are further complicated by the fact that the eyes are mounted in what is, in general, a moving rather than a stationary head. Movements of the eyes therefore need to be linked to information derived from the vestibular system or labyrinth of the inner ear, which signals rotational and translational accelerations of the head. The compensatory *vestibulo-ocular responses* take place automatically (i.e., they are reflex movements) whereas the fixational changes required to foveate a new object are voluntary responses. Details of the subtle physiological mechanisms which have evolved to meet these requirements will be found in Refs. 521–524.

### Basics

Each eye is moved in its orbit by the action of three pairs of extraocular muscles attached to the outside of the globe. Their action rotates the eye about an approximate center of rotation lying some 13.5 mm behind the cornea, although there is in fact no unique fixed point within the eye or orbit around which the eye can rotate to all possible positions that it can assume.<sup>521,523</sup> In particular, the “center of rotation” for vertical eye movements lies about 2 mm nearer the cornea than that for horizontal eye movements (mean values 12.3 mm and 14.9 mm, respectively).<sup>525</sup>

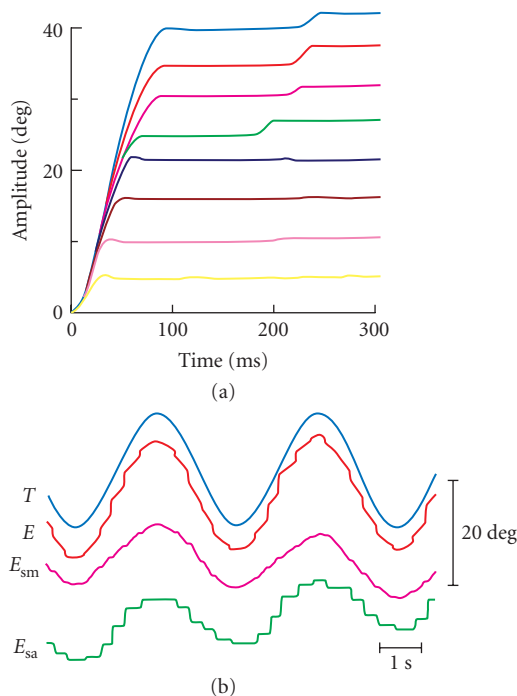
Although the two eyes can scan a field extending about 45 deg in all directions from the straight-ahead or *primary* position, in practice eye movements rarely exceed about 20 deg, fixation on more peripheral objects usually being achieved by a combination of head and eye movements.

If the angle between the two visual axes does not change during the movement, the latter is described as a *version* (or conjugate) movement. However, the lateral separation of the eyes in the head implies the need for an additional class of movements to cope with the differing convergence requirements of objects at different distances. These movements, which involve a change in the angle between the visual axes, are called *vergence* (or disjunctive) movements. Fixational changes may in general involve both types of movement, which appear to be under independent neurological control.<sup>526</sup>

## Characteristics of the Movements

The version movements involved in bringing a new part of the visual field onto the fovea (saccades) are very rapid, being completed in around 100 ms with angular velocities often reaching more than 700 deg/s, depending upon the amplitude of the movement<sup>527,528</sup> (see Fig. 29a): the saccadic latency is about 200 ms.<sup>529</sup> Note that the larger saccades tend initially to be inaccurate. Interestingly, it appears that during the saccade, when the image is moving very rapidly across the retina, vision is largely, although not completely, suppressed. This saccadic suppression (or perhaps, more appropriately, saccadic attenuation) results in, for example, the retinal thresholds for brief flashes of light being elevated, the elevation commencing some 30 to 40 ms before the actual saccadic movement starts. The subject is normally unaware of this temporary impairment of vision and the exact mechanisms responsible for it remain controversial,<sup>522,529</sup> although an explanation may lie in the masking effect of the clear images available before and after the saccade.<sup>530</sup>

Smooth voluntary pursuit movements of small targets which are moving sinusoidally in a horizontal direction are accurate at temporal frequencies up to a few hertz. Their peak velocities range up to about 25 deg/s. In practice, when a small moving target is tracked, the following movements usually consist of a mixture of smooth movements and additional corrective saccades (e.g., Ref. 531; see Fig. 29b). With repetitive or other predictable stimuli, tracking accuracy tends to improve markedly with experience, largely through the reduction of the phase lag between target and eye.



**FIGURE 29** (a) Time course of saccades of different sizes. (After Robinson.<sup>527</sup>) The traces have been superimposed so that the beginning of each saccade occurs at time zero. (b) Separation of smooth ( $E_{sm}$ ) and saccadic ( $E_{sa}$ ) components of foveal tracking. (After Collewijn and Tamminga.<sup>531</sup>)  $T$  is the target position and  $E$  is the eye position.



When the image of a substantial part of the visual field moves uniformly across the retina, the eyes tend to rotate in a following movement to approximately stabilize the retinal image, until eventually the gaze is carried too far from the primary position, when a rapid anticompensatory flick-back occurs. Thus a continuously moving visual scene, such as the view from a train, causes a series of slow following movements and fast recoveries, so that a record of the eye movements takes a quasi-regular sawtooth form. At slow object speeds, the angular velocity of the slow phase is close to that of the field, but as the field velocity rises, greater lags in eye velocity occur, until at about 100 deg/s the following movements break down. Although this *optokinetic nystagmus* is basically a reflex, it can be influenced by any instructions given to the subject.

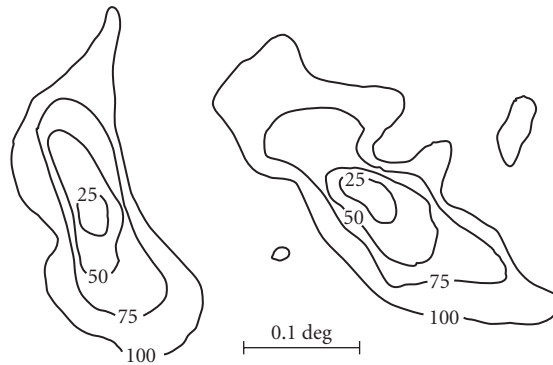
The angular speeds of the vergence eye movements (about 5 to 10 deg/s per degree of vergence<sup>532</sup>) are usually described as being much lower than those for version movements. However, studies with more natural 3-D targets suggest that it is possible for speeds to be much higher (up to 200 deg/s for 35-deg vergence movements<sup>533,534</sup>). Vergence movements typically follow a quasi-exponential course with a time constant of about 300 ms. Their latency is about 200 ms. Although the primary stimulus for vergence movements is disparity, that is, the difference in position of the images in the two eyes with respect to their foveas, vergence movements can also be driven by accommodation (see “The Accommodation Response”) and by perceptual factors such as perspective in line drawings.<sup>535</sup>

The vergence system is primarily designed to cope with the differences in the directions of the visual axes of the two eyes which arise from their horizontal separation in the head. However, some vertical vergence movements and relative torsional movements about the approximate direction of the visual axis can also occur. Such fusional movements can compensate for corresponding small relative angular misalignments in the eyes or in the optical systems of binocular or biocular instruments. The maximum amplitude of these movements is normally small (about 1 deg for vertical movements and a few degrees for cyclofusional movements, see, e.g., Ref. 336). They also take longer to complete than horizontal vergence movements (some 8 to 10 s as compared to about 1 s). Due to the limited effectiveness of vertical and torsional vergence movements, it is typically recommended that in instruments where the two eyes view an image through separate optical systems a 10 min arc tolerance be set for vertical misalignment and a 30 min arc limit be set for rotational differences between the two images (see “Tolerances in Binocular Instrumentation”).

## Stability of Fixation

When an observer attempts to maintain fixation on a stationary target, it is found that the eyes are not steady, but that a variety of small-amplitude movements occur. These *miniature eye movements* can be broken down into three basic components: tremor, drift, and microsaccades. The frequency spectrum of the tremor falls essentially linearly with the logarithm of the frequency above 10 Hz, extending to about 200 Hz.<sup>537</sup> The amplitude is small (probably less than the angle subtended at the nodal point of the eye by the diameter of the smallest foveal cones, i.e., about 24 sec arc). Drift movements are much larger and slower, with amplitudes of about 2 to 5 min arc at velocities around 4 min/s.<sup>538</sup> The errors in fixation brought about by the slow drifts (which are usually disassociated in the two eyes) are corrected microsaccades: these are correlated in the two eyes. There are large intersubject differences in both mean microsaccade amplitude (from 1 to 23 min arc) and intersaccadic interval (from about 300 ms to 5 s),<sup>539</sup> which probably reflect voluntary control under experimental conditions.

The overall stability of fixation can be illustrated by considering the statistical variation in the point of fixation (Fig. 30).<sup>540,541</sup> For most of the time the point of regard lies within a few minutes of arc of the target. Although it has been suggested that these small eye movements could have some role in reducing potential aliasing problems,<sup>542</sup> experiments measuring contrast sensitivity for briefly presented interference fringes on the retina suggest that this is unlikely.<sup>543</sup> Interestingly, when a suitable optical arrangement is used to counteract these small changes in fixation and stabilize the image on the retina, the visual image may fragment or even disappear completely,<sup>538</sup> so that these small movements are important for normal visual perception. Fincham’s suggestion<sup>381</sup> that small eye movements are of importance to the ability of the eye to respond correctly to a change in accommodation stimulus has never been properly explored.



**FIGURE 30** Stability of fixation for two subjects. The contours define areas within which the point of fixation was to be found for 25, 50, 75, and 100 percent of the time. (After Bennet-Clark.<sup>541</sup>)

## 1.11 CONCLUSION

Recent years have seen considerable advances in our understanding of the aberrations of the eye and their dependence on factors such as pupil size, age, and field angle. The refractive index distribution of the lens remains to be fully elucidated, although substantial progress has been made. In general, it appears that the optical characteristics of the eye are such that the natural retinal image quality under different conditions of lighting, field, and age is well matched to the corresponding needs of the neural parts of the visual system. Those aberrations which are present may well be useful in such roles as guidance of the accommodation response, expansion of the depth-of-focus, and control of aliasing.

## 1.12 REFERENCES

1. M. Rynders, B. Lidkea, W. Chisholm, and L. N. Thibos, "Statistical Distribution of Foveal Transverse Aberration, Pupil Centration and Angle  $\psi$  in a Population of Young Adult Eyes," *J. Opt. Soc. Am. A* **12**:2348–2357 (1995).
2. J. Tabernero, A. Benito, E. Alc3n, and P. Artal, "Mechanism of Compensation of Aberrations in the Human Eye," *J. Opt. Soc. Am. A* **24**:3274–3283 (2007).
3. S. Guidarelli, "Off-axis Imaging in the Human Eye," *Atti d. Fond. G. Ronchi* **27**:449–460 (1972).
4. J. S. Larsen, "The Sagittal Growth of the Eye IV: Ultrasonic Measurements of the Axial Growth of the Eye from Birth to Puberty," *Acta Ophthalmologica* **49**:873–886 (1971).
5. S. J. Isenberg, D. Neumann, P. Y. Cheong, L. Ling, L. C. McCall and A. J. Ziffer, "Growth of the Internal and External Eye in Term and Preterm Infants," *Ophthalmology* **102**:827–830 (1995).
6. F. C. Pennie, I. C. J. Wood, C. Olsen, S. White, and W. N. Charman, "A Longitudinal Study of the Biometric and Refractive Changes in Full-Term Infants During the First Year of Life," *Vision Res.* **41**:2799–2810 (2001).
7. A. Sorsby, B. Benjamin, M. Sheridan, and J. M. Tanner, "Emmetropia and Its Aberrations," *Spec. Rep. Ser. Med. Res. Coun.* no. 293, HMSO, London, 1957.
8. A. Sorsby, B. Benjamin, and M. Sheridan, "Refraction and Its Components during Growth of the Eye from the Age of Three," *Spec. Rep. Ser. Med. Res. Coun.* no. 301, HMSO, London, 1961.

9. A. Sorsby and G. A. Leary, "A Longitudinal Study of Refraction and Its Components During Growth," *Spec. Rep. Ser. Med. Res. Coun.* no. 309, HMSO, London, 1970.
10. R. A. Weale, *A Biography of the Eye*, H. K. Lewis, London, 1982, pp. 94–120.
11. T. Grosvenor, "Reduction in Axial Length with Age: an Emmetropizing Mechanism for the Adult Eye?" *Am. J. Optom. Physiol. Opt.* **64**:657–663 (1987).
12. F. A. Young and G. A. Leary, "Refractive Error in Relation to the Development of the Eye," *Vision and Dysfunction, Vol 1, Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 29–44.
13. J. F. Koretz, P. C. Kaufman, M. W. Neider, and P. A. Geckner, "Accommodation and Presbyopia in the Human Eye—Aging of the Anterior Segment," *Vision Res.* **29**:1685–1692 (1989).
14. A. Glasser, M. A. Croft, and P. L. Kaufman, "Aging of the Human Crystalline Lens," *Int. Ophthalmol. Clin.* **41**(2):1–15 (2001).
15. J. E. Koretz, C. A. Cook, and P. L. Kaufman, "Aging of the Human Lens: Changes in Lens Shape upon Accommodation and with Accommodative Loss," *J. Opt. Soc. Am. A* **19**:144–151 (2002).
16. B. K. Pierscionek, D. Y. C. Chan, J.P. Ennis, G. Smith, and R. C. Augusteyn. "Non-destructive Method of Constructing Three-dimensional Gradient Index Models for Crystalline Lenses: 1 Theory and Experiment," *Am. J. Optom. Physiol. Opt.* **65**:481–491 (1988).
17. B. K. Pierscionek, "Variations in Refractive Index and Absorbance of 670 nm Light with Age and Cataract Formation in the Human Lens," *Exp. Eye Res.* **60**:407–414 (1995).
18. C. E. Jones, D. A. Atchison, R. Meder, and J. M. Pope, "Refractive Index Distribution and Optical Properties of the Isolated Human Lens Measured Using Magnetic Resonance Imaging," *Vision Res.* **45**:2352–2366 (2005).
19. C. E. Jones, D. A. Atchison, and J. M. Pope, "Changes in Lens Dimensions and Refractive Index with Age and Accommodation," *Optom. Vis. Sci.* **84**:990–995 (2007).
20. A. Steiger, *Die Entstehung der sphaerischen Refraktionen des menschlichen Auges*, Karger, Berlin, 1913.
21. S. Stenström, "Untersuchungen über die Variation und Kovariation der optischen Elements des menschlichen Auges," *Acta Ophthalmologica* suppl. 26 (1946) [Translated by D. Woolf as: "Investigation of the Variation and Covariation of the Optical Elements of Human Eyes," *Am. J. Optom. Arch. Am. Acad. Optom.* **25**:218–232; 286–299; 340–350; 388–397; 438–449; 496–504 (1948)].
22. P. M. Kiely, G. Smith, and L. G. Carney, "The Mean Shape of the Human Cornea," *Optica Acta* **29**:1027–1040 (1982).
23. M. Guillon, D. P. M. Lydon, and C. Wilson, "Corneal Topography: a Clinical Model," *Ophthal. Physiol. Opt.* **6**:47–56 (1986).
24. D. A. Atchison and G. Smith, "Continuous Gradient Index and Shell Models of the Human Lens," *Vision Res.* **35**:2529–2538 (1995).
25. L. F. Garner and G. Smith, "Changes in Equivalent and Gradient Refractive Index of the Crystalline Lens with Accommodation," *Optom. Vis. Sci.* **74**:114–119 (1997).
26. G. Smith and B. K. Pierscionek, "The Optical Structure of the Lens and its Contribution to the Refractive Status of the Eye," *Ophthal. Physiol. Opt.* **18**:21–29 (1998).
27. H. T. Kasprzak, "New Approximation for the Whole Profile of the Human Lens," *Ophthal. Physiol. Opt.* **20**:31–43 (2000).
28. R. Navarro, F. Palos, and L. M. González, "Adaptive Model of the Gradient Index of the Human Lens. I. Formulation and Model of Aging *ex vivo* Lenses," *J. Opt. Soc. Am. A* **24**:2175–2185 (2007).
29. R. Navarro, F. Palos, and L. M. González, "Adaptive Model of the Gradient Index of the Human Lens. II. Optics of the Accommodating Aging Lens," *J. Opt. Soc. Am. A* **24**:2911–2920 (2007).
30. J. A. Diaz, C. Pizarro, and J. Arasa, "Single Dispersive Gradient-index Profile for the Aging Human Lens," *J. Opt. Soc. Am. A* **25**:250–261 (2008).
31. P. Rosales and S. Marcos, "Phakometry and Lens Tilt and Decentration using a Custom-developed Purkinje Imaging Apparatus: Validation and Measurements," *J. Opt. Soc. Am. A* **23**:509–520 (2006).
32. C. F. Wildsoet, "Structural Correlates of Myopia," *Myopia and Nearwork*, M. Rosenfield and B. Gilmartin (eds.), Oxford, Butterworth-Heinemann, 1998, pp. 31–56.
33. C. F. Wildsoet, "Active Emmetropization—Evidence for its Existence and Ramifications for Clinical Practice," *Ophthal. Physiol. Opt.* **17**:279–290 (1997).

34. W. M. Lyle, "Changes in Corneal Astigmatism with Age," *Am. J. Optom. Arch. Am. Acad. Optom.* **48**:467–478 (1971).
35. R. J. Farrell and J. M. Booth, *Design Handbook for Imagery Interpretation Equipment*, Boeing Aerospace Co., Seattle, Sec. 3. 2, p. 8 (1984).
36. B. Winn, D. Whitaker, D. Elliott, and N. J. Phillips, "Factors Affecting Light-adapted Pupil Size in Normal Human Subjects," *Invest. Ophthalmol. Vis. Sci.* **35**:1132–1137 (1994).
37. I. E. Loewenfeld, *The Pupil: Anatomy, Physiology and Clinical Applications*, vols. I and II, Butterworth-Heinemann, Oxford, 1999.
38. P. Reeves, "Rate of Pupillary Dilation and Contraction," *Psychol. Rev.* **25**:330–340 (1918).
39. R. A. Weale, *Focus on Vision*, Hodder and Stoughton, London, 1982, p. 133.
40. G. Walsh, "The Effect of Mydriasis on the Pupillary Centration of the Human Eye," *Ophthalm. Physiol. Opt.* **8**:178–182 (1988).
41. M. A. Wilson, M. C. W. Campbell, and P. Simonet, "Change of Pupil Centration with Change of Illumination and Pupil Size," *Optom. Vis. Sci.* **69**:129–136 (1992).
42. J. M. Woodhouse, "The Effect of Pupil Size on Grating Detection at Various Contrast Levels," *Vision Res.* **15**:645–648 (1975).
43. J. M. Woodhouse and F. W. Campbell "The Role of the Pupil Light Reflex in Aiding Adaptation to the Dark" *Vision Res.* **15**:649–65 (1975).
44. G. Wyszecki and W. S. Stiles, *Colour Science*, Wiley, New York, 1967, pp. 214–219
45. E. Ludvigh and E. F. McCarthy, "Absorption of Visible Light by the Refractive Media of the Human Eye," *Arch. Ophthalmol.* **20**:37–51 (1938).
46. E. A. Boerttner and J. R. Wolter, "Transmission of Ocular Media," *Invest. Ophthalmol.* **1**:776–783 (1962).
47. W. J. Geeraets and E. R. Berry, "Ocular Spectral Characteristics as Related to Hazards for Lasers and Other Sources," *Am. J. Ophthalmol.* **66**:15–20 (1968).
48. T. J. T. P. van den Berg and H. Spekreijse, "Near Infrared Light Absorption in the Human Eye Media," *Vision Res.* **37**:249–253 (1997).
49. G. L. Savage, C. A. Johnson, and D. I. Howard, "A Comparison of Noninvasive Objective and Subjective Measurements of the Optical density of Human Ocular Media," *Optom. Vis. Sci.* **78**:386–395 (2001).
50. D. G. Pitts, "The Ocular Effects of Ultraviolet Radiation," *Am. J. Optom. Physiol. Opt.* **55**:19–53 (1978).
51. J. Pokorny, V. C. Smith, and M. Lutze, "Aging of the Human Lens," *Applied Opt.* **26**:1437–1440 (1987).
52. P. A. Sample, F. D. Esterson, R. N. Weinreb, and R. M. Boynton, "The Aging Lens: *In vivo* Assessment of Light Absorption in 84 Human Eyes," *Invest. Ophthalmol. Vis. Sci.* **29**:1306–1311 (1988).
53. R. A. Weale, "Age and the Transmittance of the Human Crystalline Lens," *J. Physiol. (London)* **395**:577–587 (1988).
54. N. P. A. Zagers and D. van Norren, "Absorption of the Eye Lens and Macular Pigment Derived from the Reflectance of Cone Receptors," *J. Opt. Soc. Am. A* **21**:2257–2268 (2004).
55. S. Lerman, *Radiant Energy and the Eye*, Balliere Tindall, London, 1980.
56. K. Knoblauch, F. Saunders, M. Kasuda, R. Hynes, M. Podgor, K. E. Higgins, and F. M. de Monasterio, "Age and Illuminance Effects in the Farnsworth-Munsell 100-hue Test," *Applied Opt.* **26**:1441–1448 (1987).
57. K. Sagawa and Y. Takahashi, "Spectral Luminous Efficiency as a Function of Age," *J. Opt. Soc. Am. A* **18**:2659–2667 (2001).
58. J. Mellerio, "Yellowing of the Human Lens: Nuclear and Cortical Contributions," *Vision Res.* **27**:1581–1587 (1987).
59. C. Schmidt, J. Schmidt, A. Wegener, and O. Hockwin, "Ultraviolet Radiation as a Risk Factor in Cataractogenesis," *Risk Factors in Cataract Development*, *Dev. Ophthalmol.* **17**:Karger, Basel, 1989, pp. 169–172.
60. J. Marshall, "Radiation and the Ageing eye," *Ophthalm. Physiol. Opt.* **5**:241–263 (1985).
61. A. Stanworth and E. J. Naylor, "The Measurement and Clinical Significance of the Haidinger Effect," *Trans. Ophthalmol. Soc. UK* **75**:67–79 (1955).
62. P. E. Kilbride, K. B. Alexander, M. Fishman, and G. A. Fishman. "Human Macular Pigment Assessed by Imaging Fundus Reflectometry," *Vision Res.* **26**:663–674 (1989).

63. D. M. Snodderly, J. D. Auran, and F. C. Delori, "The Macular Pigment. II Spatial Distribution in Primate Retinas," *Invest. Ophthalmol. Vis. Sci.* **5**:674–685 (1984).
64. B. R. Hammond, B. R. Wooten, and D. M. Snodderly, "Individual Variations in the Spatial Profile of Human Macular Pigment," *J. Opt. Soc. Am. A* **14**:1187–1196 (1997).
65. V. M. Reading and R. A. Weale, "Macular Pigment and Chromatic Aberration," *J. Opt. Soc. Am.* **64**:231–234 (1974).
66. G. Haegerstrom-Portnoy, "Short-Wavelength Cone Sensitivity Loss with Aging: A Protective Role for Macular Pigment?," *J. Opt. Soc. Am. A* **5**:2140–2145 (1988).
67. L. J. Bour, "Polarized Light and the Eye," *Vision and Visual Dysfunction*, vol. 1, *Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 310–325.
68. W. S. Stiles and B. H. Crawford, "The Luminous Efficiency of Rays Entering the Eye Pupil at Different Points," *Proc. Roy. Soc. London B* **112**:428–450 (1933).
69. J. A. van Loo and J. M. Enoch, "The Scotopic Stiles-Crawford Effect," *Vision Res.* **15**:1005–1009 (1975).
70. J. M. Enoch and H. E. Bedell, "The Stiles-Crawford Effects," *Vertebrate Photoreceptor Optics*, J. Enoch and F. L. Tobey (eds.), Springer-Verlag, Berlin, 1981, pp. 83–126.
71. J. M. Enoch and V. Lakshminarayanan, "Retinal Fibre Optics," *Vision and Visual Dysfunction*, vol. 1, *Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 280–309.
72. W. S. Stiles, "The Luminous Efficiency of Monochromatic Rays Entering the Eye Pupil at Different Points and a New Colour Effect," *Proc. R. Soc. London B* **123**:90–118 (1937).
73. P. Moon and D. E. Spencer, "On the Stiles-Crawford Effect," *J. Opt. Soc. Am.* **34**:319–329 (1944).
74. M. Alpern, "The Stiles-Crawford Effect of the Second Kind (SCII): A Review," *Perception* **15**:785–799 (1986).
75. D. van Norren and L. F. Tiemeijer, "Spectral Reflectance of the Human Eye," *Vision Res.* **26**:313–330 (1986).
76. F. C. Delori and K. P. Pflibsen, "Spectral Reflectance of the Human Fundus," *Applied Opt.* **28**:1061–1077 (1989).
77. A. Elsner, S. A. Burns, J. J. Weiter, and F. C. Delori, "Infrared Imaging of Sub-retinal Structures in the Human Ocular Fundus," *Vision Res.* **36**:191–205 (1996).
78. F. C. Delori and S. A. Burns, "Fundus Reflectance and the Measurement of Crystalline Lens Density," *J. Opt. Soc. Am. A* **13**:215–226 (1996).
79. F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *J. Physiol (London)* **186**:558–578 (1966).
80. W. N. Charman and J. A. M. Jennings, "Objective Measurements of the Longitudinal Chromatic Aberration of the Human Eye," *Vision Res.* **16**:999–1005 (1976).
81. J. van de Kraats, T. T. J. M. Berenschot, and D. van Norren, "The Pathways of Light Measured in Fundus Reflectometry," *Vision Res.* **36**:2229–2247 (1996).
82. P. J. Delint, T. T. J. M. Berenschot, and D. van Norren, "Local Photoreceptor Alignment Measured with a Scanning Laser Ophthalmoscope," *Vision Res.* **37**:243–248 (1997).
83. J. C. He, S. Marcos, and S. A. Burns, "Comparison of Cone Directionality Determined by Psychophysical and Reflectometric Techniques," *J. Opt. Soc. Am. A* **16**:2363–2369 (1999).
84. N. P. A. Zagers, J. van de Kraats, T. T. J. M. Berenschot, and D. van Norren, "Simultaneous Measurement of Foveal Spectral Reflectance and Cone-photoreceptor Directionality," *Applied Opt.* **41**:4686–4696 (2002).
85. S. S. Choi, N. Doble, J. Lin, J. Christou, and D. R. Williams, "Effect of Wavelength on *in vivo* Images of the Human Cone Mosaic," *J. Opt. Soc. Am. A* **22**:2598–2605 (2005).
86. B. S. Jay, "The Effective Pupillary Area at Varying Perimetric Angles," *Vision Res.* **1**:418–428. (1962).
87. D. A. Atchison and G. Smith, *Optics of the Human Eye*, Butterworth-Heinemann, Oxford, 2000, pp. 25–27.
88. H. E. Bedell and L. M. Katz, "On the Necessity of Correcting Peripheral Target Luminance for Pupillary Area," *Am. J. Optom. Physiol. Opt.* **59**:767–769 (1982).
89. W. N. Charman, "Light on the Peripheral Retina," *Ophthal. Physiol. Opt.* **9**:91–92 (1989).
90. K. P. Pflibsen, O. Pomarentzeff, and R. N. Ross, "Retinal Illuminance Using a Wide-angle Model of the Eye," *J. Opt. Soc. Am. A* **5**:146–150 (1988).
91. A. C. Kooijman and F. K. Witmer, "Ganzfeld Light Distribution on the Retina of Humans and Rabbit Eyes: Calculations and *in vitro* Measurements," *J. Opt. Soc. Am. A* **3**:2116–2120 (1986).

92. W. D. Wright, *Photometry and the Eye*, Hatton Press, London, 1949.
93. G. Airy, "On the Diffraction of an Object Glass with Circular Aperture," *Trans. Camb. Philos. Soc.* **5**:283–291 (1835).
94. G. M. Byram, "The Physical and Photochemical Basis of Resolving Power. I. The Distribution of Illumination in Retinal Images," *J. Opt. Soc. Am.* **34**:571–591 (1944).
95. U. Hallden, "Diffraction and Visual Resolution. I. The Resolution of Two Point Sources of Light," *Acta Ophthalmol.* **51**:72–79 (1973).
96. L. A. Riggs, "Visual Acuity," *Vision and Visual Perception*, C. H. Graham (ed.), Wiley, New York, 1966, pp. 321–349.
97. E. Lommel, "Die Beugungerscheinungen einer Kreisrunden Oeffnung und eines runden Schirmchens," *Abh der K. Bayer Akad. D. Wissenschaft* **15**:229–328 (1884).
98. E. H. Linfoot and E. Wolf, "Phase Distribution Near Focus in an Aberration-Free Diffraction Image," *Proc. Phys. Soc. B* **69**:823–832 (1956).
99. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon, Oxford, 1993, pp. 435–449.
100. H. Struve, "Beitrag zur der Theorie der Diffraction an Fernrohren," *Ann. Physik. Chem.* **17**:1008–1016 (1882).
101. H. H. Hopkins, "The Frequency Response of a Defocused Optical System," *Proc. Roy. Soc. London A* **231**:91–103 (1955).
102. L. Levi, *Handbook of Tables for Applied Optics*, CRC Press, Cleveland, 1974.
103. W. N. Charman, "Effect of Refractive Error in Visual Tests with Sinusoidal Gratings," *Brit. J. Physiol. Opt.* **33**(2):10–20 (1979).
104. G. E. Legge, K. T. Mullen, G. C. Woo, and F. W. Campbell, "Tolerance to Visual Defocus," *J. Opt. Soc. Am A* **4**:851–863 (1987).
105. G. Westheimer, "Pupil size and Visual Resolution," *Vision Res.* **4**:39–45 (1964).
106. F. W. Campbell and D. G. Green, "Optical and Retinal Factors Affecting Visual Resolution," *J. Physiol. (London)* **181**:576–593 (1965).
107. F. W. Campbell and R. W. Gubisch, "The effect of Chromatic Aberration on Visual Acuity," *J. Physiol. (London)* **192**:345–358 (1967).
108. W. N. Charman and J. Tucker, "Dependence of Accommodation Response on the Spatial Frequency Spectrum of the Observed Object," *Vision Res.* **17**:129–139 (1977).
109. W. N. Charman and J. Tucker, "Accommodation as a Function of Object Form," *Am. J. Optom. Physiol. Opt.* **55**:84–92 (1978).
110. W. N. Charman and J. A. M. Jennings, "The Optical Quality of the Retinal Image as a Function of Focus" *Br. J. Physiol. Opt.* **31**:119–134 (1976).
111. W. N. Charman and G. Heron, "Spatial Frequency and the Dynamics of the Accommodation Response," *Optica Acta* **26**:217–228 (1979).
112. H. H. Hopkins, "Geometrical Optical treatment of Frequency Response," *Proc. Phys. Soc. B* **70**:1162–1172 (1957).
113. G. A. Fry, "Blur of the Retinal Image" *Progress in Optics*, vol. 8, E. Wolf (ed.), North Holland, Amsterdam, 1970, pp. 53–131.
114. G. Smith, "Ocular defocus, Spurious Resolution and Contrast Reversal," *Ophthalm. Physiol. Opt.* **2**:5–23 (1982).
115. C. Chan, G. Smith, and R. J. Jacobs, "Simulating Refractive Errors: Source and Observer Methods," *Am. J. Optom. Physiol. Opt.* **62**:207–216 (1985).
116. S. Duke-Elder and D. Abrams, *System of Ophthalmology*, vol. V, *Ophthalmic Optics and Refraction*, Kimpton, London, 1970, pp. 134–139.
117. R. B. Rabbetts, *Clinical Visual Optics*, 3rd ed., Butterworth-Heinemann, Oxford, 1998, pp. 220–221.
118. N. Thibos and A. Bradley, "Modelling the Refractive and Neurosensor Systems of the Eye," P. Mouroulis, (ed.), *Visual Instrumentation: Optical design and Engineering Principles*, McGraw Hill, New York, 1999, pp. 101–159.
119. R. R. Krueger, R. A. Applegate, and S. M. Macrae, (eds.), *Wavefront Customized Visual Correction: the Quest for Supervision II*, Slack Inc., Thorofare, (2004).



120. D. A. Atchison, "Recent Advances in the Measurement of Monochromatic Aberrations in Human Eyes," *Clin. Exper. Optom.* **88**:5–26 (2005).
121. D. A. Atchison, "Recent Advances in the Representation of Monochromatic Aberrations of Human Eyes," *Clin. Exper. Optom.* **87**:138–148 (2004).
122. W. N. Charman, "Wavefront Technology: Past, Present and Future," *Contact Lens Ant. Eye* **28**:75–92 (2005).
123. L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb, "Standards for Reporting the Optical Aberrations of Eyes," *J. Refract. Surg.* **18**:S652–S660 (2002).
124. ANSI, American Standards Institute, *American National Standards for Ophthalmics—Methods for Reporting Optical Aberrations of Eyes*, ANSI Z. 80. 28–2004 (2004).
125. M. K. Smolek and S. D. Klyce, "Zernike Polynomial Fitting Fails to Represent All Visually Significant Corneal Aberrations," *Invest. Ophthalmol. Vis. Sci.* **44**:4676–4681 (2003).
126. S. D. Klyce, M. D. Karon, and M. K. Smolek, "Advantages and Disadvantages of the Zernike Expansion for Representing Wave Aberration of the Normal and Aberrated Eye," *J. Refract. Surg.* **20**:S537–541 (2004).
127. R. Iskander, M. J. Collins, B. Davis, and L. G. Carney, "Monochromatic Aberrations and Characteristics of Retinal Image Quality," *Clin. Exp. Optom.* **83**:315–322 (2000).
128. J. Porter, A. Guirao, I. G. Cox, and D. R. Williams, "The Human Eye's Monochromatic Aberrations in a Large Population," *J. Opt. Soc. Am. A* **18**:1793–1803 (2001).
129. L. N. Thibos, X. Hong, A. Bradley, and X. Cheng, "Statistical Variation of Aberration Structure and Image Quality in a Normal Population of Healthy Eyes," *J. Opt. Soc. Am. A* **19**:2329–2348 (2002).
130. J. F. Castejón-Mochón, N. Lopez-Gil, A. Benito, and P. Artal, "Ocular Wavefront Statistics in a Normal Young Population," *Vision Res.* **42**:1611–1617 (2002).
131. L. Wang and D. D. Koch, "Ocular Higher-Order Aberrations in Individuals Screened for Refractive Surgery," *J. Cataract Refract. Surg.* **29**:1896–1903 (2003).
132. N. V. Netto, R. Abrosio, T. T. Shen, and S. E. Wilson, "Wavefront Analysis in Normal Refractive Surgery Candidates," *J. Refract. Surg.* **21**:332–338 (2005).
133. T. O. Salmon and C. van de Pol, "Normal-eye Zernike Coefficients and Root-Mean-Square Wavefront Errors," *J. Cataract Refract. Surg.* **32**:2064–2074 (2006).
134. R. A. Applegate, W. J. Donnelly, J. D. Marsack, and D. E. Koenig, "Three-Dimensional Relationship between Higher-Order Root-Mean-Square Wavefront Error, Pupil Diameter and Aging," *J. Opt. Soc. Am. A* **24**:578–587 (2007).
135. R. I. Calver, M. J. Cox, and D. B. Elliott, "Effects of Aging on the Monochromatic Aberrations of the Human Eye," *J. Opt. Soc. Am. A* **16**:2069–2078 (1999).
136. J. S. McLellan, S. Marcos, and S. A. Burns, "Age-Related Changes in Monochromatic Wave Aberrations of the Human Eye," *Invest. Ophthalmol. Vis. Sci.* **42**:1390–1395 (2001).
137. T. Kuroda, T. Fujikado, S. Ninomiya, N. Maeda, Y. Hirohara, and T. Mihashi, "Effect of Aging on Ocular Light Scatter and Higher order Aberration," *J. Refract. Surg.* **18**:S598–S602 (2002).
138. I. Brunette, J. M. Bueno, M. Parent, H. Hamam, and P. Simonet, "Monochromatic Aberrations as a Function of Age, from Childhood to Advanced Age," *Invest. Ophthalmol. Vis. Sci.* **44**:5438–5446 (2003).
139. D. Atchison, M. J. Collins, C. F. Wildsoet, J. Christensen, and M. D. Waterworth, "Measurement of Monochromatic Ocular Aberrations of Human Eyes as a Function of Accommodation by the Howland Aberroscope Technique," *Vision Res.* **35**:313–323 (1995).
140. S. Ninomiya, T. Fujikado, T. Kuroda, N. Maeda, Y. Tano, T. Oshika, Y. Hirohara, and T. Mihashi, "Changes in Ocular Aberration with Accommodation," *Am. J. Ophthalmol.* **134**:924–926 (2002).
141. J. C. He, S. A. Burns, and S. Marcos, "Monochromatic Aberrations in the Accommodated Human Eye," *Vision Res.* **40**:41–48 (2000).
142. C. A. Hazel, M. J. Cox, and N. C. Strang, "Wavefront Aberration and Its Relationship to the Accommodative Stimulus-Response Function in Myopic Subjects," *Optom. Vis. Sci.* **80**:151–158 (2003).
143. S. Plainis, H. S. Ginis, and A. Pallikaris, "The Effect of Ocular Aberrations on Steady-State Errors of Accommodative Response," *J. Vision* **5**(5):466–477 (2005).
144. H. Cheng, J. K. Barnett, A. S. Vilupuru, J. D. Marsack, S. Kasthurirangan, R. A. Applegate, and A. Roorda, "A Population Study on Changes in Wavefront Aberration with Accommodation," *J. Vision* **4**(4):272–280 (2004).

145. H. Radhakrishnan and W. N. Charman, "Age-Related Changes in Ocular Aberrations with Accommodation," *J. Vision* **7**(7):1–21 (2007).
146. M. P. Paquin, H. Hamam, P. Simonet, "Objective Measurement of Optical Aberrations in Myopic Eyes," *Optom. Vis. Sci.* **79**:285–291 (2002).
147. X. Cheng, A. Bradley, X. Hong, and L. N. Thibos, "Relationship between Refractive Error and Monochromatic Aberrations of the Eye," *Optom. Vis. Sci.* **80**:43–49 (2003).
148. H. Hofer, P. Artal, B. Singer, J. Laragon, and D. R. Williams, "Dynamics of the Eye's Aberration," *J. Opt. Soc. Am. A* **18**:497–506 (2001).
149. L. Diaz-Santana, C. Torti, I. Munro, P. Gasson, and C. Dainty, "Benefit of Higher Closed-Loop Bandwidth in Ocular Adaptive Optics," *Opt. Express* **11**:2597–2605 (2003).
150. T. Nirmaier, G. Pudasaini, and J. Bille, "Very Fast Wavefront Measurements at the Human Eye with a Custom CMOS-based Hartmann-Shack Sensor," *Opt. Express* **11**:2704–2716. (2003).
151. M. Zhu, M. J. Collins, and R. Iskander, "Microfluctuations of Wavefront Aberration of the Eye," *Ophthalm. Physiol. Opt.* **24**:562–571 (2004).
152. K. M. Hampson, I. Munro, C. Paterson, and C. Dainty, "Weak Correlation between the Aberration Dynamics of the Human Eye and the Cardiopulmonary System," *J. Opt. Soc. Am. A* **22**:1241–1250 (2005).
153. A. Maréchal, "Etude des Effets Combinés de la Diffraction et des Aberrations Géométriques sur l'image d'un Point Lumineux," *Rev. d'Optique* **26**:257–277 (1947).
154. J. Perrigin, D. Perrigin, and T. Grosvenor, "A Comparison of Clinical Refractive Data Obtained by Three Examiners," *Am. J. Optom. Physiol. Opt.* **59**:515–519 (1982).
155. M. A. Bullimore, R. E. Fusaro, and C. W. Adams, "The Repeatability of Automated and Clinical Refraction," *Optom. Vis. Sci.* **75**:617–622 (1998).
156. R. A. Applegate, C. Ballentine, H. Gross, E. J. Sarver, and C. A. Sarver, "Visual Acuity as a Function of Zernike Mode and Level of Root Mean Square Error," *Optom. Vis. Sci.* **80**:97–105 (2003).
157. R. A. Applegate, J. D. Marsack, R. Ramos, and E. J. Sarver, "Interaction between Aberrations to Improve or Reduce Visual Performance," *J. Cataract Refract. Surg.* **29**:1487–1495 (2003).
158. S. G. El Hage and F. Berny, "Contribution of the Crystalline Lens to the Spherical Aberration of the Eye," *J. Opt. Soc. Am.* **63**:205–211 (1973).
159. P. Artal and A. Guirao, "Contribution of the Cornea and Lens to the Aberrations of the Human Eye," *Opt. Lett.* **23**:1713–1715 (1998).
160. P. Artal, A. Guirao, E. Berrio, and D. R. Williams, "Compensation of Corneal Aberrations by the Internal Optics in the Human Eye," *J. Vision* **1**:1–8 (2001).
161. M. Mrochen, M. Jankov, M. Bueeler, and T. Seiler, "Correlation between Corneal and Total Wavefront Aberrations in Myopic Eyes," *J. Refract. Surg.* **19**:104–112 (2003).
162. J. E. Kelly, T. Mihashi, and H. C. Howland, "Compensation of Cornea Horizontal/Vertical Astigmatism, Lateral Coma and Spherical Aberration by the Internal Optics of the Eye," *J. Vision* **4**:262–271 (2004).
163. J. C. He, J. Gwiwzda, F. Thorn, and R. Held, "Wave-front Aberrations in the Anterior Corneal Surface and the Whole Eye," *J. Opt. Soc. Am. A* **20**:1155–1163 (2003).
164. P. Artal, E. Berrio, and A. Guirao, "Contribution of the Cornea and Internal Surfaces to the Change of Ocular Aberrations with Age," *J. Opt. Soc. Am. A* **19**:137–143 (2002).
165. P. Artal, A. Benito, J. Tabernero, "The Human Eye is an Example of Robust Optical Design," *J. Vision* **6**:1–7 (2006).
166. P. Artal, L. Chen, E. J. Fernández, B. Singer, S. Manzanera, and D. R. Williams, "Neural Compensation for the Eye's Optical Aberrations," *J. Vision* **4**:281–287 (2004).
167. W. N. Charman, "Aberrations and Myopia," *Ophthalm. Physiol. Opt.* **25**:285–301 (2005).
168. L. Llorente, S. Barbero, D. Cano, C. Dorronsoro, and S. Marcos, "Myopic versus Hyperopic Eyes: Axial Length, Corneal Shape and Optical Aberrations," *J. Vision* **4**(4):288–298 (2004).
169. R. Montés-Mico, J. L. Alió, G. Muñoz, J. J. Pérez-Santonja, and W. N. Charman, "Postblink Changes in Total and Corneal Optical Aberrations," *Ophthalmology* **111**:758–767, (2004).
170. K. Y. Li and G. Yoon, "Changes in Aberrations and Retinal Image Quality due to Tear Film Dynamics," *Optics Express* **14**:12552–12559 (2006).



171. T. Buehren, M. J. Collins, and L. Carney, "Corneal Aberrations and Reading," *Optom. Vis. Sci.* **80**:159–166 (2003).
172. T. Buehren, M. J. Collins, and L. Carney, "Near Work Induced Wavefront Aberrations in Myopia," *Vision Res.* **45**:1297–1312 (2005).
173. W. Han, W. Kwan, J. Wang, S. P. Yip, and M. Yap, "Influence of Eyelid Position on Wavefront Aberration," *Ophthalm. Physiol. Opt.* **27**:66–75 (2007).
174. C. E. Ferree, G. Rand, and C. Hardy, "Refraction for the Peripheral Field of Vision," *Arch. Ophthalmol. (Chicago)* **5**:717–731 (1931).
175. T. C. A. Jenkins, "Aberrations of the Eye and Their Effects on Vision, Parts I and II" *Br. J. Physiol. Opt.* **20**:50–91 and 161–201 (1963).
176. F. Rempt, J. F. Hoogerheide, and W. P. H. Hoogenboom, "Peripheral Retinoscopy and the Skiagram," *Ophthalmologica* **162**:1–10 (1971).
177. J. F. Hoogerheide, F. Rempt, and W. P. H. Hoogenboom, "Acquired Myopia in Young Pilots," *Ophthalmologica* **163**:209–215 (1971).
178. M. Millodot, "Effect of Ametropia on Peripheral Refraction," *Am. J. Physiol. Opt.* **58**:691–695 (1981).
179. M. Millodot, "Peripheral Refraction in Aphakic Eyes," *Am. J. Optom. Physiol. Opt.* **61**:586–589 (1984).
180. D. R. Williams, P. Artal, R. Navarro, M. J. McMahon, and D. H. Brainard, "Off-Axis Optical Quality and Retinal Sampling in the Human Eye," *Vision Res.* **36**:1103–1114 (1996).
181. G. Smith, M. Millodot, and N. McBrien, "The Effect of Accommodation on Oblique Astigmatism and Field Curvature of the Human Eye," *Clin. Exp. Optom.* **71**:119–125 (1988).
182. A. Guirao and P. Artal, "Off-Axis Monochromatic Aberrations Estimated from Double-Pass Measurements in the Human Eye," *Vision Res.* **39**:207–217 (1999).
183. D. O. Mutti, R. I. Scholtz, N. E. Friedman, and K. Zadnik, "Peripheral Refraction and Ocular Shape in Children," *Invest. Ophthalmol. Vis. Sci.* **41**:1022–1030 (2000).
184. J. Gustafsson, E. Terenius, J. Buckheister, and P. Unsbo, "Peripheral Astigmatism in Emmetropic Eyes," *Ophthalm. Physiol. Opt.* **21**:393–400 and 491 (2001).
185. A. Seidemann, F. Schaeffel, A. Guirao, N. Lopez-Gil, and P. Artal, "Peripheral Refractive Errors in Myopic, Emmetropic, and Hyperopic Young Subjects," *J. Opt. Soc. Am. A* **19**:2363–2373 (2002).
186. D. A. Atchison, N. Pritchard, S. D. White, and A. M. Griffiths, "Influence of Age on Peripheral Refraction," *Vision Res.* **45**:715–720 (2005).
187. W. N. Charman and J. A. M. Jennings, "Longitudinal Changes in Peripheral Refraction with Age," *Ophthalm. Physiol. Opt.* **26**:447–455 (2006).
188. R. Navarro, E. Moreno, and C. Dorronsoro, "Monochromatic Aberrations and Point-Spread Functions of the Human Eye across the Visual Field," *J. Opt. Soc. Am. A* **15**:2522–2529 (1998).
189. D. A. Atchison and D. H. Scott, "Monochromatic Aberrations of Human Eyes in the Horizontal Visual Field," *J. Opt. Soc. Am. A* **19**:2180–2184 (2002).
190. D. A. Atchison, S. D. Lucas, R. Ashman, and M. A. Huynh, "Refraction and Aberration across the Horizontal Central 10° of the Visual Field," *Optom. Vis. Sci.* **83**:213–221 (2006).
191. D. A. Atchison, "Higher Order Aberrations Across the Horizontal Visual Field," *J. Biomed Optics* **11**(3): article 034026 (2006).
192. D. A. Atchison, "Anterior Corneal and Internal Contributions to Peripheral Aberrations of Human Eyes," *J. Opt. Soc. Am. A* **21**:355–359 (2004).
193. D. A. Atchison, D. H. Scott, and W. N. Charman, "Hartmann-Shack Technique and Refraction Across the Horizontal Visual Field," *J. Opt. Soc. Am. A* **20**:965–973 (2003).
194. L. Lundström and P. Unsbo, "Transformation of Zernike Coefficients:Scaled, Translated, and Rotated Wavefronts with Circular and Elliptical Pupils," *J. Opt. Soc. Am. A* **24**:569–577 (2007).
195. D. A. Atchison, D. H. Scott, and W. N. Charman, "Measuring Ocular Aberrations in the Peripheral Field Using Hartmann-Shack Aberrometry," *J. Opt. Soc. Am. A* **24**:2963–2973 (2007).
196. Y. Le Grand, *Form and Space Vision*, M. Millodot and G. C. Heath (trans.), Indiana UP, Bloomington, 1967, pp. 5–23.
197. J. G. Sivak and T. Mandelman, "Chromatic Dispersion of the Ocular Media," *Vision Res.* **22**:997–1003 (1982).

198. T. Mandelman and J. G. Sivak, "Longitudinal Chromatic Aberration of the Vertebrate Eye," *Vision Res.* **23**:1555–1559 (1983).
199. D. A. Atchison and G. Smith, "Chromatic Dispersions of the Ocular Media of Human Eyes," *J. Opt. Soc. Am. A* **22**:29–37 (2005).
200. L. N. Thibos, A. Bradley, F. D. L. Still, X. Zhang, and P. A. Howarth, "Theory and Measurement of Ocular Chromatic Aberration," *Vision Res.* **30**:33–49 (1990).
201. L. N. Thibos, A. Bradley, and X. Zhang, "Effect of Ocular Chromatic Aberration on Monocular Visual Performance," *Optom. Vis. Sci.* **68**:599–607 (1991).
202. G. Wald and D. R. Griffin, "The Change of Refractive Power of the Human Eye in Dim and Bright Light," *J. Opt. Soc. Am.* **37**:321–336 (1947).
203. A. Ivanoff, *Les Aberrations de l'Oeil*, Revue d'Optique, Paris, 1953.
204. R. E. Bedford and G. Wyszecki, "Axial Chromatic Aberration of the Human Eye," *J. Opt. Soc. Am.* **47**:564–565 (1957).
205. A. Howarth and A. Bradley, "The Longitudinal Chromatic Aberration of the Human Eye, and Its Correction," *Vision Res.* **26**:361–366 (1986).
206. H. Hartridge, "The Visual Perception of Fine Detail," *Phil. Trans. R. Soc. A* **232**:519–671 (1947).
207. J. S. McLellan, S. Marcos, P. M. Prieto, and S. A. Burns, "Imperfect Optics may be the Eye's Defence Against Chromatic Blur," *Nature* **417**:174–176 (2002).
208. P. A. Howarth, "The Lateral Chromatic Aberration of the Human Eye," *Ophthalm. Physiol. Opt.* **4**:223–236 (1984).
209. A. van Meeteren, "Calculations on the Optical Modulation Transfer Function of the Human Eye for White Light," *Optica Acta* **21**:395–412 (1974).
210. L. N. Thibos, "Calculation of the Influence of Lateral Chromatic Aberration on Image Quality across the Visual Field," *J. Opt. Soc. Am. A* **4**:1673–1680 (1987).
211. P. Simonet and M. C. W. Campbell, "The Optical Transverse Chromatic Aberration on the Fovea of the Human Eye," *Vision Res.* **30**:187–206 (1990).
212. J. J. Vos, "Some New Aspects of Color Stereoscopy," *J. Opt. Soc. Am.* **50**:785–790 (1960).
213. B. N. Kishito, "The Colour Stereoscopic Effect," *Vision Res.* **5**:3131–329 (1965).
214. J. M. Sundet, "The Effect of Pupil Size Variation on the Colour Stereoscopic Phenomenon," *Vision Res.* **12**:1027–1032 (1972).
215. R. C. Allen and M. L. Rubin, "Chromostereopsis," *Survey Ophthalmol.* **26**:22–27 (1981).
216. D. A. Owens and H. W. Leibowitz, "Chromostereopsis and Small Pupils," *J. Opt. Soc. Am.* **65**:358–359 (1975).
217. M. Ye, A. Bradley, L. N. Thibos, and X. X. Zhang, "Interocular Differences in Transverse Chromatic Aberration Determine Chromostereopsis for Small Pupils," *Vision Res.* **31**:1787–1796 (1991).
218. J. Faubert, "Seeing Depth in Colour: More Than Just What Meets the Eyes," *Vision Res.* **34**:1165–1186 (1994).
219. L. N. Thibos, F. E. Cheney, and D. J. Walsh, "Retinal Limits to the Detection and Resolution of Gratings," *J. Opt. Soc. Am. A* **4**:1524–1529 (1987).
220. Y. U. Ogboso and H. E. Bedell, "Magnitude of Lateral Chromatic Aberration Across the Retina of the Human Eye," *J. Opt. Soc. Am. A* **4**:1666–1672 (1987).
221. L. L. Holladay, "The Fundamentals of Glare and Visibility," *J. Opt. Soc. Am.* **12**:271–319 (1926).
222. W. S. Stiles, "The Effect of Glare on the Brightness Threshold," *Proc. R. Soc. Lond. B* **104**:322–351 (1929).
223. J. J. Vos, J. Walraven, and A. van Meeteren, "Light Profiles of the Foveal Image of a Point Source," *Vision Res.* **16**:215–219 (1976).
224. M. J. Allen and J. J. Vos, "Ocular Scattered Light and Visual Performance as a Function of Age," *Am. J. Optom. Arch. Am. Acad. Optom.* **44**:717–727 (1967).
225. A. Spector, S. Li, and J. Sigelman, "Age-Dependent Changes in the Molecular Size of Human Lens Proteins and Their Relationship to Light Scatter," *Invest. Ophthalmol.* **13**:795–798 (1974).
226. R. P. Hemenger, "Intraocular Light Scatter in Normal Visual Loss with Age," *Applied Opt.* **23**:1972–1974 (1984).

227. R. A. Weale, "Effects of Senescence," *Vision and Visual Dysfunction*, vol. 5, *Limits to Vision*, J. J. Kulikowski, V. Walsh, and J. Murray (eds.), Macmillan, London, 1991, pp. 277–285.
228. W. Adrian and A. Bhanji, "A Formula to Describe Straylight in the Eye as a Function of the Glare Angle and Age," *1<sup>st</sup> Int. Symp. on Glare*, Orlando Fla., Oct 24 and 25, 1991, The Lighting Research Institute, New York, 1991, pp. 185–191.
229. J. J. Vos and J. Boogaard, "Contribution of the Cornea to Entoptic Scatter," *J. Opt. Soc. Am.* **53**:869–873 (1963).
230. R. M. Boynton and F. J. J. Clarke, "Sources of Entoptic Scatter in the Human Eye," *J. Opt. Soc. Am.* **54**:110–119, 717–719 (1964).
231. J. J. Vos, "Contribution of the Fundus Oculi to Entoptic Scatter," *J. Opt. Soc. Am.* **53**:1449–1451 (1963).
232. J. J. Vos and M. A. Bouman, "Contribution of the Retina to Entoptic Scatter," *J. Opt. Soc. Am.* **54**:95–100 (1964).
233. R. P. Hemenger, "Small-angle Intraocular Scattered Light: A Hypothesis Concerning its Source," *J. Opt. Soc. Am. A* **5**:577–582 (1988).
234. J. H. Ohzu and J. M. Enoch, "Optical Modulation by the Isolated Human Fovea," *Vision Res.* **12**:245–251 (1972).
235. D. A. Williams, "Visibility of Interference Fringes Near the Resolution Limit," *J. Opt. Soc. Am. A* **2**:1087–1093.
236. D. I. A. MacLeod, D. R. Williams, and W. Makou, "A Visual Non-linearity Fed by Single Cones," *Vision Res.* **32**:347–363 (1992).
237. D. A. Palmer, "Entopic Phenomena," *Vision and Visual Dysfunction*, vol. 1, *Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 345–370.
238. R. B. Rabbetts, "Entoptic Phenomena," *Clinical Visual Optics*, 3rd ed., Butterworth-Heinemann, Oxford, 1998, pp. 421–429.
239. A. Caldecott and W. N. Charman, "Diffraction Haloes Resulting from Corneal Oedema and Epithelial Cell Size," *Ophthalm. Physiol. Opt.* **22**:209–213 (2002).
240. T. J. T. P. van den Berg, M. P. J. Hagenouw, and J. E. Coppens, "The Ciliary Corona: Physical Model and Simulation of the Fine Needles Radiating from Points Light Sources," *Invest. Ophthalmol. Vis. Sci.* **46**:2627–2632 (2005).
241. R. Navarro and M. A. Losada, "Shape of Stars and Optical Quality of the Human Eye," *J. Opt. Soc. Am. A* **14**:353–359 (1997).
242. D. B. Henson, *Optometric Instrumentation*, 2nd ed., Butterworth-Heinemann, Oxford, 1996.
243. S. Siik, P. J. Airaksinen, A. Tuulonen, and H. Nieminem, "Autofluorescence in Cataractous Human Lens and its Relationship to Light Scatter," *Acta Ophthalmol.* **71**:388–392 (1993).
244. D. B. Elliott, K. C. H. Yang, K. Dumbleton, and A. P. Cullen, "Ultra-Violet Induced Lenticular Fluorescence: Intraocular Straylight Affecting Visual Function," *Vision Res.* **33**:1827–1833 (1993).
245. T. C. D. Whiteside, *Problems of Vision in Flight at High Altitudes*, Butterworths, London, 1957.
246. H. H. Hopkins, "The Application of Frequency Response Techniques in Optics," *Proc. Phys. Soc.* **79**:889–919 (1962).
247. H. Metcalf, "Stiles-Crawford Apodization," *J. Opt. Soc. Am.* **55**:72–74 (1965).
248. D. A. Atchison, A. Joblin, and G. Smith, "Influence of Stiles-Crawford Apodization on Spatial Visual Performance," *J. Opt. Soc. Am. A* **15**:2545–2550 (1998).
249. D. A. Atchison, D. H. Scott, N. C. Strang, and P. Artal, "Influence of Stiles-Crawford Apodization on Visual Acuity," *J. Opt. Soc. Am. A* **19**:1073–1083 (2002).
250. X. Zhang, A. Bradley, and L. N. Thibos, "Apodization by the Stiles-Crawford Effect Moderates the Visual Impact of Retinal Image Defocus," *J. Opt. Soc. Am. A* **16**:812–820 (1999).
251. M. J. Cox, D. A. Atchison, and D. H. Scott, "Scatter and Its Implications for the Measurement of Optical Image Quality in Human Eyes," *Optom. Vis. Sci.* **80**:56–68 (2003).
252. G. Walsh, W. N. Charman, and H. C. Howland, "Objective Technique for the Determination of Monochromatic Aberrations of the Human Eye," *J. Opt. Soc. Am. A* **1**:987–992 (1984).
253. L. Diaz-Santana, G. Walker, and S. X. Bara, "Sampling Geometries for Ocular Aberrometry: a Model for Evaluation of Performance," *Opt. Express* **13**:8801–8818 (2005).

254. L. Llorente, S. Marcos, C. Dorronsoro, and S. A. Burns, "Effect of Sampling on Real Ocular Aberration Measurements," *J. Opt. Soc. Am. A* **24**:2783–2796 (2007).
255. X. Hong, L. N. Thibos, A. Bradley, R. L. Woods, and R. A. Applegate, "Comparison of Monochromatic Ocular Aberrations Measured with an Objective Crossed-Cylinder Aberroscope and a Shack-Hartmann Aberrometer," *Optom. Vis. Sci.* **80**:15–25 (2003).
256. Y. Le Grand, "Sur une Mode de Vision Eliminant les Défaits Optiques de l'Oeil," *Rev. d'Optique Theor. Instrum.* **15**:6–11 (1936).
257. W. N. Charman and P. Simonet, "Yves Le Grand and the Assessment of Retinal Acuity Using Interference Fringes," *Ophthal. Physiol. Opt.* **17**:164–168 (1997).
258. G. Westheimer, "Modulation Thresholds for Sinusoidal Light Distributions on the Retina," *J. Physiol. (London)* **152**:67–74 (1960).
259. A. Arnulf and O. Dupuy, "La Transmission des Contrastes par la Système Optique de l'Oeil et les Seuls de Modulation Rétiens," *C. r. hebd. Seanc. Acad. Paris* **250**:2757–2759 (1960).
260. F. W. Campbell and D. G. Green, "Optical and Retinal Factors Affecting Visual Resolution," *J. Physiol. (London)* **181**:576–593 (1965).
261. S. Berger-Lheureux, "Mesure de la Fonction de Transfert de Modulation du Système Optique de l'Oeil et des Seuls de Modulation Rétiens," *Rev. Opt. Theor. instrum.* **44**:294–323 (1965).
262. L. J. Bour, "MTF of the Defocused Optical System of the Human Eye for Incoherent Monochromatic Light," *J. Opt. Soc. Am.* **70**:321–328 (1980).
263. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Aberration-Free Measurement of the Visibility of Isoluminant Gratings," *J. Opt. Soc. Am. A* **10**:2105–2116 (1993).
264. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Efficiency in Detection of Isoluminant and Isochromatic Interference Fringes," *J. Opt. Soc. Am. A* **10**:2118–2133 (1993).
265. J. Rovamo, J. Mustonen, and R. Näsänen, "Two Simple Methods for Determining the Optical Transfer Function of the Human Eye," *Vision Res.* **34**:2493–2502 (1994).
266. F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *J. Physiol. (London)* **186**:5589–578 (1966).
267. F. Flamant, "Etude de la Repartition de la Lumière dans l'Image Rétinienne d'une Fente," *Rev. Opt. Theor. Instrum.* **34**:433–459 (1955).
268. J. Krauskopf, "Light distribution in Human Retinal Images," *J. Opt. Soc. Am.* **52**:1046–1050 (1962).
269. J. Krauskopf, "Further Measurements in Human Retinal Images," *J. Opt. Soc. Am.* **54**:715–716 (1964).
270. G. Westheimer and F. W. Campbell, "Light Distribution in the Image Formed by the Living Human Eye," *J. Opt. Soc. Am.* **52**:1040–1045 (1962).
271. R. Rohler, U. Miller, and M. Aberl, "Zür Messung der Modulationsübertragungs-Funktion des lebenden menschlichen Augen in reflektieren Licht," *Vision Res.* **9**:407–428 (1969).
272. J. A. M. Jennings and W. N. Charman, "Off-Axis Image Quality in the Human Eye," *Vision Res.* **21**:445–455 (1981).
273. J. Santamaria, P. Artal, and J. Bescos, "Determination of the Point Spread Function of Human Eyes Using a Hybrid Optical-Digital Method," *J. Opt. Soc. Am. A* **4**:109–114 (1987).
274. P. Artal, S. Marcos, R. Navarro, and D. Williams, "Odd Aberrations and Double Pass Measurements of Retinal Image Quality," *J. Opt. Soc. Am. A* **12**:195–201 (1995).
275. R. Navarro and M. A. Losada, "Phase Transfer and Point-Spread Function of the Human Eye Determined by a New Asymmetric Double-pass Method," *J. Opt. Soc. Am. A* **12**:2385–2392 (1995).
276. P. Artal, I. Iglesias, N. López-Gil, and D. G. Green, "Double-Pass Measurements of the Retinal Image Quality with Unequal Entrance and Exit Pupil Size and the Reversibility of the Eye's Optical System," *J. Opt. Soc. Am. A* **12**:2358–2366 (1995).
277. L. Diaz-Santana and J. C. Dainty, "Effects of Retinal Scattering in the Ocular Double-Pass Procedure," *J. Opt. Soc. Am. A* **18**:1437–1444 (2001).
278. P. Artal, M. Ferro, I. Miranda, and R. Navarro, "Effects of Aging in Retinal Image Quality," *J. Opt. Soc. Am. A* **10**:1656–1662 (1993).
279. N. López-Gil, I. Iglesias, and P. Artal, "Retinal Image Quality in the Human Eye as Function of the Accommodation," *Vision Res.* **38**:2897–2907 (1998).
280. G. Westheimer and J. Liang, "Evaluating Diffusion of Light in the Eye by Objective Means," *Invest. Ophthalmol. Vis. Sci.* **35**:2652–2657 (1994).

281. J. F. Simon and P. M. Denieul, "Influence of the Size of Test Field Employed in Measurements of Modulation Transfer Function of the Eye," *J. Opt. Soc. Am.* **63**:894–896 (1973).
282. P. Artal and R. Navarro, "Simultaneous Measurement of Two-point-spread Functions at Different Locations across the Human Fovea," *Applied Opt.* **31**:3646–3656 (1992).
283. S. C. Choi, N. Doble, J. Lion, J. Christou, and D. R. Williams, "Effect of Wavelength on *in vivo* Images of the Human Cone Mosaic," *J. Opt. Soc. Am. A* **22**:2598–2605 (2005).
284. D. R. Williams, D. H. Brainard, M. J. McMahon, and R. Navarro, "Double-Pass and Interferometric Measures of the Optical Quality of the Eye," *J. Opt. Soc. Am. A* **11**:3123–3135 (1994).
285. N. López-Gil and P. Artal, "Comparison of Double-pass Estimates of the Retinal-image Quality Obtained with Green and Near-Infrared Light," *J. Opt. Soc. Am. A* **14**:961–971 (1997).
286. R. W. Gubisch, "Optical Performance of the Human Eye," *J. Opt. Soc. Am.* **57**:407–415 (1967).
287. P. Artal, "Calculations of Two-dimensional Foveal Retinal Images in Real Eyes," *J. Opt. Soc. Am. A* **7**:1374–1381 (1990).
288. W. T. Welford, *Aberrations of Optical Systems*, Adam Hilger, Bristol, 1986, pp. 243–244.
289. J. Liang and D. R. Williams, "Aberrations and Retinal Image Quality in the Normal Human Eye," *J. Opt. Soc. Am. A* **14**:2873–2883 (1997).
290. J. M. Bueno and P. Artal, "Polarization and Retinal Image Quality Estimates in the Human Eye," *J. Opt. Soc. Am. A* **18**:489–496 (2001).
291. A. C. S. van Heel, "Correcting the Spherical and Chromatic Aberrations of the Eye," *J. Opt. Soc. Am.* **36**:237–239 (1946).
292. H. Hartridge, "Visual Acuity and Resolving Power of the Eye," *J. Physiol. (London)* **57**:57–62 (1962).
293. S. A. Klein, "Optimal Corneal Ablation for Eyes with Arbitrary Hartmann-Shack Aberrations," *J. Opt. Soc. Am. A* **15**:2580–2588 (1998).
294. J. Schweigerling and R. W. Snyder, "Custom Photorefractive Keratectomy Ablations for the Correction of Spherical and Cylindrical Refractive Error and Higher-Order Aberrations," *J. Opt. Soc. Am. A* **15**:2572–2579 (1998).
295. J. Schweigerling, "Theoretical Limits to Visual Performance," *Surv. Ophthalmol.* **45**:139–146 (2000).
296. T. Seiler, M. Mrochen, and M. Kaemmerer, "Operative Correction of Ocular Aberrations to Improve Visual Acuity," *J. Refract. Surg.* **16**:S619–S622 (2000).
297. S. M. Macrae, R. R. Krueger, and R. A. Applegate (eds.), *Customized Corneal Ablation: the Quest for Supervision*, Slack Inc., Thorofare, N.J., 2001.
298. R. R. Krueger, R. A. Applegate, S. M. Macrae (eds.), *Wavefront Customized Visual Correction: the Quest for Supervision II*, Slack Inc., Thorofare, N.J., 2004.
299. J. Liang, D. R. Williams, and D. T. Miller, "Supernormal Vision and High-Resolution Retinal Imaging through Adaptive Optics," *J. Opt. Soc. Am. A* **14**:2884–2892 (1997).
300. G.-Y. Yoon and D. R. Williams, "Visual Performance after Correcting Monochromatic and Chromatic Aberrations of the Eye," *J. Opt. Soc. Am. A* **19**:286–275, (2002).
301. D. C. Chen, S. M. Jones, D. A. Silva, and S. S. Olivier, "High-Resolution Adaptive Optics Scanning Laser Ophthalmoscope with Dual Deformable Mirrors," *J. Opt. Soc. Am. A* **24**:1305–1312 (2007).
302. N. Lopez-Gil, J. F. Castejon-Mochon, A. Benito, J. M. Marin, G. Loe-a-Foe, G. Marin, B. Fermiger, D. Renard, D. Joyeux, N. Chateau, and P. Artal, "Aberration Generation by Contact Lenses with Aspheric and Asymmetric Surfaces," *J. Refract. Surg.* **18**:603–609 (2002).
303. D. A. Chernyak, and C. E. Campbell, "System for the Design, Manufacture, and Testing of Custom Lenses with Known Amounts of High-Order Aberrations," *J. Opt. Soc. Am. A* **20**:2016–2021 (2002).
304. S. Bara, T. Mancebo, and E. Moreno-Barrusio, "Positioning Tolerances for Phase Plates Compensating Aberrations of the Human Eye," *Applied Opt.* **39**:3413–3420 (2000).
305. A. Guirao, D. R. Williams, and I. G. Cox, "Effect of Rotation and Translation on the Expected Benefit of an Ideal Method to Correct the Eye's Higher-order Aberrations," *J. Opt. Soc. Am. A* **18**:1003–1015 (2001).
306. G. Yoon and T. M. Jeong, "Effect of the Movement of Customized Contact Lens on Benefit in Abnormal Eyes," *J. Vision* **3**(12):38a (2003).
307. J. De Brabander, N. Chateau, G. Marin, N. Lopez-Gil, E. van der Worp, and A. Benito, "Simulated Optical Performance of Custom Wavefront Soft Contact Lenses for Keratoconus," *Optom. Vis. Sci.* **80**:637–643 (2003).

308. N. Lopez-Gil, N. Chateau, J. Castejon-Monchon, and A. Benito, "Correcting Ocular Aberrations by Soft Contact Lenses," *S. Afr. Optom.* **62**:173–177 (2003).
309. W. N. Charman and N. Chateau, "The Prospects for Super-Acuity: Limits to Visual Performance after Correction of Monochromatic Ocular Aberration," *Ophthalm. Physiol. Opt.* **23**:479–493 (2003).
310. W. N. Charman, "Ablation Design in Relation to Spatial Frequency, Depth-of-focus and Age," *J. Refract. Surg.* **20**:S572–S579 (2004).
311. L. C. Thomson and W. D. Wright, "The Colour Sensitivity of the Retina Within the Central Fovea of Man," *J. Physiol. (London)* **105**:316–331 (1947).
312. I. Powell, "Lenses for Correcting Chromatic Aberration of the Eye," *Applied Opt.* **20**:4152–4155 (1981).
313. A. L. Lewis, M. Katz, and C. Oehrlein, "A Modified Achromatizing Lens," *Am. J. Optom. Physiol. Opt.* **59**:909–911 (1982).
314. J. A. Díaz, M. Irlbauer, and J. A. Martínez, "Diffractive-Refractive Hybrid Doublet to Achromatize the Human Eye," *J. Mod. Opt.* **51**:2223–2234 (2004).
315. Y. Benny, S. Manzanera, P. M. Prieto, E. N. Ribak, and P. Artal, "Wide-Angle Chromatic Aberration Corrector for the Human Eye," *J. Opt. Soc. Am. A* **24**:1538–1544 (2007).
316. X. Zhang, A. Bradley, and L. N. Thibos, "Achromatizing the Human Eye: the Problem of Chromatic Parallax," *J. Opt. Soc. Am. A* **8**:686–691 (1991).
317. J. A. M. Jennings and W. N. Charman, "Optical Image Quality in the Peripheral Retina," *Am. J. Optom. Physiol. Opt.* **55**:582–590 (1978).
318. R. Navarro, P. Artal, and D. R. Williams, "Modulation Transfer of the Human Eye as a Function of Retinal Eccentricity," *J. Opt. Soc. Am. A* **10**:201–212 (1993).
319. D. R. Williams, P. Artal, R. Navarro, M. J. McMahon, and D. H. Brainard, "Off-Axis Optical Quality and Retinal Sampling in the Human Eye," *Vision Res.* **36**:1103–1114 (1996).
320. J. A. M. Jennings and W. N. Charman, "Analytic Approximation of the Off-Axis Modulation Transfer Function of the Eye," *Vision Res.* **37**:697–704 (1997).
321. P. B. DeVelis and G. B. Parrent, "Transfer Function for Cascaded Optical Systems," *J. Opt. Soc. Am.* **57**:1486–1490 (1967).
322. I. Overington, "Interaction of Vision with Optical Aids," *J. Opt. Soc. Am.* **63**:1043–1049 (1973).
323. I. Overington, "The Importance of Coherence of Coupling When Viewing Through Visual Aids," *Opt. Laser Technol.* **52**:216–220 (1973).
324. I. Overington, "Some Considerations of the Role of the Eye as a Component of an Imaging System," *Optica Acta* **22**:365–374 (1975).
325. W. N. Charman and H. Whitefoot, "Astigmatism, Accommodation and Visual Instrumentation," *Applied Opt.* **17**:3903–3910 (1978).
326. P. Mouroulis, "On the Correction of Astigmatism and Field Curvature in Telescopic Systems," *Optica Acta* **29**:1133–1159 (1982).
327. G. J. Burton and N. D. Haig, "Criteria for Testing of Afocal Instruments," *Proc. Soc. Photo-Opt. Instrum. Eng.* **274**:191–201 (1981).
328. N. D. Haig and G. J. Burton, "Effects of Wavefront Aberration on Visual Instrument Performance, and a Consequential Test Technique," *Applied Opt.* **26**:492–500.
329. J. Eggert and K. J. Rosenbruch, "Vergleich der visuell und der photolektrisch gemessenen Abbildungs-guete von Fernrohren," *Optik* **48**:439–450 (1977).
330. G. J. Burton and N. D. Haig, "Effects of the Seidel Aberrations on Visual Target Discrimination," *J. Opt. Soc. Am. A* **1**:373–385 (1984).
331. R. Legras, N. Chateau, and W. N. Charman, "Assessment of Just Noticeable Differences for Refractive Errors and Spherical Aberration Using Visual Simulation," *Optom. Vis. Sci.* **81**:718–728 (2004).
332. P. Mouroulis and H. Zhang, "Visual Instrument Image Quality Metrics and the Effects of Coma and Astigmatism," *J. Opt. Soc. Am. A* **9**:34–42 (1992).
333. P. Mouroulis and G. C. Woo, "Chromatic Aberration and Accommodation in Visual Instruments," *Optik* **80**:161–166 (1989).
334. P. Mouroulis, T. G. Kim, and G. Zhao, "Transverse Color Tolerances for Visual Optical Systems," *Applied Opt.* **32**:7089–7094 (1993).



335. P. Mouroulis (ed.), *Visual Optical Systems*, McGraw-Hill, New York, 1999.
336. M. Koomen, R. Skolnik, and R. Tousey, "A Study of Night Myopia," *J. Opt. Soc. Am.* **41**:80–90 (1951).
337. D. G. Green and F. W. Campbell, "Effect of Focus on the Visual Response to a Sinusoidally Modulated Spatial Stimulus," *J. Opt. Soc. Am.* **55**:1154–1157 (1965).
338. W. N. Charman and J. A. M. Jennings, "The Optical Quality of the Retinal Image as a Function of Focus," *Br. J. Physiol. Opt.* **31**:119–134 (1976).
339. W. N. Charman, J. A. M. Jennings, and H. Whitefoot, "The Refraction of the Eye in Relation to Spherical Aberration and Pupil Size," *Br. J. Physiol. Opt.* **32**:78–93 (1978).
340. L. N. Thibos, "Unresolved Issues in the Prediction of Subjective Refraction from Wavefront Aberration Maps," *J. Refract. Surg.* **20**:S533–S536 (2004).
341. X. Cheng, A. Bradley, and L. N. Thibos, "Predicting Subjective Judgement of Best Focus with Objective Image Quality Metrics," *J. Vision* **4**:310–321 (2004).
342. L. N. Thibos, X. Hong, A. Bradley, and R. A. Applegate, "Accuracy and Precision of Objective Refraction from Wavefront Aberrations," *J. Vision* **4**:329–351 (2004).
343. D. A. Atchison, S. W. Fisher, C. A. Pedersen, and P. G. Ridall, "Noticeable, Troublesome and Objectionable Limits of Blur," *Vision Res.* **45**:1967–1974 (2005).
344. K. J. Ciuffreda, A. Selenow, B. Wang, B. Vasudevan, G. Zikos, and S. R. Ali, "Bothersome Blur: A Functional Unit of Blur Perception," *Vision Res.* **46**:895–901 (2006).
345. B. Wang and K. J. Ciuffreda, "Depth-of-Focus of the Human Eye: Theory and Clinical Implications," *Surv. Ophthalmol.* **51**:75–85 (2006).
346. D. G. Green and F. W. Campbell, "Effect of Focus on the Visual Response to a Sinusoidally Modulated Spatial Stimulus," *J. Opt. Soc. Am.* **55**:1154–1157 (1965).
347. D. G. Green, M. K. Powers, and M. S. Banks, "Depth of Focus, Eye Size and Visual Acuity," *Vision Res.* **20**:827–835 (1980).
348. F. L. van Nes and M. A. Bouman, "Spatial Modulation Transfer in the Human Eye," *J. Opt. Soc. Am.* **57**:401–407 (1967).
349. F. W. Campbell, "The Depth-of-Field of the Human Eye," *Optica Acta* **4**:157–164 (1957).
350. J. Tucker and W. N. Charman, "Depth-of-Focus and Accommodation for Sinusoidal Gratings as a Function of Luminance," *Am. J. Optom. Physiol. Opt.* **63**:58–70 (1986).
351. K. N. Ogle and J. T. Schwartz, "Depth-of-focus for the Human Eye," *J. Opt. Soc. Am.* **49**:273–280 (1959).
352. J. Tucker and W. N. Charman, "The Depth-of-Focus of the Human Eye for Snellen Letters," *Am. J. Optom. Physiol. Opt.* **52**:3–21 (1975).
353. W. N. Charman and H. Whitefoot, "Pupil Diameter and Depth-of-Field of the Human Eye as Measured by Laser Speckle," *Optica Acta* **24**:1211–1216 (1977).
354. D. A. Atchison, W. N. Charman, and R. L. Woods, "Subjective Depth-of-Focus of the Eye," *Optom. Vis. Sci.* **74**:511–520 (1997).
355. D. A. Goss and T. Grosvenor, "Reliability of Refraction—A Literature Review," *J. Am. Optom. Assoc.* **67**:619–630 (1996).
356. A. D. Miller, M. J. Kris, and A. C. Griffiths, "Effect of Small Focal Errors on Vision," *Optom. Vis. Sci.* **74**:521–526 (1997).
357. I. E. Loewenfeld, *The Pupil: Anatomy, Physiology and Clinical Applications*, Butterworth-Heinemann, London, 1999, pp. 295–317.
358. M. Stakenburg, "Accommodation without Pupillary Constriction," *Vision Res.* **31**:267–273 (1991).
359. N. J. Phillips, B. Winn, and B. Gilmartin, "Absence of Pupil Response to Blur-driven Accommodation," *Vision Res.* **32**:1775–1779 (1992).
360. F. Schaeffel, H. Wilhelm, and E. Zrenner, "Inter-Individual Variability in the Dynamics of Natural Accommodation in Humans: Relation to Age and Refractive Errors," *J. Physiol. (London)* **462**:301–320 (1993).
361. S. Kasthurirangan and A. Glasser, "Age Related Changes in the Characteristics of the Near Pupil Response," *Vision Res.* **46**:1393–1403 (2006).
362. H. Radhakrishnan and W. N. Charman, "Age-related Changes in Static Accommodation and Accommodative Miosis," *Ophthal. Physiol. Opt.* **27**:342–352 (2007).

363. F. W. Campbell, "Correlation of Accommodation between the Two Eyes," *J. Opt. Soc. Am.* **50**:738 (1960).
364. G. Heron, B. Winn, J. R. Pugh, and A. S. Eadie, "Twin Channel Infrared Optometer for Recording Binocular Accommodation," *Optom. Vis. Sci.* **66**:123–129 (1989).
365. F. W. Campbell, "The Minimum Quantity of Light Required to Elicit the Accommodation Reflex in Man," *J. Physiol. (London)* **123**:357–366 (1954).
366. C. A. Johnson, "Effects of Luminance and Stimulus Distance on Accommodation and Visual Resolution," *J. Opt. Soc. Am.* **66**:138–142 (1976).
367. D. A. Atchison, "Accommodation and Presbyopia," *Ophthalm. Physiol. Opt.* **15**:255–2272 (1995)
368. K. J. Ciuffreda, "Accommodation, the Pupil and Presbyopia," *Borish's Clinical Refraction*, W. J. Benjamin (ed.), Saunders, Philadelphia, 1998, pp. 77–120.
369. A. Glasser and P. I. Kaufman, "The Mechanism of Accommodation in Primates," *Ophthalmology* **106**:863–872 (1999).
370. W. N. Charman, "The Eye in Focus: Accommodation and Presbyopia," *Clin. Exp. Optom.* **91**:207–225 (2008).
371. H. J. Wyatt, "Application of a Simple Mechanical Model of Accommodation to the Aging Eye," *Vision Res.* **33**:731–738 (1993).
372. A. P. A. Beers and G. L. van der Heijde, "In vivo Determination of the Biomechanical Properties of the Component Elements of the Accommodation Mechanism," *Vision Res.* **34**:2897–2905 (1994).
373. S. J. Judge and H. J. Burd, "Modelling the Mechanics of Accommodation and Presbyopia," *Ophthalm. Physiol. Opt.* **22**:397–400 (2002).
374. H. Martin, R. Guthoff, T. Terwee, and K-P Schmitz, "Comparison of the Accommodation Theories of Coleman and Helmholtz by Finite Element Simulations," *Vision Res.* **45**:2910–2915 (2005).
375. F. W. Campbell and G. Westheimer, "Dynamics of the Focussing Response of the Human Eye," *J. Physiol. (London)* **151**:285–295 (1960).
376. S. D. Phillips, D. Shirachi, and L. Stark, "Analysis of Accommodation Times Using Histogram Information," *Am. J. Optom. Arch. Am. Acad. Optom.* **49**:389–401 (1972).
377. D. Shirachi, J. Liu, M. Lee, J. Jang, J. Wong, and L. Stark, "Accommodation Dynamics. I Range Non-Linearity," *Am. J. Optom. Physiol. Opt.* **55**:531–541 (1978).
378. J. Tucker and W. N. Charman, "Reaction and Response Times for Accommodation," *Am. J. Optom. Physiol. Opt.* **56**:490–503 (1979).
379. S. Kasthurirangan, A. S. Vilupuru, and A. Glasser, "Amplitude Dependent Accommodative Dynamics in Humans," *Vision Res.* **43**:2945–2956 (2003).
380. S. Kasthurirangan and A. Glasser, "Influence of Amplitude and Starting Point on Accommodative Dynamics in Humans," *Invest. Ophthalmol. Vis. Sci.* **46**:3463–3472 (2005).
381. E. F. Fincham, "The Accommodation Reflex and its Stimulus," *Br. J. Ophthalmol.* **35**:381–393 (1951).
382. F. W. Campbell and G. Westheimer, "Factors Influencing Accommodation Responses of the Human Eye," *J. Opt. Soc. Am.* **49**:568–571 (1959).
383. A. Troelstra, B. L. Zuber, D. Miller, and L. Stark, "Accommodative Tracking: a Trial and Error Function," *Vision Res.* **4**:585–594 (1964).
384. L. Stark, *Neurological Control Systems: Studies in Bioengineering*, Plenum Press, New York, 1968, Sect. III.
385. W. N. Charman and G. Heron, "On the linearity of Accommodation Dynamics," *Vision Res.* **40**:2057–2066 (2000).
386. E. Marg E. "An Investigation of Voluntary as Distinguished from Reflex Accommodation," *Am. J. Optom. Arch. Am. Acad. Optom.* **28**:347–356 (1951).
387. T. N. Cornsweet and H. D. Crane, "Training the Visual Accommodation System," *Vision Res.* **13**:713–715 (1973).
388. R. R. Provine and J. M. Enoch, "On Voluntary Ocular Accommodation," *Percept. Psychophys.* **17**:209–212 (1975).
389. R. J. Randle and M. R. Murphy, "The Dynamic Response of Visual Accommodation over a Seven-Day Period," *Am. J. Optom. Physiol. Opt.* **51**:530–540 (1974).
390. G. J. van der Wildt, M. A. Bouman, and J. van der Kraats, "The Effect of Anticipation on the Transfer Function of the Human Lens System," *Optica Acta* **21**:843–860 (1974).



391. P. B. Kruger and J. Pola, "Changing Target Size is a Stimulus for Accommodation," *J. Opt. Soc. Am. A* **2**:1832–1835 (1985).
392. T. Takeda, K. Hashimoto, N. Hiruma, and Y. Fukui, "Characteristics of Accommodation toward Apparent Depth," *Vision Res.* **39**:2087–2097 (1999).
393. W. N. Charman and G. Heron, "Fluctuations in Accommodation: A Review," *Ophthalm. Physiol. Opt.* **8**:153–164 (1988).
394. B. Winn and B. Gilmartin, "Current Perspective on Microfluctuations of Accommodation," *Ophthalm. Physiol. Opt.* **12**:252–256 (1992).
395. K. Toshida, F. Okuyama, and T. Tokoro, "Influences of the Accommodative Stimulus and Aging on the Accommodative Microfluctuations," *Optom. Vis. Sci.* **75**:221–226 (1998).
396. L. S. Gray, B. Winn, and B. Gilmartin, "Effect of Target Luminance on Microfluctuations of Accommodation," *Ophthalm. Physiol. Opt.* **13**:258–265 (1993).
397. L. R. Stark and D. A. Atchison, "Pupil Size, Mean Accommodation Response and the Fluctuations of Accommodation," *Ophthalm. Physiol. Opt.* **17**:316–323 (1997).
398. F. W. Campbell, J. G. Robson and G. Westheimer, "Fluctuations in Accommodation under Steady Viewing Conditions," *J. Physiol. (London)* **145**:579–594.
399. J. C. Kotulak and C. M. Schor, "Temporal Variations in Accommodation during Steady-State Conditions," *J. Opt. Soc. Am. A* **3**:223–227 (1986).
400. B. Winn, J. R. Pugh, B. Gilmartin, and H. Owens, "Arterial Pulse Modulates Steady-State Accommodation," *Curr. Eye Res.* **9**:971–975 (1990).
401. M. J. Collins, B. Davis, and J. Wood, "Microfluctuations of Steady-State Accommodation and the Cardiopulmonary System," *Vision Res.* **35**:2491–2502 (1995).
402. G. L. van der Heijde, A. P. A. Beers, and M. Dubbelman, "Microfluctuations of Steady-State Accommodation Measured with Ultrasonography," *Ophthalm. Physiol. Opt.* **16**:216–221 (1996).
403. P. Denieul, "Effects of Stimulus Vergence on Mean Accommodation Response, Microfluctuations of Accommodation and Optical Quality of the Human Eye," *Vision Res.* **22**:561–569 (1982).
404. M. Zhu, M. J. Collins, and D. R. Iskander, "The Contribution of Accommodation and the Ocular Surface to the Microfluctuations of Wavefront Aberration of the Eye," *Ophthalm. Physiol. Opt.* **26**:439–446 (2006).
405. G. Heron and C. Schor, "The Fluctuations of Accommodation and Ageing," *Ophthalm. Physiol. Opt.* **15**:445–449 (1995).
406. L. Stark and Y. Takahashi, "Absence of an Odd-Error Signal Mechanism in Human Accommodation," *IEEE Trans. Biomed. Eng.* **BME-12**:138–146 (1965).
407. M. Alpern, "Variability of Accommodation during Steady Fixation at Various Levels of Illuminance," *J. Opt. Soc. Am.* **48**:193–197 (1958).
408. H. D. Crane, *A Theoretical Analysis of the Visual Accommodation System in Humans*, Report NASA CR-606, NASA, Washington, D.C., 1966.
409. G. K. Hung, J. L. Semmlow, and K. J. Ciuffreda, "Accommodative Oscillation Can Enhance Average Accommodation Response: A Simulation Study," *IEEE Trans. Syst. Man. Cybern.* **SMC-12**:594–598 (1982).
410. W. N. Charman, "Accommodation and the Through-Focus Changes of the Retinal Image," *Accommodation and Vergence Mechanisms in the Visual System*, O. Franzén, H. Richter, and L. Stark (eds.), Birkhäuser Verlag, Basel, 2000, pp. 115–127.
411. B. Winn, "Accommodative Microfluctuations: a Mechanism for Steady-State Control of Accommodation," *Accommodation and Vergence Mechanisms in the Visual System*, O. Franzén, H. Richter, and L. Stark (eds.), Birkhäuser Verlag, Basel, 2000, pp. 129–140.
412. M. Millodot, "Effet des Microfluctuations de l'Accommodation sur l'Acuité Visuelle," *Vision Res.* **8**:73–80 (1968).
413. G. Walsh and W. N. Charman, "Visual sensitivity to Temporal Change in Focus and its Relevance to the Accommodation Response," *Vision Res.* **28**:1207–1221 (1988).
414. B. Winn, W. N. Charman, J. R. Pugh, G. Heron, and A. S. Eadie, "Perceptual Detectability of Ocular Accommodation Microfluctuations," *J. Opt. Soc. Am. A* **6**:459–462 (1989).

415. M. W. Morgan, "Accommodation and Its Relationship to Convergence," *Am. J. Optom. Arch. Am. Acad. Optom.* **21**:183–185 (1944).
416. H. Krueger, "Schwannkingen der Akkommodation des menschlichen Auges bei mon- und binokularer Beobachtung," *Albrecht v. Graefes Arch. Ophthalmol.* **205**:129–133 (1978).
417. M. C. Nadell and H. A. Knoll, "The Effect of Luminance, Target Configuration and Lenses upon the Refractive State of the Eye. Parts I and II," *Am. J. Optom. Arch. Am. Acad. Optom.* **33**:24–42 and 86–95 (1956).
418. G. C. Heath, "Influence of Visual Acuity on Accommodative Responses of the Eye," *Am. J. Optom. Physiol. Opt.* **33**:513–534 (1956).
419. J. Tucker, W. N. Charman and P. A. Ward, "Modulation Dependence of the Accommodation Response to Sinusoidal Gratings," *Vision Res.* **26**:1693–1707 (1986).
420. D. A. Owens, "A Comparison of Accommodative Responsiveness and Contrast Sensitivity for Sinusoidal Gratings," *Vision Res.* **20**:159–167 (1980).
421. L. J. Bour, "The Influence of the Spatial Distribution of a Target on the Dynamic Response and Fluctuations of the Accommodation of the Human Eye," *Vision Res.* **21**:1287–1296 (1981).
422. J. Tucker and W. N. Charman, "Effect of Target Content at Higher Spatial Frequencies on the Accuracy of the Accommodation Response," *Ophthalm. Physiol. Opt.* **7**:137–142 (1987).
423. K. J. Ciuffreda, M. Dul, and S. K. Fisher, "Higher-Order Spatial Frequency Contribution to Accommodative Accuracy in Normal and Amblyopic Observers," *Clin. Vis. Sci.* **1**:219–229 (1987).
424. J. C. Kotulak and C. M. Schor, "The Effects of Optical Vergence, Contrast, and Luminance on the Accommodative Response to Spatially Bandpass Filtered Targets," *Vision Res.* **27**:1797–1806 (1987).
425. K. J. Ciuffreda, M. Rosenfield, J. Rosen, A. Azimi, and E. Ong, "Accommodative Responses to Naturalistic Stimuli," *Ophthalm. Physiol. Opt.* **10**:168–174 (1990).
426. K. J. Ciuffreda, "Accommodation to Gratings and More Naturalistic Stimuli," *Optom. Vis. Sci.* **68**:243–260 (1991).
427. H. Ripps, N. B. Chin, I. M. Siegel, and G. M. Breinin, "The Effect of Pupil Size on Accommodation, Convergence and the AC/A ratio," *Invest Ophthalmol.* **1**:127–135 (1962).
428. R. T. Hennessy, R. Iida, K. Shiina, and H. W. Leibowitz, "The Effect of Pupil Size on Accommodation," *Vision Res.* **16**:587–589 (1976).
429. P. A. Ward and W. N. Charman, "Effect of Pupil Size on Steady-State Accommodation," *Vision Res.* **25**:1317–1326 (1985).
430. G. C. Heath, "Accommodative Responses of Totally Color Blind Observers," *Am. J. Optom. Arch. Am. Acad. Optom.* **33**:457–465 (1956).
431. J. Otto and D. Safra, "Ergebnisse objectiver Akkommodationsmessungen an Augen mit organisch bedingtem Zentralskotomn," *Albrecht v. Graefes Arch. Klin. Ophthalmol.* **192**:49–56 (1974).
432. I. C. J. Wood and A. T. Tomlinson, "The Accommodative Response in Amblyopia," *Am. J. Optom. Physiol. Opt.* **52**:243–247 (1975).
433. K. J. Ciuffreda and D. Rumpf, "Contrast and Accommodation in Amblyopia," *Vision Res.* **25**:1445–1447 (1985).
434. W. N. Charman "Static Accommodation and the Minimum Angle of Resolution," *Am. J. Optom. Physiol. Opt.* **63**:915–921 (1986).
435. H. W. Leibowitz and D. A. Owens, "Anomalous Myopias and the Intermediate Dark Focus of Accommodation," *Science* **189**:1121–11128 (1975).
436. H. W. Leibowitz and D. A. Owens, "Night Myopia and the Intermediate Dark Focus of Accommodation," *J. Opt. Soc. Am.* **65**:133–147 (1975).
437. H. W. Leibowitz and D. A. Owens, "New Evidence for the Intermediate Position of Relaxed Accommodation," *Doc. Ophthalmol.* **46**:1121–1128 (1978).
438. G. Smith, "The Accommodative Resting States, Instrument Accommodation and Their Measurement," *Optica Acta* **30**:347–359 (1983).
439. N. A. McBrien and M. Millodot, "The Relationship Between Tonic Accommodation and Refractive Error," *Invest. Ophthalmol. Vis. Sci.* **28**:997–1004 (1987).
440. F. M. Toates, "Accommodation Function of the Human Eye," *Physiol. Rev.* **52**:828–863 (1972).

441. M. Rosenfield, K. J. Ciuffreda, G. K. Hung, and B. Gilmartin, "Tonic Accommodation—a Review. 1. Basic Aspects," *Ophthalm. Physiol. Opt.* **13**:266–284 (1993).
442. M. Rosenfield, K. J. Ciuffreda, G. K. Hung, and B. Gilmartin, "Tonic Accommodation—a Review. 2. Accommodative Adaptation and Clinical Aspects," *Ophthalm. Physiol. Opt.* **14**:265–277 (1994).
443. H. W. Leibowitz, K. W. Gish, and J. B. Sheehy, "Role of Vergence Accommodation in Correcting for Night Myopia," *Am. J. Optom. Physiol. Opt.* **65**:383–386 (1988).
444. J. L. Semmlow and G. K. Hung, "The Near Response: Theories of Control," *Vergence Eye Movements: Basic and Clinical Concepts*, C. M. Schor and K. J. Ciuffreda (eds.), Butterworths, Boston, 1983, pp. 175–195.
445. G. K. Hung, K. J. Ciuffreda, and M. Rosenfield, "Proximal Contribution to a Linear Static Model of Accommodation and Vergence," *Ophthalm. Physiol. Opt.* **16**:31–41 (1996).
446. C. M. Schor and S. R. Bharadwaj, "Pulse-Step Models of Control Strategies for Dynamic Ocular Accommodation and Disaccommodation," *Vision Res.* **46**:242–258 (2006).
447. W. N. Charman and J. Tucker, "Accommodation and Color," *J. Opt. Soc. Am.* **68**:459–471 (1978).
448. J. V. Lovasik and H. Kergoat, "Accommodative Performance for Chromatic Displays," *Ophthalm. Physiol. Opt.* **8**:443–440 (1988).
449. W. N. Charman, "Accommodation Performance for Chromatic Displays," *Ophthalm. Physiol. Opt.* **9**:459–463 (1989).
450. W. R. Bobier, M. C. W. Campbell, and M. Hinch, "The Influence of Chromatic Aberration on the Static Accommodative Response," *Vision Res.* **32**:823–832 (1992).
451. D. A. Atchison, N. C. Strang, and L. R. Stark, "Dynamic Accommodation Responses to Stationary Colored Targets," *Optom. Vis. Sci.* **81**:699–711 (2004).
452. R. Home and J. Poole, "Measurement of the Preferred Binocular Dioptric Settings at High and Low Light Level," *Optica Acta* **24**:97 (1977).
453. M. F. Wesner and R. J. Miller, "Instrument Myopia Conceptions, Misconceptions, and Influencing Factors," *Doc. Ophthalmol.* **62**:281–308 (1986).
454. G. G. Heath, "Components of Accommodation," *Am. J. Optom. Arch. Am. Acad. Optom.* **33**:569–579 (1956).
455. S. C. Hokoda and K. J. Ciuffreda, "Theoretical and Clinical Importance of Proximal Vergence and Accommodation," *Vergence Eye Movements: Basic and Clinical Concepts*, C. M. Schor and K. K. Ciuffreda (eds.), Butterworths, Boston, 1983, pp. 75–97.
456. G. Smith, K. C. Tan, and M. Letts, "Binocular Optical Instruments and Binocular Vision," *Clin. Exper. Optom.* **69**:137–144 (1986).
457. A. Duane, "Studies in Monocular and Binocular Accommodation with Their Clinical Applications," *Am. J. Ophthalmol. Ser. 3* **5**:865–877 (1922).
458. D. Hamasaki, J. Ong, and E. Marg, "The Amplitude of Accommodation in Presbyopia," *Am. J. Optom. Arch. Am. Acad. Optom.* **33**:3–14 (1956).
459. H. W. Hofstetter, "A Longitudinal Study of Amplitude Changes in Presbyopia," *Am. J. Optom. Arch. Am. Acad. Optom.* **42**:3–8 (1965).
460. C. Ramsdale and W. N. Charman, "A Longitudinal Study of the Changes in Accommodation Response," *Ophthalm. Physiol. Opt.* **9**:255–263 (1989).
461. W. N. Charman, "The Path to Presbyopia: Straight or Crooked?," *Ophthalm. Physiol. Opt.* **9**:424–430 (1989).
462. J. A. Mordi and K. J. Ciuffreda, "Static Aspects of Accommodation: Age and Presbyopia," *Vision Res.* **38**:1643–1653 (1998).
463. M. Kalsi, G. Heron, and W. N. Charman, "Changes in the Static Accommodation Response with Age," *Ophthalm. Physiol. Opt.* **21**:77–84 (2001).
464. G. Heron and W. N. Charman, "Accommodation as a function of age and the linearity of the Response Dynamics," *Vision Res.* **44**:3119–3130 (2004).
465. G. Heron, W. N. Charman, and C. Schor, "Dynamics of the Accommodation Response to Abrupt Changes in Target Vergence as a Function of Age," *Vision Res.* **41**:507–519 (2001).
466. J. A. Mordi and K. J. Ciuffreda, "Dynamic Aspects of Accommodation: Age and Presbyopia," *Vision Res.* **44**:591–601 (2004).
467. S. Kasthurirangan and A. Glasser, "Age Related Changes in Accommodation Dynamics in Humans," *Vision Res.* **46**:1507–1519 (2006).

468. G. Smith, "Schematic Eyes: History, Description and Applications," *Clin. Exp. Optom.* **78**:176–189 (1995).
469. D. A. Atchison and G. Smith, *Optics of the Human Eye*, Butterworth-Heinemann, Oxford, 2000, pp. 39–47 and 160–179.
470. R. B. Rabbetts, *Clinical Visual Optics*, 3rd ed., Butterworth-Heinemann, Oxford, 1998, pp. 207–229.
471. A. E. A. Ridgway, "Intraocular Lens Implants," *Vision and Visual Dysfunction*, vol. 1, *Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 120–137.
472. D. Sliney and M. Wolbarsht, *Safety with Lasers and Other Optical Sources*, Plenum, New York, 1980.
473. D. H. Sliney, "Measurement of Light and the Geometry of Exposure of the Human Eye," *Vision and Visual Dysfunction*, vol. 16, *The Susceptible Visual Apparatus*, J. Marshall (ed.), Macmillan, London 1991, pp. 23–29.
474. Y. Le Grand and S. G. El Hage, *Physiological Optics*, Springer-Verlag, Berlin, 1980, pp. 64–66.
475. H. H. Emsley, *Visual Optics*, vol. 1, 5th ed., Hatton Press, London, 1953.
476. J. W. Blaker, "Toward an Adaptive Model of the Human Eye," *J. Opt. Soc. Am.* **70**:220–223 (1980).
477. J. W. Blaker, "A Comprehensive Model of the Aging, Accommodative, Adult Eye," *Technical Digest on Ophthalmic and Visual Optics*, vol 2, Optical Society of America, Washington, D.C., 1991, pp. 28–31.
478. A. Popielek-Masajada and H. T. Kasprzak, "A New Schematic Eye Model Incorporating Accommodation," *Optom. Vis. Sci.* **76**:720–727 (1999).
479. W. Lotmar, "Theoretical Eye Model with Aspherics," *J. Opt. Soc. Am.* **61**:1522–1529 (1971).
480. A. C. Kooijman (1983) "Light Distribution of the Retina of a Wide-Angle Theoretical Eye," *J. Opt. Soc. Am.* **73**:1544–1550 (1983).
481. R. Navarro, J. Santamaria, and J. Bescós, "Accommodation-Dependent Model of the Human Eye with Aspherics," *J. Opt. Soc. Am. A* **2**:1273–1281 (1985).
482. M. C. M. Dunne and D. A. Barnes, "Schematic Modelling of Peripheral Astigmatism in Real Eyes," *Ophthalm. Physiol. Opt.* **7**:235–239 (1987).
483. S. Patel, J. Marshall, and F. W. Fitzke, "Model for Predicting the Optical Performance of the Eye in Refractive Surgery," *Refract. Corneal Surg.* **9**:366–375 (1993).
484. H-L Liou and N. A. Brennan, "Anatomically Accurate, Finite Model Eye for Optical Modeling," *J. Opt. Soc. Am. A* **14**:1684–1695 (1997).
485. I. Escudero-Sanz and R. Navarro, "Off-Axis Aberrations of a Wide-angle Schematic Eye Model," *J. Opt. Soc. Am. A* **16**:1881–1891 (1999).
486. Y-J. Liu, Z. Q. Wang, L. -P. Song, and G. -G. Mu, "An Anatomically Accurate Eye Model with a Shell-Structure Lens," *Optik* **116**:241–246 (2005).
487. A. V. Goncharov and C. Dainty, "Wide-Field Schematic Eye Models with Gradient-Index Lens," *J. Opt. Soc. Am. A* **24**:2157–2174 (2007).
488. L. N. Thibos, M. Ye, X. Zhang, and A. Bradley, "The Chromatic Eye: a New Reduced-eye Model of Ocular Chromatic Aberration in Humans," *Applied Opt.* **32**:3594–3600 (1992).
489. L. N. Thibos, M. Ye, X. Zhang, and A. Bradley, "Spherical Aberration of the Reduced Schematic Eye with Elliptical Refracting Surface," *Optom. Vis. Sci.* **74**:548–556 (1997).
490. Y. -Z. Wang and L. N. Thibos, "Oblique (Off-axis) Astigmatism of the Reduced Schematic Eye with Elliptical Refracting Surface," *Optom. Vis. Sci.* **74**:557–562 (1997).
491. D. A. Atchison, "Oblique Astigmatism of the Indiana Eye," *Optom. Vis. Sci.* **75**:247–248 (1998).
492. F. W. Campbell and D. G. Green, "Monocular versus Binocular Visual Acuity," *Nature* **208**:191–192 (1965).
493. M. Lombardo, G. Lombardo, and S. Serrao, "Interocular High-Order Corneal Wavefront Aberration Symmetry," *J. Opt. Soc. Am. A* **23**:777–787 (2006).
494. S. Marcos and S. A. Burns, "On the Symmetry between Eyes of Wavefront Aberration and Cone Directionality," *Vision Res.* **40**:2437–2447 (2000).
495. M. J. Collins and A. S. Bruce, "Factors Influencing Performance with Monovision," *J. Brit. Contact Lens Assoc.* **17**:83–89 (1994).
496. J. Meyler, "Presbyopia," *Contact Lens Practice*, N. Efron (ed.), Butterworth-Heinemann, Oxford, 2002, pp. 261–274.

497. B. J. W. Evans, "Monovision: A Review," *Ophthal. Physiol. Opt.* **27**:417–439 (2007).
498. C. M. Schor and M. C. Flom, "The Relative Values of Stereopsis as a Function of Viewing Distance," *Am. J. Optom. Arch. Am. Acad. Optom.* **46**:805–809 (1969).
499. R. S. Harvey, "Some Statistics of Interpupillary Distance," *Optician* **184**(4766):29 (1982).
500. R. Sekuler and R. Blake, *Perception*, 5th ed., McGraw-Hill, New York, 2005.
501. N. A. Valyus, *Stereoscopy*, Focal Press, London, 1966.
502. A. Lit, "Depth Discrimination Thresholds as a Function of Binocular Differences of Retinal Illuminance at Scotopic and Photopic Levels," *J. Opt. Soc. Am.* **49**:746–752 (1959).
503. S. C. Rawlings and T. Shipley, "Stereoscopic Acuity and Horizontal Angular Distance from Fixation," *J. Opt. Soc. Am.* **59**:991–993 (1969).
504. W. P. Dwyer and A. Lit, "Effect of Luminance-matched Wavelength on Depth Discrimination at Scotopic and Photopic Levels of Target Illuminance," *J. Opt. Soc. Am.* **60**:127–131 (1970).
505. A. Arditi, "Binocular Vision," *Handbook of Perception and Human Performance*, vol. 1, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), Wiley, New York, 1986, Chapter 23.
506. R. R. Fagin and J. R. Griffin, "Stereoacuity Tests: Comparison of Mathematical Equivalents," *Am. J. Optom. Physiol. Opt.* **59**:427–438 (1982).
507. R. B. Rabbetts, *Clinical Visual Optics*, 4th ed., Elsevier, Oxford, 2007, Chapter 11.
508. G. Heron, S. Dholakia, D. E. Collins, and H. McLaughlan, "Stereoscopic Thresholds in Children and Adults," *Am. J. Optom. Physiol. Opt.* **62**:505–515 (1985).
509. H. J. Howard, "A Test for Judgement of Distance," *Am. J. Ophthalmol.* **2**:656–675 (1919).
510. G. Westheimer, "Effect of binocular Magnification Devices on Stereoscopic Depth Resolution," *J. Opt. Soc. Am.* **45**:278–280 (1956).
511. W. N. Charman and J. A. M. Jennings, "Binocular Vision in Relation to Stereoscopic Instruments and Three-Dimensional Displays," *Vision and Visual Dysfunction*, vol. 1, *Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 326–344.
512. D. B. Diner and D. H. Fender, *Human Engineering in Stereoscopic Viewing Devices*, Plenum Press, New York, 1993, pp. 49–65.
513. G. Smith and D. A. Atchison, *The Eye and Visual Optical Instruments*, Cambridge UP, Cambridge, 1997, pp. 727–746.
514. E. Peli, "Optometric and Perceptual Issues with Head-Mounted Displays," *Visual Instrumentation: Optical Design and Engineering Principles*, P. Mouroulis (ed.), McGraw-Hill, New York, 1999, pp. 205–276.
515. T. L. Williams, "Testing of Visual Instrumentation," *Visual Instrumentation: Optical Design and Engineering Principles* P. Mouroulis (ed.), McGraw-Hill, New York, 1999, pp. 353–421.
516. J. A. M. Jennings, "Binocular Vision through Correcting Lenses: Aniseikonia," *Vision and Visual Dysfunction*, vol. 1, *Visual Optics and Instrumentation*, W. N. Charman (ed.), Macmillan, London, 1991, pp. 163–182.
517. M. A. Taylor Kulp, T. W. Raasch, and M. Polasky, "Patients with Anisometropia and Aniseikonia," *Clinical Refraction*, W. J. Benjamin (ed.), Saunders, Philadelphia, 1998, pp. 1134–1159.
518. B. Winn, R. G. Ackerley, C. A. Brown, F. K. Murray, J. Prais, and M. F. St John, "Reduced Aniseikonia in Axial Ametropia with Contact Lens Correction," *Ophthal. Physiol. Opt.* **8**:341–344 (1988).
519. A. L. Yarbus, *Eye Movements and Vision*, Plenum, New York, 1967.
520. M. Land, N. Mennie, and J. Rusted, "The Roles of Vision and Eye Movements in the Control of Activities of Daily Living," *Perception* **28**:1311–1328 (1999).
521. M. A. Alpern, "Movements of the Eyes," *The Eye*, vol. 3, *Muscular Mechanisms*, H. Davson (ed.), Academic Press, New York, 1969, pp. 5–214.
522. R. H. S. Carpenter, *Movements of the Eyes*, 2nd ed., Pion, London, 1988.
523. R. H. S. Carpenter (ed.), *Vision and Visual Dysfunction*, vol. 8, *Eye Movements*, Macmillan, London, 1991.
524. R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*, 4th ed., Oxford UP, Oxford, 2006.
525. G. Fry and W. W. Hill, "The Mechanics of Elevating the Eye," *Am. J. Optom. Arch. Am. Acad. Optom.* **40**:707–716 (1963).

526. C. Rashbass and G. Westheimer, "Independence of Conjunctive and Disjunctive Eye Movements," *J. Physiol. (London)* **159**:361–364 (1961).
527. D. A. Robinson, "The Mechanics of Human Saccadic Eye Movements," *J. Physiol. (London)* **174**:245–264 (1964).
528. A. T. Bahill, A. Brockenbrough, and B. T. Troost, "Variability and Development of a Normative Data Base for Saccadic Eye Movements," *Invest. Ophthalmol.* **21**:116–125 (1981).
529. R. H. S. Carpenter, "The Neural Control of Looking," *Current Biology* **10**(8):R291–R293 (2000).
530. F. W. Campbell and R. H. Wurtz, "Saccadic Omission: Why We Do Not See a Grey Out during a Saccadic Eye Movement," *Vision Res.* **18**:1297–1303 (1978).
531. H. Collewijn and E. P. Tamminga, "Human Smooth Pursuit and Saccadic Eye Movements during Voluntary Pursuit of Different Target Motions on Different Backgrounds," *J. Physiol. (London)* **351**:217–250 (1984).
532. C. Rashbass and G. Westheimer, "Disjunctive Eye Movements," *J. Physiol. (London)* **159**:339–360 (1961).
533. C. J. Erkelers, J. van der Steen, R. M. Steinman, and H. Collewijn, "Ocular Vergence under Natural Conditions. I Continuous Changes of Target Distance Along the Median Plane," *Proc. R. Soc. Lond.* **B236**:417–440 (1989).
534. C. J. Erkelers, R. M. Steinman, and H. Collewijn, "Ocular Vergence Under Natural Conditions. II Gaze Shifts between Real Targets Differing in Distance and Direction," *Proc. R. Soc. Lond.* **B236**:441–465 (1989).
535. J. T. Enright, "Perspective Vergence: Oculomotor Responses to Line Drawings," *Vision Res.* **27**:1513–1526 (1987).
536. A. E. Kertesz, "Vertical and Cyclofusional Disparity Vergence," *Vergence Eye Movements: Clinical and Applied*, C. M. Schor and K. J. Ciuffreda (eds.), Butterworths, Boston, 1983, pp. 317–348.
537. J. M. Findlay, "Frequency Analysis of Human Involuntary Eye Movement," *Kybernetik* **8**:207–214 (1971).
538. R. W. Ditchburn, *Eye Movements and Visual Perception*, Clarendon Press, Oxford, 1973.
539. J. Nachmias, "Determinants of the Drift of the Eye During Monocular Fixation," *J. Opt. Soc. Am.* **51**:761–766 (1961).
540. J. Nachmias, "Two-dimensional motion of the retinal image during monocular fixation," *J. Opt. Soc. Am.* **49**:901–908 (1959).
541. H. C. Bennet-Clark, "The Oculomotor Response to Small Target Displacements," *Optica Acta* **11**:301–314 (1964).
542. W. H. Marshall and S. A. Talbot, "Recent Evidence for Neural Mechanisms in Vision Leading to a General Theory of Sensory Acuity," *Biol. Symp.* **7**:117–164 (1942).
543. O. Packer and D. R. Williams, "Blurring by Fixational Eye Movements," *Vision Res.* **32**:1931–1939 (1992).

*This page intentionally left blank*

---

# VISUAL PERFORMANCE

---

## Wilson S. Geisler

*Department of Psychology  
University of Texas  
Austin, Texas*

## Martin S. Banks

*School of Optometry  
University of California  
Berkeley, California*

---

### 2.1 GLOSSARY

---

$A$	amplitude
$a$	interpupillary distance
$A_p$	effective area of the entrance pupil
$C$	contrast
$c_o$	maximum concentration of photopigment
$d$	horizontal disparity
$d_e$	distance from the image plane to the exit pupil of an optical system
$d_\theta$	average disparity between two points and the convergence point
$E(\lambda)$	photopic spectral illuminance distribution
$E_e(\lambda)$	spectral irradiance distribution
$f$	spatial frequency of a sinusoid
$I_o$	half-bleaching constant
$J_o$	maximum photocurrent
$l$	length of outer segment
$L(\lambda)$	photopic spectral luminance distribution
$L_e(\lambda)$	spectral radiance distribution
$m$	magnification of the exit pupil relative to the actual pupil
$N$	total effective photons absorbed per second
$n_r$	index of refraction of the media where the image plane is located
$n(\lambda)$	spectral photon-flux irradiance distribution
$p$	proportion of unbleached photopigment
$t_o$	time constant of photopigment regeneration
$t(\lambda)$	transmittance function of the ocular media
$V(\lambda)$	standard photopic spectral sensitivity function of the human visual system



$\alpha(\lambda)$	absorptance spectrum
$\Delta C$	contrast increment or decrement
$\Delta f$	frequency increment or decrement
$\Delta z$	distance between any pair of points in the depth dimension
$\varepsilon$	retinal eccentricity
$\varepsilon(\lambda)$	extinction spectrum
$\theta$	convergence angle of the eyes
$\theta$	orientation of a sinusoid
$\kappa$	collection area, or aperture, of a photoreceptor
$\lambda$	wavelength of the light in a vacuum
$\xi$	isomerization efficiency
$\Sigma$	covariance matrix for a gaussian noise process
$\sigma_o$	half-saturation constant
$\tau$	time interval
$\tau_{\text{opt}}$	optimum time interval
$\Phi(\cdot)$	cumulative standard normal probability function
$\phi$	phase of a sinusoid

## 2.2 INTRODUCTION

---

Physiological optics concerns the study of (1) how images are formed in biological eyes, (2) how those images are processed in the visual parts of the nervous system, and (3) how the properties of image formation and neural processing manifest themselves in the perceptual performance of the organism. The previous chapter reviewed image formation; this chapter briefly describes the neural processing of visual information in the early levels of the human visual system, and summarizes, somewhat more extensively, what is known about human visual performance.

An enormous amount of information about the physical environment is contained in the light reaching the cornea of the eye. This information is critical for many of the tasks the human observer must perform, including identification of objects and materials, determination of the three-dimensional structure of the environment, navigation through the environment, prediction of object trajectories, manipulation of objects, and communication with other individuals. The performance of a human observer in a given visual task is limited by the amount of information available in the light at the cornea and by the amount (and type) of information encoded and transmitted by the successive stages of visual processing. This chapter presents a concise description of visual performance in a number of fundamental visual tasks. It also presents a concise description of the physiological and psychological factors believed to underlie performance in those tasks. Two major criteria governed the selection of the material to be presented. First, we attempted to focus on quantitative data and theories that should prove useful for developing rigorous models or characterizations of visual performance. Second, we attempted to focus on data and theories that have a firm empirical basis, including at least some knowledge of the underlying biological mechanisms.

## 2.3 OPTICS, ANATOMY, PHYSIOLOGY OF THE VISUAL SYSTEM

---

### Image Formation

The processing of visual information begins with the optics of the eye, which consists of three major components: the cornea, pupil, and lens. The optical components are designed to form a sharp image at the layer of the photoreceptors in the retina. Neil Charman (Chap. 1) discusses many of

the details concerning how these optical components affect image quality. We briefly describe a few of the most useful formulas and methods for computing the approximate size, location, quality, and intensity of images formed at the photoreceptors. Our aim is to provide descriptions of image formation that might prove useful for developing models or characterizations of performance in perceptual tasks.

The sizes and locations of retinal images can be found by projecting points on the objects along straight lines through the image (posterior) nodal point until they intersect the retinal surface. The intersections with the retinal surface give the image locations corresponding to the points on the objects. The angles between pairs of projection lines are *visual angles*. Image locations are usually described in terms of the visual angles that projection lines make with respect to a reference projection line (the *visual axis*), which passes through the nodal point and the center of the fovea. The radial visual angle between a projection line from an object and the visual axis is the object's *eccentricity*. Image sizes are often described by the visual angles between key points on the object. For purposes of computing the size and location of images, it is usually sufficient to use a simplified model of the eye's optics, such as the *reduced eye*, which consists of a single spherical refracting surface (radius of curvature = 5.5 mm) and a retinal surface located 16.7 mm behind the nodal point (see Fig. 2c, Chap. 1).

A general method for computing the quality of retinal images is by convolution with a point-spread function  $h(x, y)$ . Specifically, if  $o(x, y)$  is the luminance (or radiance) of the object, and  $i(x, y)$  is the image illuminance (or irradiance), then

$$i(x, y) = o(x, y) ** h(x, y) \quad (1)$$

where  $**$  represents the two-dimensional convolution operator. The shape of the point-spread function varies with wavelength and with retinal location. The precise way to deal with wavelength is to perform a separate convolution for each wavelength in the spectral luminance distribution of the object. In practice, it often suffices to convolve with a single point-spread function, which is the weighted average, across wavelength, of the monochromatic point-spread functions, where the weights are given by the shape of the spectral luminance distribution. To deal with retinal location, one can make use of the fact that the human point-spread function changes only gradually with retinal eccentricity out to about 20 deg;<sup>1</sup> thus, a large proportion of the visual field can be divided into a few annular regions, each with a different point-spread function.

Calculation of the point-spread function is normally accomplished by finding the transfer function,  $H(u, v)$ .<sup>2,3</sup> (The point-spread function can be obtained, if desired, by an inverse Fourier transform.) The transfer function is given by the autocorrelation of the generalized pupil function followed by normalization to a peak value of 1.0:

$$T(u, v) = p(x, y) e^{iW(x, y, \lambda)} \otimes p(x, y) e^{-iW(x, y, \lambda)} \Big|_{x=(\lambda d_e / mn), u, y=(\lambda d_e / mn), v} \quad (2)$$

$$H(u, v) = \frac{T(u, v)}{T(0, 0)} \quad (3)$$

where  $\lambda$  is the wavelength of the light in a vacuum,  $n_e$  is the index of refraction of the media where the image plane is located,  $d_e$  is the distance from the image plane to the exit pupil of the optical system, and  $m$  is the magnification of the exit pupil relative to the actual pupil. The generalized pupil function is the product of the simple pupil function,  $p(x, y)$  (transmittance as a function of position within the actual pupil), and the aberration function,  $e^{iW(x, y, \lambda)}$ . The exit pupil is the apparent pupil, when viewed from the image plane. The size of and distance to the exit pupil can be found by ray-tracing a schematic eye. For the Le Grand eye, the relevant parameters are approximately as follows:  $m = 1.03$ ,  $d_e = 20.5$  mm, and  $n = 1.336$ . Average values of the monochromatic and chromatic aberrations are available (see Chap. 1) and can be used as estimates of  $W(x, y, \lambda)$ .

Equation (3) can be used to compute approximate point-spread functions (and hence image quality) for many stimulus conditions. However, for some conditions, direct measurements (or psychophysical measurements) of the point-spread function are also available (see Chap. 1), and are easier to deal with. For broadband (white) light and a well-accommodated eye, the axial point-spread functions directly measured by Campbell and Gubisch<sup>4</sup> are representative. A useful set

of monochromatic point-spread functions was measured at various eccentricities by Navarro et al.<sup>1</sup> These point-spread functions can be used directly in Eq. (1) to compute approximate image quality.

The approximate retinal irradiance for extended objects is given by the following formula:

$$E_e(\lambda) = \frac{A_p}{278.3} L_e(\lambda) t(\lambda) \quad (4)$$

where  $E_e(\lambda)$  is the retinal spectral irradiance distribution ( $\text{watts} \cdot \text{m}^{-2} \cdot \text{nm}^{-1}$ ),  $L_e(\lambda)$  is the spectral radiance distribution of the object ( $\text{watts} \cdot \text{m}^{-2} \cdot \text{sr}^{-1} \cdot \text{nm}^{-1}$ ),  $t(\lambda)$  is the transmittance of the ocular media (see Chap. 1), and  $A_p$  is the effective area of the entrance pupil ( $\text{mm}^2$ ). [Note,  $A_p = \iint p(x/m', y/m') dx dy$ , where  $m'$  is magnification of the entrance pupil relative to the actual pupil; the entrance pupil is the apparent size of the pupil when viewed from outside the eye.]

Photopic retinal illuminance,  $E(\lambda)$  ( $\text{candelas} \cdot \text{nm}^{-1}$ ), is computed by an equivalent formula where the spectral radiance distribution,  $L_e(\lambda)$ , is replaced by the spectral luminance distribution,  $L(\lambda)$  ( $\text{candelas} \cdot \text{m}^{-2} \cdot \text{nm}^{-1}$ ), defined by

$$L(\lambda) = 683 V(\lambda) L_e(\lambda) \quad (5)$$

where  $V(\lambda)$  is the standard photopic spectral sensitivity function of the human visual system.<sup>5</sup>

In theoretical calculations, it is often useful to express light levels in terms of photon flux rather than in terms of watts. The photon-flux irradiance on the retina,  $n(\lambda)$  ( $\text{quanta} \cdot \text{sec}^{-1} \cdot \text{deg}^2 \cdot \text{nm}^{-1}$ ) is computed by multiplying the retinal irradiance,  $E_e(\lambda)$ , by  $8.4801 \times 10^{-8}$  which converts  $\text{m}^2$  to  $\text{deg}^2$  (based upon the *reduced eye*), and by  $\lambda/ch$  which converts watts into quanta/sec (where  $c$  is the speed of light in a vacuum, and  $h$  is Planck's constant). Thus,

$$n(\lambda) = 1.53 \times 10^6 A_p L_e(\lambda) t(\lambda) \lambda \quad (6)$$

and, by substitution of Eq. (5) into Eq. (6),

$$n(\lambda) = 2.24 \times 10^3 A_p \frac{L(\lambda)}{V(\lambda)} t(\lambda) \lambda \quad (7)$$

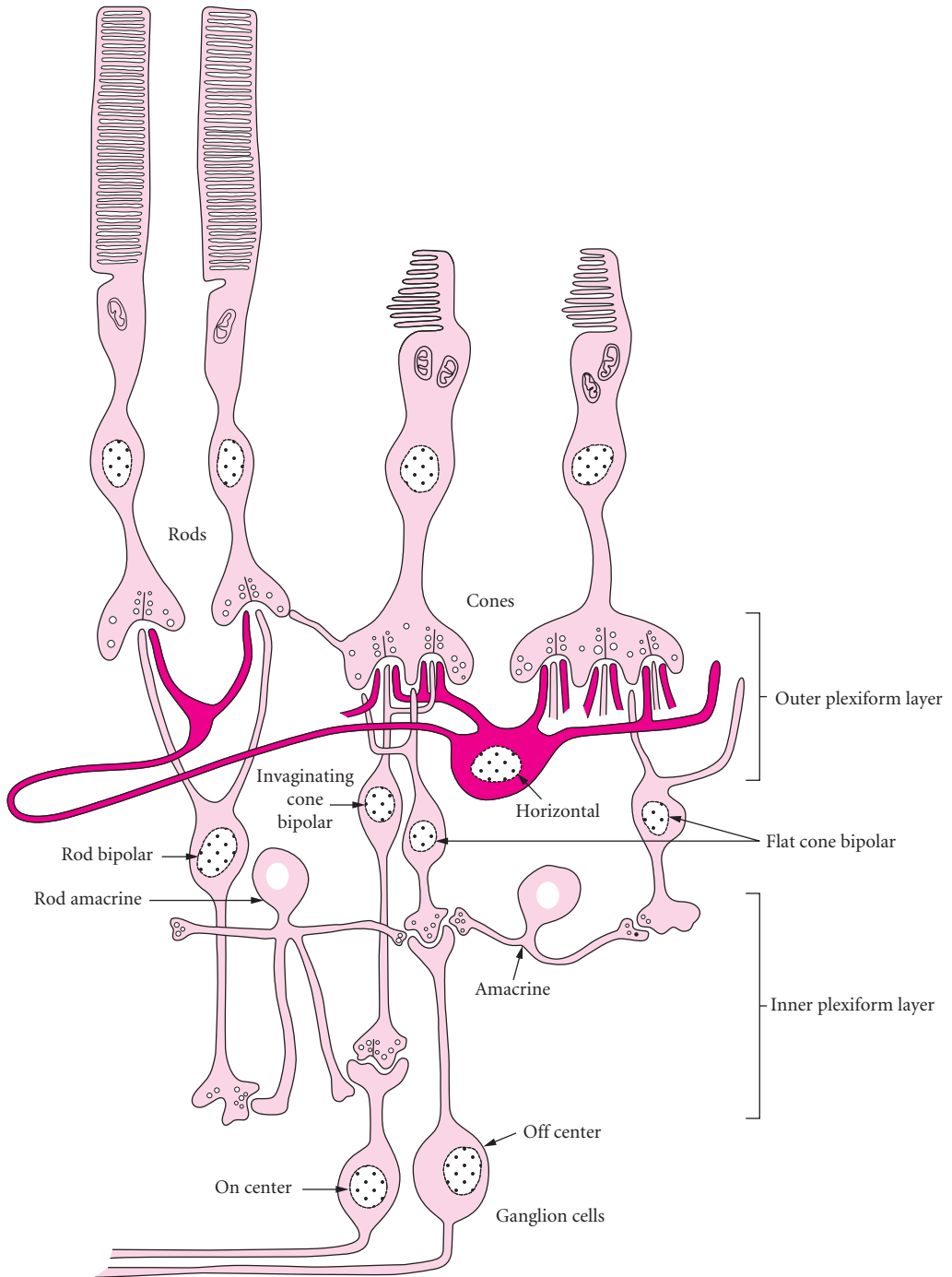
Most light-measuring devices report radiance,  $L_e(\lambda)$ , or luminance,  $L(\lambda)$ ; Eqs. (6) and (7) allow conversion to retinal photon-flux irradiance,  $n(\lambda)$ . For more details on the calculation of retinal intensity, see Wyszecki and Stiles.<sup>5</sup>

## Image Sampling by the Photoreceptors

The image formed at the receptor layer is described by a four-dimensional function  $n(x, y, t, \lambda)$ , which gives the mean photon-flux irradiance ( $\text{quanta} \cdot \text{sec}^{-1} \cdot \text{deg}^2 \cdot \text{nm}^{-1}$ ) as a function of space ( $x, y$ ), time ( $t$ ), and wavelength ( $\lambda$ ). This four-dimensional function also describes the photon noise in the image. Specifically, photon noise is adequately described as an inhomogeneous Poisson process; thus, the variance in the number of photons incident in a given interval of space, time, and wavelength is equal to the mean number of photons incident in that same interval.

The photoreceptors encode the (noisy) retinal image into a discrete representation in space and wavelength, and a more continuous representation in time. The image sampling process is a crucial step in vision that can, and often does, result in significant information loss. The losses occur because physical and physiological constraints make it impossible to sample all four dimensions with sufficiently high resolution.

As shown in the schematic diagram in Fig. 1, there are two major types of photoreceptors: rods and cones. They play very different functional roles in vision; rods subservise vision at low light levels and cones at high light levels. There are three types of cones, each with a different spectral sensitivity (which is the result of having different photopigments in the outer segment). The "long" (L), "middle" (M), and "short" (S) wavelength cones have peak spectral sensitivities at wavelengths of approximately 570, 540, and 440 nm, respectively. Information about the spectral wavelength distribution of

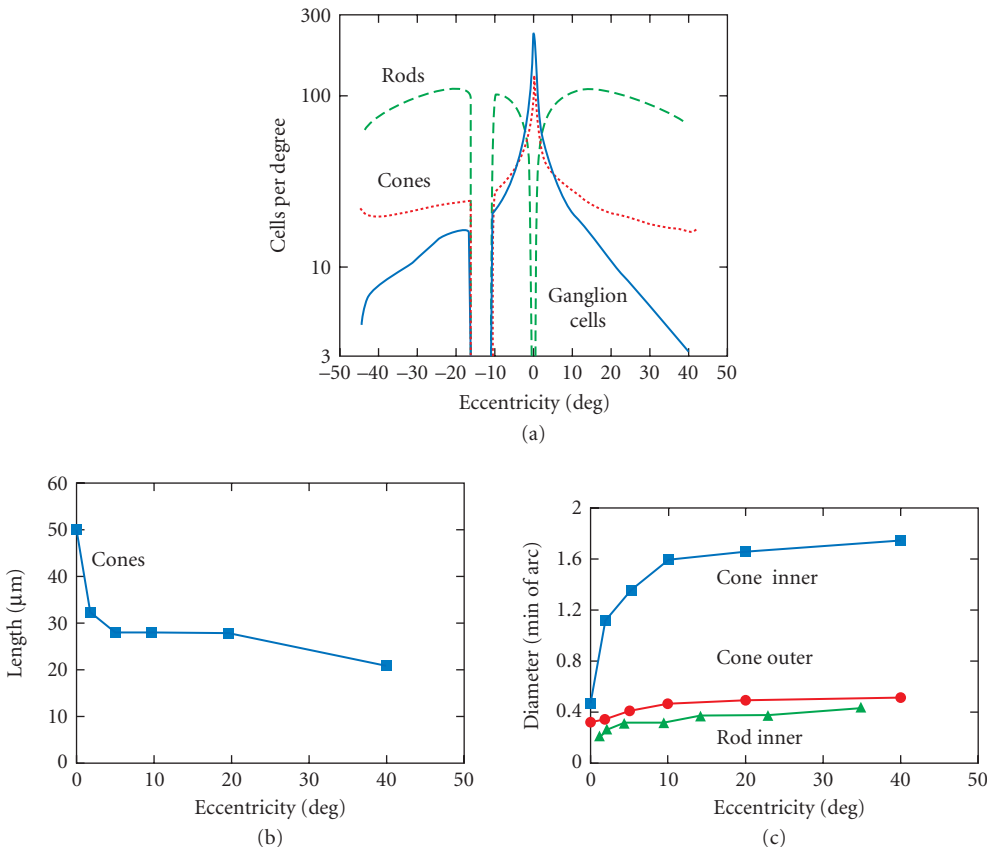


**FIGURE 1** Schematic diagram of the retinal neurons and their major synaptic connections. Note that rods, rod bipolar cells, and rod amacrine cells are absent in the fovea. (From Ref. 263.)

the light falling on the retina is encoded by the relative activities of the L, M, and S cones. All rods have the same spectral sensitivity (and the same photopigment), peaking at about 500 nm.

The quality of spatial, temporal, and wavelength information encoded by the photoreceptors depends upon: (1) the spatial distribution of the photoreceptors across the retina, (2) the efficiency with which individual photoreceptors absorb light at different wavelengths (the absorbance spectrum), (3) the area over which the individual photoreceptors collect light (the receptor aperture), and (4) the length of time over which the individual photoreceptors integrate light.

The spatial distribution of cones and rods is highly nonuniform. Figure 2a shows the typical density distribution of rod and cone photoreceptors across the retina (although there are individual differences; e.g., see Ref. 6). Cone density decreases precipitously with eccentricity; rods are absent in the fovea and reach a peak density at 20 deg. If the receptor lattice were perfectly regular, the highest unambiguously resolvable spatial frequency (the Nyquist limit), would be half the linear density (in cells  $\cdot$  deg $^{-1}$ ). Under normal viewing conditions, the Nyquist limit does not affect vision in the fovea because the eye's optics eliminate spatial frequencies at and above the limit. However, the



**FIGURE 2** (a) Linear density of cones, rods, and ganglion cells as a function of eccentricity in the human retina. (The data were modified from Refs. 6 and 32.) Conversion from cells/mm $^2$  to cells/deg $^2$  was computed assuming a posterior nodal point 16.68 mm from the retina, and a retinal radius of curvature of 12.1 mm. Conversion to cells/deg was obtained by taking the square root of areal density. Ganglion cell density in the central 10 deg was derived assuming a 3:1 ratio of ganglion cells to cones in the fovea.<sup>32</sup> (b) Human cone outer segment length. (Modified from Ref. 125.) (c) Human cone inner segment, cone outer segment, and rod diameter as a function of eccentricity. (Modified from Ref. 125.)

presentation of interference fringes (which avoid degradation by the eye's optics) yields visible spatial aliasing for spatial frequencies above the Nyquist limit.<sup>7-9</sup>

The densities and retinal distributions of the three cone types are quite different. The S cones form a rather regular lattice comprising less than 2 percent of cones in the central fovea and somewhat less than 10 percent of the cones elsewhere;<sup>10-12</sup> they may be absent in the central 20' to 25' of the fovea.<sup>13</sup> It is much more difficult to distinguish individual L and M cones anatomically, so their densities and distributions are less certain. Psychophysical evidence indicates that the ratio of L to M cones is approximately 2:1,<sup>14,15</sup> but the available physiological data in monkey suggest a ratio closer to 1:1.<sup>16,17</sup>

From the Beer-Lambert law, the absorbance spectrum of a receptor depends upon the concentration of the photopigment,  $c_o p$ , in the receptor outer segment, the length of the outer segment,  $l$ , and the extinction spectrum,  $\epsilon(\lambda)$ , of the photopigment,

$$\alpha(\lambda) = 1 - 10^{-lc_o p \epsilon(\lambda)} \quad (8)$$

where  $c_o$  is the concentration of photopigment in the dark-adapted eye and  $p$  is proportion of unbleached photopigment. The specific absorbance spectra for the different classes of receptor are described in Chap. 11 by Andrew Stockman and David H. Brainard. At the peak wavelength, the photoreceptors absorb approximately 50 percent of the incident photons (although there are some variations with eccentricity).

Because of photopigment bleaching and regeneration processes within the photoreceptor, the absorption spectra of the photopigments change depending upon the history of photon absorptions. For example, prolonged exposure to high intensities depletes a significant fraction of the available photopigment, reducing the overall optical density. Reflection densitometry measurements<sup>18-20</sup> have shown, to first approximation, that the proportion  $p$  of available photopigment at a given point in time is described by a first-order differential equation:

$$\frac{dp}{dt} = \frac{n(\lambda)\xi(1 - 10^{-lc_o \epsilon(\lambda)p})}{lc_o} + \frac{1-p}{t_o} \quad (9)$$

where  $n(\lambda)$  is the photon-flux irradiance distribution ( $\text{quanta} \cdot \text{sec}^{-1} \cdot \text{deg}^{-2} \cdot \text{nm}^{-1}$ ),  $t_o$  is the exponential time constant of regeneration, and  $\xi$  is the isomerization efficiency (which is believed to be near 1.0). For broadband light, Eq. (9) simplifies and can be expressed in terms of retinal illumination:<sup>19</sup>

$$\frac{dp}{dt} = -\frac{Ip}{Q_e} + \frac{(1-p)}{t_o} \quad (10)$$

where  $I$  is the retinal illumination (in trolands\*) and  $Q_e$  is the energy of a flash (in troland · sec) required to reduce the proportion of unbleached photopigment to  $1/e$ . For the cone photopigments, the time constant of regeneration is approximately 120 sec, and  $Q_e$  (for broadband light) is approximately  $2.4 \times 10^6$  troland · sec. For rods, the time constant of regeneration is approximately 360 sec, and  $Q_e$  is approximately  $1.0 \times 10^7$  scotopic troland · sec. Equation (10) implies that the steady-state proportion of available photopigment is approximately

$$p(I) = \frac{I_o}{I_o + I} \quad (11)$$

where  $I_o = Q_e/t_o$ . ( $I_o$  is known as the *half-bleaching constant*.) Equation (11) is useful when computing photon absorptions at high ambient light levels; however, photopigment bleaching is not significant at low to moderate light levels. Equation (7) also implies that bleaching and regeneration can produce changes in the shapes of the absorbance spectra (see Chap. 10).

The lengths of the photoreceptor outer segments (and hence their absorbances) change with eccentricity. Figure 2b shows that the cone outer segments are longer in the fovea. Rod outer segments are approximately constant in length. If the dark-adapted photopigment concentration,  $c_o$ , is constant

\*The troland is defined to be the retinal illumination produced by viewing a surface with a luminance of 1 cd/m<sup>2</sup> through a pupil with an area of 1 mm<sup>2</sup>.

in a given class of photoreceptor (which is likely), then increasing outer segment length increases the amount of light collected by the receptor (which increases signal-to-noise ratio).

The light collection area of the cones is believed to be approximately 70 to 90 percent of the cross-sectional area of the inner segment at its widest point (see Fig. 1), although direct measurements are not available (see Refs. 21 and 22). The collection area of the rod receptors is equal to the cross-sectional area of the inner and outer segments (which is the same). Figure 2c plots inner-segment diameter as a function of eccentricity for cones and rods. As can be seen, the cone aperture increases with eccentricity, while the rod aperture is fairly constant. Increasing the cone aperture increases the light collected (which increases signal-to-noise ratio), but slightly reduces contrast sensitivity at high spatial frequencies (see later).

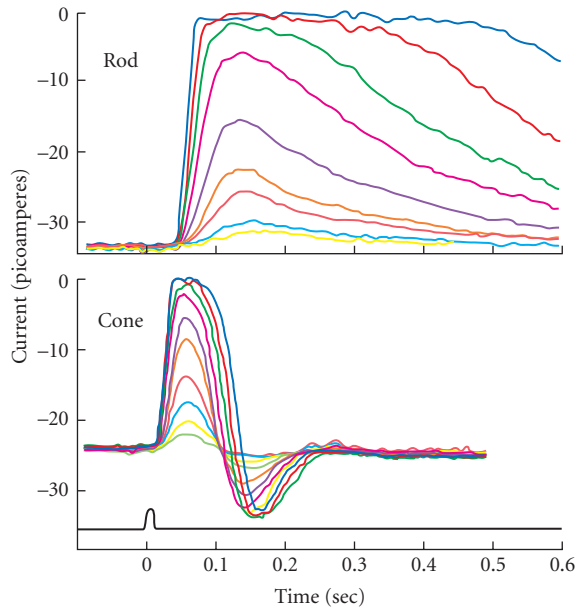
The data and formulas above [Eqs. (6) to (10)] can be used to compute (approximately) the total effective photons absorbed per second,  $N$ , in any receptor, at any eccentricity:

$$N = \int \kappa \xi \alpha(\lambda) n(\lambda) d\lambda \quad (12)$$

where  $\kappa$  is the collection area, or aperture, of the receptor (in  $\text{deg}^2$ ).

Photons entering through the pupillary center have a greater chance of being absorbed in the cones than photons entering through the pupillary margins. This phenomenon, known as the Stiles-Crawford effect, can be included in the above calculations by modifying the pupil function [Eqs. (3) and (4); also see Chaps. 1 and 8]. We also note that the cones are oriented toward the exit pupil.

The temporal response properties of photoreceptors are more difficult to measure than their spatial integration properties. Figure 3 shows some physiological measurements of the photocurrent



**FIGURE 3** Photocurrent responses of macaque rod and cone photoreceptors to flashes of light. Each trace represents the response to a different intensity level. The flash intensities were varied by factors of two. (From Ref. 23. Copyright © 1989 by Scientific American, Inc. All rights reserved.)

responses of primate (macaque) rods and cones at low and moderate light levels.<sup>23</sup> Rods integrate photons over a substantially longer duration than cones do. In addition, rods produce reliable responses to single photon absorptions, but cones do not.

Photoreceptors, like all neurons, have a limited dynamic range. Direct recordings from receptors in macaque have shown that the peak responses of photoreceptors to brief flashes of light are adequately described by either a modified Michaelis-Menton function

$$f(z) = \frac{J_o z^n}{z^n + \sigma_o^n} \quad (13)$$

or an exponential function

$$f(z) = J_o (1 - 2^{-z^n / \sigma_o^n}) \quad (14)$$

where  $J_o$  is the maximum photocurrent,  $\sigma_o$  is the half-saturation constant (the value of  $z$  that produces exactly half the maximum response), and  $n$  is an exponent that has a value near 1.0.<sup>17,24,25</sup> As the above equations imply, the photoreceptors respond approximately linearly to transient intensity changes at low to moderate light levels, but respond nonlinearly (ultimately reaching saturation) at higher light levels. Recent electroretinogram (ERG) measurements suggest that the flash response of human photoreceptors is similar to those of the macaque.<sup>26</sup>

The nonlinear response saturation suggests that the receptors should not be able to encode high image intensities accurately (because any two intensities that saturate receptor response will be indistinguishable). However, the cones avoid saturation effects under many circumstances by adjusting their gain ( $\sigma_o$ ) depending upon the ambient light level; the rods do not adjust their gain nearly as much. The gain adjustment is accomplished in a few ways. Photopigment depletion (see earlier) allows multiplicative gain changes, but only operates at very high ambient light levels. Faster-acting mechanisms in the phototransduction sequence are effective at moderate to high light levels.<sup>27</sup> At this time, there remains some uncertainty about how much gain adjustment (adaptation) occurs within the cones.<sup>17,24,28</sup>

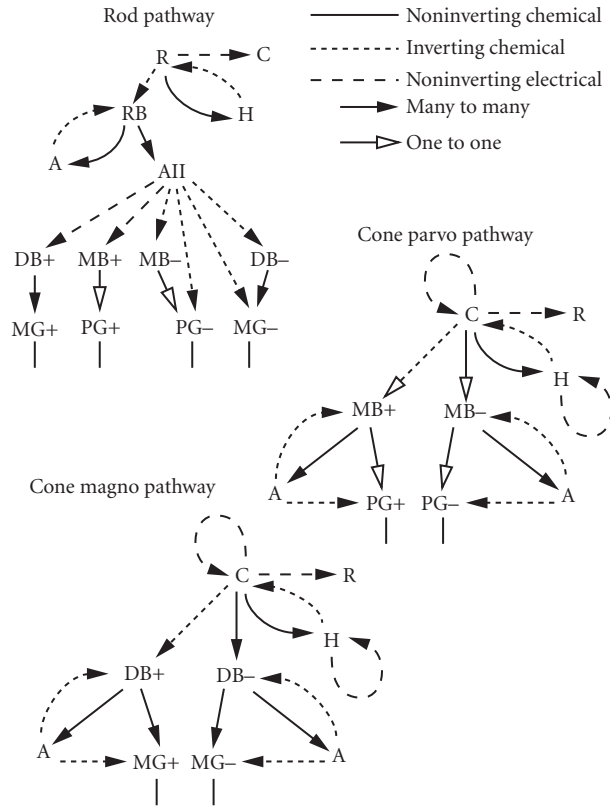
## Retinal Processing

The information encoded by the photoreceptors is processed by several layers of neurons in the retina. The major classes of retinal neuron in the primate (human) visual system and their typical interconnections are illustrated schematically in Fig. 1. Although the retina has been studied more extensively than any other part of the nervous system, its structure and function are complicated and not yet fully understood. The available evidence suggests that the primate retina is divided into three partially overlapping neural pathways: a *rod pathway*, a *cone parvo pathway*, and a *cone magno pathway* (for reviews see Refs. 29–31). These pathways, illustrated in Fig. 4, carry most of the information utilized in high-level visual processing tasks. There are other, smaller pathways (which will not be described here) that are involved in functions such as control of the pupil reflex, accommodation, and eye movements. Although Fig. 4 is based upon the available evidence, it should be kept in mind that there remains considerable uncertainty about some of the connections.

**Retinal Anatomy** As indicated in Fig. 4, the photoreceptors (R, C) form electrical synapses (gap-junctions) with each other, and chemical synapses with *horizontal cells* (H) and *bipolar cells* (RB, MB+, MB–, DB+, DB–). The electrical connections between receptors are noninverting\* and hence produce simple spatial and temporal summation. The dendritic processes and the axon-terminal processes of the horizontal cells provide spatially extended, negative-feedback connections from cones onto cones and from rods onto rods, respectively. It is likely that the horizontal cells play an important role in creating the surround response of ganglion cells. Rod responses probably do not influence cone responses (or vice versa) via the horizontal cells.

\*By *noninverting* we mean that changes in the response of the presynaptic neuron, in a given direction, produce changes in the response of the postsynaptic neuron in the same direction.





**FIGURE 4** Schematic diagram of neural connections in the primate retina. R (rod), C (cone), H (horizontal cell), RB (rod bipolar), MB+ (midget on-bipolar), MB- (midget off-bipolar), DB+ (diffuse on-bipolar), DB- (diffuse off-bipolar), A (amacrine), AII (rod amacrine), PG+ (P or midget on-center ganglion cell), PG- (P or midget off-center ganglion cell), MG+ (M or parasol on-center ganglion cell), MG- (M or parasol off-center ganglion cell).

In the rod pathway, rod bipolar cells (RB) form noninverting chemical synapses with *amacrine* cells (AII, A). The rod amacrine cells (AII) form electrical (noninverting) synapses with the on-center bipolar cells (MB+, DB+), and form chemical (inverting) synapses with off-center bipolar cells (MB-, DB-) and with off-center ganglion cells (PG-, MG-). Other amacrine cells (A) form reciprocal negative-feedback connections to the rod bipolar cells.

In the cone pathways, on-bipolar cells (MB+, DB+) form chemical synapses with amacrine cells (A) and with on-center *ganglion* cells (PG+, MG+).<sup>\*</sup> Off-bipolar cells (FMB, FDB) form chemical synapses with amacrine cells and off-center ganglion cells (PG-, MG-).

The ganglion cells are the output neurons of the retina; their myelinated axons form the optic nerve. In the fovea and parafovea, each P ganglion cell (PG+, PG-) forms synapses with only one midget bipolar cell (MB+, MB-), and each midget bipolar cell forms synapses with only one cone;

<sup>\*</sup>The parvo ganglion cells and magno ganglion cells are also referred to as *midget ganglion cells* and *parasol ganglion cells*, respectively.

however, each cone makes contact with a midget on-bipolar and a midget-off bipolar. Thus, the P pathway (in the fovea) is able to carry fine spatial information. Current evidence also suggests that the P pathway carries most of the chromatic information (see Chap. 11). The M ganglion cells (MG+, MG−) form synapses with diffuse bipolar cells (DB+, DB−), and each diffuse bipolar cell forms synapses with many cones. Thus, the M pathway is less able to carry fine spatial information. However, the larger spatial summation (and other boosts in gain) give the M pathway greater sensitivity to changes in contrast. Furthermore, the larger sizes of neurons in the M pathway provide somewhat faster responses and hence somewhat faster transmission of information.

Figure 2a shows the density of ganglion cells as a function of eccentricity in the human retina;<sup>32</sup> for macaque retina, see Ref. 33. At each eccentricity, the P cells (PG+, PG−) comprise approximately 80 percent of the ganglion cells and the M cells (MG+, MG−) approximately 10 percent. Half the P and M cells are on-center (PG+ and MG+) and half are off-center (PG− and MG−). Thus, the approximate densities of the four major classes of ganglion cells can be obtained by scaling the curve in Fig. 2a. As can be seen, ganglion cell density decreases more quickly with eccentricity than does cone density, but in the fovea there are at least 2 to 3 ganglion cells per cone (a reasonable assumption is that there are two P cells for every cone). The Nyquist limit of the retina would appear to be set by the cone density in the fovea and by the ganglion cell density in the periphery.

**Retinal Physiology** Ganglion cells transmit information to the brain via action potentials propagating along axons in the optic nerve. The other retinal neurons transmit information as graded potentials (although amacrine cells generate occasional action potentials). Much remains to be learned about exactly how the computations evident in the responses of ganglion cells are implanted within the retinal circuitry.\*

The receptive fields<sup>†</sup> of ganglion cells are approximately circular and, based upon their responses to spots of light, can be divided into a center region and an antagonistic, annular, surround region.<sup>34</sup> In on-center ganglion cells, light to the center increases the response, whereas light to the surround suppresses the response; the opposite occurs in off-center ganglion cells.

For small to moderate contrast modulations around a steady mean luminance, most P and M cells respond approximately linearly, and hence their spatial and temporal response properties can be usefully characterized by a spatiotemporal transfer function. The spatial components of the transfer function are adequately described as a difference of gaussian functions with separate space constants representing the center and surround;<sup>35–37</sup> see Fig. 5a. The available data suggest that the surround diameter is typically 3 to 6 times the center diameter.<sup>37</sup> The temporal components of the transfer function have been described as a cascade of simple feedback and feed-forward linear filters (e.g., Ref. 38; see Fig. 5b). There also appears to be a small subset of M cells that is highly nonlinear, similar to the Y cells in cat.<sup>39</sup>

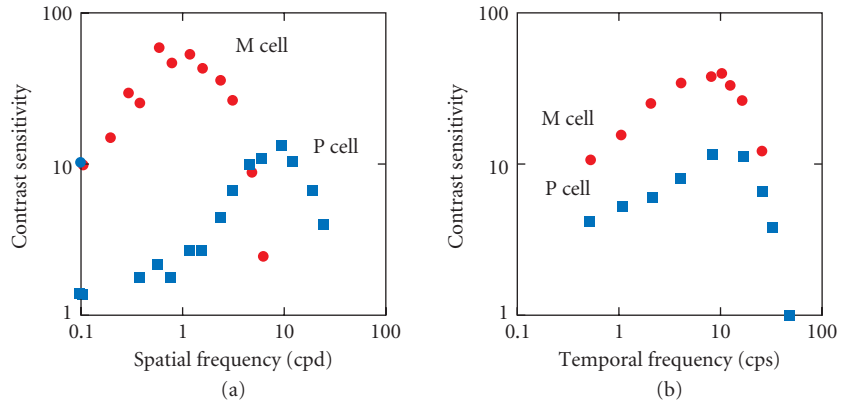
The response amplitude of M and P cells as a function of sinewave contrast is reasonably well described by a Michaelis-Menton function [i.e., by Eq. (14)], but where  $z$  and  $\sigma_0$  are contrasts rather than intensities. As indicated by Fig. 5, the M ganglion cells are (over most spatial and temporal frequencies) about 4 to 10 times more sensitive than the P ganglion cells; in other words,  $\sigma_0$  is 4 to 10 times smaller in M cells than P cells.

When mean (ambient) light level increases, the high-frequency falloffs of the temporal transfer functions of M and P cells shift toward higher frequencies, corresponding to a decrease in the time constants of the linear filters.<sup>38,40</sup> The effect is reduction in gain (an increase in  $\sigma_0$ ), and an increase in temporal resolution.

Relatively little is known about how the spatial transfer functions of primate ganglion cells change with mean light level. In cat, the relative strength of the surround grows with mean luminance, but the space constants (sizes) of the center and surround components appear to change relatively little.<sup>41,42</sup>

\*In this subsection we describe the electrophysiology of M and P cells. This description is based on a composite of data obtained from ganglion cells and geniculate cells. The response properties of M and P ganglion cells and M and P geniculate cells are very similar.

†The receptive field of a neuron is defined to be the region of the visual field (or, equivalently, of the receptor array) where light stimulation has an effect on the response of the neuron.



**FIGURE 5** Typical spatial and temporal contrast sensitivity functions of P and M ganglion cells. (a) Spatial contrast sensitivity functions for sinusoidal gratings with a mean intensity of 1400 td, drifting at 5.2 cps. The symbols are modulation thresholds for a uniform field flickering sinusoidally at 5.2 cps. The threshold criterion was approximately 10 extra impulses per second. (Adapted from Ref. 37.) (b) Temporal contrast sensitivity functions for drifting sinusoidal gratings with mean intensities of 4600 td (P cell) and 3600 td (M cell), and spatial frequencies of 3 cpd (P cell) and 1.6 cpd (M cell). The threshold criterion was approximately 10 extra impulses per second. (Adapted from Ref. 38.)

The major effect of increasing adaptation luminance is an increase in the magnitude of the low-frequency falloff of the spatial transfer function.

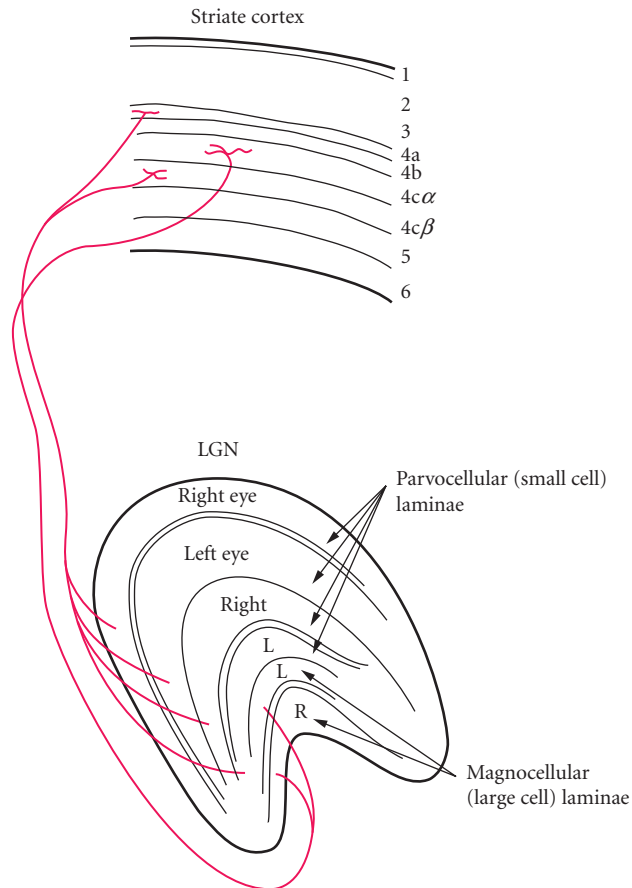
The center diameters of both P and M cells increase with eccentricity,<sup>42,43</sup> roughly in inverse proportion to the square root of ganglion-cell density (cf., Fig. 2a). However, the precise relationship has not been firmly established because of the difficulty in measuring center diameters of ganglion cells near the center of the fovea, where the optical point-spread function is likely to be a major component of center diameter.<sup>37</sup>

## Central Visual Processing

Most of the information encoded in the retina is transmitted via the optic nerve to the lateral geniculate nucleus (LGN) in the thalamus. The neurons in the LGN then relay the retinal information to the primary visual cortex (V1). Neurons in V1 project to a variety of other visual areas. Less is known about the structure and function of V1 than of the retina or LGN, and still less is known about the subsequent visual areas.

**Central Anatomy** The LGN is divided into six layers (see Fig. 6); the upper four layers (the parvocellular laminae) receive synaptic input from the P ganglion cells, the lower two layers (the magnocellular laminae) receive input from the M ganglion cells. Each layer of the LGN receives input from one eye only, three layers from the left eye and three from the right eye. The LGN also receives input from other brain areas including projections from the reticular formation and a massive projection (of uncertain function) from layer 6 of V1. The total number of LGN neurons projecting to the visual cortex is slightly larger than the number of ganglion cells projecting to the LGN.

The segregation of M and P ganglion cell afferents into separate processing streams at the LGN is preserved to some extent in subsequent cortical processing. The magnocellular neurons of the LGN project primarily to neurons in layer 4c $\alpha$  in V1, which project primarily to layer 4b. The neurons in layer 4b project to several other cortical areas including the middle-temporal area (MT) which appears to play an important role in high-level motion processing (for reviews see Refs. 44–46).



**FIGURE 6** Schematic diagram of a vertical section through the lateral geniculate nucleus and through striate cortex (V1) of the primate visual system. (Reproduced from Ref. 49.)

The parvocellular neurons of the LGN project to layers 4a and 4c $\beta$  in V1. The neurons in layers 4a and 4c $\beta$  project to the superficial layers of V1 (layers 1, 2, and 3), which project to areas V2 and V3. Areas V2 and V3 send major projections to area V4 which sends major projections to IT (infero-temporal cortex). The superficial layers of V1 also send projections to layers 5 and 6.

It has been hypothesized that the magno stream subserves crucial aspects of motion and depth perception and that the parvo stream subserves crucial aspects of form and color perception. However, current understanding of the areas beyond V1 is very limited; it is likely that our views of their functional roles in perception will change substantially in the near future. For reviews of cortical functional anatomy see, for example, Refs. 46–49.

**Central Physiology** The receptive field properties of LGN neurons are quite similar to those of their ganglion-cell inputs.<sup>50–52</sup> LGN neurons have, on average, a lower spontaneous response rate than ganglion cells, and are more affected by changes in alertness and anesthetic level, but otherwise display similar center/surround organization, and similar spatial, temporal, and chromatic response properties (however, see Ref. 53).

The receptive-field properties of neurons in the primary visual cortex are substantially different from those in the retina or LGN; V1 neurons have elongated receptive fields that display a substantial degree of selectivity to the size (spatial frequency), orientation, direction of motion of retinal stimulation, and binocular disparity.<sup>54–56</sup> Thus, the probability that a cortical neuron will be activated by an arbitrary retinal image is much lower than for retinal and LGN neurons.

Each region of the visual field is sampled by cortical neurons selective to the full range of sizes, orientations, directions of motion, and binocular disparities. When tested with sinewave gratings, cortical neurons have a spatial-frequency bandwidth (full width at half height) of 1 to 1.5 octaves and an orientation bandwidth of 25 to 35 degrees.<sup>57,58</sup> The number of neurons in area V1 is more than two orders of magnitude greater than the number of LGN neurons;<sup>59</sup> thus, there are sufficient numbers of cortical neurons in each region of the visual field for their receptive fields to tile the whole spatial-frequency plane transmitted from retina to cortex several times. However, the number of cortical cells encoding each region of the spatial-frequency plane is still uncertain. The temporal frequency tuning of V1 neurons is considerably broader than the spatial-frequency tuning with peak sensitivities mostly in the range of 5 to 10 cps.<sup>60</sup> The direction selectivity of V1 neurons varies from 0 to 100 percent with an average of 50 to 60 percent.<sup>57\*</sup> Early measurements of disparity tuning in primate suggested that cortical neurons fall into three major categories: one selective to crossed disparities, one selective to uncrossed disparities, and one tuned to disparities near zero.<sup>56</sup> Recent evidence<sup>61</sup> in agreement with evidence in cat cortex<sup>62,63</sup> suggests a more continuous distribution of disparity tuning.

The chromatic response properties of V1 neurons are not yet fully understood.<sup>49</sup> There is some evidence that discrete regions in the superficial layers of V1, called cytochrome oxidase “blobs,” contain a large proportion of neurons responsive to chromatic stimuli;<sup>64</sup> however, see Ref. 65.

The spatial-frequency tuning functions of cortical cells have been described by a variety of simple functions including Gabor filters, derivatives of gaussian filters, and log Gabor filters (e.g., Ref. 66). The temporal-frequency tuning functions have been described by difference of gamma filters. The full spatiotemporal tuning (including the direction-selective properties of cortical cells) has been described by quadrature (or near quadrature) pairs of separable spatiotemporal filters with spatial and temporal components drawn from the types listed above (e.g., Ref. 67).

Essentially all V1 neurons display nonlinear response characteristics, but can be divided into two classes based upon their nonlinear behavior. *Simple cells* produce approximately half-wave rectified responses to drifting or counterphase sinewave gratings; *complex cells* produce a large unmodulated (DC) response component with a superimposed half-wave or full-wave rectified response component.<sup>68</sup> Simple cells are quite sensitive to the spatial phase or position within the receptor field; complex cells are relatively insensitive to position within the receptive field.<sup>54</sup>

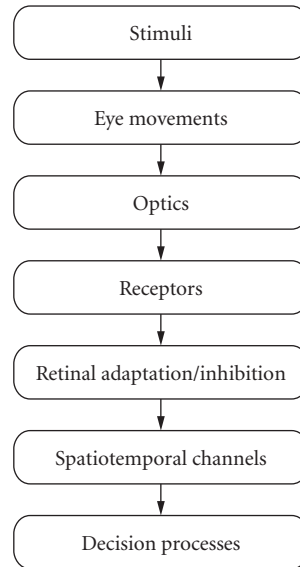
Most simple and complex cells display an accelerating response nonlinearity at low contrasts and response saturation at high contrasts, which can be described by a Michaelis-Menton function [Eq. (11)] with an exponent greater than 1.0.<sup>69</sup> These nonlinearities impart the important property that response reaches saturation at approximately the same physical contrast independent of the spatial frequency, orientation, or direction of motion of the stimulus. The nonlinearities sharpen the spatiotemporal tuning of cortical neurons while maintaining that tuning largely independent of contrast, even though the neurons often reach response saturation of 10 to 20 percent contrast.<sup>70,71</sup>

Unlike retinal ganglion cells and LGN neurons, cortical neurons have little or no spontaneous response activity. When cortical cells respond, however, they display noise characteristics similar to retinal and LGN neurons; specifically, the variance of the response is approximately proportional to the mean response;<sup>72,73</sup> unlike Poisson noise, the proportionality constant is usually greater than 1.0.

## 2.4 VISUAL PERFORMANCE

A major goal in the study of human vision is to relate performance—for example, the ability to see fine detail—to the underlying anatomy and physiology. In the sections that follow, we will discuss some of the data and theories concerning the detection, discrimination, and estimation of contrast,

\*Direction selectivity is defined as  $100 \times (R_p - R_n)/R_p$ , where  $R_p$  is the magnitude of response in the preferred direction and  $R_n$  is the magnitude of response in the nonpreferred direction.



**FIGURE 7** Simple information-processing model of the visual system representing the major factors affecting contrast detection and discrimination.

position, shape, motion, and depth by human observers. In this discussion, it will be useful to use as a theoretical framework the information-processing model depicted in Fig. 7. The visual system is represented as a cascade of processes, starting with eye movements, that direct the visual axis toward a point of interest, proceeding through a number of processing stages discussed in the previous sections and in Chaps. 1 and 11, and ending with a decision. When a human observer makes a decision about a set of visual stimuli, the accuracy of this decision depends on all the processes shown in Fig. 7. Psychophysical experiments measure the performance of the system as a whole, so if one manipulates a particular variable and finds it affects observers' performance in a visual task, one cannot readily pinpoint the responsible stage(s). The effect could reflect a change in the amount of information the stimulus provides, in the fidelity of the retinal image, in representation of the stimulus among spatiotemporal channels, or in the observer's decision strategy. Obviously, it is important to isolate effects due to particular processing stages as well as possible.

We begin by discussing the decisions a human observer might be asked to make concerning visual stimuli. Humans are required to perform a wide variety of visual tasks, but they can be divided into two general categories: *identification tasks and estimation tasks*.<sup>\*</sup> In an identification task, the observer is required to identify a visual image, sequence of images, or a part of an image, as belonging to one of a small number of discrete categories. An important special case is the *discrimination task* in which there are only two categories (e.g., Is the letter on the TV screen an *E* or an *F*?). When one of the two categories is physically uniform, the task is often referred to as a *detection task* (e.g., Is there a letter *E* on the screen or is the screen blank?). In the estimation task, the observer is required to estimate the value of some property of the image (e.g., What is the height of the letter?). The distinction between estimation and identification is quantitative, not qualitative; strictly speaking, an estimation task can

<sup>\*</sup>Visual tasks can also be categorized as *objective* or *subjective*. Objective tasks are those for which there is, in principle, a precise physical standard against which performance can be evaluated (e.g., a best or correct performance). The focus here is on objective tasks because performance in those tasks is more easily related to the underlying physiology.

be regarded as an identification task with a large number of categories. There are two fundamental measures of performance in visual tasks: the *accuracy* with which the task is performed, and the *speed* at which the task is performed. For further discussion of methods for measuring visual performance, see Chap. 3 by Denis G. Pelli and Bart Farell.

Referring again to the information-processing model of Fig. 7, we note that it is difficult to isolate the decision strategy; the best one can do is to train observers to adhere as closely as possible to the same decision strategy across experimental conditions. So, whether the experiment involves simple identification tasks, such as discrimination or detection procedures, or estimation tasks, observers are generally trained to the point of using a consistent decision strategy.

The stimulus for vision is distributed over space, time, and wavelength. As mentioned earlier, we are primarily concerned here with spatial and temporal variations in intensity; variations in wavelength are taken up in Chap. 11.

Before considering human visual performance, we briefly describe *ideal-observer theory* (e.g., Refs 74 and 75), which has proven to be a useful tool for evaluating and interpreting visual performance.

## Ideal-Observer Theory

In attempting to understand human visual performance, it is critically important to quantify performance limitations imposed by the information contained in the stimulus. There is a well-accepted technique, based on the theory of ideal observers,<sup>75</sup> for so doing. As applied to vision (see Refs. 76–79), this approach is based on comparing human observers' performance to that of an ideal observer for the same stimuli. Ideal observers have three main elements: (1) a precise description of the stimuli, including any random variation or noise; (2) descriptions of how the stimuli are modified by the visual stages of interest (e.g., optics, receptors); and (3) an optimal decision rule. Such observers specify the best possible level of performance (e.g., the smallest detectable amount of light) given the variability in the stimulus and information losses in the incorporated visual stages. Because ideal observer performance is the best possible, the thresholds obtained are a measure of the information available in the stimulus (to perform the task) once processed by the incorporated stages. Poorer performance by the human observer must be due to information losses in the unincorporated stages. Thus, ideal-observer theory provides the appropriate physical benchmark against which to evaluate human performance.<sup>75,80</sup> One can, of course, use models of anatomical and physiological mechanisms to incorporate various processing stages into an ideal observer (as shown in Fig. 7); comparisons of human and ideal observer performance in such cases can allow one to compute precisely the information transmitted and lost by mechanisms of interest. This allows assessment of the contributions of anatomical and physiological mechanisms to overall visual performance.<sup>78</sup>

It is important to recognize that ideal-observer theory is not a theory of human performance (humans generally perform considerably below ideal); thus, the theory is not a substitute for psychophysical and physiological modeling. However, the theory is crucial for understanding the stimuli and the task. In many ways, measuring and reporting the information content of stimuli with an ideal observer is as fundamental as measuring and reporting the basic physical dimensions of stimuli.

To illustrate the concepts of ideal-observer theory, consider an identification task with response alternatives (categories)  $\alpha_1$  through  $\alpha_m$ , where the probability of a stimulus from category  $\alpha_j$  equals  $q_j$ . Suppose, further, that the stimuli are static images, with an onset at some known point in time. (By a static image, we mean there is no temporal variation within a stimulus presentation except for photon noise.) Let  $Z$  be the list of random values (e.g., photon counts) at each sample location (e.g., photoreceptor) in some time interval  $\tau$  for a single presentation of a stimulus:

$$Z = (Z_1, \dots, Z_n) \quad (15)$$

where  $Z_p$  is the sample value at the  $i$ th location.

Overall accuracy in an identification task is always optimized (on average) by picking the most probable stimulus category given the sample data and the prior knowledge. Thus, if the goal is to

maximize overall accuracy in a given time interval  $\tau$ , then the maximum average percent correct ( $PC_{\text{opt}}$ ) is given by the following sum:

$$PC_{\text{opt}}(\tau) = \sum_{\mathbf{z}} q_* p_{\tau}(\mathbf{z} | \alpha_*) \quad (16)$$

where  $p_{\tau}(\mathbf{z} | \alpha_*)$  is the probability density function associated with  $\mathbf{Z}$  for stimulus category  $\alpha_j$ , and the subscript  $*$  represents the category  $j$  for which the product  $q_* p_{\tau}(\mathbf{z} | \alpha_*)$  is maximum. (Note,  $\mathbf{Z}$  is a random vector and  $\mathbf{z}$  is a simple vector.)

Suppose, instead, that the goal of the task is to optimize speed. Optimizing speed at a given accuracy level is equivalent to finding the minimum stimulus duration,  $\tau_{\text{opt}}(\epsilon)$ , at a criterion error rate  $\epsilon$ . The value of  $\tau_{\text{opt}}(\epsilon)$  is obtained by setting the left side of Eq. (16) to the criterion accuracy level  $(1 - \epsilon)$  and then solving for  $\tau$ .

For the detection and discrimination tasks, where there are two response categories, Eq. (16) becomes

$$PC_{\text{opt}}(\tau) = \frac{1}{2} + \frac{1}{2} \sum_{\mathbf{z}} [q_1 p_{\tau}(\mathbf{z} | \alpha_1) - (1 - q_1) p_{\tau}(\mathbf{z} | \alpha_2)] \quad (17)$$

[The summation signs in Eqs. (16) and (17) are replaced by an integral sign if the probability density functions are continuous.]

Because the sums in Eqs. (16) and (17) are overall possible vectors  $\mathbf{z} = (z_1, \dots, z_n)$ , they are often not of practical value in computing optimal performance. Indeed, in many cases there is no practical analytical solution for ideal-observer performance, and one must resort to Monte Carlo simulation.

Two special cases of ideal-observer theory have been widely applied in the analysis of psychophysical discrimination and detection tasks. One is the ideal observer for discrimination or detection tasks where the only source of stimulus variability is photon noise.<sup>77,81,82</sup> In this case, optimal performance is given, to close approximation, by the following formulas:<sup>83,84</sup>

$$PC_{\text{opt}}(\tau) = q + 1(1 - q) \Phi\left(\frac{c}{d'} + \frac{d'}{2}\right) - q \Phi\left(\frac{c}{d'} - \frac{d'}{2}\right) \quad (18)$$

where  $c = \ln [(1 - q)/q]$ ,  $\Phi(\cdot)$  is the cumulative standard normal probability distribution, and

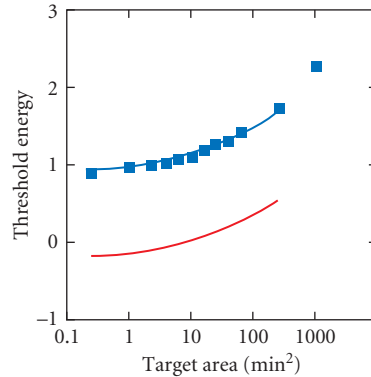
$$d' = \frac{\tau^{1/2} \sum_{i=1}^n (b_i - a_i) \ln(b_i/a_i)}{\left[ \sum_{i=1}^n (b_i + a_i) \ln^2(b_i/a_i) \right]^{1/2}} \quad (19)$$

In Eq. (19),  $a_i$  and  $b_i$  are the average numbers of photons per unit time at the  $i$ th sample location, for the two alternative stimuli ( $\alpha_1$  and  $\alpha_2$ ). For equal presentation probabilities ( $q = 0.5$ ), these two equations are easily solved to obtain  $\tau_{\text{opt}}(\epsilon)$ .<sup>85,86</sup> In signal-detection theory<sup>75</sup> the quantity  $d'$  provides a criterion-independent measure of signal detectability or discriminability.

Figure 8 provides an example of the use of ideal-observer theory. It compares human and ideal performance for detection of square targets of different areas presented briefly on a uniform background. The symbols show the energy required for a human observer to detect the target as a function of target area. The lower curve shows the absolute performance for an ideal observer operating at the level of photon absorption in the receptor photopigments; the upper curve shows the same curve shifted vertically to match the human data. The differences between real and ideal performance represent losses of information among neural processes. The data show that neural efficiency,  $[d'_{\text{real}}]^2 / [d'_{\text{ideal}}]^2$ , for the detection of square targets is approximately 1/2 percent for all but the largest target size.

Another frequently applied special case is the ideal observer for detection and discrimination tasks where the signal is known exactly (SKE) and white or filtered gaussian noise (i.e., image or pixel noise) has been added. Tasks employing these stimuli have been used to isolate and measure central mechanisms that limit discrimination performance<sup>80,87,88</sup> to evaluate internal (neural) noise levels in





**FIGURE 8** Comparison of real and ideal performance for detection of square targets on a uniform background of  $10 \text{ cd/m}^2$  (viewed through a 3-mm artificial pupil). The symbols represent human thresholds as a function of target area. Threshold energy is in units of  $\text{cd/m}^2 \times \text{min}^2 \times \text{sec}$ . The lower curve is ideal performance at the level of the photoreceptors; the upper curve is the ideal performance shifted vertically to match the human data. (Adapted from Ref. 264.)

the visual system,<sup>79,89</sup> and to develop a domain of applied psychophysics relevant to the perception of the noisy images, such as those created by radiological devices and image enhancers.<sup>90-92</sup> For the gaussian-noise-limited ideal discriminator, Eq. (18) still applies, but  $d'$  is given by the following:

$$d' = \frac{E(L|\alpha_2) - E(L|\alpha_1)}{\sqrt{\text{VAR}(L)}} \quad (20)$$

where

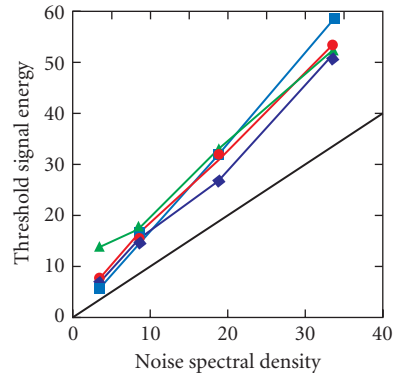
$$L = [\mathbf{m}_2 - \mathbf{m}_1]' \Sigma^{-1} \mathbf{Z} \quad (21)$$

In Eq. (21),  $\mathbf{Z}$  is the column vector of random values from the sample locations (e.g., pixels),  $[\mathbf{m}_2 - \mathbf{m}_1]'$  is a row vector of the differences in the mean values at each sample point (i.e.,  $b_1 - a_1, \dots, b_n - a_n$ ), and  $\Sigma$  is the covariance matrix resulting from the gaussian noise (i.e., the element  $\sigma_{ij}$  of  $\Sigma$  is the covariance between the  $i$ th and  $j$ th sample location). In the case of white noise,  $\Sigma$  is a diagonal matrix with diagonal elements equal to  $\sigma^2$ , and thus,\*

$$d' = \frac{1}{\sigma} \sqrt{\sum (b_i - a_i)^2} \quad (22)$$

As an example, Fig. 9 compares human and ideal performance for amplitude discrimination of small targets in white noise as a function of noise spectral power density (which is proportional to  $\sigma^2$ ). The solid black line of slope 1.0 shows the absolute performance of an ideal observer operating at the level of the cornea (or display screen); thus, the difference between real and ideal performance represents all losses of information within the eye, retina, and central visual pathways. For these conditions, efficiency for amplitude discrimination of the targets ranges from about 20 to 70 percent, much

\*It should be noted that the performance of the photon-noise-limited observer [Eq. (19)] becomes equivalent to an SKE white noise observer [Eq. (22)] for detection of targets against intense uniform backgrounds.



**FIGURE 9** Comparison of real (symbols) and ideal (solid black line) performance for contrast discrimination of spatially localized targets in white noise backgrounds with a mean luminance of  $154 \text{ cd/m}^2$ . Noise spectral density and target signal energy are in units of  $10^{-7} \text{ deg}^2$  (the arrow indicates a noise standard deviation of 26 percent contrast per pixel). (■) Gaussian target (standard deviation =  $0.054 \text{ deg}$ ); (●) Gaussian damped sinewave target ( $9.2 \text{ cpd}$ , standard deviation =  $0.109 \text{ deg}$ ); (▲) Gaussian damped sinewave target ( $4.6 \text{ cpd}$ , standard deviation =  $0.217 \text{ deg}$ ); (◆) Sinewave target ( $4.6 \text{ cpd}$ ,  $0.43 \times 0.43 \text{ deg}$ ). (Adapted from Ref. 87.)

higher than for detecting targets in uniform backgrounds (cf., Fig. 8). Efficiencies are lower in Fig. 8 than in Fig. 9 primarily because, with uniform backgrounds, internal (neural) noise limits human, but not ideal, performance; when sufficient image noise is added to stimulus, it begins to limit both real and ideal performance.

The following sections are a selective review of human visual performance and underlying anatomical and physiological mechanisms. The review includes spatial and temporal contrast perception, light adaptation, visual resolution, motion perception, and stereo depth perception, but due to space limitations, excludes many other active areas of research such as color vision (partially reviewed Chaps. 10 and 11), eye movements,<sup>93</sup> binocular vision,<sup>94,95</sup> spatial orientation and the perception of layout,<sup>96,97</sup> pattern and object recognition,<sup>96-99</sup> visual attention,<sup>100</sup> visual development,<sup>101,102</sup> and abnormal vision.<sup>103</sup>

## Contrast Detection

The study of contrast detection in the human visual system has been dominated for the last 20 years by methods derived from the concepts of linear systems analysis. The rationale for applying linear systems analysis in the study of visual sensitivity usually begins with the observation that a complete characterization of a visual system would describe the output resulting from any arbitrary input. Given that the number of possible inputs is infinite, an exhaustive search for all input-output relationships would never end. However, if the system under study is linear, then linear systems analysis provides a means for characterizing the system from the measurement of a manageably small set of input-output relationships. Of course, the visual system has important nonlinearities, so strictly speaking, the assumptions required by linear systems analysis are violated in general. Nonetheless, there have been several successful applications of linear systems analysis in vision.

In the case of spatial vision, linear systems analysis capitalizes on the fact that any spatial retinal illumination distribution,  $I(x, y)$  can be described exactly by the sum of a set of basis functions, such as the set of sinusoids. Sinusoids are eigenfunctions of a linear system, which implies that the system response to a sinusoidal input can be completely characterized by just two numbers; an amplitude change and a phase change. Indeed, it is commonly assumed (usually with justification) that spatial phase is unaltered in processing, so only one number is actually required to describe the system response to a sinusoid. Linear systems analysis provides a means for predicting, from such characterizations of system responses to sinusoids, the response to any arbitrary input. For these reasons, the measurement of system responses to spatial sinusoids has played a central role in the examination of human spatial vision. We will begin our discussion with the spatial sinusoid and the system response to it.

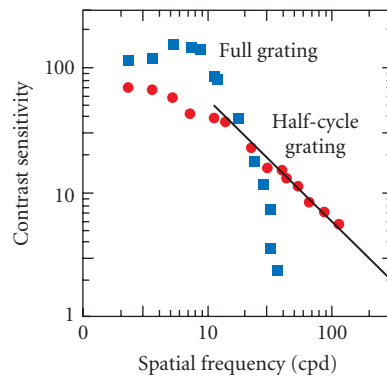
A spatial sinusoid can be described by

$$I(x, y) = A \sin[2\pi f(x \cos(\theta) + y \sin(\theta)) + \phi] + \bar{I} \quad (23)$$

where  $\bar{I}$  is the space-average intensity of the stimulus field (often expressed in trolands),  $A$  is the amplitude of the sinusoid,  $f$  is the spatial frequency (usually expressed in cycles  $\cdot$  deg $^{-1}$  or cpd),  $\theta$  is the orientation of the pattern relative to axes  $x$  and  $y$ , and  $\phi$  is the phase of the sinusoid with respect to the origin of the coordinate system. The contrast  $C$  of a spatial sinewave is defined as  $(I_{\max} - I_{\min}) / (I_{\max} + I_{\min})$ , where  $I_{\max}$  is the maximum intensity of the sinusoid and  $I_{\min}$  is the minimum; thus  $C = A / \bar{I}$ . When the sinusoid is composed of vertical stripes,  $\theta$  is zero and Eq. (23) reduces to

$$I(x, y) = A \sin[2\pi f x + \phi] + \bar{I} \quad (24)$$

When the contrast  $C$  of a sinusoidal grating (at a particular frequency) is increased from zero (while holding  $\bar{I}$  fixed), there is a contrast at which it first becomes reliably detectable, and this value defines the contrast detection threshold. It is now well-established that contrast threshold varies in a characteristic fashion with spatial frequency. A plot of the reciprocal of contrast at threshold as a function of spatial frequency constitutes the *contrast sensitivity function* (CSF). The CSF for a young observer with good vision under typical indoor lighting conditions is shown in Fig. 10 (blue squares).



**FIGURE 10** Contrast sensitivity as a function of spatial frequency for a young observer with good vision under typical indoor lighting conditions. The blue squares represent the reciprocals of contrast at detection threshold for sinusoidal grating targets. The red circles represent the reciprocals of contrast thresholds for half-cycle sinusoids. The solid line represents the predicted threshold function for half-cycle gratings for a linear system whose CSF is given by the blue squares. (Adapted from Ref. 193.)

The function is bandpass with a peak sensitivity at 3 to 5 cpd. At those spatial frequencies, a contrast of roughly 1/2 percent can be detected reliably (once the eye has adapted to the mean luminance). At progressively higher spatial frequencies, sensitivity falls monotonically to the so-called high-frequency cutoff at about 50 cpd; this is the finest grating an observer can detect when the contrast  $C$  is at its maximum of 1.0. At low spatial frequencies, sensitivity falls as well, although (as we will see) the steepness of this low-frequency rolloff is quite dependent upon the conditions under which the measurements are made. It is interesting that most manipulations affect the two sides of the function differently, a point we will expand later.

The CSF is not an invariant function; rather the position and shape of the function are subject to significant variations depending on the optical quality of the viewing situation, the average luminance of the stimulus, the time-varying and chromatic qualities of the stimulus, and the part of the retina on which the stimulus falls. We will examine all of these effects because they reveal important aspects of the visual processes that affect contrast detection.

Before turning to the visual processes that affect contrast detection, we briefly describe how CSF measurements and linear systems analysis can be used to make general statements about spatial contrast sensitivity. Figure 10 plots contrast sensitivity to targets composed of extended sinusoids and to targets composed of one half-cycle of a cosinusoid. Notice that sensitivity to high spatial frequencies is much higher with the half-cycle cosinusoids than it is with extended sinusoids.

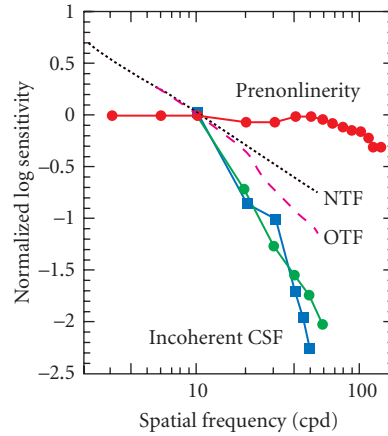
The half-cycle wave forms can be described as cosinusoids multiplied by rectangular functions of widths equal to one-half period of the cosinusoid. The truncating rectangular function causes the frequencies in the pattern to “splatter” to higher and lower values than the nominal target frequency. Multiplication of the Fourier transforms of the half-cycle targets by the CSF obtained with extended sinusoids yields an estimate of the visual system’s output response to the half-cycle targets. From there, use of a simple decision rule allows one to derive predicted half-cycle contrast sensitivities for a linear system. One expects to find greater visibility to high spatial frequencies with half-cycle gratings. The quantitative predictions are represented by the solid line in Fig. 10 and they match the observations rather well. There are numerous cases in the literature in which the CSF yields excellent predictions of the visibility of other sorts of patterns, but there are equally many cases in which it does not. We will examine some of the differences between those two situations later.

All of the information-processing stages depicted in Fig. 7 affect the CSF, and most of those effects have now been quantified. We begin with eye movements.

**Eye Movements** Even during steady fixation, the eyes are in constant motion, causing the retina to move with respect to the retinal images.<sup>104</sup> Such motion reduces retinal image contrast by smearing the spatial distribution of the target, but it also introduces temporal variation in the image. Thus, it is not obvious whether eye position jitter should degrade or improve contrast sensitivity. It turns out that the effect depends on the spatial frequency of the target. At low frequencies, eye movements are beneficial. When the image moves with respect to the retina, contrast sensitivity improves for spatial frequencies less than about 5 cpd.<sup>105,106</sup> The effect of eye movements at higher spatial frequencies is less clear, but one can measure it by presenting sinusoidal interference fringes (which bypass optical degradations due to the eye’s aberrations) at durations too short to allow retinal smearing and at longer durations at which smearing might affect sensitivity.<sup>107</sup> There is surprisingly little attenuation due to eye position jitter; for 100-ms target presentations, sensitivity at 50 cpd decreases by only 0.2 to 0.3 log units relative to sensitivity at very brief durations. Interestingly, sensitivity improves at long durations of 500 to 2000 ms. This may be explained by noting that the eye is occasionally stationary enough to avoid smearing due to the retina’s motion.

Thus, the eye movements that occur during steady fixation apparently improve contrast sensitivity at low spatial frequencies and have little effect at high frequencies. This means that with steady fixation there is little need for experimenters to monitor or eliminate the effects of eye movements in measurements of spatial contrast sensitivity at high spatial frequencies.

**Optics** The optical transfer function (OTF) of an optical system describes the attenuation in contrast (and the phase shift) that sinusoidal gratings undergo when they pass through the optical system. As shown in Chap. 1 the human OTF (when the eye is well accommodated) is a low-pass function



**FIGURE 11** The contrast sensitivity function (CSF) and the transfer functions for various visual processing stages. All functions have been normalized to a log sensitivity of 0.0 at 10 cpd. The blue squares represent the CSF measured in the conventional way using incoherent white light.<sup>22</sup> The red circles represent the transfer function of the prenonlinearity filter measured.<sup>22</sup> The dotted line represents the neural transfer function that results from a bank of photon-noise-limited spatial filters of constant bandwidth.<sup>122</sup> The dashed line represents the optical transfer function for a 3.8-mm pupil.<sup>4</sup> The green circles are the product of the three transfer functions.

whose shape depends on pupil size and eccentricity. The dashed curve and the blue squares in Fig. 11 show the foveal OTF and the foveal CSF, respectively, for a pupil diameter of 3.8 mm. At high spatial frequencies, the foveal CSF declines at a faster rate than the foveal OTF. Thus, the OTF can only account for part of the high-frequency falloff of the CSF. Also, it is obvious that the OTF (which is a low-pass function) cannot account for the falloff in the CSF at low spatial frequencies.

**Receptors** Several receptor properties affect spatial contrast sensitivity and acuity. As noted earlier, the receptors in the fovea are all cones, and they are very tightly packed and arranged in a nearly uniform hexagonal lattice. With increasing retinal eccentricity, the ratio of cones to rods decreases steadily (see Fig. 2a) and the regularity of the lattice diminishes.

Several investigators have noted the close correspondence between the Nyquist frequency and the highest-detectable spatial frequency<sup>108–111</sup> and have argued that the geometry of the foveal cone lattice sets a fundamental limit to grating acuity. However, more importance has been placed on this relationship than is warranted. This can be shown in two ways. First, grating acuity varies with several stimulus parameters including space-average luminance, contrast, and temporal duration. For example, grating acuity climbs from about 5 cpd at a luminance of 0.001 cd/m<sup>2</sup> to 50 to 60 cpd at luminances of 100 cd/m<sup>2</sup> or higher.<sup>112</sup> Obviously, the cone lattice does not change its properties in conjunction with changes in luminance, so this effect has nothing to do with the Nyquist limit. The only conditions under which the Nyquist limit might impose a fundamental limit to visibility would be at high luminances and contrasts, and relatively long target durations. Second, and more importantly, several investigators have shown that observers can actually detect targets whose frequency components are well above the Nyquist limit. For example, Williams<sup>113</sup> presented high-frequency sinusoidal gratings using laser inter-ferometry to bypass the contrast reductions that normally occur in passage through the eye's optics. He found that observers could detect gratings at frequencies as high as 150 cpd, which

is 2.5 times the Nyquist limit. Observers did so by detecting the gratings' "aliases," which are Moiré-like distortion products created by undersampling the stimulus.<sup>114</sup> The CSF under these conditions is smooth with no hint of a loss of visibility at the Nyquist limit. The observer simply needs to switch strategies above and below the Nyquist limit; when the targets are below the limit, the observer can detect the undistorted targets, and when the targets are above the limit, the observer must resort to detecting the targets' lower frequency aliases.<sup>7</sup> On the other hand, when the task is to resolve the stripes of a grating (as opposed to detecting the grating), performance is determined by the Nyquist limit over a fairly wide range of space-average luminances.<sup>115</sup>

In addition to receptor spacing, one must consider receptor size in the transmission of signals through the early stages of vision. As noted earlier, the prevailing model of cones holds that the inner segment offers an aperture to incoming photons. In the fovea, the diameter of the inner segment is roughly  $0.5 \text{ min}^{6,21}$  (see Fig. 2c). Modeling the cone aperture by a cylinder function of this diameter, one can estimate the low-pass filtering due to such an aperture from the cylinder's Fourier transform, which is a first-order Bessel function whose first zero occurs at 146 cpd. However, the entrance to a structure only a few wavelengths of light wide cannot be described accurately by geometric optics,<sup>116</sup> so the receptor aperture is generally modeled by a gaussian function whose full width at half height is somewhat smaller than the anatomical estimates of the aperture diameter.<sup>22</sup>

The modulation transfer of the neural stages prior to a compressive nonlinearity (thought to exist in the bipolar cells of the retina or earlier) has been measured in the following way.<sup>22</sup> Laser interferometry was used to bypass the contrast reductions that normally occur as light passes through the eye's optics. The stimuli were two high-contrast sinusoids of the same spatial frequency but slightly different orientations; these created a spatial "beat" of a different orientation and frequency than the component gratings. Passing two sinusoids through a compressive nonlinearity, like the ones known to exist early in the visual system, creates a number of distortion products at various frequencies and orientations. MacLeod et al.<sup>22</sup> varied the orientations and frequencies of the components so as to create a "beat" at 10 cpd, and then measured contrast sensitivity to the "beat." Because the "beat" was always at the same spatial frequency, the filtering properties of postnonlinearity stages could not affect the measurements and, therefore, any effect of component frequency on contrast sensitivity had to be due to the filtering properties of prenonlinearity stages. Observers were able to detect the "beat" produced by component gratings of frequencies as high as 140 cpd. This implies that the prenonlinearity filter has exceptionally large bandwidth. By inverse Fourier transformation of these data, one can estimate the spatial extent of the filter; it appears to have a full width at half height of about 16 arcsec, which is a bit smaller than current estimates of the cone aperture. Thus, the early neural stages are able to pass information up to extremely high spatial frequencies once the optics are bypassed.

The red circles in Fig. 11 display the estimate of the transfer function of the prenonlinearity neural filter for foveal viewing. Notice that it is virtually flat up to 140 cpd. Also shown (blue squares) is the CSF obtained in the conventional manner with incoherent white light from a CRT display. This function represents the product of the optics and the prenonlinearity and postnonlinearity spatial filters. As can be seen, the bandwidth of the conventional CSF is much smaller than the bandwidth of the prenonlinearity filter. Thus, optics and postnonlinearity filtering are the major constraints to contrast sensitivity at higher spatial frequencies.

As mentioned earlier, the transfer function of the optics is shown as the dashed line. The product of the optical transfer function and the prenonlinearity transfer function would be very similar to the dashed line because the prenonlinearity function is virtually flat. Thus, the product of optics and prenonlinearity filter does not match the observed CSF under conventional viewing conditions at all; the postnonlinearity filter must also contribute significantly to the shape of the CSF. The mismatch between the dashed line and the CSF is obvious at low and high spatial frequencies. We consider the high-frequency part first; the low-frequency part is considered later under "Adaptation and Inhibition."

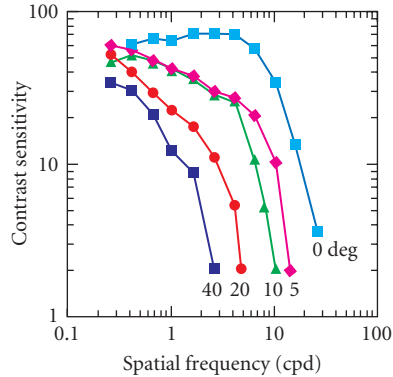
**Spatial Channels** Spatial filters with narrow tuning to spatial frequency have been demonstrated in a variety of ways.<sup>117,118</sup> For example, the visibility of a sinusoidal target is reduced by the presence of a narrowband noise masker whose center frequency corresponds to the spatial frequency of the sinusoid, but is virtually unaffected by the presence of a masker whose center frequency is more than

1.5 octaves from the frequency of the sinusoid (Ref. 119; see “Contrast Discrimination and Contrast Masking” later in chapter). These spatial mechanisms, which have been described mathematically by Gabor functions and other functions,<sup>120</sup> correspond to a first approximation with the properties of single neurons in the visual cortex.<sup>58</sup> They seem to have roughly constant bandwidths, expressed in octaves or log units, regardless of preferred frequency.<sup>58,117</sup> As a consequence, mechanisms responding to higher spatial frequencies are smaller in vertical and horizontal spatial extent than mechanisms responding to lower frequencies.<sup>121,122</sup> Thus, for a given stimulus luminance, a higher-frequency mechanism receives fewer photons than does a lower-frequency mechanism. The mean number of photons delivered to such a mechanism divided by the standard deviation of the number of photons—the signal-to-noise ratio—follows a square root relation,<sup>123</sup> so the signal-to-noise ratio is inversely proportional to the preferred frequency of the mechanism. The transfer function one expects for a set of such mechanisms should be proportional to  $1/f$  where  $f$  is spatial frequency.<sup>122</sup> This function is represented by the dotted line in Fig. 11. If we use that relation for describing the postnonlinearity filter, and incorporate the measures of the optical and prenonlinearity filters described above, the resultant is represented by the green circles. The fit between the observed CSF and the resultant is good, so we can conclude that the filters described here are sufficient to account for the shape of the high-frequency limb of the human CSF under foveal viewing conditions.

By use of the theory of ideal observers (see earlier), one can calculate the highest possible contrast sensitivity a system could have if limited by the optics, photoreceptor properties, and constant bandwidth spatial channels described above. As can be seen (green circles), the high-frequency limb of the CSF of such an ideal observer is similar in shape to that of human observers,<sup>122,124</sup> although the rate of falloff at the highest spatial frequencies is not quite as steep as that of human observers.<sup>124</sup> More importantly, the absolute sensitivity of the ideal observer is significantly higher. Some of the low performance of the human compared to the best possible appears to be the consequence of processes that behave like noise internal to the visual system and some appears to be the consequence of employing detecting mechanisms that are not optimally suited for the target being presented.<sup>79,89</sup> The high detection efficiencies observed for sinewave grating patches in the presence of static white noise (e.g., Ref. 87; Fig. 9) suggest that poor spatial pooling of information is not the primary factor responsible for the low human performance (at least when the number of cycles in the grating patches are kept low).

**Optical/Retinal Inhomogeneity** The fovea is undoubtedly the most critical part of the retina for numerous visual tasks such as object identification and manipulation, reading, and more. But, the fovea occupies less than 1 percent of the total retinal surface area, so it is not surprising to find that nonfoveal regions are important for many visual skills such as selecting relevant objects to fixate, maintaining posture, and determining one’s direction of self-motion with respect to environmental landmarks. Thus, it is important to characterize the visual capacity of the eye for nonfoveal loci as well.

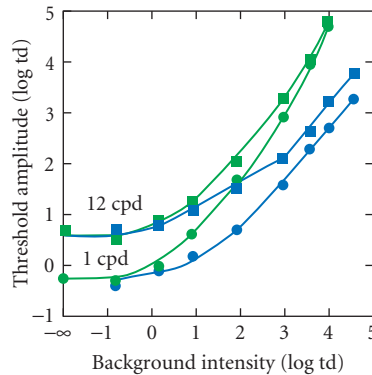
The quality of the eye’s optics diminishes slightly from 0 to 20 deg of retinal eccentricity and substantially from 20 to 60 deg;<sup>1</sup> see Chap. 1. As described earlier and displayed in Fig. 2, the dimensions and constituents of the photoreceptor lattice and the other retinal neural elements vary significantly with retinal eccentricity. With increasing eccentricity, cone and retinal ganglion cell densities fall steadily and individual cones become broader and shorter. Given these striking variations in optical and receptor properties across the retina, it is not surprising that different aspects of spatial vision vary greatly with retinal eccentricity. Figure 12 shows CSFs at eccentricities of 0 to 40 deg. With increasing eccentricity, contrast sensitivity falls off at progressively lower spatial frequencies. The high-frequency cutoffs range from approximately 2 cpd at 40 deg to 50 cpd in the fovea. Low-frequency sensitivity is similar from one eccentricity to the next. There have been many attempts to relate the properties of the eye’s optics, the receptors, and postreceptor mechanisms to contrast sensitivity and acuity. For example, models that incorporate eccentricity-dependent variations in optics and cone lattice properties, plus the assumption of fixed bandwidth filters (as in Ref. 122), have been constructed and found inadequate for explaining the observed variations in contrast sensitivity; specifically, high-frequency sensitivity declines more with eccentricity than can be explained by information losses due to receptor lattice properties and fixed bandwidth filters alone.<sup>125</sup> Adding a low-pass filter to the model representing the convergence of cones onto retinal ganglion cells<sup>32</sup> (i.e., by incorporating variation in receptive field center diameter) yields a reasonably accurate account of eccentricity-dependent variations in contrast sensitivity.<sup>125</sup>



**FIGURE 12** CSFs at different retinal eccentricities. (Adapted from Ref. 125.)

**Adaptation and Inhibition** Contrast sensitivity is also dependent upon the space-average intensity (i.e., the mean or background luminance) and the prior level of light adaptation. The blue symbols in Fig. 13 show how threshold amplitudes for spatial sinewaves vary as a function of background luminance for targets of 1 and 12 cpd.<sup>126</sup> [Recall that sinewave amplitude is contrast multiplied by background intensity; see discussion preceding Eq. (24).] The backgrounds were presented continuously and the observer was allowed to become fully light-adapted before thresholds were measured. For the low-spatial-frequency target (blue circles), threshold increases linearly with a slope of 1.0 above the low background intensities (1 log td). A slope of 1.0 in log-log coordinates implies that

$$A_T = k\bar{I} \quad (25)$$



**FIGURE 13** Amplitude threshold for gaussian-damped sinusoidal targets with 0.5-octave bandwidths as a function of background intensity. The targets were presented for 50 ms. The green symbols are thresholds measured at the onset of backgrounds flashed for 500 ms in the dark-adapted eye. The blue symbols are thresholds measured against steady backgrounds after complete adaptation to the background intensity. The difference between the green and blue symbols shows the effect of adaptation on contrast sensitivity. (Adapted from Ref. 126.)



where  $A_T$  is the amplitude threshold (or equivalently, it implies that contrast threshold,  $C_T$ , is constant). This proportional relationship between threshold and background intensity is known as *Weber's law*. For the higher spatial-frequency target (blue squares), threshold increases linearly with a slope of 0.5 at intermediate background intensities (0 to 3 log td) and converges toward a slope of 1.0 at the highest background intensities. A slope of 0.5 implies that

$$A_T = k\sqrt{I} \quad (26)$$

(or equivalently, that contrast threshold,  $C_T$ , decreases in inverse proportion to the square root of background intensity). This relationship between amplitude threshold and the square root of background intensity is known as the *square root law* or the *DeVries-Rose law*. As suggested by the data in Fig. 13, the size of the square root region gradually grows as spatial frequency is increased. The effect on the CSF is an elevation of the peak and of the high-frequency limb relative to the low-frequency limb as mean luminance is increased, thereby producing a shift of the peak toward higher spatial frequencies.<sup>127,128</sup> This result explains why humans prefer bright environments when performing tasks that require high spatial precision.

The visual system is impressive in its ability to maintain high sensitivity to small contrasts (as produced, for example, by small changes in surface reflectance) over the enormous range of ambient light levels that occur in the normal environment. In order to have high sensitivity to small contrasts, the response gain of visual neurons must be kept high, but high gain leaves the visual system vulnerable to the effects of response saturation (because neurons are noisy and have a limited response range). The visual system's solution to this "dynamic-range problem" is threefold: (1) use separate populations of receptors (rods and cones) to detect contrasts in the lower and upper ranges of ambient intensity, (2) adjust the retinal illumination via the pupil reflex, and (3) directly adjust the sensitivity of individual receptors and neurons (e.g., Refs. 129, 130). All three components are important, but the third is the most significant and is accomplished within the receptors and other neurons through photochemical and neural adaptation. The combined effect of the photochemical and neural adaptation mechanisms on foveal contrast detection is illustrated by the difference between the open and solid data points in Fig. 13 (all the data were obtained with a fixed pupil size). The green symbols show cone detection threshold for the 1- and 12-cpd targets on a background flashed in the dark-adapted eye (i.e., there was no chance for light adaptation) as a function of background intensity. The difference between the green and blue symbols shows the improvement in detection sensitivity due to photochemical and neural adaptation within the cone system.

The pupil reflex and photopigment depletion (see "Image Sampling by the Photoreceptors") are *multiplicative adaptation* mechanisms; they adjust the sensitivity of the eye by effectively scaling the input intensity by a multiplicative gain factor that decreases with increasing average intensity. The pupil reflex operates over most of the intensity range (see Chap. 1); photopigment depletion is only effective in the cone system above about 4 log td (see Fig. 13) and is ineffective in the rod system, at least over the range of relevant intensities (see earlier discussion). There is considerable psychophysical evidence that a substantial component of the remaining improvement in sensitivity illustrated in Fig. 12 is due to multiplicative neural adaptation mechanisms.<sup>129,131–134</sup> However, multiplicative adaptation alone cannot reduce threshold below a line of slope 1.0, tangent to the threshold curve in dark-adapted eye.<sup>129,134</sup> The remaining improvements in sensitivity (the difference between the tangent lines and the blue symbols) can be explained by neural *subtractive adaptation* mechanisms.<sup>134,135</sup> There is evidence that the multiplicative and subtractive adaptation mechanisms each have fast and slow components.<sup>134–137</sup> Threshold functions, such as those in Fig. 13, can be predicted by simple models consisting of a compressive (saturating) nonlinearity [Eq. (13)], a multiplicative adaptation mechanism, a subtractive adaptation mechanism, and either constant additive noise or multiplicative (Poisson) neural noise.

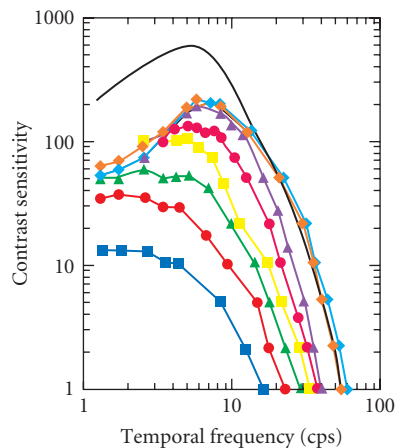
A common explanation of the low-frequency falloff of the CSF (e.g., Fig. 12) is based on center/surround mechanisms evident in the responses of retinal and LGN neurons. For example, compare the CSFs measured in individual LGN neurons in Fig. 5 with the human CSF in Fig. 12. The center and surround mechanisms are both low-pass spatial filters, but the cutoff frequency of the surround is much lower than that of the center. The center and surround responses are subtracted neurally, so

at low spatial frequencies, to which center and surround are equally responsive, the neuron's response is small. With increasing spatial frequency, the higher resolution of the center mechanism yields a greater response relative to the surround, so the neuron's response increases.

While the center/surround mechanisms are undoubtedly part of the explanation, they are unlikely to be the whole story. For one thing, the time course of the development of the low-frequency falloff is slower than one expects from physiological measurements.<sup>105</sup> Second, the surround strength in retinal and LGN neurons is on average not much more than half the center strength;<sup>37</sup> this level of surround strength is consistent with the modest low-frequency falloffs observed under transient presentations, but inconsistent with the steep falloff observed for long-duration, steady-fixation conditions. Third, there is considerable evidence for slow subtractive and multiplicative adaptation mechanisms, and these ought to contribute strongly to the shape of the low-frequency limb under steady fixation conditions.<sup>137-139</sup> Thus, the strong low-frequency falloff of the CSF under steady viewing conditions may reflect, to some degree, slow subtractive and multiplicative adaptation mechanisms whose purpose is to prevent response saturation while maintaining high contrast sensitivity over a wide range of ambient light levels. The weaker low-frequency falloff seen for brief presentations may be the result of the fast-acting surround mechanisms typically measured in physiological studies; however, fast-acting local multiplicative and subtractive mechanisms could also play a role.

**Temporal Contrast Detection** Space-time plots of most contrast-detection stimuli (plots of intensity as a function of  $x, y, t$ ) show intensity variations in both space and time; in other words, most contrast detection stimuli contain both spatial and temporal contrast.

The discussion of contrast detection has, so far, emphasized spatial dimensions (e.g., spatial frequency), but there is, not surprisingly, a parallel literature emphasizing temporal dimensions (for a recent review see Ref. 140). Figure 14 shows de Lange's<sup>141</sup> measurements of temporal contrast sensitivity



**FIGURE 14** Temporal contrast sensitivity functions measured in the fovea at several mean luminances for a  $2^\circ$  circular test field. The test field was modulated sinusoidally and was imbedded in a large, uniform background field of the same mean luminance. Mean luminance: (■) 0.375 td, (●) 1 td, (▲) 3.75 td, (■) 10 td, (●) 37.5 td, (▲) 100 td, (◆) 1000 td, (◆) 10,000 td. (Data from Ref. 141.) The black curve is the MTF derived from the impulse response function of a single macaque cone reported by Schnapf et al.<sup>17</sup> The cone MTF was shifted vertically (in log-log coordinates) for comparison with the shape of the high-frequency falloff measured psychophysically.

functions at several mean luminances for uniform discs of light modulated in intensity sinusoidally over time. The major features of these results have been confirmed by other researchers.<sup>142,143</sup> Like spatial CSFs, temporal CSFs typically have a bandpass shape. At high mean luminances, the peak contrast sensitivity occurs at approximately 8 cycles per second (cps) and the cutoff frequency (also known as the critical flicker frequency) at approximately 60 cps. Contrast sensitivity increases with increasing mean luminance, larger increases (in log-log coordinates) being observed at middle and high temporal frequencies. For low temporal frequencies (below 6 cps), contrast sensitivity is nearly constant (i.e., Weber's law is observed) for mean intensities greater than 10 td or so.

The entire temporal CSF can be fit by a weighted difference of linear low-pass filters, where one of the filters has a relatively lower cutoff frequency, and/or where one of the filters introduces a relatively greater time delay (see Ref. 140). However, to fit the temporal CSFs obtained under different stimulus conditions, the time constants and relative weights of the filters must be allowed to vary. The low-frequency falloff may be partly due to the biphasic response of the photoreceptors (see Fig. 3), which appears to involve a fast subtractive process.<sup>17</sup> Surround mechanisms, local subtractive mechanisms, and local multiplicative mechanisms also probably contribute to the low-frequency falloff.

The factors responsible for the high-frequency falloff are not fully understood, but there is some evidence that much of the high-frequency falloff is the result of temporal integration within the photoreceptors.<sup>144,145</sup> The black curve in Fig. 14 is the MTF of the macaque cone photocurrent response derived from measured impulse-response functions<sup>17</sup> (see Fig. 3). The impulse responses were obtained in the linear response range by presenting brief, dim flashes on a dark background. Although the background was dark, there appears to be little effect of light adaptation on cone responses until background intensity exceeds at least a few hundred trolands;<sup>17,24</sup> thus, the cone impulse-response functions measured in the dark ought to be valid up to moderate background intensities. Following Boynton and Baron,<sup>144</sup> we have shifted the MTF vertically in order to compare the shape of the cone high-frequency falloff to that of the temporal CSF. The shapes are similar, lending support to the hypothesis that cone temporal integration is responsible for the shape of the high-frequency limb of the temporal CSF.

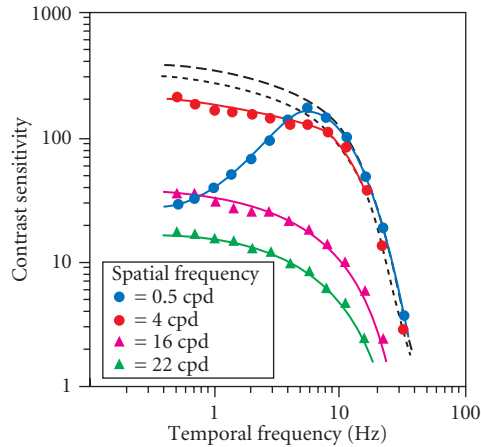
It is well established that a pulse of light briefer than some critical duration will be just visible when the product of its duration and intensity is constant. This is *Bloch's law*, which can be written:

$$I_T T = k \quad \text{for} \quad T < T_c \quad (27)$$

where  $k$  is constant,  $T_c$  is the critical duration and  $I_T$  is the threshold intensity at duration  $T$ . Naturally, Bloch's law can also be stated in terms of contrasts by dividing the intensities in the above equation by the background intensity. Below the critical duration, log-log plots of threshold intensity versus duration have a slope of  $-1$ . Above the critical duration, threshold intensity falls more slowly with increasing duration than predicted by Bloch's law (slope between  $-1$  and  $0$ ) and, in some cases, it even becomes independent of duration (slope of  $0$ ).<sup>76,146</sup> It is sometimes difficult to determine the critical duration because the slopes of threshold-versus-duration plots can change gradually, but under most conditions Bloch's critical duration declines monotonically with increasing background intensity from about 100 ms at 0 log td to about 25 ms at 4 log td.

Bloch's law is an inevitable consequence of a linear filter that only passes temporal frequencies below some cutoff value. This can be seen by computing the Fourier transform of pulses for which the product  $I_T T$  is constant. The amplitude spectrum is given by  $I_T \sin(T\pi f)/\pi f$  where  $f$  is temporal frequency; this function has a value  $I_T T$  at  $f=0$  and is quite similar over a range of temporal frequencies for small values of  $T$ . Indeed, to a first approximation, one can predict the critical duration  $T_c$  from the temporal CSFs shown in Fig. 14, as well as the increase in  $T_c$  with decreasing background intensity.<sup>140</sup>

The shape of the temporal sensitivity function depends upon the spatial frequency of the stimulus. Figure 15 shows temporal CSFs measured for spatial sinuswave gratings whose contrast is modulated sinusoidally in time.<sup>147</sup> With increasing spatial frequency, the temporal CSF evolves from a bandpass function with high sensitivity (as in Fig. 14) to a low-pass function of lower sensitivity.<sup>128,147</sup> The solid curves passing through the data for the high spatial-frequency targets and the dashed curves passing through the data for the low spatial-frequency targets are identical except for vertical shifting. Similar behavior occurs for spatial CSFs at high spatial frequencies: changes in temporal frequency cause a vertical sensitivity shift (in log coordinates). Such behavior shows that the spatiotemporal CSF is separable at high temporal and spatial frequencies; that is, sensitivity can be predicted from the product



**FIGURE 15** Temporal contrast sensitivity functions measured for several different spatial frequencies. The contrast of spatial sinewave gratings were modulated sinusoidally in time. The solid curves passing through the data for the high spatial-frequency targets and the dashed curves passing through the data for the low-frequency targets are identical except for vertical shifting. (*Reproduced from Ref. 147.*)

of the spatial and temporal CSFs at the appropriate frequencies. This finding suggests that the same anatomical and physiological mechanisms that determine the high-frequency slope of spatial and temporal CSFs determine sensitivity for the high-frequency regions of the spatiotemporal CSF.

The spatiotemporal CSF is, however, clearly not separable at low spatial and temporal frequencies, so an explanation of the underlying anatomical and physiological constraints is more complicated. The lack of spatiotemporal separability at low spatiotemporal frequencies has been interpreted by some as evidence for separate populations of visual neurons tuned to different temporal frequencies<sup>143,148</sup> and by others as evidence for different spatiotemporal properties of center and surround mechanisms within the same populations of neurons.<sup>147,149</sup> It is likely that a combination of the two explanations is the correct one. For a review of the evidence concerning this issue prior to 1984, see Ref. 140; for more recent evidence, see Refs. 150, 151, and 152.

The available evidence indicates that the temporal CSF varies little with retinal eccentricity once differences in spatial resolution are taken into account.<sup>153,154</sup>

**Chromatic Contrast Detection** The discussion has, so far, concerned luminance contrast detection (contrast detection based upon changes in luminance over space or time); however, there is also a large literature concerned with chromatic contrast detection (contrast detection based upon changes in wavelength distribution over space or time). Although the focus of this chapter is on achromatic vision, it is appropriate to say a few words about chromatic contrast sensitivity functions. A chromatic spatial CSF is obtained by measuring thresholds for the detection of sinewave gratings that modulate spatially in wavelength composition, without modulating in luminance.\* This is typically accomplished by adding two sinewave gratings of different wavelength distributions but the

\*The standard definition of luminance is based upon “the  $V(\lambda)$  function” [see Eq. (5)] which is an average of spectral sensitivity measurements on young human observers. Precision work in color vision often requires taking individual differences into account. To do this a psychophysical procedure, such as flicker photometry, is used to define luminance separately for each observer in the study. For more details, see Ref. 5 and Chap. 11.

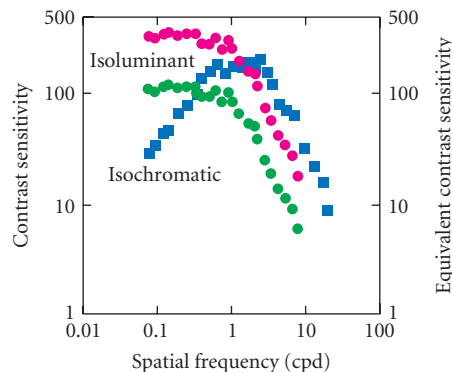
same luminance amplitude. The two gratings are added in opposite spatial phase so that the mean luminance is constant across the pattern; such a pattern is said to be *isoluminant*. Formally, an isoluminant sinewave grating can be defined as follows:

$$I(x, y) = (A \sin(2\pi fx) + \bar{I}_1) + (-A \sin(2\pi fx) + \bar{I}_2) \quad (28)$$

where the first term on the right represents the grating for one of the wavelength distributions, and the second term on the right for the other distribution. The terms  $\bar{I}_1$  and  $\bar{I}_2$  are equal in units of luminance. Contrast sensitivity at spatial frequency  $f$  is measured by varying the common amplitude,  $A$ . (Note that  $I(x, y)$  is constant independent of  $A$ , and that when  $A = 0.0$  the image is a uniform field of constant wavelength composition.) Chromatic contrast is typically defined as  $C = 2A/(\bar{I}_1 + \bar{I}_2)$ .

The green circles in Fig. 16 are an isoluminant CSF for a combination of red and green gratings.<sup>155</sup> The squares are a luminance (isochromatic) CSF measured with the same procedure, on the same subjects. As can be seen, there are three major differences between luminance and chromatic CSFs:<sup>156</sup> (1) the high-frequency falloff of the chromatic CSFs occurs at lower spatial frequencies (acuity is worse); (2) the chromatic CSF has a much weaker (or absent) low-frequency falloff, even under steady viewing conditions; and (3) chromatic contrast sensitivity is greater than luminance contrast sensitivity at low spatial frequencies.

Two of the factors contributing to the lower high-frequency limb of the chromatic (red/green) CSF are the large overlap in the absorption spectra of the L (red) and M (green) cones, and the overlap (when present) of the wavelength distributions of the component gratings. Both of these factors reduce the effective contrast of the grating at the photoreceptors. A precise way to quantify these effects (as well as those of the other preneural factors) is with an ideal-observer analysis (e.g., Refs. 78,124,157). For example, one can physically quantify the equivalent luminance contrast of a chromatic grating by (1) computing the detection accuracy of an ideal observer, operating at the level of the photoreceptor photopigments, for the chromatic grating at the given chromatic contrast, and then (2) computing the equivalent luminance contrast required to bring the ideal observer to the same accuracy level.<sup>124,158</sup> The magenta circles in Fig. 16 are the chromatic CSF data replotted in terms



**FIGURE 16** Isoluminant and isochromatic contrast sensitivity functions. The blue squares are isochromatic (luminance) CSFs for a green grating (526 nm). The green circles are isoluminant (chromatic) CSFs for gratings modulating between red (602) and green (526 nm). The magenta circles are the isoluminant data replotted as equivalent luminance contrast sensitivity (right axis). The isochromatic CSF plots the same on the left and right axes. (Data from Ref. 155.)

of equivalent luminance contrast. (The luminance CSF data are, of course, unaffected and hence are the same on both scales.) If the differences between the high-frequency falloffs of the luminance and chromatic CSFs were due to preneural factors alone, then the high-frequency limbs (magenta circles and blue squares) would superimpose. The analysis suggests that there are also neural factors contributing to the difference between high-frequency limbs.

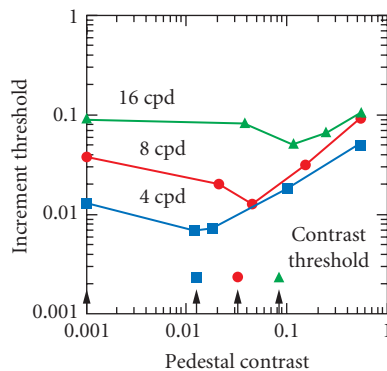
One difficulty in interpreting the high-frequency limb of chromatic CSFs is the potential for luminance artifacts due to chromatic aberration.<sup>156</sup> Recently Sekiguchi et al.<sup>124</sup> eliminated chromatic aberration artifacts by producing isoluminant gratings with a laser interferometer which effectively bypasses the optics of the eye. Their results are similar to those in Fig. 16.

Comparison of the magenta circles and blue squares in Fig. 16 shows that the neural mechanisms are considerably more efficient at detecting chromatic contrast than luminance contrast at spatial frequencies below 2 cpd. One possible explanation is based upon the receptive-field properties of neurons in the retina and LGN (e.g., see Refs. 48 and 159). Many retinal ganglion cells (and presumably bipolar cells) are chromatically opponent in the sense that different classes of photoreceptors dominate the center and surround responses. The predicted CSFs of linear receptive fields with chromatically opponent centers and surrounds are qualitatively similar to those in Fig. 16.

Chromatic temporal CSFs have been measured in much the same fashion as chromatic spatial CSFs except that the chromatic sinewaves were modulated temporally rather than spatially [i.e.,  $x$  represents time rather than space in Eq. (28)]. The pattern of results is also quite similar.<sup>94,156,160</sup> (1) the high-frequency falloff of chromatic CSFs occurs at lower temporal frequencies, (2) the chromatic CSF has a much weaker low-frequency falloff, and (3) chromatic contrast sensitivity is greater than luminance contrast sensitivity at low temporal frequencies. Again, the general pattern of results is qualitatively predicted by a combination of preneural factors and the opponent receptive field properties of retinal ganglion cells (e.g., see Ref. 48).

## Contrast Discrimination and Contrast Masking

The ability to discriminate on the basis of contrast is typically measured by presenting two sinusoids of the same spatial frequency, orientation, and phase, but differing contrasts. The common contrast  $C$  is referred to as the *pedestal contrast* and the additional contrast  $\Delta C$  in one of the sinusoids as the *increment contrast*. Figure 17 shows typical *contrast discrimination functions* obtained from such measurements. The increment contrast varies as a function of the pedestal contrast. At low



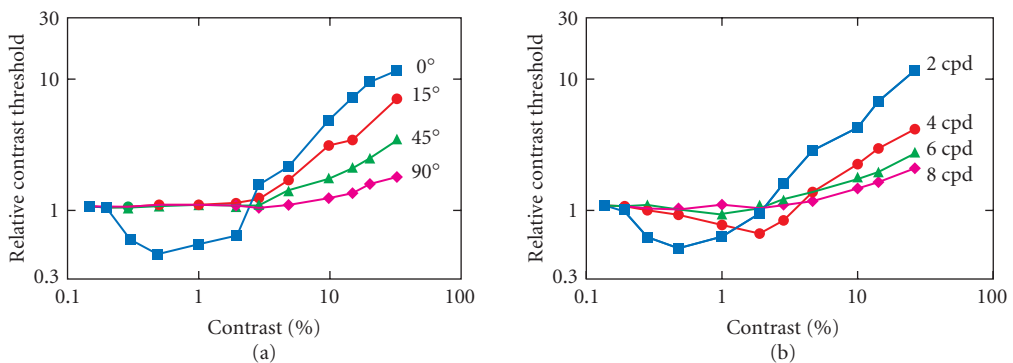
**FIGURE 17** Contrast discrimination functions at different spatial frequencies. The just-discriminable contrast increment is plotted as a function of the contrast common to the two targets. The arrows indicate contrast thresholds at the indicated spatial frequencies. (Adapted from Ref. 78.)

pedestal contrasts,  $\Delta C$  falls initially and then at higher contrasts becomes roughly proportional to  $C^{0.6}$ . Because the exponent is generally less than 1.0, Weber's law does not hold for contrast discrimination. Ideal observer calculations show that pedestal contrast has no effect on discrimination information; therefore, the variations in threshold with pedestal contrast must be due entirely to neural mechanisms. The dip at low pedestal contrasts has been modeled as a consequence of an accelerating nonlinearity in the visual pathways (e.g., Ref. 161) or the consequence of observer uncertainty (e.g., Ref. 162). The evidence currently favors the former.<sup>89</sup> Contrast discrimination functions at different spatial frequencies<sup>78,163</sup> and at different eccentricities<sup>164</sup> are, to a first approximation, similar in shape; that is to say, the functions can be superimposed by shifting the data along the log pedestal and log increment axes by the ratio of detection thresholds.

A generalization of the contrast-discrimination experiment is the contrast-masking experiment in which the increment and pedestal differ from one another in spatial frequency, orientation and/or phase. In contrast-masking experiments, the increment and pedestal are usually referred to as the *target* and *masker*, respectively. Figure 18a shows typical contrast-masking functions obtained for sinewave maskers of different orientations. The orientation of the sinewave target is vertical (0 deg); thus, when the orientation of the masker is also vertical (squares), the result is just a contrast discrimination function, as in Fig. 17. For suprathreshold maskers (in this case, maskers above 2 percent contrast), the threshold elevation produced by the masker decreases as the difference between target and masker orientation increases; indeed, when the target and masker are perpendicular (diamonds), there is almost no threshold elevation even at high masker contrasts. Also, the slope of the contrast-masking function (in log-log coordinates) decreases as the difference in orientation increases. Finally, notice that threshold enhancement at low masking contrasts occurs only when the masker has approximately the same orientation as the target.

Figure 18b shows a similar set of masking functions obtained for sinewave maskers of different spatial frequencies. The effect of varying masker spatial frequency is similar to that of varying masker orientation: with greater differences in spatial frequency between masker and target, masking is reduced, the slope of the contrast-masking function is reduced, and the threshold enhancement at low masker contrasts becomes less evident.

The substantial variations in threshold seen as a function of masker spatial frequency and orientation provide one of the major lines of evidence for the existence of multiple orientation and spatial-frequency-tuned channels in the human visual system;<sup>119,165</sup> for a review, see Ref. 166. Masking data (such as those in Fig. 18) have been analyzed within the framework of specific multiple-channel models in order to estimate the tuning functions of the spatial channels (e.g., Ref. 167). In one popular model, each channel consists of an initial linear filter followed by a static (zero memory) nonlinearity



**FIGURE 18** Contrast masking functions for sinewave-grating maskers of various orientations and spatial frequencies. The target was a 2-cpd sinewave grating. (a) Contrast-masking functions for 2-cpd maskers of various orientations. (b) Contrast-masking functions for 0° (vertical) maskers of various spatial frequencies. (Adapted from Ref. 265.)



that is accelerating at low contrasts and compressive at high contrasts (e.g., Refs. 161, 167, 168). Wilson and colleagues have shown that the spatial-channel tuning functions estimated within the framework of this model can predict discrimination performance in a number of different tasks (for a summary, see Ref. 169). However, recent physiological evidence<sup>70,71</sup> and psychophysical evidence<sup>169,170</sup> suggests that the compressive component of the nonlinearity is a broadly tuned, multiplicative gain mechanism, which is fundamentally different from the accelerating component. The full implications of these findings for the psychophysical estimation of channel-tuning functions and for the prediction of discrimination performance are not yet known.

Masking paradigms have also been used to estimate the tuning of temporal channels.<sup>151,171</sup>

## Contrast Estimation

We have argued that the differential visibility of coarse and fine patterns can be understood from an analysis of information losses in the early stages of vision: optics, receptor sampling, the inverse relation between the preferred spatial frequency and size of tuned spatial mechanisms, and retinal adaptation mechanisms. It is important to note, however, that the differential visibility described by the CSF in Fig. 11 does not relate directly to perceptual experience. For example, if you hold this page up close and then at arm's length, the apparent contrast of the text does not change appreciably even though the spatial frequency content (in cycles per degree) does. This observation has been demonstrated experimentally by asking observers to adjust the contrast of a sinusoidal grating of one spatial frequency until it appeared to match the contrast of a sinusoid of a different frequency.<sup>172</sup> For example, consider two gratings, one at 5 cpd and another at 20 cpd; the former is near the peak of the CSF and the latter is well above the peak so it requires nearly a log unit more contrast to reach threshold. The 5-cpd target was set to a fixed contrast and the observer was asked to adjust the contrast of the 20-cpd target to achieve an apparent contrast match. When the contrast of the 5-cpd target was set near threshold, observers required about a log unit more contrast in the higher-frequency target before they had the same apparent contrast. The most interesting result occurred when the contrast of the 5-cpd grating was set to a value well above threshold. Observers then adjusted the contrast of the 20-cpd grating to the same physical value as the contrast of the lower-frequency target. This is surprising because, as described above, two gratings of equal contrast but different spatial frequencies produce different retinal image contrasts (see Chap. 1). In other words, when observers set 5- and 20-cpd gratings to equal physical contrasts, they were accepting as equal in apparent contrast two gratings whose retinal image contrast differed substantially. This implies that the visual system compensates at suprathreshold contrasts for the defocusing effects of the eye's optics and perhaps for the low-pass filtering effects of early stages of processing. This phenomenon has been called *contrast constancy*; the reader might recognize the similarity to deblurring techniques used in aerial and satellite photography (e.g., Ref. 173).

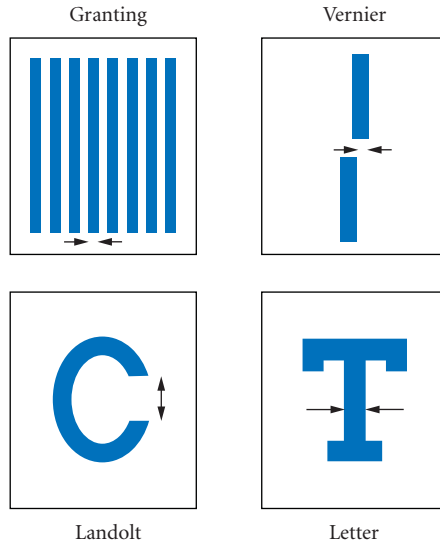
## Visual Acuity

The ability to perceive high-contrast spatial detail is termed *visual acuity*. Measurements of visual acuity are far and away the most common means of assessing ocular health<sup>174</sup> and suitability for operating motor vehicles.<sup>175</sup> The universal use of visual acuity is well justified for clinical assessment,<sup>176</sup> but there is evidence that it is unjustified for automobile licensing (e.g., Ref. 177).

As implied by the discussion of contrast sensitivity, eye movements, optics, receptor properties, and postreceptor neural mechanisms all conspire to limit acuity; one factor may dominate in a given situation, but they all contribute. To assess acuity, high-contrast patterns of various sizes are presented at a fixed distance. The smallest pattern or smallest critical pattern element that can be reliably detected or identified is taken as the threshold value and is usually expressed in minutes of arc. Countless types of stimuli have been used to measure visual acuity, but the four illustrated in Fig. 19 are most common.

Gratings are used to measure minimum angle of resolution (MAR); the spatial frequency of a high-contrast sinusoidal grating is increased until its modulation can no longer be perceived. The





**FIGURE 19** Targets commonly used to assess visual acuity. The dimension of the target that is taken as the acuity threshold is indicated by the arrows.

reader should note that this measurement is equivalent to locating the high-frequency cutoff of the spatial CSFs shown in previous sections. Under optimal conditions, the finest detectable grating bars are about 30 arcsec wide.<sup>112</sup>

The Landolt ring target is a high-contrast ring with a gap appearing in one of four positions. Threshold is defined as the smallest gap that can be correctly located. Under optimal conditions, threshold is again about 30 arcsec or slightly smaller.<sup>112</sup>

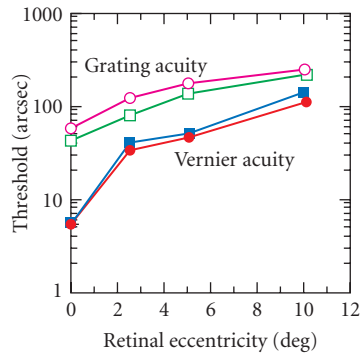
The most common test in clinical and licensing situations is the letter acuity task. The stimuli are a series of solid, high-contrast letters, varying in size. Threshold is defined by the smallest identifiable stroke width which, under optimal conditions, is nearly 30 arcsec.

Finally, the vernier acuity task involves the presentation of two nearly aligned line segments. Threshold is defined by the smallest visible offset from colinearity. Under optimal conditions, vernier acuity is about 5 arcsec.<sup>178,179</sup> The special label, *hyperacuity*, has been provided for spatial thresholds, like vernier acuity, that are smaller than the intercone distance.<sup>180</sup>

In practice, testing environments can vary widely along dimensions that affect visual acuity. These dimensions include illumination, the care chosen to correct refractive errors, and more. The National Academy of Sciences (1980) has provided a set of recommended standards for use in acuity measurements including use of the Landolt ring.

Although these various acuity tasks are clearly related, performance on the four tasks is affected differently by a variety of factors and, for this reason, one suspects that the underlying mechanisms are not the same.<sup>110,181</sup> Figure 20 shows how grating and vernier acuities vary with retinal eccentricity. It is obvious from this figure that no single scaling factor can equate grating and vernier acuity across retinal eccentricities; a scaling factor that would equate grating acuity, for example, would not be large enough to offset the dependence of vernier acuity on retinal eccentricity.

Indeed, there have been several demonstrations that the two types of acuity depend differently on experimental conditions. For example, grating acuity appears more susceptible to optical defects and contrast reductions, whereas vernier, Landolt, and letter acuities are more susceptible to a visual condition called *amblyopia*.<sup>182</sup> There are numerous models attempting to explain why different types



**FIGURE 20** Grating and vernier acuity as a function of retinal eccentricity. The open symbols represent the smallest detectable bar widths for a high-contrast grating. The filled symbols represent the smallest detectable offset of one line segment with respect to another. Circles and squares are the data from two different observers. (*Adapted from Ref. 110.*)

of acuity are affected dissimilarly by experimental conditions, particularly eccentric viewing (e.g., Refs. 183 and 184).

The explanation of spatial discrimination thresholds smaller than the distance between photoreceptors, such as vernier acuity, has become a central issue in contemporary models of spatial vision. One view is that the fine sensitivity exhibited by the hyperacuity<sup>179</sup> reveals a set of local spatial primitives or features that are encoded particularly efficiently in the early stages of vision.<sup>185–187</sup> The other view is that hyperacuity can be understood from an analysis of information processing among the spatially tuned mechanisms described in the earlier sections.<sup>183,188,189</sup> Evidence favoring the second view is manifold. It has been shown, for example, that offset thresholds smaller than the grain of the photoreceptor lattice are predicted from measurements of the linear filtering and signal-to-noise properties of retinal ganglion cells.<sup>188</sup> Likewise, human vernier thresholds can be predicted in many situations from measurements of contrast discrimination thresholds.<sup>190</sup> Also, ideal-observer analysis has demonstrated that the information available at the level of the photoreceptors predicts better performance in hyperacuity tasks than in acuity (resolution) tasks,<sup>191</sup> suggesting that any relatively complete set of spatially tuned mechanisms should produce better performance in the hyperacuity tasks. The remaining puzzle for proponents of the second view, however, is why vernier acuity and the other hyperacuity diminish so strikingly with increasing retinal eccentricity.

## Pattern Discrimination

An important aspect of spatial vision is the ability to distinguish one pattern from another. In this section, we consider the ability to distinguish simple, suprathreshold patterns that vary along a single dimension. The dimensions considered are orientation, size or spatial frequency, and position or phase.

The detectability of a pattern varies to some degree with orientation. For most observers, for example, targets are most visible when oriented vertically or horizontally and least when oriented obliquely.<sup>192</sup> This *oblique effect* is demonstrable for periodic and aperiodic targets and is largest for fine-detail targets. The cause of the effect appears to be differential sensitivity among orientation-tuned mechanisms, presumably in the visual cortex.<sup>193</sup> The ability to discriminate gratings differing in orientation depends on several stimulus parameters including target length, contrast, and

spatial frequency, but observers can generally discriminate extended, high-contrast targets that differ by 1 deg or less.

The ability to discriminate gratings differing in spatial frequency or aperiodic stimuli differing in size depends critically on the reference frequency or size. That is to say, the ratio of the frequency discrimination threshold,  $\Delta f$ , to the reference frequency,  $f$ , is roughly constant at about 0.03 for a wide range of reference frequencies;<sup>194</sup> similarly, the ratio of size discrimination threshold to reference size is roughly constant for a variety of reference sizes, except when the reference is quite small.<sup>179</sup>

The encoding of spatial phase is crucial to the identification of spatial patterns. Its importance is demonstrated quite compellingly by swapping the amplitude and phase spectra of two images; the appearance of such hybrid images corresponds to a much greater degree with their phase spectra than with their amplitude spectra.<sup>195</sup> The conventional explanation is that the phase spectrum determines the spatial structure of an image, but this is perhaps too simplistic because the amplitude spectra of most natural images are much more similar to one another than are the phase spectra (e.g., Refs. 196 and 197). Also, given that the result was obtained using global Fourier transforms, it does not have direct implications about the relative importance of phase coding within spatially localized channels.<sup>198</sup> Nonetheless, such demonstrations illustrate the important relationship between the encoding of spatial phase and pattern identification.

In phase discrimination tasks, the observer distinguishes between patterns—usually periodic patterns—that differ only in the phase relationships of their spatial frequency components. The representation of spatial phase in the fovea is very precise: observers can, for instance, discriminate relative phase shifts as small as 2 to 3 deg (of phase angle) in compound gratings composed of a fundamental and third harmonic of 5 cpd;<sup>199</sup> this is equivalent to distinguishing a positional shift of about 5 arcsec. Phase discrimination is much less precise in extrafoveal vision.<sup>200–202</sup> There has been speculation that the discrimination anomalies observed in the periphery underlie the diminished ability to distinguish spatial displacements,<sup>110</sup> to segment textures on the basis of higher-order statistics,<sup>203,204</sup> and to identify complex patterns such as letters,<sup>205</sup> but detailed hypotheses linking phase discrimination to these abilities have not been developed.

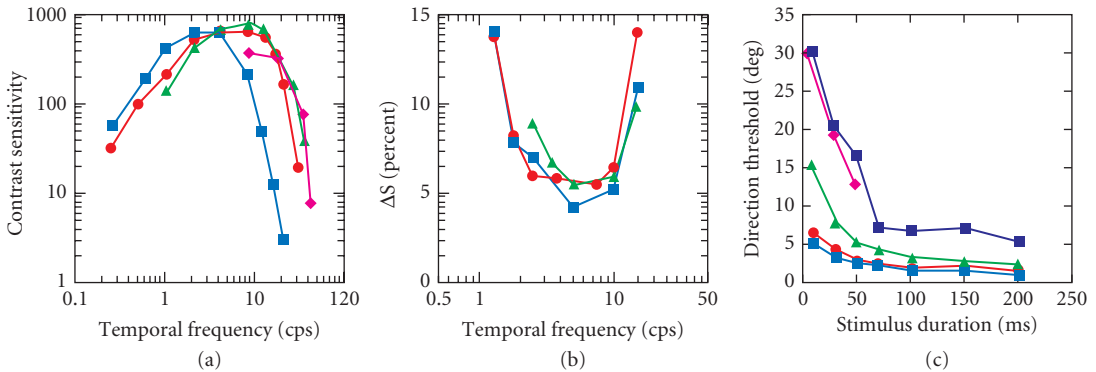
One model of phase discrimination<sup>206</sup> holds that local energy computations are performed on visual inputs and that local energy peaks are singled out for further analysis. The energy computation is performed by cross-correlating the waveform with even- and odd-symmetric spatial mechanisms of various preferred spatial frequencies. The relative activities of the even- and odd-symmetric mechanisms are used to represent the sort of image feature producing the local energy peak. This account is supported by the observation that a two-channel model, composed of even- and odd-symmetric mechanisms, predicts many phase discrimination capabilities well (e.g., Refs. 207–209). The model also offers an explanation for the appearance of illusory Mach bands at some types of edges and not others.<sup>206</sup> For this account to work, however, one must assume that odd-symmetric, and not even-symmetric, mechanisms are much less sensitive in the periphery than in the fovea because some relative phase discriminations are extremely difficult in extrafoveal vision and others are quite simple<sup>202,209,210</sup> (but see Ref. 211).

## Motion Detection and Discrimination

The ability to discriminate the form and magnitude of retinal image motion is critical for many fundamental visual tasks, including navigation through the environment, shape estimation, and distance estimation. There is a vast literature on motion perception (some representative reviews from somewhat different perspectives can be found in Refs. 45, 212, 213). Here, we briefly discuss simple motion discrimination in the frontoparallel plane, and (as an example of more complicated motion processing) observer heading discrimination.

Detection and/or discrimination of motion in the frontoparallel plane is influenced by a number of factors, including contrast, wavelength composition, spatial-frequency content, initial speed, and eccentricity. For simplicity, and for consistency with the other sections in this chapter, we consider here only experiments employing sinewave-grating or random-dot stimuli.

Contrast sensitivity functions for detection of moving (drifting) sinewave-grating targets have the interesting property that they are nearly shape invariant on a log spatial-frequency axis; furthermore,



**FIGURE 21** Motion detection and discrimination thresholds. (a) Contrast sensitivity for drifting sinewave gratings plotted as a function of temporal frequency. Each curve is for a different drift speed; (■) 1 deg/s, (●) 10 deg/s, (▲) 100 deg/s; (◆) 800 deg/s. (Adapted from Ref. 214.) (b) Speed discrimination thresholds plotted as a function of temporal frequency. Each curve is for a different drift speed; (●) 5 deg/s, (■) 10 deg/s, (▲) 20 deg/s. (Adapted from Ref. 215.) (c) Direction threshold for random dot patterns as a function of stimulus duration. Each curve is for a different drift speed; (■) 1 deg/s, (▲) 4 deg/s, (●) 16 deg/s, (■) 64 deg/s, (◆) 256 deg/s. (Adapted from Ref. 216.)

as shown in Fig. 21a, they are nearly superimposed at high speeds and low spatial frequencies, when plotted as a function of temporal frequency.<sup>106,214\*</sup> This latter result corresponds to the fact (mentioned earlier) that spatial CSFs are relatively flat at low spatial frequencies and high temporal frequencies.<sup>128,147</sup> In interpreting Fig. 21a, it is useful to note that the velocity ( $V$ ) of a drifting sinewave grating is equal to the temporal frequency ( $f_t$ ) divided by the spatial frequency ( $f_s$ ):

$$V = \frac{f_t}{f_s} \quad (29)$$

Measurements of CSFs for moving stimuli provide relatively little insight into how the visual system extracts or represents motion information. For example, in one common paradigm<sup>106</sup> it is possible for subjects to perform accurately without “seeing” any motion at all. Greater insight is obtained by requiring the observer to discriminate between different aspects of motion such as speed or direction.

Representative measurements of speed discrimination for drifting sinewave grating targets are shown in Fig. 21b. The figure shows the just-detectable change in speed as a function of temporal frequency; each curve is for a differential initial or base speed. Similar to the motion CSFs, speed discrimination is seen to be largely dependent upon the temporal frequency and relatively independent of the speed.<sup>215</sup> As can be seen, the smallest detectable changes in speed are approximately 5 percent, and they occur in the temporal frequency range of 5 to 10 cps, which is similar to the temporal frequency where contrast sensitivity for drifting gratings is greatest (cf., Fig. 21a). Interestingly, 5 to 10 cps is the temporal frequency range of strongest response for most neurons in the macaque’s primary visual cortex.<sup>60</sup> For random-dot patterns, the smallest detectable change in speed is also approximately 5 percent.<sup>216</sup>

Representative measurements of direction discrimination for drifting random-dot patterns are shown in Fig. 21c, which plots direction threshold in degrees as a function of stimulus duration for a wide range of speeds. Direction discrimination improves with duration and is a U-shaped function of dot speed. Under optimal conditions, direction discrimination thresholds are in the neighborhood of 1 to 2°.<sup>216,217</sup>

The available evidence suggests that direction and speed discrimination improve quickly with contrast at low contrasts but show little improvement when contrast exceeds a few percent.<sup>215,218,219</sup> Motion discrimination is relatively poor at isoluminance.<sup>220,221</sup> The variations in motion discrimination with eccentricity can be largely accounted for by changes in spatial resolution.<sup>222</sup>

\*The temporal frequency of a drifting grating is the number of stripes passing a fixed reference point per unit time (seconds).

Psychophysical evidence suggests that there are at least two mechanisms that can mediate motion discrimination—a *short-range mechanism* which appears to be closely associated with motion-selective neurons in the early levels of cortical processing, and a *long-range mechanism* which appears to be associated with more “inferential” calculations occurring in later levels of cortical processing. Clear evidence for this view was first observed in an apparent motion paradigm in which observers were required to judge the shape of a region created by laterally displacing a subset of random dots in a larger random-dot pattern.<sup>223</sup> (These patterns are examples of *random-dot kinematograms*, the motion analog to random-dot stereograms.) Observers could accurately judge the shape of the region only if the displacements (1) were less than approximately 15 minarc ( $D_{\max} = 15 \text{ min}$ ); (2) occurred with a delay of less than approximately 100 ms; and (3) occurred within the same eye (i.e., dichoptic presentations failed to produce reliable discrimination). These data suggest the existence of motion-sensitive mechanisms that can only detect local, monocular correlations in space-time.

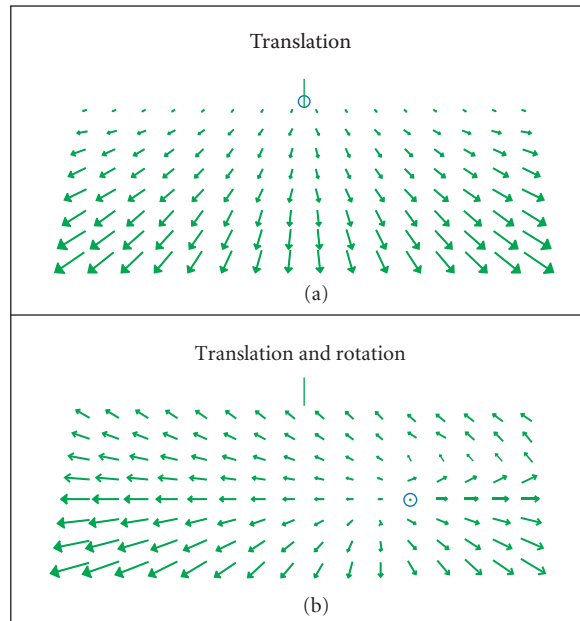
The fact that motion-selective neurons in the early levels of the cortex can also only detect local spatiotemporal correlations suggests that they may be the primary source of neural information used by the subsequent mechanisms that extract shape and distance information in random-dot kinematograms. The receptive fields of motion-selective neurons decrease in size as a function of optimal spatial frequency, and they increase in size as a function of retinal eccentricity. Thus, the hypothesis that motion-selective neurons are the substrate for shape extraction in random-dot kinematograms is supported by the findings that  $D_{\max}$  increases as spatial-frequency content is decreased via low-pass filtering,<sup>224</sup> and that  $D_{\max}$  increases as total size of the kinematogram is increased.<sup>225</sup>

Accurate shape and motion judgments can sometimes be made in apparent-motion displays where the stimulation is dichoptic and/or the spatiotemporal displacements are large. However, in these cases the relevant shape must be clearly visible in the static frames of the kinematogram. The implication is that higher-level, inferential analyses are used to make judgments under these circumstances. A classic example is that we can infer the motion of the hour hand on a clock from its long-term changes in position even though the motion is much too slow to be directly encoded by motion-selective cortical neurons.

The evidence for short-term and long-term motion mechanisms raises the question of what mechanisms actually mediate performance in a given motion discrimination task. The studies described above (e.g., Fig. 21) either randomized stimulus parameters (such as duration and spatial frequency) and/or used random-dot patterns; hence, they most likely measured properties of the short-range mechanisms. However, even with random-dot patterns, great care must be exercised when interpreting discrimination experiments. For example, static pattern-processing mechanisms may sometimes contribute to motion discrimination performance even though performance is at chance when static views are presented at great separations of space or time. This might occur because persisting neural responses from multiple stimulus frames merge to produce “virtual” structured patterns (e.g., Glass patterns<sup>226</sup>). Similar caution should be applied when interpreting discrimination performance with random-dot stereograms (see below).

The discussion of motion perception to this point has focused on the estimation of velocities of various stimuli. Such estimation is importantly involved in guiding locomotion through and estimating the three-dimensional layout of cluttered environments. Motion of an observer through a rigid visual scene produces characteristic patterns of motion on the retina called the *optic flow field*.<sup>227,228</sup> Observers are able to judge the direction of their own motion through the environment based upon this optical flow information alone.<sup>229,230</sup> In fact, accurate motion perception is possible even for random-dot flow fields, suggesting that the motion-selective neurons in the early levels of the visual cortex are a main source of information in perceiving self-motion. But, how is the local motion information provided by the motion-selective neurons used by subsequent mechanisms to compute observer motion?

Gibson<sup>227,228</sup> proposed that people identify their direction of self-motion with respect to obstacles by locating the source of flow: the focus of expansion. Figure 22a depicts the flow field on the retina as the observer translates while fixating ahead. Flow is directed away from the focus of expansion and this point corresponds to the direction of translation. Not surprisingly, observers can determine their heading to within  $\pm 1$  deg from such a field of motion.<sup>230</sup>



**FIGURE 22** Optic flow fields resulting from forward translation across a rigid ground plane. (a) Flow field in the retinal image when the observer translates straight ahead while maintaining constant eye and head position; the heading is indicated by the small vertical line. (b) Retinal flow field when the observer translates straight ahead while making an eye movement to maintain fixation on the circle; again the heading is indicated by the small vertical line. (Adapted from Ref. 231.)

The situation becomes much more complicated once eye/head movements are considered. Figure 22*b* illustrates this by portraying the flow resulting from forward translation while the observer rotates the eyes to maintain fixation on a point off to the right. This motion does not produce a focus of expansion in the image corresponding to the heading; the only focus corresponds to the point of fixation. Consequently, heading cannot be determined by locating a focus (or singularity) in the retinal flow field. Nonetheless, human observers are able to locate their heading to within  $\pm 1.5$  deg.<sup>231</sup>

Recent theoretical efforts have concentrated on this problem of computing heading in the presence of rotations due to eye/head movements. There are two types of models. One holds that observers measure the rotational flow components due to eye/head movements by means of an extraretinal signal; that is, the velocity of rotations is signaled by proprioceptive feedback from the extraocular and/or neck muscles or by efferent information.<sup>232–236</sup> The rotational flow components are then subtracted from the flow field as a whole to estimate the heading. The other type of model holds that people determine heading in the presence of rotations from retinal image information alone (e.g., Refs. 233–236). These models hypothesize that the visual system decomposes the flow field into rotational and translational components by capitalizing on the fact that flows due to translation and rotation depend differently on the scene geometry. Once the decomposition is performed, the heading can be estimated from the translational components (see Fig. 22*a*).

Current experimental evidence favors the extraretinal signal model,<sup>237,238</sup> but there is also evidence supporting the retinal-image model when the eye-head rotations are slow.<sup>231</sup>

The optic flow field also contains information specifying the relative depths of objects in the visual scene: The velocities of retinal image motions are, to a first approximation, proportional to the inverse depths of the corresponding objects.<sup>239</sup> Human observers are quite sensitive to this depth

cue as evidenced by the fact that they perceive depth variation for differential motions of as small as  $\frac{1}{2} - \frac{1}{3}$  arcsec per s.<sup>240</sup>

## Binocular Stereoscopic Discrimination

Light from the three-dimensional environment is imaged onto the two-dimensional array of photoreceptors, so the dimension associated with distance from the eye is collapsed in the two-dimensional retinal image. The images in the two eyes generally differ, however, because the eyes are separated by 6 to 7 cm. The differences between the images, which are called *binocular disparities*, are used by the visual system to recover information about the distance dimension. The depth percept that results from binocular disparity is called *stereopsis*. Binocular disparity (stereo) information is formally equivalent to the monocular information that results from a step translation of one eye by 6 to 7 cm. This observation implies a close theoretical connection between the computation of distance from motion and distance from binocular disparity. As indicated in an earlier section, the extraction of stereo and motion information presumably begins in the early stages of cortical processing (i.e., in V1 of the macaque), where many neurons are selective for disparity and direction of motion.

The geometry of stereopsis is described in Chap. 1, Fig. 25. Here, we briefly consider the nature of the information available for computing distance from disparity. The distance between any pair of points in the depth dimension ( $\Delta z$ ) is related to the horizontal disparity ( $d$ ) and the horizontal convergence angle of the eyes ( $\theta$ ) by the following formula:<sup>\*</sup>

$$\Delta z = \frac{ad}{(\theta + d_\theta + d/2)(\theta + d_\theta - d/2)} \quad (30)$$

where  $d_\theta$  is the average disparity between the two points and the convergence point, and  $a$  is the interpupillary distance (the distance between the image nodal points of the eyes). If  $d_\theta$  is zero, then the points are centered in depth about the convergence point. The locus of all points where the disparity  $d$  equals zero is known as the (Vieth-Müller) *horopter*. As Eq. (30) indicates, the computation of absolute distance between points from horizontal disparity requires knowledge of the “eye parameters” (the interpupillary distance and the angle of convergence). However, relative distance information is obtained even if the eye parameters are unknown (in other words,  $\Delta z$  is monotonic with  $d$ ). Furthermore, absolute distance can be recovered in principle without knowing the angle of convergence from use of horizontal and vertical binocular disparities (e.g., Refs. 241 and 242).

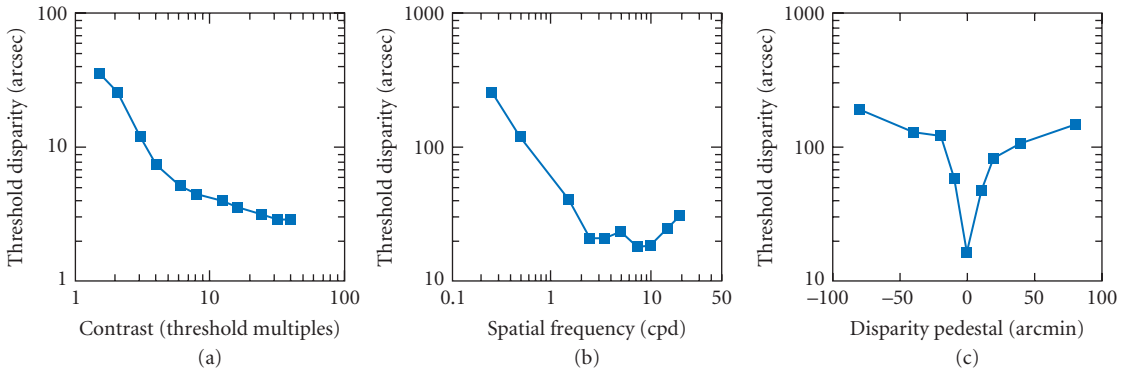
Under optimal conditions, the human visual system is capable of detecting very small binocular disparities and hence rather small changes in distance. In the fovea, disparities of roughly 10 arcsec can be detected if the test objects contain sharp, high-contrast edges and if they are centered in depth about a point on or near the horopter ( $d_\theta = 0.0$ ).<sup>178,243</sup> At a viewing distance of 50 cm, 10 s corresponds to a distance change of about 0.02 cm, but at 50 m, it corresponds to a distance change of about 2 m.

In measuring stereoscopic discrimination performance, it is essential that the task cannot be performed on the basis of monocular information alone. This possibility can be assessed by measuring monocular and binocular performance with the same stimuli; if monocular performance is substantially worse than binocular performance, then binocular mechanisms per se are used in performing the stereoscopic task.<sup>243,244</sup> Random-dot stereograms<sup>244,245</sup> are particularly effective at isolating binocular mechanisms because discriminations cannot be performed reliably when such stereograms are viewed monocularly.<sup>244,246</sup>

Stereopsis is affected by a number of factors. For example, the contrast required to produce a stereoscopic percept varies with the spatial frequency content of the stimulus. This result is quantified by the contrast sensitivity function for stereoscopic discrimination of spatially filtered random-dot stereograms; the function is similar in shape to CSFs measured for simple contrast detection,<sup>247</sup> but

<sup>\*</sup>This formula is an approximation based on the assumption of relatively small angles and object location near the midsagittal plane; it is most accurate for distances under a few meters.





**FIGURE 23** Disparity discrimination thresholds as a function of contrast, spatial frequency, and pedestal disparity. (a) Disparity threshold as a function of luminance contrast (in contrast threshold units) for dynamic random-dot stereograms. (Adapted from Ref. 256.) (b) Disparity threshold as a function of spatial frequency for sinewave-grating stereograms. (Adapted from Ref. 255.) (c) Disparity threshold as a function of the disparity pedestal (distance from the convergence plane) for difference-of-gaussian (DOG) stereograms. (Adapted from Ref. 258.).

detection occurs at lower contrasts than stereopsis occurs. In other words, there is a range of contrasts that are clearly detectable, but insufficient to yield a stereoscopic percept.<sup>247</sup>

The fact that the CSFs for contrast detection and stereoscopic discrimination have similar shapes suggests that common spatial-frequency mechanisms are involved in the two tasks. This hypothesis receives some support from masking and adaptation experiments demonstrating spatial-frequency tuning<sup>248–251</sup> and orientation tuning<sup>252</sup> for stereo discrimination, and from electrophysiological studies demonstrating that cortical cells selective for binocular disparity are also usually selective for orientation<sup>55,56,62</sup> and spatial frequency.<sup>253</sup>

The smallest detectable disparity, which is called *stereoacuity*, improves as luminance contrast is increased.<sup>254–256</sup> As shown in Fig. 23a, stereoacuity improves approximately in inverse proportion to the square of contrast at low contrasts, and in inverse proportion to the cube root of contrast at high contrasts. Stereoacuity is also dependent upon spatial frequency; it improves in inverse proportion to target spatial frequency over the low spatial-frequency range, reaching optimum near 3 cpd (Fig. 23b).<sup>255</sup> Both the inverse-square law at low contrasts and the linear law at low spatial frequencies are predicted by signal-detection models that assume independent, additive noise and simple detection rules.<sup>255,256</sup>

Stereoacuity declines precipitously as a function of the distance of the test objects from the convergence plane (Fig. 23c).<sup>257,258</sup> For example, adding a disparity pedestal of 40 minarc to test objects reduces acuity by about 1 log unit. This loss of acuity is not the result of losses of information due to the geometry of stereopsis; it must reflect the properties of the underlying neural mechanisms. [Note in Eq. (30) that adding a disparity to both objects is equivalent to changing the value of  $d_p$ .]

Much like the spatial channels described earlier under “Spatial Channels,” there are multiple channels tuned to different disparities, but it is unclear whether there are a small number of such channels—“near,” “far,” and “tuned”<sup>259</sup>—or a continuum of channels.<sup>56,63,260,261</sup> Models that assume a continuum of channels with finer tuning near the horopter predict both the sharp decline in stereoacuity away from the horopter<sup>261,262</sup> and the shapes of tuning functions that have been estimated from adaptation and probability summation experiments.<sup>260,261</sup>

## 2.5 ACKNOWLEDGMENTS

This project was supported in part by NIH grant EY02688 and AFOSR grant F49620-93-1-0307 to WSG and by NIH grant HD 19927 to MSB.



## 2.6 REFERENCES

1. R. Navarro, P. Artal, and D. R. Williams, "Modulation Transfer Function of the Human Eye as a Function of Retinal Eccentricity," *Journal of the Optical Society of America A* **10**:201–212 (1993).
2. J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, Wiley, New York, 1978.
3. J. W. Goodman, *Introduction to Fourier Optics*, Physical and Quantum Electronics Series, H. Heffner and A. E. Siegman (eds.), McGraw-Hill, New York, 1968.
4. F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *Journal of Physiology* **186**:558–578 (London, 1966).
5. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2d ed., Wiley, New York, 1982.
6. C. A. Curcio, et al., "Human Photoreceptor Topography," *Journal of Comparative Neurology* **292**:497–523 (1990).
7. D. R. Williams, "Aliasing in Human Foveal Vision," *Vision Research* **25**:195–205 (1985).
8. L. N. Thibos, F. E. Cheney, and D. J. Walsh, "Retinal Limits to the Detection and Recognition of Gratings," *Journal of the Optical Society of America* **4**:1524–1529 (1987).
9. D. R. Williams, "Topography of the Foveal Cone Mosaic in the Living Human Eye," *Vision Research* **28**:433–454 (1988).
10. R. E. Marc and H. G. Sperling, "Chromatic Organization of Primate Cones," *Science* **196**:454–456 (1977).
11. F. M. de Monasterio, et al., "Density Profile of Blue-Sensitive Cones Along the Horizontal Meridian of Macaque Retina," *Investigative Ophthalmology and Visual Science* **26**:283–288 (1985).
12. P. Ahnelt, C. Keri, and H. Kolb, "Identification of Pedicles of Putative Blue-Sensitive Cones in the Human Retina," *Journal of Comparative Neurology* **293**:39–53 (1990).
13. D. R. Williams, D. I. A. MacLeod, and M. M. Hayhoe, "Foveal Tritanopia," *Vision Research* **21**:1341–1356 (1981).
14. J. L. Nerger and C. M. Cicerone, "The Ratio of L Cones to M Cones in the Human Parafoveal Retina," *Vision Research* **32**:879–888 (1992).
15. C. M. Cicerone, "Color Appearance and the Cone Mosaic in Trichromacy and Dichromacy," *Color Vision Deficiencies*, Y. Ohta (ed.) Kugler & Ghedini, Amsterdam, 1990, pp. 1–12.
16. J. K. Bowmaker, "Visual Pigments and Colour in Primates," *From Pigments to Perception: Advances in Understanding Visual Processes*, A. Valberg and B. L. Lee (eds.) Plenum Press, New York, 1991.
17. J. L. Schnapf et al., "Visual Transduction in Cones of the Monkey Macaca Fascicularis," *Journal of Physiology* **427**:681–713 (London, 1990).
18. W. A. H. Rushton, "The Difference Spectrum and the Photosensitivity of Rhodopsin in the Living Human Retina," *Journal of Physiology* **134**:11–297 (London, 1956).
19. W. A. H. Rushton and G. H. Henry, "Bleaching and Regeneration of Cone Pigments in Man," *Vision Research* **8**:617–631 (1968).
20. M. Alpern and E. N. Pugh, "The Density and Photosensitivity of Human Rhodopsin in the Living Retina," *Journal of Physiology* **237**:341–370 (London, 1974).
21. W. H. Miller and G. D. Bernard, "Averaging over the Foveal Receptor Aperture Curtails Aliasing," *Vision Research* **23**:1365–1369 (1983).
22. D. I. A. MacLeod, D. R. Williams, and W. Makous, "A Visual Nonlinearity Fed by Single Cones," *Vision Research* **32**:347–363 (1992).
23. J. L. Schnapf and D. A. Baylor, "How Photoreceptor Cells Respond to Light," *Scientific American* **256**(4):40–47 (1987).
24. J. M. Valetton and D. van Norren, "Light-Adaptation of Primate Cones: An Analysis Based on Extracellular Data," *Vision Research* **23**:1539–1547 (1982).
25. D. A. Baylor, B. J. Nunn, and J. L. Schnapf, "The Photocurrent, Noise and Spectral Sensitivity of Rods of the Monkey Macaca Fascicularis," *Journal of Physiology* **357**:576–607 (London, 1984).
26. D. C. Hood and D. G. Birch, "A Quantitative Measure of the Electrical Activity of Human Rod Photoreceptors Using Electroretinography," *Visual Neuroscience* **5**:379–387 (1990).

27. E. N. Pugh and T. D. Lamb, "Cyclic GMP and Calcium: The Internal Messengers of Excitation and Adaptation in Vertebrate Photoreceptors," *Vision Research* **30**(12):1923–1948 (1990).
28. D. C. Hood and D. G. Birch, "Human Cone Receptor Activity: The Leading Edge of the A-Wave and Models of Receptor Activity," *Visual Neuroscience* **10**:857–871 (1993).
29. R. W. Rodieck, "The Primate Retina," *Comparative Primate Biology, Neurosciences*, H. D. Steklis and J. Erwin (eds.), Liss, New York, 1988, pp. 203–278.
30. P. Sterling, "Retina," *The Synaptic Organization of the Brain*, G. M. Sheperd (ed.), Oxford University Press, New York, 1990, pp. 170–213.
31. H. Wässle and B. B. Boycott, "Functional Architecture of the Mammalian Retina," *Physiological Reviews*, **71**(2):447–480 (1991).
32. C. A. Curcio and K. A. Allen, "Topography of Ganglion Cells in the Human Retina," *Journal of Comparative Neurology* **300**:5–25 (1990).
33. H. Wässle et al., "Retinal Ganglion Cell Density and Cortical Magnification Factor in the Primate," *Vision Research* **30**(11):1897–1911 (1990).
34. S. W. Kuffler, "Discharge Patterns and Functional Organization of the Mammalian Retina," *Journal of Neurophysiology* **16**:37–68 (1953).
35. R. W. Rodieck, "Quantitative Analysis of Cat Retinal Ganglion Cell Response to Visual Stimuli," *Vision Research* **5**:583–601 (1965).
36. C. Enroth-Cugell and J. G. Robson, "The Contrast Sensitivity of Retinal Ganglion Cells of the Cat," *Journal of Physiology* **187**:517–552 (London, 1966).
37. A. M. Derrington and P. Lennie, "Spatial and Temporal Contrast Sensitivities of Neurones in Lateral Geniculate Nucleus of Macaque," *Journal of Physiology* **357**:219–240 (London, 1984).
38. K. Purpura et al., "Light Adaptation in the Primate Retina: Analysis of Changes in Gain and Dynamics of Monkey Retinal Ganglion Cells," *Visual Neuroscience* **4**:75–93 (1990).
39. R. M. Shapley and H. V. Perry, "Cat and Monkey Retinal Ganglion Cells and Their Visual Functional Roles," *Trends in Neuroscience* **9**:229–235 (1986).
40. B. B. Lee et al., "Luminance and Chromatic Modulation Sensitivity of Macaque Ganglion Cells and Human Observers," *Journal of the Optical Society of America* **7**(12):2223–2236 (1990).
41. B. G. Cleland and C. Enroth-Cugell, "Quantitative Aspects of Sensitivity and Summation in the Cat Retina," *Journal of Physiology* **198**:17–38 (London, 1968).
42. A. M. Derrington and P. Lennie, "The Influence of Temporal Frequency and Adaptation Level on Receptive Field Organization of Retinal Ganglion Cells in Cat," *Journal of Physiology* **333**:343–366 (London, 1982).
43. J. M. Cook et al., "Visual Resolution of Macaque Retinal Ganglion Cells," *Journal of Physiology* **396**:205–224 (London, 1988).
44. J. H. R. Maunsell and W. T. Newsome, "Visual Processing in Monkey Extrastriate Cortex," *Annual Review of Neuroscience* **10**:363–401 (1987).
45. K. Nakayama, "Biological Image Motion Processing: A Review," *Vision Research* **25**(5):625–660 (1985).
46. W. H. Merigan and J. H. R. Maunsell, "How Parallel Are the Primate Visual Pathways," *Annual Review of Neuroscience* **16**:369–402 (1993).
47. D. C. Van Essen, "Functional Organization of Primate Visual Cortex," *Cerebral Cortex*, A. A. Peters and E. G. Jones (eds.), Plenum, New York, 1985, pp. 259–329.
48. R. L. DeValois and K. K. DeValois, *Spatial Vision*, Oxford University Press, New York, 1988.
49. P. Lennie et al., "Parallel Processing of Visual Information," *Visual Perception: The Neurophysiological Foundations*, L. Spillman and J. S. Werner (eds.), Academic Press, San Diego, 1990.
50. R. L. DeValois, I. Abramov, and G. H. Jacobs, "Analysis of Response Patterns of LGN Cells," *Journal of the Optical Society of America* **56**:966–977 (1966).
51. T. N. Wiesel and D. H. Hubel, "Spatial and Chromatic Interactions in the Lateral Geniculate Body of the Rhesus Monkey," *Journal of Neurophysiology* **29**:1115–1156 (1966).
52. E. Kaplan and R. Shapley, "The Primate Retina Contains Two Types of Ganglion Cells, with High and Low Contrast Sensitivity," *Proceedings of the National Academy of Sciences* **83**:125–143 (U.S.A., 1986).
53. E. Kaplan, K. Purpura, and R. M. Shapley, "Contrast Affects the Transmission of Visual Information through the Mammalian Lateral Geniculate Nucleus," *Journal of Physiology* **391**:267–288 (London, 1987).

54. D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *Journal of Physiology* **195**:215–243 (London, 1968).
55. D. Hubel and T. Wiesel, "Cells Sensitive to Binocular Depth in Area 18 of the Macaque Monkey Cortex," *Nature* **225**:41–42 (1970).
56. G. F. Poggio and B. Fischer, "Binocular Interaction and Depth Sensitivity in Striate and Prestriate Cortex of Behaving Rhesus Monkey," *Journal of Neurophysiology* **40**:1392–1405 (1977).
57. R. L. DeValois, E. W. Yund, and N. Hepler, "The Orientation and Direction Selectivity of Cells in Macaque Visual Cortex," *Vision Research* **22**:531–544 (1982).
58. R. L. DeValois, D. G. Albrecht, and L. G. Thorell, "Spatial Frequency Selectivity of Cells in Macaque Visual Cortex," *Vision Research* **22**:545–559 (1982).
59. H. B. Barlow, "Critical Limiting Factors in the Design of the Eye and Visual Cortex," *Proceedings of the Royal Society of London*, series B **212**:1–34 (1981).
60. K. H. Foster et al., "Spatial and Temporal Frequency Selectivity of Neurones in Visual Cortical Areas V1 and V2 of the Macaque Monkey," *Journal of Physiology* **365**:331–363 (London, 1985).
61. G. F. Poggio, F. Gonzalez, and F. Krause, "Stereoscopic Mechanisms in Monkey Visual Cortex: Binocular Correlation and Disparity Selectivity," *Journal of Neuroscience* **8**:4531–4550 (1988).
62. H. B. Barlow, C. Blakemore, and J. D. Pettigrew, "The Neural Mechanism of Binocular Depth Discrimination," *Journal of Physiology* **193**:327–342 (London, 1967).
63. S. LeVay and T. Voigt, "Ocular Dominance and Disparity Coding in Cat Visual Cortex," *Visual Neuroscience* **1**:395–414 (1988).
64. M. S. Livingstone and D. H. Hubel, "Segregation of Form, Color, Movement and Depth: Anatomy, Physiology and Perception," *Science* **240**:740–750 (1988).
65. P. Lennie, J. Krauskopf, and G. Sclar, "Chromatic Mechanisms in Striate Cortex of Macaque," *Journal of Neuroscience* **10**:649–669 (1990).
66. M. J. Hawken and A. J. Parker, "Spatial Properties of Neurons in the Monkey Striate Cortex," *Proceedings of the Royal Society of London*, series B **231**:251–288 (1987).
67. D. B. Hamilton, D. G. Albrecht, and W. S. Geisler, "Visual Cortical Receptive Fields in Monkey and Cat: Spatial and Temporal Phase Transfer Function," *Vision Research* **29**:1285–1308 (1989).
68. B. C. Skottun et al., "Classifying Simple and Complex Cells on the Basis of Response Modulation," *Vision Research* **31**(7/8):1079–1086 (1991).
69. D. G. Albrecht and D. H. Hamilton, "Striate Cortex of Monkey and Cat: Contrast Response Function," *Journal of Neurophysiology* **48**(1):217–237 (1982).
70. D. G. Albrecht and W. S. Geisler, "Motion Selectivity and the Contrast-Response Function of Simple Cells in the Visual Cortex," *Visual Neuroscience* **7**:531–546 (1991).
71. D. J. Heeger, "Nonlinear Model of Neural Responses in Cat Visual Cortex," *Computational Models of Visual Processing*, M. S. Landy and A. Movshon (eds.), MIT Press; Cambridge, Mass., 1991, pp. 119–133.
72. D. J. Tolhurst, J. A. Movshon, and A. F. Dean, "The Statistical Reliability of Signals in Single Neurons in the Cat and Monkey Visual Cortex," *Vision Research* **23**:775–785 (1983).
73. W. S. Geisler et al., "Discrimination Performance of Single Neurons: Rate and Temporal-Pattern Information," *Journal of Neurophysiology* **66**:334–361 (1991).
74. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, New York, 1968.
75. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Krieger, New York, 1974.
76. H. B. Barlow, "Temporal and Spatial Summation in Human Vision at Different Background Intensities," *Journal of Physiology* **141**:337–350 (London, 1958).
77. A. Rose, "The Sensitivity Performance of the Human Eye on an Absolute Scale," *Journal of the Optical Society of America* **38**:196–208 (1948).
78. W. S. Geisler, "Sequential Ideal-Observer Analysis of Visual Discrimination," *Psychological Review* **96**:267–314 (1989).
79. D. G. Pelli, "The Quantum Efficiency of Vision," *Vision: Coding and Efficiency*, C. Blakemore (ed.), Cambridge University Press, Cambridge, 1990, pp. 3–24.
80. H. B. Barlow, "The Efficiency of Detecting Changes of Density in Random Dot Patterns," *Vision Research* **18**:637–650 (1978).

81. H. de Vries, "The Quantum Character of Light and Its Bearing upon Threshold of Vision, the Differential Sensitivity and Visual Acuity of the Eye," *Physica* **10**:553–564 (1943).
82. H. B. Barlow, "Increment Thresholds at Low Intensities Considered as Signal/Noise Discriminations," *Journal of Physiology* **136**:469–488 (London, 1957).
83. H. Helstrom, "The Detection and Resolution of Optical Signals," *IEEE Transactions on Information Theory* **IT-10**:275–287 (1964).
84. W. S. Geisler, "Physical Limits of Acuity and Hyperacuity," *Journal of the Optical Society of America A* **1**:775–782 (1984).
85. W. S. Geisler and K. Chou, "Separation of Low-Level and High-Level Factors in Complex Tasks: Visual Search," *Psychological Review* 1994 (in press).
86. M. E. Rudd, "Quantal Fluctuation Limitations on Reaction Time to Sinusoidal Gratings," *Vision Research* **28**:179–186 (1988).
87. A. E. Burgess et al., "Efficiency of Human Visual Signal Discrimination," *Science* **214**:93–94 (1981).
88. D. Kersten, "Statistical Efficiency for the Detection of Visual Noise," *Vision Research* **27**:1029–1040 (1987).
89. G. E. Legge, D. Kersten, and A.E. Burgess, "Contrast Discrimination in Noise," *Journal of the Optical Society of America A* **4**:391–404 (1987).
90. A. E. Burgess, R. F. Wagner, and R. J. Jennings, "Human Signal Detection Performance for Noisy Medical Images," *Proceedings of IEEE Computer Society International Workshop on Medical Imaging*, 1982.
91. K. J. Myers et al., "Effect of Noise Correlation on Detectability of Disk Signals in Medical Imaging," *Journal of the Optical Society of America A* **2**:1752–1759 (1985).
92. H. H. Barrett et al., "Linear Discriminants and Image Quality," *Image and Vision Computing* **10**:451–460 (1992).
93. R. H. S. Carpenter, "Eye Movements," *Vision and Visual Dysfunction*, J. Cronley-Dillon (ed.), Macmillan Press, London, 1991.
94. D. Regan and C. W. Tyler, "Some Dynamic Features of Colour Vision," *Vision Research* **11**:1307–1324 (1971).
95. G. E. Poggio and T. Poggio, "The Analysis of Stereopsis," *Annual Reviews of Neuroscience* **7**:379–412 (1984).
96. I. P. Howard, *Human Visual Orientation*, Wiley, New York, 1982.
97. I. P. Howard, "The Perception of Posture, Self Motion, and the Visual Vertical," *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), Wiley, New York, 1986.
98. I. Rock, "The Description and Analysis of Object and Event Perception," *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), Wiley, New York, 1986.
99. R. J. Watt, "Pattern Recognition by Man and Machine," *Vision and Visual Dysfunction*, J. Cronley-Dillon (ed.), Macmillan Press, London, 1991.
100. A. Treisman, "Properties, Parts, and Objects," *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), Wiley, New York, 1986.
101. M. S. Banks and P. Salapatek, "Infant Visual Perception," *Handbook of Child Psychology*, M. M. Haith and J. J. Campos (eds.), Wiley, New York, 1983.
102. K. Simons, *Normal and Abnormal Visual Development*, Springer, New York, 1993.
103. J. Marshall, "The Susceptible Visual Apparatus," *Vision and Visual Dysfunction*, J. Cronley-Dillon (ed.), Macmillan Press, London, 1991.
104. F. Ratliff and L. A. Riggs, "Involuntary Motions of the Eye During Monocular Fixation," *Journal of Experimental Psychology* **46**:687–701 (1950).
105. L. E. Arend, "Response of the Human Eye to Spatially Sinusoidal Gratings at Various Exposure Durations," *Vision Research* **16**:1311–1315 (1976).
106. D. H. Kelly, "Motion and Vision II. Stabilized Spatio-Temporal Threshold Surface," *Journal of the Optical Society of America* **69**:1340–1349 (1979).
107. O. Packer and D. R. Williams, "Blurring by Fixational Eye Movements," *Vision Research* **32**:1931–1939 (1992).
108. D. G. Green, "Regional Variations in the Visual Acuity for Interference Fringes on the Retina," *Journal of Physiology* **207**:351–356 (London, 1970).

109. J. Hirsch and R. Hylton, "Quality of the Primate Photoreceptor Lattice and Limits of Spatial Vision," *Vision Research* **24**:347–355 (1984).
110. D. M. Levi, S. A. Klein, and A. P. Aitsebaomo, "Vernier Acuity, Crowding and Cortical Magnification," *Vision Research* **25**:963–977 (1985).
111. D. S. Jacobs and C. Blakemore, "Factors Limiting the Postnatal Development of Visual Acuity in Monkeys," *Vision Research* **28**:947–958 (1987).
112. S. Shlaer, "The Relation Between Visual Acuity and Illumination," *Journal of General Physiology* **21**:165–188 (1937).
113. D. R. Williams, "Visibility of Interference Fringes Near the Resolution Limit," *Journal of the Optical Society of America A* **2**:1087–1093 (1985).
114. R. N. Bracewell, *The Fourier Transform and Its Applications: Networks and Systems*, S. W. Director (ed.), McGraw-Hill, New York, 1978.
115. N. J. Coletta and H. K. Clark, "Change in Foveal Acuity with Light Level: Optical Factors," *Ophthalmic and Visual Optics*, Optical Society of America, Monterey, Calif., 1993.
116. J. M. Enoch and V. Lakshminarayanan, "Retinal Fibre Optics," *Vision and Visual Dysfunctions*, J. Cronly-Dillon (ed.), Macmillan Press, London, 1991.
117. C. B. Blakemore and F. W. Campbell, "On the Existence of Neurones in the Human Visual System Selectively Sensitive to the Orientation and Size of Retinal Images," *Journal of Physiology* **203**:237–260 (London, 1969).
118. H. R. Wilson and J. R. Bergen, "A Four Mechanism Model for Threshold Spatial Vision," *Vision Research* **19**:19–32 (1979).
119. C. F. Stromeyer and B. Julesz, "Spatial Frequency Masking in Vision: Critical Bands and the Spread of Masking," *Journal of the Optical Society of America* **62**:1221–1232 (1972).
120. J. G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters," *Journal of the Optical Society of America* **2**:1160–1169 (1985).
121. E. R. Howell and R. F. Hess, "The Functional Area for Summation to Threshold for Sinusoidal Gratings," *Vision Research* **18**:369–374 (1978).
122. M. S. Banks, W. S. Geisler, and P. J. Bennett, "The Physical Limits of Grating Visibility," *Vision Research* **27**:1915–1924 (1987).
123. A. Rose, "The Relative Sensitivities of Television Pickup Tubes, Photographic Film, and the Human Eye," *Proceedings of the Institute of Radio Engineers* **30**:293–300 (1942).
124. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Efficiency for Detecting Isoluminant and Isochromatic Interference Fringes," *Journal of the Optical Society of America* **10**:2118–2133 (1993).
125. M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral Spatial Vision: Limits Imposed by Optics, Photoreceptors, and Receptor Pooling," *Journal of the Optical Society of America* **8**:1775–1787 (1991).
126. P. T. Kortum and W. S. Geisler, "Contrast Sensitivity Functions Measured on Flashed Backgrounds in the Dark-Adapted Eye," *Annual Meeting of the Optical Society of America*, Optical Society of America, Albuquerque, NM, 1992.
127. F. L. Van Nes and M. A. Bouman, "Spatial Modulation Transfer in the Human Eye," *Journal of the Optical Society of America* **57**:401–406 (1967).
128. D. H. Kelly, "Adaptation Effects on Spatio-Temporal Sine-Wave Thresholds," *Vision Research* **12**:89–101 (1972).
129. D. C. Hood and M. A. Finkelstein, "Sensitivity to Light," *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), John Wiley and Sons, New York, 1986.
130. J. Walraven et al., "The Control of Visual Sensitivity: Receptor and Postreceptor Processes," *Visual Perception: The Neurophysiological Foundations*, L. Spillman and J. S. Werner (eds.), Academic Press, San Diego, 1990.
131. K. J. W. Craik, "The Effect of Adaptation on Subjective Brightness," *Proceedings of the Royal Society of London*, series B **128**:232–247, 1940.
132. D. C. Hood et al., "Human Cone Saturation as a Function of Ambient Intensity: A Test of Models of Shifts in the Dynamic Range," *Vision Research* **19**:983–993 (1978).
133. W. S. Geisler, "Adaptation, Afterimages and Cone Saturation," *Vision Research* **18**:279–289 (1978).

134. W. S. Geisler, "Effects of Bleaching and Backgrounds on the Flash Response of the Visual System," *Journal of Physiology* **312**:413–434 (London, 1981).
135. M. M. Hayhoe, N. E. Benimoff, and D. C. Hood, "The Time-Course of Multiplicative and Subtractive Adaptation Process," *Vision Research* **27**:1981–1996 (1987).
136. W. S. Geisler, "Mechanisms of Visual Sensitivity: Backgrounds and Early Dark Adaptation," *Vision Research* **23**:1423–1432 (1983).
137. M. M. Hayhoe, M. E. Levin, and R. J. Koshel, "Subtractive Processes in Light Adaptation," *Vision Research* **32**:323–333 (1992).
138. C. A. Burbeck and D. H. Kelly, "Role of Local Adaptation in the Fading of Stabilized Images," *Journal of the Optical Society of America A* **1**:216–220 (1984).
139. U. Tulunay-Keesey et al., "Apparent Phase Reversal During Stabilized Image Fading," *Journal of the Optical Society of America A* **4**:2166–2175 (1987).
140. A. B. Watson, "Temporal Sensitivity," *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), John Wiley and Sons, New York, 1986.
141. H. de Lange, "Research into the Dynamic Nature of the Human Fovea-Cortex Systems with Intermittent and Modulated Light. I. Attenuation Characteristics with White and Colored Light," *Journal of the Optical Society of America* **48**:777–784 (1958).
142. D. H. Kelly, "Visual Responses to Time-Dependent Stimuli. I. Amplitude Sensitivity Measurements," *Journal of the Optical Society of America* **51**:422–429 (1961).
143. J. A. J. Roufs, "Dynamic Properties of Vision-I. Experimental Relationships Between Flicker and Flash Thresholds," *Vision Research* **12**:261–278 (1972).
144. R. M. Boynton and W. S. Baron, "Sinusoidal Flicker Characteristics of Primate Cones in Response to Hererochromatic Stimuli," *Journal of the Optical Society of America* **65**:1091–1100 (1975).
145. D. H. Kelly, R. M. Boynton, and W. S. Baron, "Primate Flicker Sensitivity: Psychophysics and Electrophysiology," *Science* **194**:177–179 (1976).
146. G. E. Legge, "Sustained and Transient Mechanisms in Human Vision: Temporal and Spatial Properties," *Vision Research* **18**:69–81 (1978).
147. J. G. Robson, "Spatial and Temporal Contrast Sensitivity Functions of the Visual System," *Journal of the Optical Society of America* **56**:1141–1142 (1966).
148. J. J. Kulikowski and D. J. Tolhurst, "Psychophysical Evidence for Sustained and Transient Mechanisms in Human Vision," *Journal of Physiology*, **232**:149–163 (London, 1973).
149. C. Burbeck and D. H. Kelly, "Spatiotemporal Characteristics of Visual Mechanisms: Excitatory-Inhibitory Model," *Journal of the Optical Society of America* **70**:1121–1126 (1980).
150. M. B. Mandler and W. Makous, "A Three Channel Model of Temporal Frequency Perception," *Perception* **24**:1881–1887 (1984).
151. S. R. Lehky, "Temporal Properties of Visual Channels Measured by Masking," *Journal of the Optical Society of America A* **2**:1260–1272 (1985).
152. R. J. Snowden and R. F. Hess, "Temporal Properties of Human Visual Filters: Number, Shapes, and Spatial Covariation," *Vision Research* **32**:47–59 (1992).
153. J. J. Koenderink et al., "Perimetry of Contrast Detection Thresholds of Moving Spatial Sine Wave Patterns I. The Near Peripheral Field (0°–8°)," *Journal of the Optical Society of America* **68**:845–849 (1978).
154. V. Virsu et al., "Temporal Contrast Sensitivity and Cortical Magnification," *Vision Research* **22**:1211–1263 (1982).
155. K. T. Mullen, "The Contrast Sensitivity of Human Colour Vision to Red-Green and Blue-Yellow Chromatic Gratings," *Journal of Physiology* **359**:381–400 (London, 1985).
156. G. J. C. van der Horst and M. A. Bouman, "Spatiotemporal Chromaticity Discrimination," *Journal of the Optical Society of America* **59**:1482–1488 (1969).
157. M. S. Banks and P. J. Bennett, "Optical and Photoreceptor Immaturities Limit the Spatial and Chromatic Vision of Human Neonates," *Journal of the Optical Society of America A: Optics and Image Science* **5**:2059–2079 (1988).
158. J. R. Jordan, W. S. Geisler, and A. C. Bovik, "Color as a Source of Information in the Stereo Correspondence Process," *Vision Research* **30**:1955–1970 (1990).



159. P. Lennie and M. D'Zmura, "Mechanisms of Color Vision," *CRC Critical Reviews in Neurobiology* **3**:333–400 (1988).
160. D. H. Kelly, "Luminous and Chromatic Flickering Patterns Have Opposite Effects," *Science* **188**:371–372 (1975).
161. G. E. Legge and J. M. Foley, "Contrast Masking in Human Vision," *Journal of the Optical Society of America* **70**:1458–1470 (1980).
162. D. G. Pelli, "Uncertainty Explains Many Aspects of Visual Contrast Detection and Discrimination," *Journal of the Optical Society of America A* **2**:1508–1532 (1985).
163. A. Bradley and I. Ohzawa, "A Comparison of Contrast Detection and Discrimination," *Vision Research* **26**:991–997 (1986).
164. G. E. Legge and D. Kersten, "Contrast Discrimination in Peripheral Vision," *Journal of the Optical Society of America A* **4**:1594–1598 (1987).
165. F. W. Campbell and J. J. Kulikowski, "Orientation Selectivity of the Human Visual System," *Journal of Physiology* **187**:437–445 (London, 1966).
166. N. Graham, *Visual Pattern Analyzers*, Oxford, New York, 1989.
167. H. R. Wilson, D. K. McFarlane, and G. C. Philips, "Spatial Frequency Tuning of Orientation Selective Units Estimated by Oblique Masking," *Vision Research* **23**:873–882 (1983).
168. H. R. Wilson, "A Transducer Function for Threshold and Suprathreshold Human Vision," *Biological Cybernetics* **38**:171–178 (1980).
169. H. R. Wilson, "Psychophysics of Contrast Gain," *Annual Meeting of the Association for Research in Vision and Ophthalmology*, Investigative Ophthalmology and Visual Science, Sarasota, Fla., 1990.
170. J. M. Foley, "Human Luminance Pattern Vision Mechanisms: Masking Experiments Require a New Theory," *Journal of the Optical Society of America A* (1994), in press.
171. R. F. Hess and R. J. Snowden, "Temporal Properties of Human Visual Filters: Number, Shapes, and Spatial Covariation," *Vision Research* **32**:47–59 (1992).
172. M. A. Georgeson and G. D. Sullivan, "Contrast Constancy: Deblurring in Human Vision by Spatial Frequency Channels," *Journal of Physiology* **252**:627–656 (London, 1975).
173. D. B. Gennery, "Determination of Optical Transfer Function by Inspection of Frequency-Domain Plot," *Journal of the Optical Society of America* **63**:1571–1577 (1973).
174. D. D. Michaels, *Visual Optics and Refraction: A Clinical Approach*, Mosby, St. Louis, 1980.
175. W. N. Charman, "Visual Standards for Driving," *Ophthalmic and Physiological Optics* **5**:211–220 (1985).
176. G. Westheimer, "Scaling of Visual Acuity Measurements," *Archives of Ophthalmology* **97**:327–330 (1979).
177. C. Owsley et al., "Visual/Cognitive Correlates of Vehicle Accidents in Older Drivers," *Psychology and Aging* **6**:403–415 (1991).
178. R. N. Berry, "Quantitative Relations among Vernier, Real Depth and Stereoscopic Depth Acuities," *Journal of Experimental Psychology* **38**:708–721 (1948).
179. G. Westheimer and S. P. McKee, "Spatial Configurations for Visual Hyperacuity," *Vision Research* **17**:941–947 (1977).
180. G. Westheimer, "Visual Acuity and Hyperacuity," *Investigative Ophthalmology* **14**:570–572 (1975).
181. G. Westheimer, "The Spatial Grain of the Perifoveal Visual Field," *Vision Research* **22**:157–162 (1982).
182. D. M. Levi and S. A. Klein, "Vernier Acuity, Crowding and Amblyopia," *Vision Research* **25**:979–991 (1985).
183. S. A. Klein and D. M. Levi, "Hyperacuity Thresholds of 1 sec: Theoretical Predictions and Empirical Validation," *Journal of the Optical Society of America A* **2**:1170–1190 (1985).
184. H. R. Wilson, "Model of Peripheral and Amblyopic Hyperacuity," *Vision Research* **31**:967–982 (1991).
185. H. B. Barlow, "Reconstructing the Visual Image in Space and Time," *Nature* **279**:189–190 (1979).
186. F. H. C. Crick, D. C. Marr, and T. Poggio, "An Information-Processing Approach to Understanding the Visual Cortex," *The Organization of the Cerebral Cortex*, F. O. Smith (ed.), MIT Press, Cambridge, 1981.
187. R. J. Watt and M. J. Morgan, "A Theory of the Primitive Spatial Code in Human Vision," *Vision Research* **25**:1661–1674 (1985).
188. R. Shapley and J. D. Victor, "Hyperacuity in Cat Retinal Ganglion Cells," *Science* **231**:999–1002 (1986).

189. H. R. Wilson, "Responses of Spatial Mechanisms Can Explain Hyperacuity," *Vision Research* **26**:453–469 (1986).
190. Q. M. Hu, S. A. Klein, and T. Carney, "Can Sinusoidal Vernier Acuity Be Predicted by Contrast Discrimination?" *Vision Research* **33**:1241–1258 (1993).
191. W. S. Geisler and K. D. Davila, "Ideal Discriminators in Spatial Vision: Two-Point Stimuli," *Journal of the Optical Society of America A* **2**:1483–1497 (1985).
192. S. Appelle, "Perception and Discrimination As a Function of Stimulus Orientation: The 'Oblique Effect' in Man and Animals," *Psychological Bulletin* **78**:266–278 (1972).
193. F. W. Campbell, R. H. S. Carpenter, and J. Z. Levinson, "Visibility of Aperiodic Patterns Compared with That of Sinusoidal Gratings," *Journal of Physiology* **204**:283–298 (London, 1969).
194. F. W. Campbell, J. Nachmias, and J. Jukes, "Spatial Frequency Discrimination in Human Vision," *Journal of the Optical Society of America* **60**:555–559 (1970).
195. A. V. Oppenheim and J. S. Lim, "The Importance of Phase in Signals," *Proceedings of the IEEE* **69**:529–541 (1981).
196. D. J. Field, "Relations Between the Statistics of Natural Images and the Response Properties of Cortical Cells," *Journal of the Optical Society of America A* **4**:2379–2394 (1987).
197. J. Huang and D. L. Turcotte, "Fractal Image Analysis: Application to the Topography of Oregon and Synthetic Images," *Journal of the Optical Society of America A* **7**:1124–1129 (1990).
198. M. J. Morgan, J. Ross, and A. Hayes, "The Relative Importance of Local Phase and Local Amplitude in Patchwise Image Reconstruction," *Biological Cybernetics* **65**:113–119 (1991).
199. D. R. Badcock, "Spatial Phase or Luminance Profile Discrimination," *Vision Research* **24**:613–623 (1984).
200. I. Rentschler and B. Treutwein, "Loss of Spatial Phase Relationships in Extrafoveal Vision," *Nature* **313**:308–310 (1985).
201. R. F. Hess and J. S. Pointer, "Evidence for Spatially Local Computations Underlying Discrimination of Periodic Patterns in Fovea and Periphery," *Vision Research* **27**:1343–1360 (1987).
202. P. J. Bennett and M. S. Banks, "Sensitivity Loss Among Odd-Symmetric Mechanisms and Phase Anomalies in Peripheral Vision," *Nature* **326**:873–876 (1987).
203. B. Julesz, E. N. Gilbert, and J. D. Victor, "Visual Discrimination of Textures with Identical Third-Order Statistics," *Biological Cybernetics* **31**:137–147 (1978).
204. I. Rentschler, M. Hubner, and T. Caelli, "On the Discrimination of Compound Gabor Signals and Textures," *Vision Research* **28**:279–291 (1988).
205. G. E. Legge et al., "Psychophysics of Reading I. Normal Vision," *Vision Research* **25**:239–252 (1985).
206. M. C. Morrone and D. C. Burr, "Feature Detection in Human Vision: A Phase Dependent Energy Model," *Proceedings of the Royal Society of London, series B* **235**:221–245 (1988).
207. D. J. Field and J. Nachmias, "Phase Reversal Discrimination," *Vision Research* **24**:333–340 (1984).
208. D. C. Burr, M. C. Morrone, and D. Spinelli, "Evidence for Edge and Bar Detectors in Human Vision," *Vision Research* **29**:419–431 (1989).
209. P. J. Bennett and M. S. Banks, "The Effects of Contrast, Spatial Scale, and Orientation on Foveal and Peripheral Phase Discrimination," *Vision Research* **31**:1759–1786, 1991.
210. A. Toet and D. M. Levi, "The Two-Dimensional Shape of Spatial Interaction Zones in the Parafovea," *Vision Research* **32**:1349–1357 (1992).
211. M. C. Morrone, D. C. Burr, and D. Spinelli, "Discrimination of Spatial Phase in Central and Peripheral Vision," *Vision Research* **29**:433–445 (1989).
212. S. Anstis, "Motion Perception in the Frontal Plane: Sensory Aspects," *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas (eds.), John Wiley and Sons, New York, 1986.
213. C. C. Hildreth and C. Koch, "The Analysis of Visual Motion: From Computational Theory to Neuronal Mechanisms," *Annual Review of Neuroscience* **10**:477–533 (1987).
214. D. C. Burr and J. Ross, "Contrast Sensitivity at High Velocities," *Vision Research* **22**:479–484 (1982).
215. S. P. McKee, G. H. Silverman, and K. Nakayama, "Precise Velocity Discrimination despite Random Variations in Temporal Frequency and Contrast," *Vision Research* **26**(4):609–619 (1986).
216. B. De Bruyn and G. A. Orban, "Human Velocity and Direction Discrimination Measured with Random Dot Patterns," *Vision Research* **28**(12):1323–1335 (1988).



217. S. N. J. Watamaniuk, R. Sekuler, and D. W. Williams, "Direction Perception in Complex Dynamic Displays: The Integration of Direct Information," *Vision Research* **29**:47–59 (1989).
218. K. Nakayama and G. H. Silverman, "Detection and Discrimination of Sinusoidal Grating Displacements," *Optical Society of America* **2**(2):267–274 (1985).
219. A. Pantle, "Temporal Frequency Response Characteristics of Motion Channels Measured with Three Different Psychophysical Techniques," *Perception and Psychophysics* **24**:285–294 (1978).
220. P. Cavanagh and P. Anstis, "The Contribution of Color to Motion in Normal and Color-Deficient Observers," *Vision Research* **31**:2109–2148 (1991).
221. D. T. Lindsey and D. Y. Teller, "Motion at Isoluminance: Discrimination/Detection Ratios for Moving Isoluminant Gratings," *Vision Research* **30**(11):1751–1761 (1990).
222. S. P. McKee and K. Nakayama, "The Detection of Motion in the Peripheral Visual Field," *Vision Research* **24**:25–32 (1984).
223. O. J. Braddick, "Low-Level and High-Level Processes in Apparent Motion," *Philosophical Transactions of the Royal Society of London B* **290**:137–151 (1980).
224. J. J. Chang and B. Julesz, "Displacement Limits for Spatial Frequency Filtered Random-Dot Cinematograms in Apparent Motion," *Vision Research* **23**(12):1379–1385 (1983).
225. C. L. Baker and O. J. Braddick, "Does Segregation of Differently Moving Areas Depend on Relative or Absolute Displacement?" *Vision Research* **22**:851–856 (1982).
226. L. Glass, "Moire Effect from Random Dots," *Nature* **243**:578–580 (1969).
227. J. J. Gibson, *The Senses Considered as Perceptual Systems*, Houghton Mifflin, Boston, 1966.
228. J. J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.
229. J. E. Cutting, *Perception with an Eye to Motion*, MIT Press, Cambridge, 1986.
230. W. H. Warren, M. W. Morris, and M. Kalish, "Perception of Translational Heading from Optical Flow," *Journal of Experimental Psychology: Human Perception and Performance* **14**:646–660 (1988).
231. W. H. Warren and D. J. Hannon, "Eye Movements and Optical Flow," *Journal of the Optical Society of America A* **7**:160–169 (1990).
232. E. von Hoist, "Relations between the Central Nervous System and the Peripheral Organs," *Animal Behavior* **2**:89–94 (1954).
233. D. J. Heeger and A. D. Jepson, "Subspace Methods for Recovering Rigid Motion. I: Algorithm and Implementation," *University of Toronto Technical Reports on Research in Biological and Computational Vision*, RBCV-TR-90-35.
234. H. C. Longuet-Higgins and K. Prazdny, "The Interpretation of a Moving Retinal Image," *Proceedings of the Royal Society of London B* **208**:385–397 (1980).
235. J. A. Perrone, "Model for Computation of Self-Motion in Biological Systems," *Journal of the Optical Society of America A* **9**:177–194 (1992).
236. J. H. Rieger and D. T. Lawton, "Processing Differential Image Motion," *Journal of the Optical Society of America A* **2**:354–360 (1985).
237. C. S. Royden, M. S. Banks, and J. A. Crowell, "The Perception of Heading during Eye Movements," *Nature* **360**:583–585 (1992).
238. A. V. van den Berg, "Robustness of Perception of Heading from Optic Flow," *Vision Research* **32**:1285–1296 (1992).
239. J. J. Gibson, P. Olum, and F. Rosenblatt, "Parallax and Perspective During Aircraft Landings," *American Journal of Psychology* **68**:373–385 (1955).
240. B. J. Rogers and M. Graham, "Similarities Between Motion Parallax and Stereopsis in Human Depth Perception," *Vision Research* **22**:261–270 (1982).
241. B. J. Rogers and M. F. Bradshaw, "Vertical Disparities, Differential Perspective and Binocular Stereopsis," *Nature* **361**:253–255 (1993).
242. J. E. W. Mayhew and H. C. Longuet-Higgins, "A Computational Model of Binocular Depth Perception," *Nature* **297**:376–378 (1982).
243. G. Westheimer and S. P. McKee, "What Prior Uniocular Processing Is Necessary for Stereopsis?" *Investigative Ophthalmology and Visual Science* **18**:614–621 (1979).

244. B. Julesz, "Binocular Depth Perception of Computer-Generated Patterns," *Bell System Technical Journal* **39**:1125–1162 (1960).
245. C. M. Aschenbrenner, "Problems in Getting Information Into and Out of Air Photographs," *Photogrammetric Engineering*, **20**:398–401 (1954).
246. B. Julesz, "Stereoscopic Vision," *Vision Research* **26**:1601–1612 (1986).
247. J. P. Frisby and J. E. W. Mayhew, "Contrast Sensitivity Function for Stereopsis," *Perception* **7**:423–429 (1978).
248. C. Blakemore and B. Hague, "Evidence for Disparity Detecting Neurones in the Human Visual System," *Journal of Physiology* **225**:437–445 (London, 1972).
249. T. B. Felton, W. Richards, and R. A. Smith, "Disparity Processing of Spatial Frequencies in Man," *Journal of Physiology* **225**:349–362 (London, 1972).
250. B. Julesz and J. E. Miller, "Independent Spatial Frequency Tuned Channels in Binocular Fusion and Rivalry," *Perception* **4**:315–322 (1975).
251. Y. Yang and R. Blake, "Spatial Frequency Tuning of Human Stereopsis," *Vision Research* **31**:1177–1189 (1991).
252. J. S. Mansfield and A. J. Parker, "An Orientation-Tuned Component in the Contrast Masking of Stereopsis," *Vision Research* **33**:1535–1544 (1993).
253. R. D. Freeman and I. Ohzawa, "On the Neurophysiological Organization of Binocular Vision," *Vision Research* **30**:1661–1676 (1990).
254. D. L. Halpern and R. R. Blake, "How Contrast Affects Stereoacuity," *Perception* **17**:483–495 (1988).
255. G. E. Legge and Y. Gu, "Stereopsis and Contrast," *Vision Research* **29**:989–1004 (1989).
256. L. K. Cormack, S. B. Stevenson, and C. M. Schor, "Interocular Correlation, Luminance Contrast, and Cyclopean Processing," *Vision Research* **31**:2195–2207 (1991).
257. C. Blakemore, "The Range and Scope of Binocular Depth Discrimination in Man," *Journal of Physiology* **211**:599–622 (London, 1970).
258. D. R. Badcock and C. M. Schor, "Depth-Increment Detection Function for Individual Spatial Channels," *Journal of the Optical Society of America A* **2**:1211–1216 (1985).
259. W. Richards, "Anomalous Stereoscopic Depth Perception," *Journal of the Optical Society of America* **61**:410–414 (1971).
260. L. K. Cormack, S. B. Stevenson, and C. M. Schor, "Disparity-Tuned Channels of the Human Visual System," *Visual Neuroscience* **10**:585–596 (1993).
261. S. B. Stevenson, L. K. Cormack, and C. M. Schor, "Disparity Tuning in Mechanisms of Human Stereopsis," *Vision Research* **32**:1685–1694 (1992).
262. S. R. Lehky and T. J. Sejnowski, "Neural Model of Stereoacuity and Depth Interpolation Based on a Distributed Representation of Stereo Disparity," *Journal of Neuroscience* **10**:2281–2299 (1990).
263. C. H. Bailey and P. Gouras, "The Retina and Phototransduction," *Principles of Neural Science*, E. R. Kandel and J. H. Schwartz (eds.), Elsevier Science Publishing Co., Inc., New York, 1985, pp. 344–355.
264. K. D. Davila and W. S. Geisler, "The Relative Contributions of Pre-Neural and Neural Factors to Areal Summation in the Fovea," *Vision Research* **31**:1369–1380 (1991).
265. J. Ross and H. D. Speed, "Contrast Adaptation and Contrast Masking in Human Vision," *Proceedings of the Royal Society of London*, series B **246**:61–69 (1991).

*This page intentionally left blank*

---

# PSYCHOPHYSICAL METHODS

---

## Denis G. Pelli

*Psychology Department and Center for Neural Science  
New York University  
New York*

## Bart Farell

*Institute for Sensory Research  
Syracuse University  
Syracuse, New York*

---

### 3.1 INTRODUCTION

---

Psychophysical methods are the tools for measuring perception and performance. These tools are used to reveal basic perceptual processes, to assess observer performance, and to specify the required characteristics of a display. We are going to ignore this field's long and interesting history,<sup>1</sup> and much theory as well.<sup>2,3</sup> Here we present a formal treatment, emphasizing the theoretical concepts of psychophysical measurement. For practical advice in setting up an experiment, please turn to our user's guide.<sup>4</sup> Use the supplied references for further reading.

Consider the psychophysical evaluation of the suitability of a visual display for a particular purpose. A home television to be used for entertainment is most reasonably assessed in a "beauty contest" of subjective preference,<sup>5</sup> whereas a medical imaging display must lead to accurate diagnoses<sup>6,7</sup> and military aerial reconnaissance must lead to accurate vehicle identifications.<sup>8</sup> In our experience, the first step toward defining a psychophysically answerable question is to formulate the problem as a task that the observer must perform. One can then assess the contribution of various display parameters toward that performance. Where precise parametric assessment is desired it is often useful to substitute a simple laboratory task for the complex real-life activity, provided one can either demonstrate, or at least reasonably argue, that the laboratory results are predictive.

Psychophysical measurement is usually understood to mean measurement of behavior to reveal internal processes. The experimenter is typically not interested in the behavior itself, such as pressing a button, which merely communicates a decision by the observer about the stimulus.\* This chapter reviews the various decision tasks that may be used to measure perception and performance and evaluates their strengths and weaknesses. We begin with definitions and a brief review of visual stimuli. We then explain and evaluate the various psychophysical tasks, and end with some practical tips.

---

\*Psychophysical measurement can also be understood to include noncommunicative physiological responses such as pupil size, eye position, electrical potentials measured on the scalp and face, and even BOLD fMRI responses in the brain, which might be called "unintended" responses. (These examples are merely suggestive, not definitive. Observers can decide to move their eyes and, with feedback, can learn to control many other physiological responses. Responses are "unintended" only when they are not used for overt communication by the observer.) Whether these unintended responses are called psychophysical or physiological is a matter of taste. In any case, decisions are usually easier to measure and interpret, but unintended responses may be preferred in certain cases, as when assessing noncommunicative infants and animals.

## 3.2 DEFINITIONS

At the highest level, an *experiment* answers a question about how certain “experimental conditions” affect observer performance. *Experimental conditions* include stimulus parameters, observer instruction, and anything else that may affect the observer’s state. Experiments are usually made up of many individual measurements, called “trials,” under each experimental condition. Each trial presents a stimulus and collects a response—a decision—from the observer.

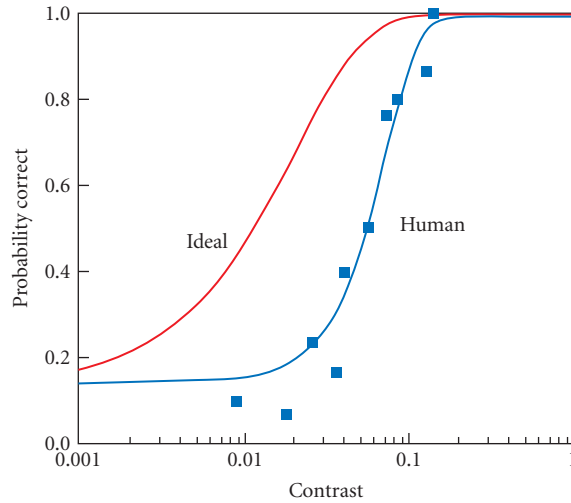
There are two kinds of decision tasks: judgments and adjustments. It is useful to think of one as the inverse of the other. In one case the experimenter gives the observer a stimulus and asks for a classification of the stimulus or percept; in the other case the experimenter, in effect, gives the observer a classification and asks for an appropriate stimulus back. Either the experimenter controls the stimulus and the observer makes a *judgment* based on the resulting percept, or the observer *adjusts* the stimulus to satisfy a perceptual criterion specified by the experimenter (e.g., match a sample). Both techniques are powerful. Adjustments are intrinsically subjective (because they depend on the observers’ understanding of the perceptual criterion), yet they can often provide good data quickly and are to be preferred when applicable. But not all questions can be formulated as adjustment tasks. Besides being more generally applicable, judgments are often easier to analyze, because the stimulus is under the experimenter’s control and the task may be objectively defined. Observers typically like doing adjustments and find judgments tedious, partly because judgment experiments usually take much longer.

An obvious advantage of adjustment experiments is that they measure physical stimulus parameters, which may span an enormous dynamic range and typically have a straightforward physical interpretation. Judgment tasks measure human performance (e.g., frequency of seeing) as a function of experimental parameters (e.g., contrast). This is appropriate if the problem at hand concerns human performance per se. For other purposes, however, raw measures of judgment performance typically have a very limited useful range, and a scale that is hard to interpret. Having noted that adjustment and judgment tasks may be thought of as inverses of one another, we hasten to add that in practice they are often used in similar ways. Judgment experiments often vary a stimulus parameter on successive trials in order to find the value that yields a criterion judgment. These “sequential estimation methods” are discussed in Sec. 3.6. The functional inversion offered by sequential estimation allows judgment experiments to measure a physical parameter as a function of experimental condition, like adjustment tasks, while retaining the judgment task’s more rigorous control and interpretation.

Distinguishing between judgment and adjustment tasks emphasizes the kind of response that the observer makes. It is also possible to subdivide tasks in a way that emphasizes the stimuli and the question posed. In a *detection* task there may be any number of alternative stimuli, but one is a blank, and the observer is asked only to distinguish between the blank and the other stimuli. Slightly more general, a *discrimination* task may also have any number of alternative stimuli, but one of the stimuli, which need not be blank, is designated as the reference, and the observer is asked only to distinguish between the reference and other stimuli. A decision that distinguishes among more than two categories is usually called an *identification* or *classification*.<sup>9</sup> All decision tasks allow for alternative responses, but two alternatives is an important special case.<sup>10</sup>

As normally used, the choice of term, *detection* or *discrimination*, says more about the experimenter’s way of thinking than it does about the actual task faced by the observer. This is because theoretical treatments of detection and discrimination usually allow for manipulation of the experimental condition by introduction of an extraneous element, often called a “mask” or “pedestal,” that is added to every stimulus. Thus, one is always free to consider a discrimination task as detection in the presence of a mask. This shift in perspective can yield new insights (e.g., Refs. 11–15). Since there is no fundamental difference between detection and discrimination,<sup>16</sup> we have simplified the presentation below by letting detection stand in for both. The reader may freely substitute “reference” for “blank” (or suppose the presence of an extraneous mask) in order to consider the discrimination paradigm.

The idea of “threshold” plays a large role in psychophysics. Originally deterministic, *threshold* once referred to the stimulus intensity above which the stimulus was always distinguishable from blank, and below which it was indistinguishable from blank. In a discrimination task one might refer to a “discrimination threshold” or a “just-noticeable difference.” Nowadays the idea is statistical; we know that the observer’s probability of correct classification rises as a continuous function of stimulus



**FIGURE 1** Probability of correctly identifying a letter in noise, as a function of letter contrast. The letters are bandpass filtered. Gaussian noise was added independently to each pixel. Each symbol represents the proportion correct in 30 trials. The solid curve through the points is a maximum likelihood fit of a Weibull function. The other curve represents a similar maximum likelihood fit to the performance of a computer program that implements the ideal letter classifier.<sup>40</sup> Efficiency, the squared ratio of threshold contrasts, is 9 percent. (Courtesy of Joshua A. Solomon.)

intensity (see Fig. 1). Threshold is defined as the stimulus intensity (e.g., contrast) corresponding to an arbitrary level of performance (e.g., 82 percent correct). However, the old intuition, now called a “high threshold,” still retains a strong hold on everyone’s thinking for the good reason that the transition from invisible to visible, though continuous, is quite abrupt, less than a factor of two in contrast.

Most psychophysical research has concentrated on measuring thresholds. This has been motivated by a desire to isolate low-level sensory mechanisms by using operationally defined tasks that are intended to minimize the roles of perception and cognition. This program is generally regarded as successful—visual detection is well understood (e.g., Ref. 17)—but leaves most of our visual experience and ability unexplained. This has stimulated a great deal of experimentation with suprathreshold stimuli and nondetection tasks in recent years.

### 3.3 VISUAL STIMULI

Before presenting the tasks, which are general to all sense modalities (not just vision), it may be helpful to briefly review the most commonly used visual stimuli. Until the 1960s most vision research used a spot as the visual stimulus (e.g., Ref. 18). Then cathode ray tube displays made it easy to generate more complex stimuli, especially sinusoidal gratings, which provided the first evidence for multiple “spatial frequency channels” in vision.<sup>19</sup> Sinusoidal grating patches have two virtues. A sinusoid at the display always produces a sinusoidal image on the retina.\* And most visual mechanisms are selective in space and in spatial frequency, so it is useful to have a stimulus that is restricted in both domains.

\*This is strictly true only within an isoplanatic patch, i.e., a retinal area over which the eye’s optical point spread function is unchanged.

Snellen,<sup>20</sup> in describing his classic eye chart, noted the virtue of letters as visual stimuli—they offer a large number of stimulus alternatives that are readily identifiable.<sup>21,22</sup> Other commonly used stimuli include annuli, lines, arrays of such elements, and actual photographs of faces, nature, and military vehicles. There are several useful texts on image quality, emphasizing signal-to-noise ratio.<sup>23–26</sup> Finally, there has been some psychophysical investigation of practical tasks such as reading,<sup>27</sup> flying an airplane,<sup>28</sup> or shopping in a supermarket.<sup>29</sup>

The stimulus alternatives used in vision experiments are usually parametric variations along a single dimension, most commonly contrast, but frequently size and position in the visual field. *Contrast* is a dimensionless ratio: the amplitude of the luminance variation within the stimulus, normalized by the background luminance. Michelson contrast (used for gratings) is the maximum minus the minimum luminance divided by the maximum plus the minimum. Weber contrast (used for spots and letters) is the maximum deviation from the uniform background divided by the background luminance. RMS contrast is the root-mean-square deviation of the stimulus luminance from the mean luminance, divided by the mean luminance.

## 3.4 ADJUSTMENTS

Adjustment tasks require that the experimenter specify a perceptual criterion to the observer, who adjusts the stimulus to satisfy the criterion. Doubts about the observer's interpretation of the criterion may confound interpretation of the results. The adjustment technique is only as useful as the criterion is clear.

### Threshold

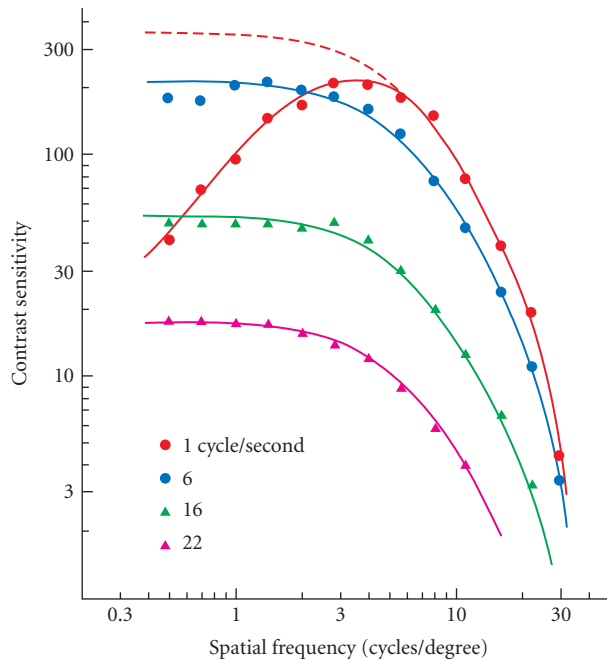
Figure 2 shows contrast sensitivity (the reciprocal of the threshold contrast) for a sinusoidal grating as a function of spatial and temporal frequency.<sup>30</sup> These thresholds were measured by what is probably the most common form of the adjustment task, which asks the observer to adjust the stimulus contrast up and down to the point where it is “just barely detectable.” While some important studies have collected their data in this way, one should bear in mind that this is a vaguely specified criterion. What should the observer understand by “barely” detectable? Seen half the time? In order to adjust to threshold, the observer must form a subjective interpretation and apply it to the changing percept. It is well known that observers can be induced (e.g., by coaching) to raise or lower their criterion, and when comparing among different observers it is important to bear in mind that social and personality factors may lead to systematically different interpretations of the same vague instructions. Nevertheless, these subjective effects are relatively small (about a factor of two in contrast) and many questions can usefully be addressed, in at least a preliminary way, by quick method-of-adjustment threshold settings. Alternatively, one might ignore the mean of the settings and instead use the standard deviation to estimate the observer's discrimination threshold.<sup>31</sup>

### Nulling

Of all the many kinds of adjustments, nulling is the most powerful. Typically, there is a simple basic stimulus that is distorted by some experimental manipulation, and the observer is given control over the stimulus and asked to adjust it so as to cancel the distortion (e.g., Ref. 32). The absence of a specific kind of distortion is usually unambiguous and easy for the observer to understand, and the observer's null setting is typically very reliable.

### Matching

Two stimuli are presented, and the observer is asked to adjust one to match the other. Sometimes the experiment can be designed so that the observer can achieve a perfect match in which the stimuli are utterly indistinguishable, which Brindley<sup>33</sup> calls a “Class A” match. Usually, however, the stimuli



**FIGURE 2** Spatial contrast sensitivity (reciprocal of threshold contrast) functions for sinusoidal gratings temporally modulated (flickered) at several temporal frequencies. The points are the means of four method-of-adjustment measurements and the curves (one with a dashed low-frequency section) differ only in their positions along the contrast-sensitivity scale. (From Robson.<sup>30</sup>)

are obviously different and the observer is asked to match only a particular aspect of the stimuli, which is called a “Class B” match. For example, the observer might be shown two grating patches, one fine and one coarse, and asked to adjust the contrast of one to match the contrast of the other.<sup>34</sup> Or the observer might see two uniform patches of different colors and be asked to match their brightnesses.<sup>35</sup> Observers (and reviewers for publication) are usually comfortable with matching tasks, but, as Brindley points out, it is amazing that observers can seemingly abstract and compare a particular parameter of the multidimensional stimuli in order to make a Class B match. Matching tasks are extremely useful, but conclusions based on Class B matches may be less secure than those based on Class A matches because our understanding of how the observer does the task is less certain.

## Magnitude Production

The observer is asked to adjust a stimulus to match a numerically specified perceptual criterion, e.g., “as bright as a 60-watt light bulb.” The number may have a scale (watts in this case) or be a pure number.<sup>2</sup> The use of pure numbers, without any scale, to specify a perceptual criterion is obviously formally ambiguous, but in practice many experimenters report that observers seem comfortable with such instructions and produce stable results that are even reasonably consistent among different observers. Magnitude production, however, is rarely used in visual psychophysics research.



## 3.5 JUDGMENTS

Judgment tasks ask the observer to classify the stimulus or percept. They differ primarily in the number of alternative stimuli that may be presented on a given trial and the number of alternative responses that the observer is allowed.

### The Ideal Observer

When the observer is asked to classify the stimulus (not the percept) it may be useful to consider the mathematically defined ideal classifier that would yield the most accurate performance using only the information (the stimuli and their probabilities) available to the observer.<sup>36–42</sup> Obviously this would be an empty exercise unless there is some known factor that makes the stimuli hard to distinguish. Usually this will be visual noise: random variations in the stimulus, random statistics of photon absorptions in the observer's eyes, or random variations in neural processes in the observer's visual system. If the stimuli plus noise can be defined statistically at some site—at the display, as an image at the observer's retinae, as a pattern of photon absorptions, or as a spatiotemporal pattern of neural activity—then one can solve the problem mathematically and compute the highest attainable level of performance. This ideal often provides a useful point of comparison in thinking about the actual human observer's results. A popular way of expressing such a comparison is to compute the human observer's efficiency, which will be a number between 0 and 1. For example, in Fig. 1 at threshold the observer's efficiency for letter identification is 9 percent. As a general rule, the exercise of working out the ideal and computing the human observer's efficiency is usually instructive but, obviously, low human efficiencies should be interpreted as a negative result, suggesting that the ideal is not particularly relevant to understanding how the human observer does the task.

### Yes-No

The best-known judgment task is yes-no. It is usually used for detection, although it is occasionally used for discrimination. The observer is either asked to classify the stimulus, "Was a nonblank stimulus present?" or classify the percept, "Did you see it?" The observer is allowed only two response alternatives: yes or no. There may be any number of alternative stimuli. If the results are to be compared with those of an ideal observer, then the kind of stimulus, blank or nonblank, must be unpredictable.

As with the method-of-adjustment thresholds discussed above, the question posed in a yes-no experiment is fundamentally ambiguous. Where is the dividing line between yes and no on the continuum of internal states between the typical percepts generated by the blank and nonblank stimuli? Theoretical considerations and available evidence suggest that observers act as if they reduced the percept to a "decision variable," a pure magnitude—a number if you like—and compared that magnitude with an internal criterion that is under their conscious control.<sup>40,43</sup> Normally we are not interested in the criterion, yet it is troublesome to remove its influence on the results, especially since the criterion may vary between experimental conditions and observers. For this reason, most investigators no longer use yes-no tasks.

As discussed next, this pesky problem of the observer's subjective criterion can be dealt with explicitly, by using "rating scale" tasks, or banished, by using unbiased "two-alternative forced choice" (2afc) tasks. Rating scale is much more work, and unless the ratings themselves are of interest, the end result of using either rating scale or 2afc is essentially the same.

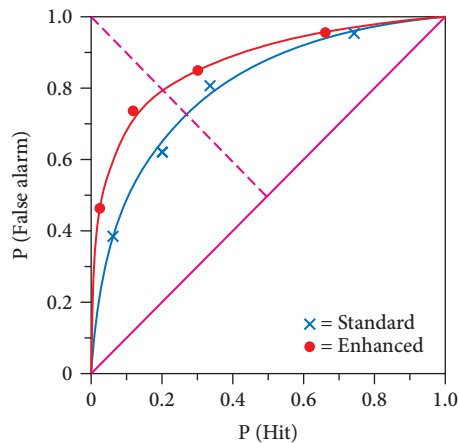
### Rating Scale

In a rating scale task the observer is asked to rate the likelihood that a nonblank stimulus was presented. There must be blank and nonblank stimulus alternatives, and there may be any number of alternative ratings—five is popular—but even a continuous scale may be allowed.<sup>44</sup> The endpoints of

the rating scale are “The stimulus was definitely blank” and “The stimulus was definitely nonblank,” with intermediate degrees of confidence in between. The results are graphed as a receiver operating characteristic, or ROC, that plots one conditional probability against another. The observer’s ratings are transformed into yes-no judgments by comparing them with an external criterion. Ratings above the criterion become “yes” and those below the criterion become “no.” This transformation is repeated for all possible values of the external criterion. Finally, the experimenter plots—for each value of the criterion—the probability of a yes when a nonblank stimulus was present (a “hit”) against the probability of a yes when a blank stimulus was present (a “false alarm”). In medical contexts the hit rate is called “sensitivity” and one minus the false alarm rate is called “specificity.” Figure 3 shows an ROC curve for a medical diagnosis;<sup>45</sup> radiologists examined mammograms and rated the likelihood that a lesion was benign or malignant.

In real-life applications the main value of ROC curves is that they can be used to optimize yes-no decisions based on ratings, e.g., whether to refer a patient for further diagnosis or treatment. However, this requires knowledge of the prior stimulus probabilities (e.g., in Fig. 3, the incidence of disease in the patient population), the benefit of a hit, and the cost of a false alarm.<sup>6,7</sup> These conditions are rarely met. One usually can estimate prior probability and assess the cost of the wasted effort caused by the false alarms, but it is hard to assign a commensurate value to the hits, which may save lives through timely treatment.

The shape of the ROC curve has received a great deal of attention in the theoretical detection literature, and there are various mathematical models of the observer’s detection process that can account for the shape.<sup>46–48</sup> However, unless the actual situation demands rating-based decisions, the ROC shape has little or no practical significance, and the general practice is to summarize the ROC curve by the area under the curve. The area is 0.5 when the observers’ ratings are independent of the stimuli (i.e., useless guessing). The area can be at most 1—when the observer makes no mistakes. The



**FIGURE 3** Example of an empirical ROC. Six radiologists attempted to distinguish between malignant and benign lesions in a set of 118 mammograms, 58 malignant and 60 benign, first when the mammograms were viewed in the usual manner (“standard”), and then—“enhanced”—when they were viewed with two aids, including a checklist of diagnostic features. The ratings were “very likely malignant,” “probably malignant,” “possibly malignant,” “probably benign,” and “very likely benign.” The areas under the curves are 0.81 and 0.87. (From Swets.<sup>45</sup>)

area can descend below 0.5 when the observer reverses the categories of blank and nonblank. We'll see in a moment that a result equivalent to ROC area can usually be obtained with much less effort by doing a two-alternative forced choice experiment instead.

## Two-Alternative Forced Choice

This task is traditionally characterized by two separate stimulus presentations, one blank and one nonblank, in random order. The two stimuli may be presented successively or side by side. The observer is asked whether the nonblank stimulus was first or second (or on the left or right). We noted above that in yes-no tasks observers seem to reduce the stimulus to a decision variable, the magnitude upon which they base their decisions. The 2afc task is said to be “unbiased” because the observer presumably chooses the presentation that generated the higher magnitude, without referring to any subjective internal criterion. At the beginning of this section we said that all judgment tasks consist of the presentation of a stimulus followed by a judgment. In this view, we might consider the two presentations in the 2afc task to be a single stimulus. The two possible composite stimuli to be discriminated are reflections of one another, either in space or time. The symmetry of the two alternatives suggests that the observer's choice between them may be unbiased.

Other related tasks are often called “two-alternative forced choice” and are similarly claimed to be unbiased. There is some confusion in the literature over which tasks should be called “2afc.” In our view, the “2afc” label is of little consequence. What matters is whether the task is unbiased, i.e., are the alternative stimuli symmetric for the observer? Thus a yes-no discrimination of blank and nonblank stimuli may be biased even though there are two response alternatives and the choice is forced, whereas it may be reasonable to say that the judgment of the orientation of a grating that is either horizontal or vertical is unbiased even though there is only a single presentation. We suggest that authors wishing to claim that their task is unbiased say so explicitly and state why. This claim might be based on a priori considerations of the symmetry between the stimuli to be discriminated, or on a post hoc analysis of relative frequencies of the observer's responses.

In theory, if we accept the assumptions that each stimulus presentation produces in the observer a unidimensional magnitude (one number, the decision variable), that the observer's ratings and 2afc decisions are based, in the proper way, on this magnitude, and that these magnitudes are stochastically independent between presentations, then the probability of a correct response on a 2afc trial must equal the area under the ROC curve.<sup>47</sup> Nachmias<sup>43</sup> compared 2afc proportion correct and ROC area empirically, finding that ROC area is slightly smaller, which might be explained by stimulus-induced variations in the observer's rating criteria.

---

## MAGNITUDE ESTIMATION

In the inverse of magnitude production, a stimulus is presented and the observer is asked to rate it numerically.<sup>2</sup> Some practitioners provide a reference (e.g., a stimulus that rates 100), and some don't, allowing observers to use their own scale. Magnitude estimation and rating scale are fundamentally the same. Magnitude estimation experiments typically test many different stimulus intensities a few times to plot mean magnitude versus intensity, and rating-scale experiments typically test few intensities many times to plot an ROC curve at each intensity.

## Response Time

In practical situations the time taken by the observer to produce a judgment usually matters, and it will be worthwhile recording it during the course of the experiment. Some psychophysical research has emphasized response time as a primary measure of performance in an effort to reveal mental processes.<sup>49</sup>

## 3.6 STIMULUS SEQUENCING

So far we have discussed a single trial yielding a single response from the observer. Most judgments are stochastic, so judgment experiments usually require many trials. An uninterrupted sequence of trials is called a *run* (or a *block*). There are two useful methods of sequencing trials within a run.

### Method of Constant Stimuli

Experimenters have to worry about small, hard-to-measure variations in the observer's sensitivity that might contaminate comparisons of data collected at different times. It is therefore desirable to run the trials for the various conditions as nearly simultaneously as possible. One technique is to interleave trials for the various conditions. This is the classic "method of constant stimuli." Unpredictability of the experimental condition and equal numbers of trials for each condition are typically both desirable. These are achieved by using a randomly shuffled list of all desired trials to determine the sequence.

### Sequential Estimation Methods

One can use the method of constant stimuli to measure performance as a function of a signal parameter—let us arbitrarily call it intensity—and determine, by interpolation, the threshold intensity that corresponds to a criterion level of performance.\* This approach requires hundreds of trials to produce a precise threshold estimate. Various methods have been devised that obtain precise threshold estimates in fewer trials, by using the observer's previous responses to choose the stimulus intensity for the current trial. The first methods were simple enough for the experimenter to implement manually, but as computers appeared and then became faster, the algorithms have become more and more sophisticated. Even so, the requisite computer programs are very short.

In general, there are three stages to threshold estimation. First, all methods, implicitly or explicitly, require that the experimenter provide a confidence interval around a guess as to where threshold may lie. (This bounds the search. Lacking prior knowledge, we would have an infinite range of possible intensities. Without a guess, where would we place the first trial? Without a confidence interval, where would we place the second trial?) Second, one must select a test intensity for each trial based on the experimenter's guess and the responses to previous trials. Third, one must use the collected responses to estimate threshold. At the moment, the best algorithm is called ZEST,<sup>50</sup> which is an improvement over the popular QUEST.<sup>51</sup> The principal virtues of QUEST are that it formalizes the three distinct stages, and implements the first two stages efficiently. The principal improvement in ZEST is an optimally efficient third stage.

## 3.7 CONCLUSION

This chapter has reviewed the practical considerations that should guide the choice of psychophysical methods to quickly and definitely answer practical questions related to perception and performance. Theoretical issues, such as the nature of the observer's internal decision process, have been de-emphasized. The question of how well we see is answerable only after we reduce the question to measurable performance of a specific task. The task will be either an adjustment—for a quick answer when the perceptual criterion is unambiguous—or a judgment—typically to find threshold by sequential estimation.

\*The best way to interpolate frequency-of-seeing data is to make a maximum likelihood fit by an S-shaped function.<sup>52</sup> Almost any S-shaped function will do, provided it has adjustable position and slope.<sup>53</sup>

The success of psychophysical measurements often depends on subtle details: the seemingly incidental properties of the visual display, whether the observers receive feedback about their responses, and the range of stimulus values encountered during a run. Decisions about these matters have to be taken on a case-by-case basis.

## 3.8 TIPS FROM THE PROS

---

We asked a number of colleagues for their favorite tips.

- Experiments often measure something quite different from what the experimenter intended. Talk to the observers. Be an observer yourself.
- Viewing distance is an often-neglected but powerful parameter, trivially easy to manipulate over a 100:1 range. Don't be limited by the length of your keyboard cable.
- Printed vision charts are readily available, offering objective measurement of visibility, e.g., to characterize the performance of a night-vision system.<sup>20,21,54,55</sup>
- When generating images on a cathode ray tube, avoid generating very high video frequencies (e.g., alternating black and white pixels along a horizontal raster line) and very low video frequencies (hundreds of raster lines per cycle) since they are typically at the edges of the video amplifier's passband.<sup>56</sup>
- Liquid crystal displays (LCD) have largely replaced cathode ray tube (CRT) displays in the market place. LCDs are fine for static images, but have complicated temporal properties that are hard to characterize. Thus, CRTs are still preferable for presentation of dynamic images, as they allow you to know exactly what you are getting.<sup>57</sup>
- Consider the possibility of aftereffects, whereby past stimuli (e.g., at high contrast or different luminance) might affect the visibility of the current stimulus.<sup>58-61</sup>
- Drift of sensitivity typically is greatest at the beginning of a run. Do a few warm-up trials at the beginning of each run. Give the observer a break between runs.
- Allow the observer to see the stimulus once in a while. Sequential estimation methods tend to make all trials just barely detectable, and the observer may forget what to look for. Consider throwing in a few high-contrast trials, or defining threshold at a high level of performance.
- Calibrate your display before doing the experiment, rather than afterward when it may be too late.

## 3.9 ACKNOWLEDGMENTS

---

Josh Solomon provided Fig. 1. Tips were contributed by Al Ahumada, Mary Hayhoe, Mary Kaiser, Gordon Legge, Walt Makous, Suzanne McKee, Eugenio Martinez-Uriega, Beau Watson, David Williams, and Hugh Wilson. Al Ahumada, Katey Burns, and Manoj Raghavan provided helpful comments on the manuscript. Supported by National Eye Institute grants EY04432 and EY06270.

## 3.10 REFERENCES

---

1. E. G. Boring, *Sensation and Perception in the History of Experimental Psychology*, Irvington Publishers, New York, 1942.
2. G. A. Gescheider, *Psychophysics: Methods, Theory, and Application*, 2d ed., Lawrence Erlbaum and Associates, Hillsdale, N.J., 1985, pp. 174–191.
3. N. A. Macmillan and C. D. Creelman, *New Developments in Detection Theory*, Cambridge University Press, Cambridge, U.K., 1991.

4. B. Farell, and D. G. Pelli, "Psychophysical Methods, or How to Measure a Threshold and Why," R. H. S. Carpenter and J. G. Robson (eds.), *Vision Research: A Practical Guide to Laboratory Methods*, Oxford University Press, New York, 1999.
5. P. Mertz, A. D. Fowler, and H. N. Christopher, "Quality Rating of Television Images," *Proc. IRE* **38**:1269–1283 (1950).
6. J. A. Swets, and R. M. Pickett, *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*, Academic Press, New York, 1982.
7. C. E. Metz, "ROC Methodology in Radiologic Imaging," *Invest. Radiol.* **21**:720–733 (1986).
8. F. Scott, "The Search for a Summary Measure of Image Quality—A Progress Report," *Photographic Sci. and Eng.* **12**:154–164 (1968).
9. F. G. Ashby, "Multidimensional Models of Categorization," *Multidimensional Models of Perception and Cognition*, F. G. Ashby (ed.), Lawrence Erlbaum Associates, Hillsdale, N.J., 1992.
10. D. G. Pelli, N. J. Majaj, N. Raizman, C. J. Christian, E. Kim, and M. C. Palomares, "Grouping in Object Recognition: The Role of a Gestalt Law in Letter Identification," *Cognitive Neuropsychology* (2009) In press.
11. F. W. Campbell, E. R. Howell, and J. G. Robson, "The Appearance of Gratings with and without the Fundamental Fourier Component," *J. Physiol.* **217**:17–18 (1971).
12. B. A. Wandell, "Color Measurement and Discrimination," *J. Opt. Soc. Am. A* **2**:62–71 (1985).
13. A. B. Watson, A. Ahumada, Jr., and J. E. Farrell, "The Window of Visibility: A Psychophysical Theory of Fidelity in Time-Sampled Visual Motion Displays," *NASA Technical Paper, 2211*, National Technical Information Service, Springfield, Va. 1983.
14. E. H. Adelson, and J. R. Bergen, "Spatiotemporal Energy Models for the Perception of Motion," *J. Opt. Soc. Am. A* **2**:284–299 (1985).
15. S. A. Klein, E. Casson, and T. Carney, "Vernier Acuity as Line and Dipole Detection," *Vision Res.* **30**:1703–1719 (1990).
16. B. Farell, and D. G. Pelli, "Psychophysical Methods," *A Practical Guide to Vision Research*, J. G. Robson and R. H. S. Carpenter (eds.), Oxford University Press, New York, 1999.
17. N. V. S. Graham, *Visual Pattern Analyzers*, Oxford University Press, Oxford, 1989.
18. H. B. Barlow, "Temporal and Spatial Summation in Human Vision at Different Background Intensities," *J. Physiol.* **141**:337–350 (1958).
19. F. W. Campbell, and J. G. Robson, "Application of Fourier Analysis to the Visibility of Gratings," *J. Physiol.* **197**:551–566 (1968).
20. H. Snellen, *Test-Types for the Determination of the Acuteness of Vision*, London: Norgate and Williams, 1866.
21. D. G. Pelli, J. G. Robson, and A. J. Wilkins, "The Design of a New Letter Chart for Measuring Contrast Sensitivity," *Clin. Vis. Sci.* **2**:187–199 (1988).
22. D. G. Pelli, and J. G. Robson, "Are Letters Better than Gratings?," *Clin. Vis. Sci.* **6**:409–411 (1991).
23. J. C. Dainty, and R. Shaw, *Image Science*, Academic Press, New York, 1974.
24. E. H. Linfoot, *Fourier Methods in Optical Image Evaluation*, Focal Press, New York, 1964.
25. D. E. Pearson, *Transmission and Display of Pictorial Information*, John Wiley & Sons, New York, 1975.
26. O. H. Schade, Sr., *Image Quality: A Comparison of Photographic and Television Systems*, RCA Laboratories, Princeton, N.J., 1975.
27. G. E. Legge, D. G. Pelli, G. S. Rubin, and M. M. Schleske, "Psychophysics of Reading—I. Normal Vision," *Vision Res.* **25**:239–252 (1985).
28. J. M. Rolf and K. J. Staples, *Flight Simulation*, Cambridge University Press, Cambridge, U.K., 1986.
29. D. G. Pelli, "The Visual Requirements of Mobility," *Low Vision: Principles and Application*, G. C. Woo (ed.), Springer-Verlag, New York, 1987, pp. 134–146.
30. J. G. Robson, "Spatial and Temporal Contrast-Sensitivity Functions of the Visual System," *J. Acoust. Soc. Am.* **56**:1141–1142 (1966).
31. R. S. Woodworth and H. Schlosberg, *Experimental Psychology*, Holt, Rinehart, and Winston, New York, 1963, pp. 199–200.
32. P. Cavanagh and S. Anstis, "The Contribution of Color to Motion in Normal and Color-Deficient Observers," *Vision Res.* **31**:2109–2148 (1991).

33. G. A. Brindley, *Physiology of the Retina and the Visual Pathways*, Edward Arnold Ltd., London, 1960.
34. M. A. Georgeson and G. D. Sullivan, "Contrast Constancy: Deblurring in Human Vision by Spatial Frequency Channels," *J. Physiol.* **252**:627–656 (1975).
35. R. M. Boynton, *Human Color Vision*, Holt Rinehart and Winston, New York, 1979, pp. 299–301.
36. W. W. Peterson, T. G. Birdsall, and W. C. Fox, "Theory of Signal Detectability," *Trans. IRE PGIT* **4**:171–212 (1954).
37. W. P. Tanner, Jr. and T. G. Birdsall, "Definitions of  $d'$  and  $\eta$  as Psychophysical Measures," *J. Acoust. Soc. Am.* **30**:922–928 (1958).
38. H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, Wiley, New York, 1968.
39. W. S. Geisler, "Sequential Ideal-Observer Analysis of Visual Discriminations," *Psychol. Rev.* **96**:267–314 (1989).
40. D. G. Pelli, "Uncertainty Explains Many Aspects of Visual Contrast Detection and Discrimination," *J. Opt. Soc. Am. A* **2**:1508–1532 (1985).
41. D. G. Pelli, "The Quantum Efficiency of Vision," *Vision: Coding and Efficiency*, C. Blakemore (ed.), Cambridge University Press, Cambridge, U.K., 1990, pp. 3–24.
42. D. G. Pelli, C. W. Burns, B. Farell, and D. C. Moore-Page, "Feature Detection and Letter Identification," *Vision Res.* **46**(28):4646–4674 (2006). See Appendix A.
43. J. Nachmias, "On the Psychometric Function for Contrast Detection," *Vision Res.* **21**:215–223 (1981).
44. H. E. Rockette, D. Gur and C. E. Metz, "The Use of Continuous and Discrete Confidence Judgments in Receiver Operating Characteristic Studies of Diagnostic-Imaging Techniques," *Invest. Radiol.* **27**:169–172 (1992).
45. J. A. Swets, "Measuring the Accuracy of Diagnostic Systems," *Science* **240**:1285–1293 (1988).
46. J. Nachmias and R. M. Steinman, "Brightness and Discriminability of Light Flashes," *Vision Res.* **5**:545–557 (1965).
47. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Krieger Press, Huntington, N.Y., 1974.
48. L. W. Nolte and D. Jaarsma, "More on the Detection of One of M Orthogonal Signals," *J. Acoust. Soc. Am.* **41**:497–505 (1967).
49. R. D. Luce, *Response Times: Their Role in Inferring Elementary Mental Organization*, Oxford University Press, New York, 1986.
50. P. E. King-Smith, S. S. Grigsby, A. J. Vingrys, S. C. Benes and A. Supowit, "Efficient and Unbiased Modifications of the QUEST Threshold Method: Theory, Simulations, Experimental Evaluation and Practical Implementation," *Vision Res.* **34**:885–912 (1994).
51. A. B. Watson and D. G. Pelli, "QUEST: A Bayesian Adaptive Psychometric Method," *Percept Psychophys.* **33**:113–120 (1983).
52. A. B. Watson, "Probability Summation Over Time," *Vision Res.* **19**:515–522 (1979).
53. D. G. Pelli, "On the Relation Between Summation and Facilitation," *Vision Res.* **27**:119–123 (1987).
54. S. Ishihara, *Tests for Color Blindness*, 11th ed., Kanehara Shuppan, Tokyo, 1954.
55. D. Regan and D. Neima, "Low-Contrast Letter Charts as a Test of Visual Function," *Ophthalmology* **90**:1192–1200 (1983).
56. D. G. Pelli and L. Zhang, "Accurate Control of Contrast on Microcomputer Displays," *Vision Res.* **31**:1337–1350 (1991).
57. D. H. Brainard, D. G. Pelli and T. Robson, "Display Characterization," In J. Hornak (ed.), *Encyclopedia of Imaging Science and Technology*, Wiley, 2002, pp. 172–188.
58. C. Blakemore and F. W. Campbell, "Adaptation to Spatial Stimuli," *J. Physiol.* **200**(1):11–13 (1969).
59. C. Blakemore and F. W. Campbell, "On the Existence of Neurons in the Human Visual System Selectively Sensitive to the Orientation and Size of Retinal Images," *J. Physiol.* **203**:237–260 (1969).
60. T. N. Cornsweet, *Visual Perception*, Academic Press, New York, 1970.
61. F. S. Frome, D. I. A. MacLeod, S. L. Buck and D. R. Williams, "Large Loss of Visual Sensitivity to Flashed Peripheral Targets," *Vision Res.* **21**:1323–1328 (1981).

---

# VISUAL ACUITY AND HYPERACUITY

---

Gerald Westheimer

*Division of Neurobiology  
University of California  
Berkeley, California*

---

## 4.1 GLOSSARY

---

**Airy disk.** Point-spread function in the image of a diffraction-limited optical instrument with a circular pupil.

**Diffraction limit.** Minimum dissipation of spatial information in the imaging of an optical system, due to the aperture restriction in the propagation of electromagnetic energy.

**Fovea.** Region in the center of the retina where receptor elements are most closely packed and resolution highest.

**Hyperacuity.** Performance in task where thresholds are substantially lower than the grain of the receiving layer.

**Light.** Visually evaluated radiant energy. In this chapter radiant and luminous energy terms are used interchangeably.

**Optical-transfer function.** Modulation in the transmitted images of spatial sinusoids, as a function of their spatial frequency; it is complex, that is, has amplitude and phase terms.

**Psychophysics.** Procedure for studying an observer's performance by relating the variables of physical stimuli to measurements of associated responses.

**Point-spread function.** Spatial distribution of energy in the image of a point object.

**Snellen letters.** Alphanumeric characters of defined size and shape used in standard clinical testing of visual acuity.

**Spatial frequency.** Number of cycles of a sinusoidal grating target per unit distance. Commonly cycles/degree visual angle.

**Superresolution.** Ability to garner knowledge of spatial details in an optical image based on previous available information, either by extrapolation or averaging.

**Vernier acuity.** Performance limit in the alignment of two abutting line segment; it is the prime example of hyperacuity.

**Visual acuity.** Performance limit in distinguishing spatial details in visual object.

**Visual angle.** Angle subtended by an object at the center of the eye's entrance pupil; it is a measure of distance in the retinal image.



**Equation (1).** Point-spread function of a purely diffraction-limited optical imaging system with a round pupil.

$\theta$  angular subtense of radius of Airy's disk

$\lambda$  wavelength of radiation

$a$  diameter of aperture

**Equation (2).** Specification of contrast in the spatial distribution.

$L_{\max}, L_{\min}$  Luminance of maximum, minimum, respectively

## 4.2 INTRODUCTION

---

Visual acuity—literally sharpness—refers to the limit of the ability to discriminate spatial partitioning in the eye's object space. As a psychophysical measure, its analysis encompasses

- The physics—in this case optics—of the stimulus situation
- The anatomical and physiological apparatus within the organism that processes the external stimulus
- The operations leading to the generation of a response

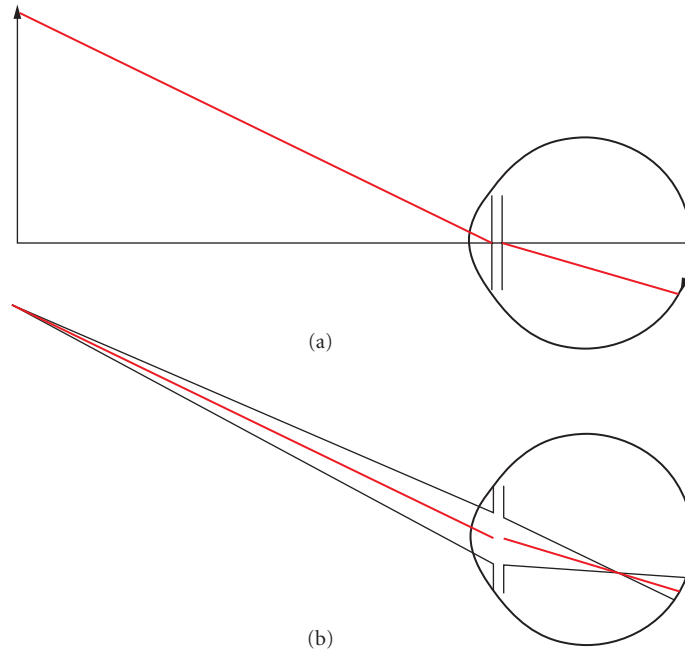
Measurement of visual acuity involves the organism as a whole, even though it is possible to identify the performance limits of only a segment of the full operation, for example, the eyeball as purely an optical instrument, or the grain of the receptor layer of the retina, or neural activity in the visual cortex. But the term acuity is reserved for the behavioral essay and therefore necessarily includes the function of all components of the arc reaching from physical object space to some indicator of the response of the whole organism. It is a psychophysical operation. The fact that it includes a component that usually involves an observer's awareness does not preclude it from being studied with any desired degree of rigor.

## 4.3 STIMULUS SPECIFICATIONS

---

Specification of the stimulus is an indispensable preliminary. For this purpose, an Euclidean object space containing visual targets is best defined by a coordinate system with its origin at the center of the eye's entrance pupil and its three axes coinciding with those of the eye. The observer is usually placed to make the vertical ( $y$ ) axis that of gravity, and the horizontal ( $x$ ) axis orthogonal and passing through equivalent points in the two eyes; distances from the observer are measured along the  $z$  axis. When an optical device is associated with the eye, its principal axes should accord, either by positioning the device or the observer. On occasions when both eyes of an observer are involved, the origin of the coordinate system is located at the midpoint of the line joining the two eyes. More details of possible coordinate systems have been described elsewhere.<sup>1</sup>

The ocular structure most relevant to the spatial dissection of an observer's object world is the retina, and the inquiry begins with the quest for the most instructive way of relating the *distal* and *proximal* stimuli, that is, the actual objects and their retinal images. Here it is achieved by using the center of the entrance pupil of the eye as the origin of the coordinate system which has at least two advantages. First, while associated with the eye's imagery, it belongs to object space and hence is objectively determinable by noninvasive means. In a typical human eye, the entrance pupil is located about 3 mm behind the corneal vertex and is not far from round. Its center is therefore an operationally definable point. The second reason for choosing it as a reference point involves the nature of the eye's image-forming apparatus. As seen in Fig. 1, the bundle of rays from a point source converging toward the retina is centered on the ray emerging from the center of the eye's exit pupil, which is the optical conjugate of the center of the entrance pupil. Regardless of the position



**FIGURE 1** Schematic diagram of imaging in human eye. (a) The retinal image size of a target is represented by the angle it subtends at the eye's entrance pupil. (b) The position of the image of a point is most effectively demarcated by the intercept of the image-sided chief ray, which is the center of the light distribution even if the eye is out of focus.

of the geometrical image with respect to the retina, that is, regardless of the state of defocus, the center of the retinal image patch, blurred or sharp, will be defined by the intersection of that ray, called *chief ray*, with the retina. When two object points are presented to the eye, the retinal distance corresponding to their separation is given by the intercept of the chief rays from these objects. In this manner, the three-dimensional object space of the eye has been collapsed into the two-dimensional one of the retinal surface. What has been lost, and needs to be specified separately, is the object's distance from the eye along the chief ray. But all objects, in or out of focus, anywhere along a given chief ray share a single retinal location, or, to rephrase it, the coordinates on the retinal surface are homologous to angular coordinates within the object-sided sheaf of rays converging on the center of the eye's entrance pupil.

Hence in the specification of retinal distances it suffices to identify corresponding angles in the eye's object space, objectively determinable measures. Units of measurement are the radian, or, degrees or minutes of arc. At a distance of 57 cm, a 1-cm object subtends 1 deg, a 0.16-mm object 1 arcmin. At the standard eye-chart distance of 6 m (20 ft) the limb of a 20/20 letter is just under 2 mm wide.

The next specification to be considered is that of the luminous intensity impinging on the eye. One starts with the most elemental stimulus: the luminous intensity of a point source is given in the internationally agreed-on unit of candela ( $\text{lumens} \cdot \text{steradian}^{-1}$ ). This again is object-sided and objectively determinable. Extended sources are measured in terms of luminance ( $\text{lumens} \cdot \text{steradian}^{-1} \cdot \text{unit area}^{-1}$ , in practice  $\text{cd} \cdot \text{m}^{-2}$ ). Hence the specification of visual acuity targets in the eye's object space requires, apart from their observation distance, their spatial extent in angular measure at the eye's entrance pupil and the luminance of the background from which or against which they are formed. The luminous energy reaching the retina differs in that it depends on the pupil area, the absorption in the eye's

media, and retinal factors considered elsewhere. Attenuation in the passage through the ocular media will differ from one eye to another and is prominently age dependent (Chap. 1). (Polarization and coherence properties of the incoming luminous energy are generally not relevant, but see special conditions analyzed in Chap. 14). How this luminous energy is distributed on the retinal surface depends on the transfer characteristics of the eye's optics which will now be considered.

## 4.4 OPTICS OF THE EYE'S RESOLVING CAPACITY

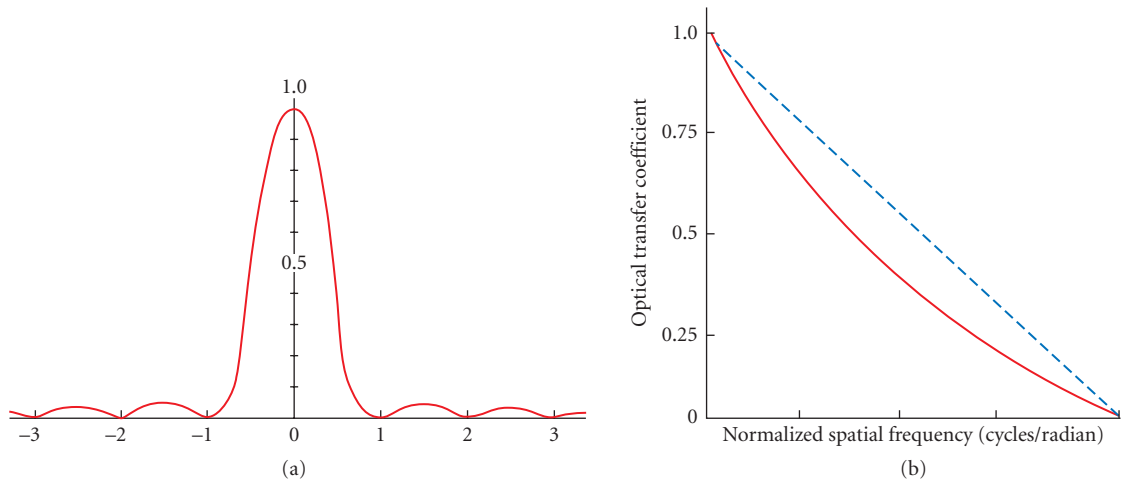
The spatial distribution of energy reaching the retina will differ from that incident on the cornea by being subject to spread produced by the eye's imaging.

### Light Spread in the Retinal Image and Limits of Resolution

The two modes of proceeding in this discussion, via the point-spread or the contrast-transfer functions, are equivalent (Fig. 2). So long as one remains in the realm of optics and does not enter that of neural and psychophysical processing where linearity is not guaranteed, it is permissible to transfer back and forth between the two.

**Point-Spread Function** When the object space is restricted to a single point, the spatial distribution of energy in the image is called the point-spread function and describes the spread introduced by passage through the eye's optics. Even in ideal focus, the spread depends on the eye's aperture and the wavelength of the electromagnetic energy; the object-sided distribution then has  $\theta$ , that is, angle subtended by the distance from its center to the first zero, given by

$$\theta = \frac{1.22\lambda}{a} \quad (1)$$



**FIGURE 2** Retinal light distribution in an idealized optical system like the eye's. (a) Diffraction-limited point-spread function (Airy's disk). Unit distance is given by  $1.22 \lambda/a$  in radians, where  $\lambda$  is the wavelength of light and  $a$  the diameter of the round pupil, both in the same units of length. (b) The optical contrast-transfer function for a square pupil (blue dashed line) and a round pupil (red solid line). It descends to zero at a spatial frequency equal to  $a/\lambda$  cycles/radian.

A point can never be imaged smaller than a patch of such a size; the light distribution for any object is the convolution of that of the points constituting it. In practice there are additional factors due to the aberrations and scattering in each particular eye.

***Spatial-Frequency Coordinates and Contrast-Transfer Function*** A fundamental property of optical imagery in its application to the eye is its linearity. Hence a permissible description of light distributions is in terms of their spatial Fourier spectra, that is, the amplitudes and phases of those spatial sinusoidal intensity distribution in terms of their spatial frequency that, when superimposed, will exactly reconstruct the original distribution. The limit here is the cutoff spatial frequency at which the optical transfer coefficient of the eye reaches zero (Fig. 2*b*).

Either of the two descriptors of the optical transfer between the eye's object and image spaces, the point-spread function, that is, light spread in the image of a point, and the contrast-transfer function, that is, the change in amplitude and phase that the component sinusoids (as a function of spatial frequency in two angular dimensions) experience as they are transferred from the eye's object space to the retinal image, is complete and the transposition between the two descriptors in uncomplicated.

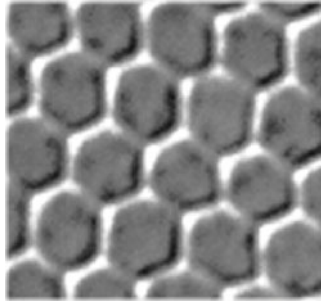
Because resolution relates to the finest detail that can be captured in an image, the interest is in the narrowness of the point-spread function (e.g., width at half-height) or, equivalently, the high-frequency end of the contrast-transfer function. The absolute limit imposed by diffraction gives a bound to these two functions but this is actually achieved only in fully corrected eyes with pupil diameters below about 3 mm, when other dioptric deficits are minimal. Concentrating on light of wavelength 555 nm, for which the visual system during daylight is most sensitive, Fig. 2, taken over directly from diffraction theory, illustrates the best possible performance that may be expected from a normal human eye under ordinary circumstances. The cutoff spatial frequency then is near 90 cycles/degree<sup>-1</sup> and the diameter of Airy's disk about 1.5 arcmin.

A great deal of effort has gone into the determination of the actual point-spread functions of eyes which include factors such as aberrations. Only under very exceptional circumstances—and nowadays also with the aid of adaptive optics—does one get better imagery than what is shown in Fig. 2. When the pupil diameter is increased, theoretical improvement due to narrowing of the Airy disk is counteracted by aberrations which become more prominent as the outer zones of the pupil are uncovered. The effect of refractive errors on imaging can also be described in theory, but then phase changes in the contrast-transfer function enter because its complex nature (in the mathematical sense) can no longer be ignored (see under "Defocus").

## 4.5 RETINAL LIMITATIONS—RECEPTOR MOSAIC AND TILING OF NEURONAL RECEPTIVE FIELDS

Also amenable to physical analysis are the limitations imposed on the resolving capacity of the eye by the structure of the retina. All information that is handed on to the neural stages of vision is in the first place partitioned by the elements of the receptor layer, each of which has an indivisible spatial signature. The spacing of the receptors is not uniform across the retina, nor is the individuality of their local sign necessarily retained in further processing. Ultimately, the information transfer to the brain is confined by the number of individual nerve fibers emerging from the retina. Nevertheless it is instructive to inquire into the grain of the retina in the foveal region where there is certainly at least one optic nerve fiber for each receptor.

Figure 3 is a cross section of the layer of a primate retina and shows an approximately hexagonal array of cones, whose average spacing in the center is of the order of 0.6 arcmin. No matter what else is in play, human visual resolution cannot be better than is allowed by this structure. The center of the fovea is only about 0.5 deg in diameter; the further one proceeds into the retinal periphery the coarser the mosaic and the lower the ratio of receptors to optic nerve fibers. This explains the reason for our highly developed oculomotor system with its quick ability to transfer foveal gaze to different eccentric and even moving targets.



**FIGURE 3** Histological cross section of the retinal mosaic in the primate fovea. Each receptor represents an object-sided angle of about 0.6 arcmin.

The neural elements of the retina are not passive transducers but actively rearrange the optical signals that reach the receptors. After transmission to the brain, processing of these neural signals involves interaction from other regions and modification as a result of such factors as attention and memory. As yet the neural circuitry interposed between the optical image on the retina and the individual's acuity response has not reached a level of understanding equivalent to that of the optical and receptor stages.

## 4.6 DETERMINATION OF VISUAL RESOLUTION THRESHOLDS

Awareness of the limitations imposed by the optics and anatomy of the eye is, of course, of value, but visual acuity is, in the end, a function of the operation of the whole organism: what are the finest spatial differences that can be distinguished? In answering this question, attention has to be paid to the manner of obtaining the measurements.

Since there is scatter in individual determinations, the number of trials will be controlled by the needed precision. The armamentarium of psychophysical procedures allows determination of a threshold with arbitrary precision. The standard optometric visual acuity chart, in use for 150 years, is a textbook case of effective employment of this approach. The ensemble of test symbols, the information content per symbol, the accepted answers, and the scaling of the steps and number of trials (letters in each row) have all been optimized for quick and reliable acuity identification.

A psychophysical threshold is a number along a scale of a variable (e.g., distance between double stars) at which a correct response is made in a predetermined proportion of trials. In a good experiment it is accompanied by a standard error. Thus if in a particular situation the two-star resolution threshold is found to be  $0.92 \pm 0.09$  arcmin and the 50-percent criterion was employed, this means that on 50 percent of occasions when the separation was 0.92 inch the observer would say "yes" (the percentage increasing with increasing separation) and that the scatter of the data and the number of observations were such that if the whole experiment were repeated many times, the value would be expected to be between 0.83 inch and 1.01 inch in 19 out of 20 runs of data.

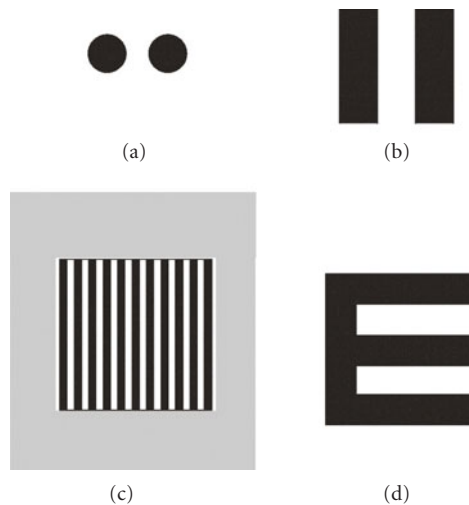
The distinction is often made between detection and discrimination thresholds. In sophisticated detection procedures, stimulus presentations are alternated randomly with blanks. A count is kept of the number of times the observer gives a "yes" answer when there was no stimulus, the so-called false positives. An elaborate analytical procedure can then be deployed to examine the internal

“noise” against which the incoming sensory signal has to compete.<sup>2</sup> This methodology is appropriate when there is a blank as one of the alternatives in the test, for example, in the measurement of the high spatial-frequency cutoff for grating resolution. In the bulk of acuity determinations the observer has to discriminate between at least two alternative configurations each presented well above detection threshold, and what has to be safeguarded against are bias errors.

It is the current practice in vision research to observe as many of these niceties of psychophysical methodology as possible (Chap. 3). However, in clinical and screening situations less time and observer engagement is available, with, as a consequence, diminished reliability and repeatability of findings.

## 4.7 KINDS OF VISUAL ACUITY TESTS

The common denominator of acuity tests is the determination of the finest detectable spatial partitioning, and this can be done in many ways. The closest to the optical concept of resolving power are the two-point or two-line patterns, whose minimum separation is measured at which they are seen as double (Fig. 4a). More popular are the two-bar experiments, most effectively implemented in a matrix of  $3 \times 3$  elements, in which either the top and bottom rows, or the right and left columns



**FIGURE 4** Patterns used in visual acuity tests and the associated response criteria: (a) Two point resolution. (b) Koenig bars in a  $3 \times 3$  matrix. (c) Grating resolution. (d) Letters, as used in the clinical Snellen chart. Observers respond to these questions: You will be shown either a single or a double star. Was it “one” or “two?” (a). You will be shown a two-line pattern. Were the lines vertical or horizontal? (b). You will be shown a field that is either blank or contains vertical stripes? Was it “blank” or “striped?” (c). You will be shown an alphanumeric character. What letter or digit was it? (d).

have a different contrast than the middle row or column (Fig. 4*b*), the observer's responses being limited to "horizontal" and "vertical." The size of the matrix elements is increased to determine the observer's threshold. Overall size is no clue and the response is based on the detection of the internal image structure.

Some brightness variables are available to the experimenter. In the old days, the lines would be black on a white background whose luminance would be specified. With the advent of oscilloscopic displays this can now be white on black or, more generally, brighter and darker than a uniform background. Contrast is then a further variable, usually defined by the Michelson formula

$$\frac{(L_{\max} - L_{\min})}{(L_{\max} + L_{\min})} \quad (2)$$

With the advent of the Fourier approach to optics, grating targets have become popular. For the purposes of acuity, the highest spatial-frequency grating is determined at which, with 100-percent modulation, the field is just seen as striped rather than uniform (Fig. 4*c*). The phenomenon of spurious resolution, described below, makes this test inadvisable when focus errors may be at play. The role of grating targets in acuity measurements differs from that in modulation sensitivity tests, where gratings with a range of spatial periods are demodulated till they are no longer detectable. This process yields the modulation sensitivity curve, discussed below.

Since they were first introduced in the second half of the 19th century, the standard for clinical visual acuity are the Snellen letters—alphanumeric characters, each drawn within a  $5 \times 5$  matrix, with limb thickness as the parameter (Fig. 4*d*). From the beginning it was accepted that the resolution limit of the human eye is 1 arcmin, and hence the overall size of the Snellen letter for normal acuity is 5 arcmin, or 9.5 mm at a distance of 6 m or 20 ft (optical infinity for practical purposes). When such letters can be read at 20 ft, visual acuity is said to be 20/20. Letters twice this size can normally be read at 40 ft; an observer who can only read such double-sized letters at 20 ft has 20/40 acuity. The charts are usually assembled in lines of about 8 letters for 20/20 and progressively fewer for the lower ratings, with just a single letter for 20/200. Because in acuity determinations the error is proportional to the size of the letters<sup>3</sup> and the sequence of letter sizes in charts is usually logarithmic,<sup>4</sup> Snellen acuity is often converted to a fraction, 20/20 becoming 1.0, 20/40 becoming 0.5, and so on. When some of the letters in a line are missed, say 2 in the 20/20 line, a score of  $20/20 - 2$  is recorded. For example, if there are 7 letters in the 20/25 (0.8) line of which 3 are missed, and the next line is 20/30 (0.67) a numerical value of  $0.74 [0.8 - (3/7)(0.8 - 0.67)]$  can be entered for statistical purposes.

Snellen charts have been made available in many alphabets, but letters may not have equal legibility, even in English. Hence a stripped-down version of letter acuity is often used. A single letter E can be shown in four or even eight orientations and the observer asked to respond, perhaps by pointing a hand with outstretched fingers. The detection of the location of a gap in an annulus, that is, the distinction between the letter O and an oriented C, called the Landolt C test after its inventor is particularly useful. Both the E and Landolt C tests can be fitted into the tradition of Snellen letters by, for the 20/20 targets, generating them with 1' line within a  $5 \times 5'$  matrix, with progressive size increases to arrive at a solid numerical acuity value. These two tests make no demands on the subjects' literacy and have the virtue that the effect of guessing is a known quantity. Even here, a minor problem arises because of the "oblique effect," a small performance deficit in oblique orientations over the horizontal and vertical.

The development of the visual system in infants and the early detection of visual anomalies has sparked the design of infant visual acuity tests, usually depending on the observation of eye movements to targets of interest whose size can be progressively diminished.<sup>5</sup> Apart from this "preferential looking" technique, optokinetic nystagmus, that is, the involuntary eye tracking of large moving fields, can be effectively utilized to measure acuity by progressively diminishing the size of details until tracking fails.

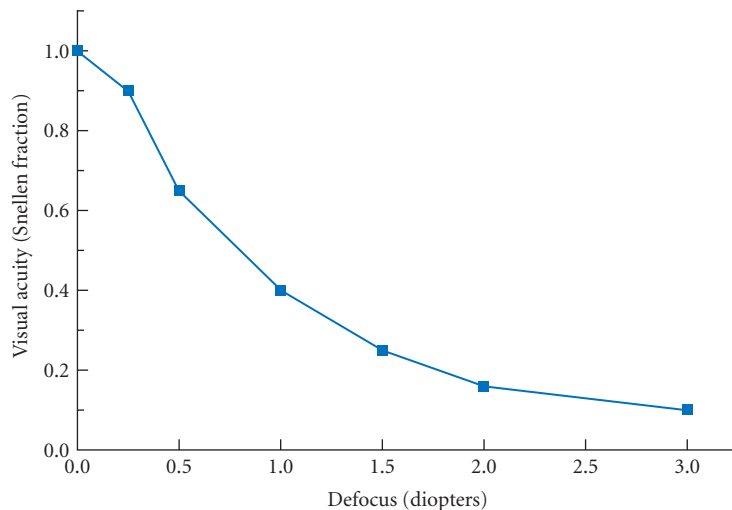
As outlined so far, all the tests have in common the need for the subject to be aware and cooperative, though not necessary literate. When these conditions are absent, other procedures have to be adopted, for example, recording the signals from the eye into the central nervous system from the scalp. These are not further described here.

## 4.8 FACTORS AFFECTING VISUAL ACUITY

Visual acuity performance will be diminished whenever any of the contributing functions have not been optimized. Initial analysis concentrates on optical and retinal factors; not enough is known about the subsequent central neural stages to differentiate all the possible ways in which their operation can be encumbered or rendered inefficient. Such factors as attention, training, and task familiarity are clearly relevant. A treatment of the subject from the clinical point of view is available in textbooks.<sup>6,7</sup>

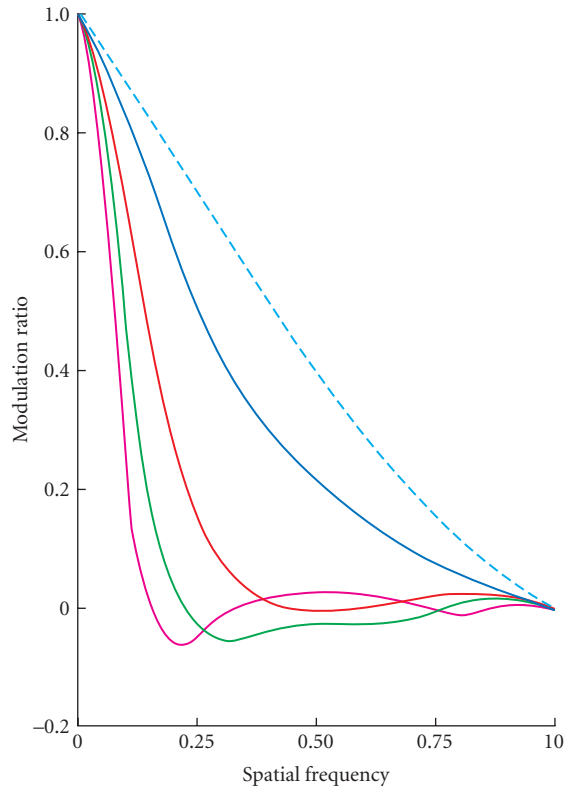
**Pupil** When the optical line-spread function has been widened for whatever reason, resolution will obviously suffer. This is the case when the pupil is too small—2 mm or less—when diffraction widens it or, in the presence of aberrated wavefronts, when it is very large—usually 6 mm or larger.

**Defocus** Focus errors are of particular interest, because one of the most ubiquitous applications of visual acuity testing is to ascertain the best refractive state of eyes for purposes of spectacle, contact lens, or surgical correction. Ever since the establishment of the current optometric routines in the late 19th century, rules of thumb have existed for the relationship between refractive error and unaided acuity. One of these is shown in Fig. 5. But this does not take into account a patient's pupil size which governs depth of focus, the possible presence of astigmatism, and higher-order aberrations, nor some complications arising from the nature of out-of-focus imagery. When a spherical wavefront entering the eye's image space does not have its center at the retina, the imagery can be described as having a phase error that increases as the square of the distance from the center of the aperture.<sup>9</sup> The contrast-transfer function, which is the Fourier transform of the complex (i.e., amplitude and phase) pupil aperture function, then does not descend to the cutoff spatial frequency monotonically, but shows oscillatory behavior (Fig. 6); in some regions in the spatial-frequency spectrum it dips below zero and the grating images have their black and white stripes reversed. This so-called spurious resolution means that if one views grating pattern under these conditions and gradually increases spatial frequency, stripes will first be visible, then disappear at the zero-crossing of the transfer function, then reappear with inverted contrast and so on. The seen image of objects like Snellen letters will undergo even more complex changes, with the possibility that "spurious" recognition is achieved with specific states of pupil size and defocus.



**FIGURE 5** Visual acuity in a typical eye as a function of uncorrected spherical refractive error in diopters. (Adapted from Laurance.<sup>8</sup>)

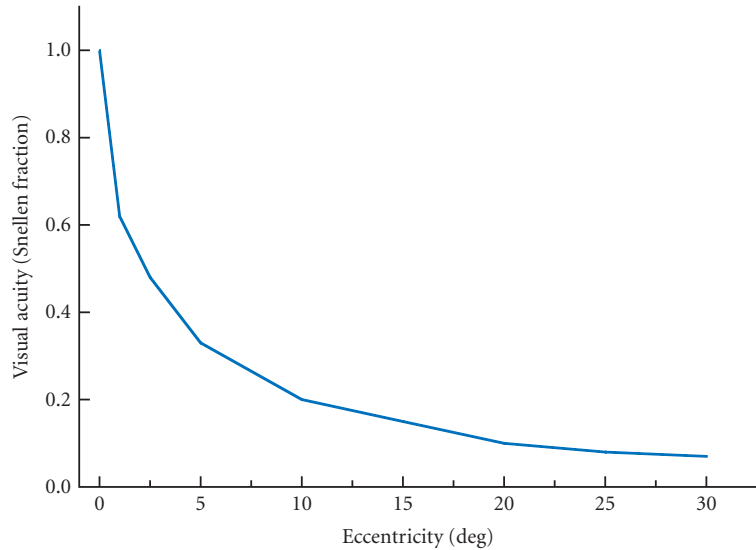




**FIGURE 6** Normalized optical-transfer function for various degrees of defocus, showing regions of the spatial-frequency spectrum in which the coefficients are negative and the contrast of grating targets is reversed in the image compared to that in the object. (Adapted from Hopkins.<sup>10</sup>) For an eye with a 3-mm round pupil and wavelength 560 nm, the cutoff spatial frequency denoted by the normalized value 1.0 on the axis of abscissas is 1.53 cycles/arcmin and the five curves show the theoretical response for 0, 0.23, 0.31, 0.62, and 1 diopters defocus.

**Color** The diffraction equations have wavelength as an explicit variable. The width of the point-spread function varies inversely with wavelength. This is, however, only a minor factor where the effect of color on visual acuity is concerned. More immediately involved is the eye's chromatic aberration, giving defocus of about 1 diopter at the extremes of the visual spectrum, where also more energy is needed to compensate for the fact that the luminous efficiency of the eye peaks as 555 nm for the photopic and 500 nm for the scotopic (rod) system. In practice, therefore, each situation will have to be handled individually depending on the wavelength distribution of the particular stimulus.

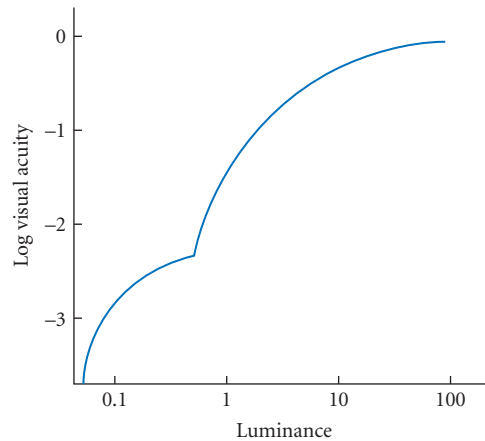
**Retinal Eccentricity** Due to the coarsening of the grain of the anatomical connections in the peripheral retina, visual acuity falls off with increasing eccentricity (Fig. 7) and this is more pronounced in the extreme nasal than temporal field of view of each eye. In binocular vision, when



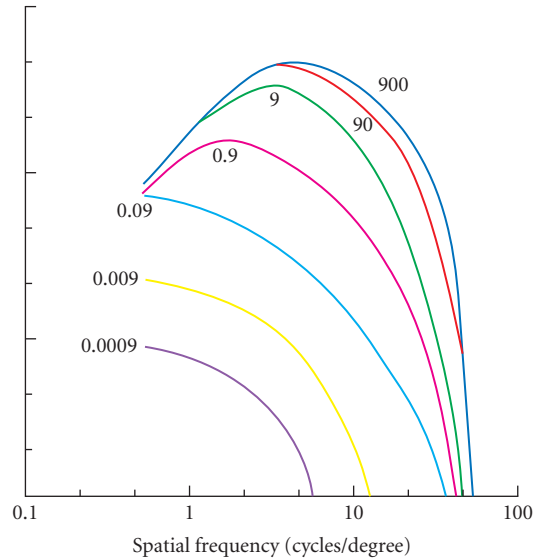
**FIGURE 7** Expected visual acuity in various locations in the peripheral visual field. (Adapted from Wertheim.<sup>11</sup>)

both eyes are in play, acuity is usually better than in monocular vision, not only because the better eye covers any possible deficiency of the other, but also because of probability summation due to the two retinas acting as independent detectors.

**Luminance** Acuity measured for black symbols against a white background, that is, for 100 percent contrast (Fig. 8) remains constant from about  $10 \text{ cd} \cdot \text{m}^{-2}$  up. Below about  $1 \text{ cd} \cdot \text{m}^{-2}$  the photopic system drops out and rods, which are absent in the fovea, take over. Their luminosity curve peaks at



**FIGURE 8** Visual acuity for targets as a function of their luminance. (Adapted from Shlaer.<sup>12</sup>) The rod-cone break occurs at a light level equivalent to that of a scene lit by full-moon light.



**FIGURE 9** Modulation sensitivity (contrast detection) curve of the human visual apparatus as a function of spatial frequency at different light levels. (From van Nes and Bouman.<sup>13</sup>) Visual acuity is equivalent to the highest spatial frequency at which a response can still be obtained (intersection with the  $x$  axis) and the data here map well on those in Fig. 8.

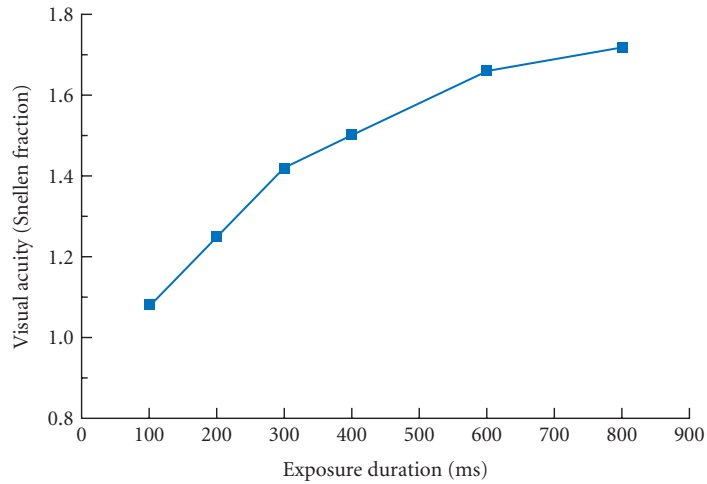
about 500 nm. Also, they are color blind, and subject to considerable spatial summation and adaptation, that is, become more sensitive with increased time in the dark, up to as much as 30 to 45 min.

**Contrast** Most clearly shown with the use of grating stimuli, there is reduction in performance at both the low and high spatial frequency ends of the spectrum (Fig. 9). Because contrast sensitivity is lowered in various ocular abnormalities, particularly scatter or absorption in the media, it has been found valuable to perform acuity measurements with low-contrast charts.<sup>14,15</sup> Reversing the contrast polarity, that is, presenting bright letters against a dark background, has virtue for eyes with a wide point-spread function and light scatter<sup>16</sup> and indeed improves acuity performance in some older eyes.<sup>17</sup>

**Time** Time is decidedly a factor in visual acuity. For very short presentations, about 20 ms or less, the eye integrates all the light flux, what matters then is the product of the intensity and the duration, in any combination. But acuity improves with duration in the several hundred millisecond range where light detection no longer depends on duration but only intensity (Fig. 10).

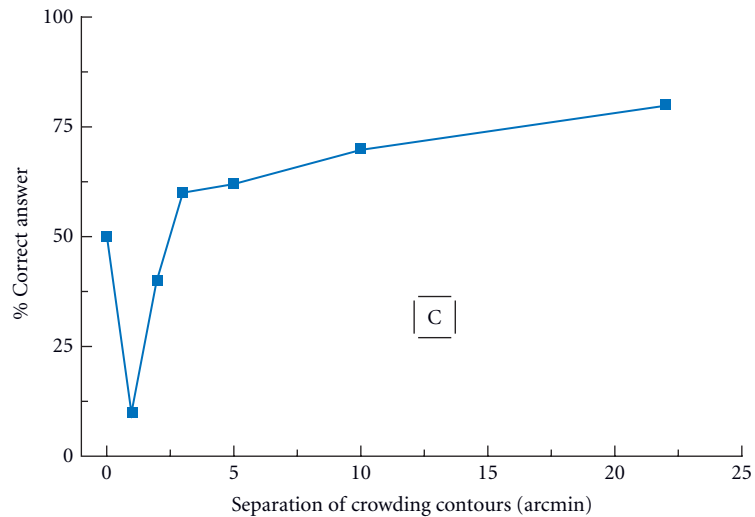
**Surround** A prominent effect in visual acuity is that of crowding, where the presence of any contour close to the resolution target interferes with performance (Fig. 11).

**Practice Effects** Surprisingly, in the fovea of normal observers, practice in the task does not confer any further advantage—it seems that optimum performance has been entrained through continuous exercise of the facility in everyday situation. Peripheral acuity can, however, be improved by training.<sup>20,21</sup>



**FIGURE 10** Visual acuity as function of exposure duration. (From Baron and Westheimer.<sup>18</sup>)

**Stage of Development and Aging** Visual acuity shows a steep increase in the first few months of life and, if a secure measurement can be obtained, is not far from normal at least by the third year.<sup>5</sup> The aging eye is subject to a large number of conditions that impair acuity (Chap. 14); it is their presence or absence that determines any individual patient's status. Consequently age decrements of acuity can range from none to severe.

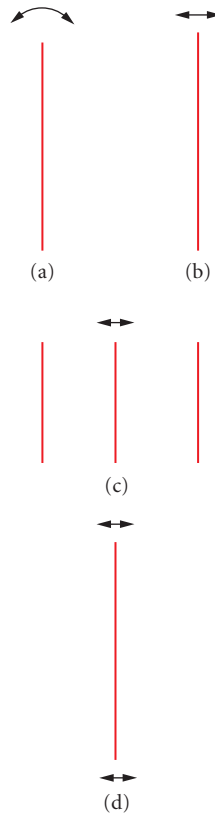


**FIGURE 11** Data showing the “crowding” effect in visual acuity. A standard letter is surrounded by bars on all four sides and the performance drops when the bars are separated from the edge of the letter by the thickness of the letter's line-width. (Adapted from Flom, Weymouth, and Kahneman.<sup>19</sup>)

## 4.9 HYPERACUITY

For a long time human spatial discriminations have been known where thresholds are markedly lower than the resolution limit. For vernier acuity, the foveal alignment threshold for two abutting lines is just a few arcsecs, as compared with the 1 arcmin or so of ordinary resolution limit.

Such high precision of localization is shared by many kinds of pattern elements, both in the direction joining them, in their alignment, and in deviations from rectilinearity (Fig. 12). The word hyperacuity is applied to this discrimination of relative position, in recognition that it surpasses by at least an order of magnitude the traditional acuity. Whereas the limitation of the latter are mainly in the resolving capacity of the eye's optics and retinal mosaic, hyperacuity depends on the neural visual system's ability to extract subtle differences within the spatial patterns of the optical image on the retina.



**FIGURE 12** Configuration in which location differences can be detected with a precision higher than the resolution limit by up to an order of magnitude. These hyperacuity tasks do not contradict any laws of optics and are the result of sophisticated neural circuits identifying the centroids of light distributions and comparing their location. Thresholds are just a few arcsec in the foveal discrimination of these patterns from their null standards: (a) orientation deviation from the vertical of a short line; (b) alignment or vernier acuity; (c) bisection of a spatial interval; and (d) deviation from straightness of a short line.

Localization discriminations in the hyperacuity range are performed by identification of the centroid of the retinal light distributions<sup>22</sup> of the involved pattern components, for example, abutting lines in the vernier task. The information is available in the image and can be described equally well in the domains of light distribution and its Fourier spectrum. It is testimony of sophisticated neural processing to arrive at the desired decision. Resolution and localization acuity share many attributes, though usually not in the same numerical measure.

Like ordinary visual acuity, hyperacuity is susceptible to crowding and to the oblique effect, but the two classes of spatial discrimination do not share all the other attributes.<sup>23</sup> Specifically, hyperacuity is more robust to reduction in exposure duration and diminishes more steeply with retinal eccentricity. It therefore follows that the neural processing apparatus is different and involves subtly recognition of differences in the excitation state of a neural population. In this respect it is similar to a whole host of fine discriminations, for example, those in the visual domains of color and stereoscopic depth.

## 4.10 RESOLUTION, SUPERRESOLUTION, AND INFORMATION THEORY

The threshold difference between ordinary visual acuity and hyperacuity raises the question whether any fundamental physical principles are being disobeyed.<sup>24</sup>

### Resolution and Superresolution

In the first instance, all of vision must satisfy the laws of physical optics according to which no knowledge can be acquired about an object that is contained in the region beyond the cutoff spatial frequency decreed by diffraction theory. Here the concepts associated with the term *superresolution* (Chaps. 3 and 4 in Vol. I) are relevant; they have been expanded since first formulated as involving an extrapolation: If the predominant features in which the spatial-frequency spectrum of two objects differ are located beyond the cutoff, but are always accompanied by a characteristic signature within it, then in principle it is possible to make the distinction between the two objects from detailed study of the transmitted spectrum and extrapolate from that to arrive at a correct decision. More recently the word has also been used to describe a procedure by which several samples of noisy transmitted spatial-frequency spectra from what is known to be the same object are superimposed and averaged. In both uses of the word, no diffraction theory limit has been broached; rather, knowledge is secured from detailed analyses of observed energy distributions on the basis of prior information—in the first case that the difference in the spatial-frequency spectrum inside the cutoff limit is always associated with those beyond it, in the second that the several spectra that are averaged arose from the same target.

### Resolution and Information

Information theory has been of help in understanding some potentially knotty problems in defining resolution.<sup>25</sup> Traditionally, the resolving power of an ideal optical instrument is regarded to have been reached when two stars are separated by the width of the Airy disk, when the images of the two stars partially overlap and the dip between the two peaks is about 19 percent. This so-called Rayleigh criterion usually allows the receptive apparatus to signal the two peaks and the trough as separable features—provided its spatial grain is fine enough and the intensity difference detectable. It is then possible to decide, without prior knowledge, that there are two stars or that a spectral line is double. But when there is prior knowledge that the source can only be either single or double, and if double, what their relative intensity is, then the decision can be made for smaller separations, because the light distribution resulting from the overlapping images of two adjoining sources differs in predictable fashion from that of a single source, even when the dip between the peaks is less than at the

Rayleigh limit and even when there is no dip at all. This specific example highlights how information theory, as first formulated by Shannon, enters the discussion: in quantifying the transmitted information, prior knowledge needs to be factored in.

The diffraction image of a point is an expression of the uncertainty principle: it describes the probability distribution of the location of the source of an absorbed photon. The larger the number of absorbed photons, the more secure the knowledge of the source's location *provided always that it is known that it has remained the same source as the photons are being accumulated*. With sufficient number of absorbed photons, and the assurance that they arose from the same source, it is possible to assign location with arbitrary precision. Hence, from the standpoint of optics, there is nothing mysterious about the precision of locating visual target that is now called hyperacuity. Rather it draws attention to the physiological and perceptual apparatus that enables this precision to be attained.

---

## 4.11 SUMMARY

The human visual system's capacity to discriminate spatial details is governed by the eye's optical imagery, by the grain of the retinal receiving layer, by the physiological apparatus of the retina and the central nervous system, and by methodological considerations of recording observers' responses. Overall, under the best conditions of ordinary viewing, the resolution limit is close to that governed by the diffraction theory for the involved optical parameters. Performance is impaired whenever any of the optical, anatomical, physiological, or perceptual components does not operate optimally. There is an additional class of visual threshold based not on resolution of spatial detail but on locating relative object position. Because thresholds then are at least one order of magnitude better, they are called hyperacuity. They do not contravene any laws of optics but are testimony to sophisticated neural processing that can identify the location of the centroid of retinal light distributions with high precision.

---

## 4.12 REFERENCES

1. G. Westheimer, "The Visual System and Its Stimuli," *The Senses. A Comprehensive Reference*, R. H. Masland, (ed.), Academic Press, Oxford, 2008.
2. N. A. Macmillan and C. D. Creelman, *Detection Theory*, 2nd ed., Erlbaum, New York, 2005.
3. G. Westheimer, "Scaling of Visual Acuity Measurements," *Arch. Ophthalmol.* **97**(2):327–330 (1979).
4. I. L. Bailey and J. E. Lovie, "New Design Principles for Visual Acuity Letter Charts," *Am. J. Optom. Physiol. Opt.* **53**:740–745 (1976).
5. D. Y. Teller, "First Glances: The Vision of Infants. The Friedenwald Lecture," *Invest. Ophthalmol. Vis. Sci.* **38**:2183–2203 (1997).
6. I. Borish, *Clinical Refraction*, Saunders, Philadelphia, 1998.
7. G. Westheimer, "Visual Acuity," *Adler's Physiology of the Eye*, 10th ed., P. L. Kaufman and A. Alm, (eds.), Mosby, St. Louis, 2003, pp. 453–469.
8. L. Laurence, *Visual Optics and Sight Testing*, 3rd ed., School of Optics, London, 1926.
9. G. Westheimer, "Optical Properties of Vertebrate Eyes," *Handbook of Sensory Physiology*, M. G. F. Fuortes, (ed.), Springer-Verlag, Berlin, 1972, pp. 449–482.
10. H. H. Hopkins, "The Application of Frequency Response Techniques in Optics," *Proc. Physical Soc.* **79**:889–918 (1962).
11. T. Wertheim, "Ueber die indirekte Sehschärfe," *Z. Psychol.* **7**:172–189 (1894).
12. S. Shlaer, "The Relation between Visual Acuity and Illumination," *J. gen. Physiol.* **21**:167–188 (1937).
13. F. L. van Nes and M. A. Bouman, "Spatial Modulation Transfer in the Human Eye," *J. Opt. Soc. Am.* **57**:401–406 (1967).

14. D. G. Pelli, J. G. Robson, and A. J. Wilkins, "The Design of a New Letter Chart for Measuring Contrast Sensitivity," *Clinical Vision Sciences* **2**:187–199 (1988).
15. D. Regan, "Low Contrast Letter Charts and Sine Wave Grating Tests in Ophthalmological and Neurological Disorders," *Clinical Vision Sciences* **2**:235–250 (1988).
16. G. Westheimer, "Visual Acuity with Reversed-Contrast Charts: I. Theoretical and Psychophysical Investigations," *Optom. Vis. Sci.* **80**:745–748 (2003).
17. G. Westheimer, P. Chu, W. Huang, T. Tran, and R. Dister, "Visual Acuity with Reversed-Contrast Charts: II. Clinical Investigation," *Optom. Vis. Sci.* **80**:749–750 (2003).
18. W. S. Baron and G. Westheimer, "Visual Acuity as a Function of Exposure Duration," *J. Opt. Soc. Am.* **63**(2):212–219 (1973).
19. M. C. Flom, F. W. Weymouth, and D. Kahneman, "Visual Resolution and Contour Interaction," *J. Opt. Soc. Am.* **53**:1026–1032 (1963).
20. B. L. Beard, D. M. Levi, and L. N. Reich, "Perceptual Learning in Parafoveal Vision," *Vision Res.* **35**:1679–1690 (1995).
21. G. Westheimer, "Is Peripheral Visual Acuity Susceptible to Perceptual Learning in the Adult?," *Vision Res.* **41**(1):47–52 (2001).
22. G. Westheimer and S. P. McKee, "Integration Regions for Visual Hyperacuity," *Vision Res.* **17**(1):89–93 (1977).
23. G. Westheimer, "Hyperacuity," *Encyclopedia of Neuroscience*, L. A. Squire, (ed.), Academic Press, Oxford, 2008.
24. G. Westheimer, "Visual Acuity and Hyperacuity: Resolution, Localization, Form," *Am. J. Optom. Physiol. Opt.* **64**(8):567–574 (1987).
25. G. Toraldo di Francia, "Resolving Power and Information," *J. Opt. Soc. Am.* **45**:497–501 (1955).



*This page intentionally left blank*

---

# OPTICAL GENERATION OF THE VISUAL STIMULUS

---

Stephen A. Burns

*School of Optometry  
Indiana University  
Bloomington, Indiana*

Robert H. Webb

*The Schepens Eye Research Institute  
Boston, Massachusetts*

---

## 5.1 GLOSSARY

---

$A$	area
$D$	distance
$E_r$	illuminance at the retina (retinal illuminance)
$f$	focal length
$f_e$	focal length of the eye
$L_s$	intrinsic luminance of the source
td	troland (the unit of retinal illuminance)
$x$	position
$\Delta$	change in position
$\Phi$	flux (general)
$\Phi_p$	flux at the pupil
$\tau$	transmittance

We have also consistently used subscripted and/or subscripted versions of  $A$  and  $S$  for area,  $D$  for distance,  $f$  for focal lengths,  $x$  for positions, and  $\Delta$  for changes in position.

---

## 5.2 INTRODUCTION

---

This chapter presents basic techniques for generating, controlling, and calibrating the spatial and temporal pattern of light on the retina (the visual stimulus). It deals with the optics of stimulus generation and the control of light sources used in the vision laboratory. Generation of stimuli by computer video displays is covered in detail in Chap. 22. Units for measuring radiation are discussed in Chap. 34, “Radiometry and Photometry,” by Edward Zalewski and Chap. 37, “Radiometry and Photometry for Vision Optics,” by Yoshi Ohno in Vol. II.

### 5.3 THE SIZE OF THE VISUAL STIMULUS

The size of a visual stimulus can be specified either in terms of the angle which the stimulus subtends at the pupil, or in terms of the physical size of the image of the stimulus formed at the retina, for most purposes vision scientists specify stimuli in terms of the angular subtense at the pupil. If an object of size  $h$  is viewed at distance  $D$  then we express its angular extent as

$$\text{Degrees visual angle} = 2 \left( \tan^{-1} \left( \frac{h}{2D} \right) \right) \quad (1)$$

or, for angles less than about 10 deg

$$\text{Degrees visual angle} \cong \frac{360h}{2\pi D} = \frac{57.3h}{D} \quad (2)$$

In this chapter we specify stimuli in terms of the actual retinal area, as well as the angular extent. Angular extent has the advantage that it is independent of the eye, that is, it can be specified totally in terms of externally measurable parameters. However, an understanding of the physical dimensions of the retinal image is crucial for understanding the interrelation of eye size, focal length, and light intensity, all of which are part of the design of optical systems for vision research.

### 5.4 FREE OR NEWTONIAN VIEWING

There are two broad classes of optical systems that are used in vision research. Free viewing, or *newtonian* viewing, forms an image of a target on the retina with minimal accessory optics (Fig. 1a).

#### Retinal Illuminance

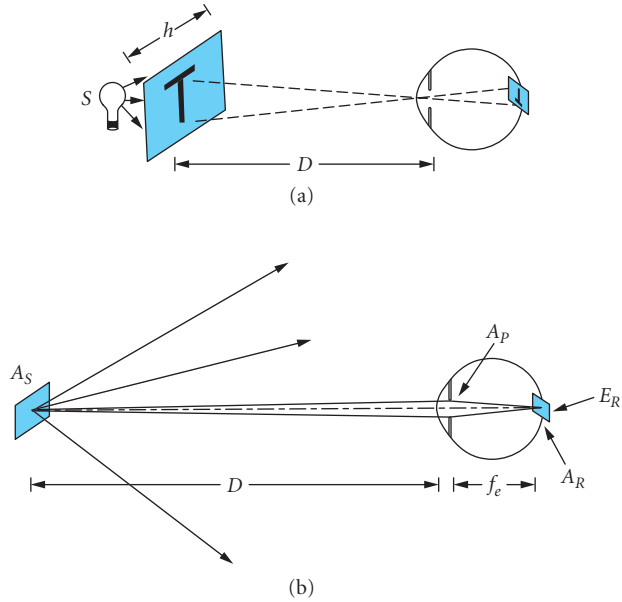
Photometric units incorporate the overall spectral sensitivity of the eye, but that is the only special property of this set of units. That is, there are no terms specific to color (for a discussion of colorimetry, see Chap. 10) and no allowances for the details of visual perception. The eye is treated as a linear detector which integrates across wavelengths. The retinal illuminance of an object in newtonian view is determined by the luminous intensity of the target and the size of the pupil of the eye. The luminous power at the pupil [dimensions are luminous power or energy per unit time, the SI units are lumens (lm)]

$$\Phi_p = L_s A_s \frac{A_p}{D^2} \quad (3)$$

where  $L_s$  is the luminance of the source [the SI units are lumens per meter squared per steradian ( $\text{lm}/\text{m}^2/\text{sr}$ ) or candelas per meter squared ( $\text{cd}/\text{m}^2$ )],  $A_s$  is the source area,  $A_p$  is the area of the pupil, and  $D$  is the distance from the pupil to the source [so  $(A_p/D^2)$  is the solid angle the pupil subtends].

The area of the image element on the retina is

$$A'_R = A_s m^2 = A_s \left( \frac{f_e}{D} \right)^2 \quad (4)$$



**FIGURE 1** (a) In free viewing or newtonian viewing the eye's optics are used to image a target onto the retina. (b) Computation of the retinal illuminance in free-viewing.  $A_s$  the area of the source,  $D$  the distance between the source and the eye,  $A_p$  the area of the pupil,  $f_e$  the optical focal length of the eye,  $A_R$  the area of the image of the source on the retina, and  $E_R$  the retinal illuminance.

where  $m$  is the magnification of the eye and  $f_e$  is the effective focal length of the eye (Fig. 1b). From this we compute the illuminance at the retina as

$$E_R = \frac{\Phi_p}{A'_R} = \frac{A_p L_s}{(f_e)^2} \quad (5)$$

(The SI units of illuminance are  $\text{lm}/\text{m}^2$ ). A typical value for  $f_e$ , is 16.67 mm.<sup>1,2</sup> Note that the retinal illuminance does not depend on the distance. That is, the retinal illuminance when viewing an extended source such as a video screen is independent of the viewing distance and is dependent only on the size of the eye's pupil, the luminance of the source, and the focal length of the eye. In most cases, the focal length of the viewer's eye is not known, but it is possible to measure the size of the pupil. For this reason a standard unit for specifying retinal illuminance was developed, the *troland*.

**The Troland (td)** The troland is a unit of illuminance (luminous power per unit area). The troland quantifies the luminous power per unit area at the retina (the retinal illuminance). One troland was defined as the illuminance at the retina when the eye observes a surface with luminance = 1  $\text{cd}/\text{m}^2$  through a pupil having an area of 1  $\text{mm}^2$ . Using Eq. (5) and a standard  $f_e$  of 16.67 mm we find that the troland is defined by

$$1 \text{ td} = 0.0035 \text{ lumens}/\text{m}^2 \quad (6)$$

This definition ties the troland to the illuminance on the retina of a standard eye assuming no transmission loss in the ocular media at 555 nm. Thus, two observers with different-size eyes, different-size pupils, or different relative losses in the ocular media, viewing the same surface, will have different retinal illuminances. Wyszecki and Stiles<sup>2</sup> have recommended that the term *troland value* be used to distinguish the trolands computed for a standard eye from the actual retinal illuminance. General usage is that the retinal illuminance is determined simply by measuring the luminance of a surface in  $\text{cd/m}^2$  and multiplying this value by the area of the pupil in  $\text{mm}^2$ .

### Limitations of Free Viewing: An Example

There are two major limitations to newtonian view systems. The first is that the retinal illuminance is limited. For instance, a 60-W frosted incandescent bulb can produce a 120,000-td field, but to obtain a uniform 20 deg field it must be placed 17 inches from the observer's eye. This requires an accommodative effort that not all observers can make. A comparable illuminance at more realistic distances, or with variable focus, requires larger light sources or more elaborate optical systems. The second limitation of free viewing is that variations in pupil size are not readily controlled. This means that the experimenter cannot specify the retinal illuminance for different stimulus conditions or for different individuals. Maxwellian view optical systems solve these problems.

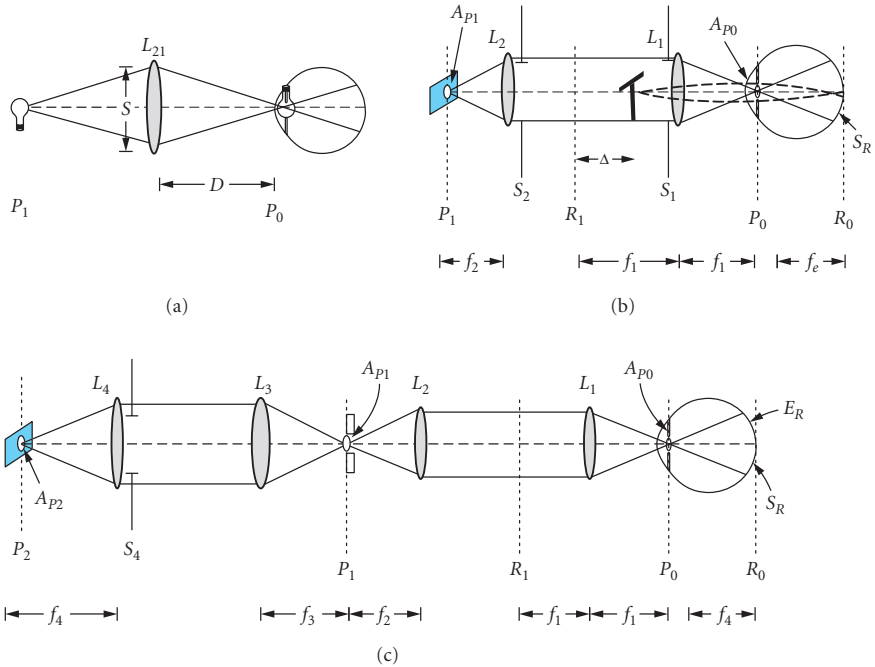
## 5.5 MAXWELLIAN VIEWING

Figure 2a shows a simple *Maxwellian* view system. The key factor that distinguishes the maxwellian view system is that the illumination source is made optically conjugate to the pupil of the eye. As a result, the target which forms the stimulus is not placed at the source, but rather at a separate plane optically conjugate to the retina. In the system of Fig. 2a the plane conjugate to the retina, where a target should be placed, lies at the focal point of lens  $L_2$ , between the source and the lens (see Chap. 6). Light from the source is diverging at this point, and slight changes in target position will cause changes in both the plane of focus and the magnification of the target. For this reason most real maxwellian view systems use multiple lenses. Figure 2b shows such a maxwellian view system where the single lens has been replaced by two lenses. The first lens collimates light from the source and the second forms an image of the source at the pupil. This places the retinal conjugate plane at the focal plane of  $L_1$ . We label the conjugate planes starting at the eye, so  $R_1$  is the first plane conjugate to the retina and  $P_1$  is the first plane conjugate to the pupil.

### Control of Focus and the Retinal Conjugate Plane

The focus of the maxwellian view system is controlled by varying the location of the target.<sup>3-5</sup> Moving the target away from  $R_1$  allows the experimenter either to adjust for ametropia or to require the observer to accommodate. To see this we compute where lens  $L_1$  (Fig. 2b) places the image of the target relative to the eye. If the target is at the focal point of  $L_1$ , the image is at infinity, which means there is a real image in focus at the retina for an emmetropic eye. If we move the target toward lens  $L_1$ , the lens of an emmetrope must shorten the focus to bring the image back into focus on the retina. We quantify this required change in focus as the change in the dioptric power of the eye (or the optical system if spectacle correction is used) where the dioptric power is simply the inverse of the focal length measured in meters. If the target is displaced by  $\Delta$  from the focal point of  $L_1$  then using the newtonian form of the lens formula ( $x'x = f^2$ ) we find that

$$\text{Change in dioptric power} = \frac{\Delta}{(f_1)^2} \quad (7)$$



**FIGURE 2** (a) In a minimal maxwellian view system a lens  $L_{21}$  images a source in the plane of the eye's pupil. The lens is shown at a distance  $D$  from the eye equal to twice its focal length. The field stop ( $S$ ) in the aperture of  $L_{21}$ . (b) A maxwellian view optical system with an accessible retinal conjugate plane ( $R_1$ ). Lens  $L_2$  collects light from the source and collimates it. Lens  $L_1$  images the source in the plane of the eye's pupil ( $P_0$ ). Pupil conjugate planes ( $P_i$ ) and retinal conjugate planes ( $R_i$ ) are labeled starting at the eye. For an emmetrope lens  $L_1$  images the retina ( $R_0$ ) at  $R_1$ . This is the plane where a visual target will appear in best focus for an emmetropic observer. Moving the target away from  $R_1$  a distance  $\Delta$  changes the plane of focus of the retinal image, requiring the emmetrope to accommodate. The maximum size of the retinal image ( $S_R$ ) is limited for this system to the aperture of lens  $L_2$  ( $S_2$ ). The pupil is imaged at position  $P_1$ , where the light source is placed. (c) Shows a more complex system where a second set of pupil conjugate and retinal conjugate planes have been added. An artificial pupil ( $A_{p1}$ ) is added at  $P_1$ . It is the image of this pupil at the source ( $A_{p2}$ ) that limits the retinal illuminance produced in this optical configuration. Other symbols as above.

If the target move toward the lens  $L_1$ , this compensating power must be positive (the focal length of the eye or eye plus spectacle shortens). If the target moves from  $f_1$  away from lens  $L_1$ , the required compensation is negative. If the eye accommodates to keep the target in focus on the retina, then the size of the retinal image is unchanged by the change in position.<sup>3,4</sup>

## Size

For this system the area of the illuminated retinal field is determined by the limiting aperture in a retinal conjugate plane between the light source and the retina (the field stop  $S_2$  in Fig. 2b) and is

$$S_R = \left( \frac{f_e}{f_1} \right)^2 S_2 \quad (8)$$

or for the schematic eye we use

$$S_R = \left( \frac{16.67}{f_1} \right)^2 S_2 \quad (9)$$

where  $f_e$  is the focal length of the eye (in mm) and  $f_1$  is the focal length of lens  $L_1$ . This formula is used to compute the physical size of an image on the retina. Note that the linear magnification is  $(f_e/f_1)$  and the areal magnification is  $(f_e/f_1)^2$ . The angular subtense of the field is the same as in Eq. (1), substituting the diameter of  $S_2$  for  $h$  and  $f_1$  for the distance  $D$ . Also note that the angular extent is independent of details of the optics of the eye. This convenience is the main reason that angular extent is the most widely used unit for specifying the stimulus in vision research.

## Retinal Illuminance

One of the *principal advantages* of a maxwellian view system is that it provides a large, uniformly bright field of view. This mode of illumination is called Kohler illumination in microscopy. The light available in a maxwellian view optical system is determined by two main factors, the luminance of the light source, and the effective pupillary aperture of the optical system being used. In Fig. 2b we see that lens  $L_2$  collects light from the source. The amount of light collected is the maximum amount of light that can be presented to the eye and is<sup>1,3</sup>

$$\Phi = A_{p1} L_s \frac{S_2}{(f_2)^2} \quad (10)$$

where  $A_{p1}$  is the area of source being considered (or, equivalently, a unit area of the source),  $L_s$  is the luminance of the source (in  $\text{lm}/\text{m}^2/\text{sr}$ ),  $S_2$  is the aperture of lens  $L_2$ , and  $f_2$  is the focal length of lens  $L_2$ . We have used the area of lens  $L_2$ , for this example, although in actual practice this area might be reduced by later field stops. This quantity will cancel out in the subsequent calculations. Finally, if all of the light collected by lens  $L_2$  is distributed across the retina, then the retinal illuminance is

$$E_R = \frac{\Phi}{S_R} = A_s L_s \left( \frac{f_1}{f_2 f_e} \right)^2 \quad (11)$$

where  $S_R$  is obtained from Eq. (8). Note that only the luminance of the source, the source area, and the focal lengths of the optical elements are important in setting the retinal illuminance. Using different diameters for lens  $L_2$  will change the amount of light collected and the size of the retinal area illuminated. Such a change will be reflected by changing the area parameter in both Eqs. (8) and (10), which cancel.

In Eq. (11) the area of the source image ( $A_{p1}$ ) at the pupil is

$$A_{p0} = A_{p1} \left( \frac{f_1}{f_2} \right)^2 \quad (12)$$

If the eye's pupil ( $A_{p0}$ ) is smaller than this source image then the eye's pupil is limiting and

$$E_R = \frac{A_{p0} L_s}{(f_e)^2} \quad (13)$$

Equation (13) for the maxwellian view system is identical to Eq. (5) which we obtained for the newtonian view system, but now the entire field is at the retinal luminance set by the source luminance. Thus, in maxwellian view a large, high retinal illuminance field can be readily obtained.

To control the size of the entry pupil rather than allowing it to fluctuate with the natural pupil, most maxwellian view systems use a pupillary stop. Figure 2c shows a system where a stop has been introduced at an intermediate pupil conjugate plane  $A_{p1}$ . This has the advantage of placing the pupillary stop of the system conjugate to the eye's pupil. The projection of  $A_{p1}$  at the source is  $A_{p2}$  and it is  $A_{p2}$  that limits the available luminous area of the source. Lenses  $L_3$  and  $L_4$  image the source onto the artificial pupil, and lenses  $L_1$  and  $L_2$  image the artificial pupil in the plane of the eye's pupil ( $A_{p0}$ ). The retinal illuminance of this more complex system can be computed as follows: the field stop over which light is collected by lens  $L_4$  is  $S_4$  and can be computed by projecting the retinal area illuminated ( $S_R$ ) back to lens  $L_4$ .

$$S_4 = S_R \left( \frac{f_1 f_3}{f_2 f_e} \right)^2 \quad (14)$$

$A_{p2}$  is the area of the source which passes through the artificial pupil:

$$A_{p2} = \left( \frac{f_4}{f_3} \right)^2 A_{p1} \quad (15)$$

Therefore the total amount of usable light collected is

$$\Phi = L_s A_{p2} \frac{S_4}{(f_4)^2} = S_R L_s A_{p1} \left( \frac{f_1}{f_2 f_e} \right)^2 \quad (16)$$

and the retinal illuminance is

$$E_R = \frac{\Phi}{S_R} = L_s A_p \left( \frac{f_1}{f_2 f_3} \right)^2 \quad (17)$$

Note that, as in Eq. (13),  $[A_{p1}(f_1/f_2)^2]$  is the size of the image of the aperture  $A_{p1}$  when measured at the exit pupil ( $A_{p0}$ ). Thus, even in this more complex case we find that the retinal illuminance is dependent only on the source luminance, the pupillary area measured at  $P_0$ , and the focal length of the eye.

## Advantages of Maxwellian Viewing: Example Revisited

The strength of a maxwellian view system is that properties of the target (focus, size, shape) can be controlled independently from retinal illuminance. If the source is larger than the limiting pupil, then the maximum retinal illuminance is the luminance of the source, scaled only by the exit pupil and eye size [Eqs. (13) and (17)]. The retinal illuminance is controlled by pupillary stops, and the target size, shape, and focus are controlled by retinal stops. Additionally, focus can be controlled independently of retinal illuminance and image size.

If we use the 60-W frosted bulb that we used in the example of newtonian viewing as a source in a maxwellian view optical system we find that we can produce the same maximum retinal illuminance for a given pupil size [Eqs. (5) and (13)]. However, with the maxwellian view system relatively inexpensive, achromat lenses will allow us to generate a field size greater than 20 deg. In addition, we can dispense with the frosted bulb (which was needed in free viewing to give a homogeneously illuminated target) and use an unfrosted tungsten halogen bulb to produce in excess of 1 million td.



## Controlling the Spatial Content of a Stimulus

The spatial frequency content of a retinal image is usually measured in cycles per degree (of visual angle). The frequencies present in the retinal image are those present in the target, cut off by the limiting aperture of the pupillary stop—either the eye's pupil or an artificial pupil in a pupillary plane *between the target and the eye*.<sup>25</sup> Apertures placed on the source side of the target do nothing to the spatial frequency content of the retinal image, leaving the pupil of the eye controlling the image, not the artificial pupil. (For a more detailed treatment of the effects of pupils on the spatial content of images see Refs. 6–10. For a discussion of the spatial resolving power of the eye's optics see Chaps. 1 and 4 in this volume of the *Handbook*).

## Positioning the Subject

One of the practical disadvantages in using maxwellian view systems is the need to position the eye's pupil at the focal point of the maxwellian lens ( $L_1$  in Fig. 2c) and to maintain its position during an experimental session. One technique for stabilizing the position of the eye is to make a wax impression of the subject's teeth (a bite bar). By attaching the bite bar to a mechanism that can be accurately positioned in three dimensions (such as a milling machine stage) the eye can be aligned to the optical system and, once aligned, the eye position is maintained by having the subject bite loosely on the bite bar. Alignment of the eye can be achieved either by observing the eye, as described at the end of this chapter, or by using the subject to guide the alignment process as described in the following paragraphs.

The first step in alignment is to position the eye at the proper distance from lens  $L_1$ . This can be accomplished by noting that, when the eye's pupil is at the plane of the source image, slight movements of the head from side to side cause the target to dim uniformly. If the eye's pupil is not at the proper distance from lens  $L_1$  then moving the head from side to side causes the target to be occluded first on one side, then the other. By systematically varying the distance of the eye from lens  $L_1$ , the proper distance of the eye can be determined. It is then necessary to precisely center the eye's pupil on the optical axis of the apparatus. There are several approaches to centering the eye.

1. The physical center of the pupil can be located by moving the translation stage such that the pupil occludes the target first on one side, then on the other. The center of these two positions is then the center (in one dimension) of the pupil. The process is then repeated for vertical translations.
2. The entry position that produces the optimum image quality of the target can be determined. One variant is to use a target that generates strong chromatic aberration (for instance, a red target on a blue background). The head can then be moved to minimize and center the chromatic fringes.<sup>11</sup> This process defines the achromatic axis of the eye.
3. The eye can be positioned to maximize the brightness of the target, which centers the Stiles-Crawford maximum in the pupil with respect to the exit pupil of the instrument.
4. A set of stimuli can be generated that enter the eye from different pupil positions. If the eye is centered, then all stimuli will be seen. If the eye is not centered, then the pupil occludes one of the pupil entry positions and part of the stimulus array disappears. In this case, the subject merely has to keep his or her head positioned such that all stimuli are visible.

## 5.6 BUILDING AN OPTICAL SYSTEM

### Alternating Source Planes and Retinal Planes in a Controlled Manner

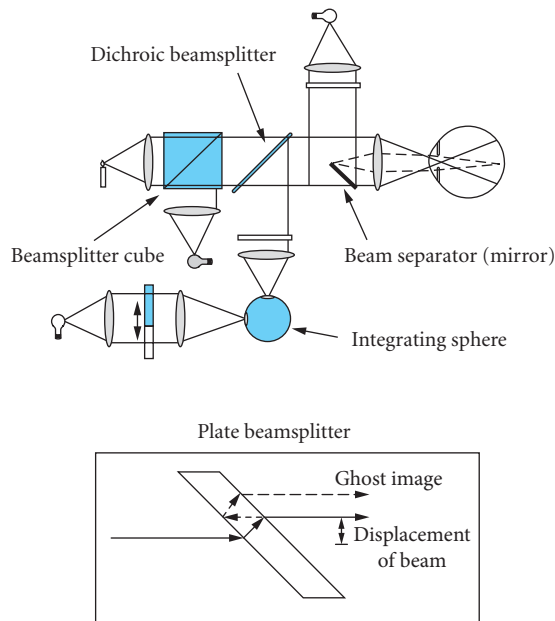
The separation of retinal and pupil conjugate planes in a maxwellian view system allows precise control over the spatial and temporal properties of the visual stimulus. By placing the light source conjugate to the pupil of the eye, every point on the source projects to every point in the retinal image

and vice versa. Thus, to control the whole retinal image, such as turning a light on and off, manipulation of a pupil conjugate plane is optimal. To control the shape of the retinal image without altering the entry pupil characteristics, variation at the retinal conjugate planes is required. However, there is an exception to this rule. Light from the edges of the image traverse the pupil conjugate plane at a higher angle than light from the center of the image. For small fields (<15-deg diameter) the angular dependence is minimal, but for larger fields it can be significant and filters should be placed in a collimated light path.

## Combining Lights in an Optical System

To control aspects of the visual stimulus independently, different light sources and targets, each with its own set of filters and shutters, can be combined. We will call these separate “channels.” Three different techniques allow for combining optical channels, beamsplitters, beam separators or reflective apertures, and diffusers. All three methods are demonstrated in Fig. 3.

**Beamsplitters** A beamsplitter both transmits and reflects a fraction of the incident light. Any surface with an index of refraction change can be a beamsplitter. In vision we generally use either a cube or a plate beamsplitter. By locating the beamsplitter in a collimated portion of the optical system (Fig. 3), but away from a retinal conjugate plane, two channels can be combined. For channels with different spectral compositions the beamsplitter can be dichroic, reflecting some wavelengths and transmitting others, usually by interference effects. Plate beamsplitters have the disadvantage

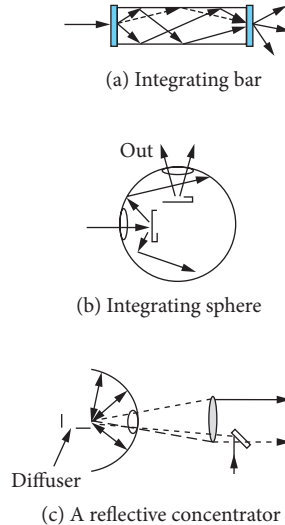


**FIGURE 3** Techniques for combining light from different light sources or optical paths. The inset illustrates potential problems that need to be considered when using a plate style of beamsplitter. These problems are eliminated by the use of a cube beamsplitter or wedge beamsplitter.

that there are secondary reflections, slightly displaced (inset, Fig. 3) from the main beam. In an image-carrying beam this produces “ghost” images, slightly displaced from the main image. The displacement decreases with decreasing thickness. A pellicle is simply a plate beamsplitter with negligible thickness. A wedge beamsplitter displaces a ghost image a long way, and Liebman has used this to design a tapered beamsplitter with no secondary reflections.<sup>12</sup> Unlike plate beamsplitters, the beamsplitter cube has the advantage that there is no lateral displacement of an image.

**Beam Separators (or Reflective Apertures)** A beam separator combines two optical channels into one, while keeping the light spatially distinct. For instance, a mirror placed halfway across the aperture at 45 deg combines two channels at 90 deg to each other. Such a separator in a retinal conjugate plane is imaged at the retina to produce a bipartite field (Fig. 3). The mirror must be oriented to hide the bevel from the subject and must have a high-quality edge. A transparent plate evenly silvered in a pattern and optically sandwiched between right angle prisms is a convenient beam separator with good edges.

**Diffusers, Integrating Spheres, and Optical Fibers** Diffusers also mix light. Diffusers are used when either the spatial uniformity of the final beam or thorough spatial mixing of light sources is important. The direction of light remitted from the diffuser is independent of the incident light, which simplifies the combination of sources. Both integrating spheres and integrating bars have been used to combine lights (Fig. 4a and b, respectively). Diffusers have been widely used in the construction of colorimeters.<sup>2</sup> In the La Jolla colorimeter<sup>13</sup> light is passed through a filter assembly with three monochromatic filters assembled edge to edge. Moving the filter assembly across the light path changes the relative proportions of the light impinging on each filter, which changes the average color of the light. An integrating sphere then completely mixes the colors producing a variable chromaticity light source without the spatial inhomogeneity of the filters. The major disadvantages of using diffusers are light loss, the need for careful optical baffling (since any light impinging on the diffuser gets mixed into the system) and the gradual deterioration of the diffuser due to dirt and dust. While the light loss from an



**FIGURE 4** An example of different types of beam combiners based on diffusers. In general it is critical that high-efficiency diffusing materials be used in all of these techniques.

integrating sphere is ideally very low, slight decreases in the efficiency of the diffusing surfaces cause large losses in the output. Collection of light from a simple diffuser can be increased by the use of a spherical mirror<sup>14</sup> (Fig. 4c). Similarly, a noncoherent fiber-optic bundle can be used to mix lights.

Mixing coherent light sources with diffusers requires movement to minimize interference (speckle) effects. Either liquid diffusers<sup>15</sup> which work by Brownian motion or moving diffusers<sup>16</sup> can be used.

## Lenses

Lens orientation affects any optical system. For instance, the most curved surface of a lens should be toward the more collimated light. This makes both sides of the lens share the work of refraction and decreases aberrations. Achromats reduce both spherical and chromatic aberrations (they have three or more refracting surfaces) and are typically designed to form a collimated beam.

Typically, the goal in aligning an optical system is to place all lenses upon a common optical axis, oriented such that a ray of light traversing that axis is not deviated. A simple prescription is

1. Define the system axis using pinholes, pointers, or a laser beam, with all lenses removed. In a multichannel optical system it is often easiest to start from the position of the eye's pupil ( $P_0$ ).
2. Introduce mirrors, beamsplitters, and other nonrefractive optics in their proper positions, being careful to keep the beam at the same height. A technique for ensuring a constant height, such as a movable pinhole, also helps.
3. Starting at the alignment source, set each lens, one at a time, such that its Boys points lie on the system axis. The Boys points are the focused reflections from the various curved surfaces of lenses (see Ref. 17 for a detailed discussion of using Boys points for alignment). When using a laser they are the "bull's eye" interference patterns. Lateral translation of the lens moves the first (brightest) Boys point and rotation moves the second, so convergence is rapid.
4. To determine the position of a lens along the optical axis, a photographic loop (a magnifying lens) and a target grid (graph paper) can be used. By placing the graph paper at a known location such as at the exit pupil, and then looking with the photographic loop into the system (turn off the laser first!) the exact position of the next pupil conjugate point can be determined. To position a lens to collimate light from a source, place a mirror in the collimated beam, reflecting the light back into the lens. When the lens is placed at one focal length from the source, the mirror causes the source to be imaged back onto itself (autocollimation).

## Field Quality

The uniformity of the illumination of the retinal image is controlled by the size and emission characteristics of the source and by the uniformity of illumination of the target. Ideally, a source emits uniformly into all angles subtended by the collection lens (lens  $L_2$  in Fig. 2). LEDs do not meet this requirement and produce inhomogeneous illumination unless the light is further diffused. Tungsten-halogen sources with coiled-coil filaments uniformly illuminate the collecting lens and can produce uniform retinal images. Problems arise if a retinal conjugate is placed too close to a pupil conjugate plane. This occurs most often when the target is placed at the end of a long collimated portion of the optical path. In this case, the structure of the source becomes visible at the retinal plane. In general, collimation lengths should be kept to less than roughly twice the focal length of the collimating lens. It is also important to control the pupillary conjugate points; in some cases multiple pupil conjugate apertures can help.<sup>18</sup>

## Controlling Wavelength

Spectral composition of a light source can be varied either by absorbing the unwanted light (in a filter), or by redirecting it (by either interference or refraction). Table 1 presents techniques for controlling wavelength. Interference filters can have blocking filters that absorb wavelengths outside the specified transmission band. However, a blocking of  $10^{-3}$  may not be sufficient for the human visual

**TABLE 1** Controlling Wavelength

Technique	Properties	Advantages	Disadvantages/Comments
Absorption filters: aka dye, film, gel, or pigment filters	Available in both gels and glass forms. Generally broadband.	Can be extremely stable over time. Wide range of colors available. Reasonably priced.	Broadband, can be damaged by too much light, as encountered with lasers and focused beams. Most fluoresce.
Narrowband interference	Half-width about 10 nm, usually bonded to an absorption filter for blocking distant wavelengths.	Convenient, readily available, and moderately priced.	If not fully sealed, humidity can cause long-term degradation. Wavelength depends on angle of incidence. Need good blockers.
Broadband interference (see also dichroic beam splitters) Grating monochromators	Half-width 40–100 nm. Tunable, bandwidth can be set by slits.	Same as narrowband interference filters, less need for blocking filters. Can have low stray light.	Must match aperture of system to the interior grating to minimize stray light. Expensive.
Prism monochromators Interference wedges (spatially varying interference filters)	Produces a spectrum by refraction. Allow continuous change of wavelength by changing the position of light incidence.	High throughput. Small, easy to use.	Bulky. Some have more leakage than a monochromator, may not be adequately blocked for higher-order transmission. Must be placed at a pupil conjugate point to avoid chromatic wedging on the retina.
Special sources (Na, Cd, etc.)	Discharge sources that produce a few exactly specified spectral lines.	Many lines are available. Very pure wavelengths. Larger aperture sources are available.	Mostly replaced by lasers which are stable, cheap, and easier to use but are point sources (see Table 5).

system. For instance, a narrow 670-nm filter may transmit only  $10^{-3}$  of the intensity at 555 nm, but the human visual system is about 30 times as sensitive to 555 nm as to 670 nm and can integrate over the entire visible spectrum. Thus, the visual effectiveness of the “blocked” light can be considerable. Monochromatic and narrowband light sources and lasers can also be used both to provide a specific wavelength and to calibrate the wavelength scale of monochromators.

Wavelength selection filters are best placed in a collimated beam to avoid changing the passband to shorter wavelengths. This shift is asymmetric, and the maximum transmission changes as the angle of incidence decreases from 90 deg.<sup>2</sup> Differently colored lights can be combined using broadband interference filters (dichroic filters) placed at an angle, but using narrowband filters designed for normal incidence at 45 deg can introduce unwanted polarization properties.<sup>2</sup> Filters designed for use at 45 deg are readily available.

## Turning the Field On and Off

Light from an optical channel can be clocked entirely by a shutter. Shutters should generally be located at a pupil conjugate plane. At this position the source is focused to a small area and thus a smaller, faster shutter can be used. In addition, manipulation of a pupil plane changes the retinal image uniformly (see above). Parameters of interest are the speed, repetition rate (and duty cycle), and the transmission of the shutter in both its open and closed states. Table 2 summarizes details on common types of shutters. Shutters that work by means of polarization generally need to be placed in a collimated beam, and their extinction (transmission when off) and maximum transmission are wavelength-dependent.

## Controlling Intensity

The visual system operates over about 11 log units of luminance (11 orders of magnitude, or  $10^{11}$ ), and within this range can make discriminations of 0.1 percent or better. This level of visual capability requires a two-stage control of retinal illuminance. The overall retinal illuminance is set by

**TABLE 2** Types of Shutters

Shutter	Speed	Size of Aperture (mm)	T off/T on	Comments
Mechanical shutter	ms	1 mm and up	0	Larger apertures are slower. Most designs cannot run in a continuous mode, but these are ideal for low duty-cycle use.
Galvanometers	ms	A few mm for high speeds	0	Operating at high rates requires fairly careful driver design but they can be used to excellent effect and are commercially available with drivers. Can be run continuously.
Choppers	kHz	Variable	0	For continuous on-off flicker choppers are ideal. They have no attenuation when open and no transmission when closed. Small, feedback-controlled devices are available off the shelf. Stepper motors with vanes mounted on the shaft can also be used.
Acousto-optic modulator (AOM)	$\mu$ s	<2 mm	$\sim 10^{-3}$	AOMs have small aperture and high $f$ -number and thus work best for lasers. They are inherently chromatic and orientation-sensitive. Polarization effects can be important.
LCD shutters	ms	A few cm	$\sim 10^{-2}$	Speeds are increasing rapidly, works by polarization. Maximum transmission is less than 50 percent. Use in collimated beam.
LCD displays	$\mu$ s	Work as video screens; individual pixels can be $\mu$ m.	$\sim 10^{-2}$	Work by polarization. Speed is increasing with development of displays for computer use.
Kerr cells	ps	cm	$\sim 10^{-2}$	Require high voltages, work by polarization.

discrete neutral density filters, while smaller steps are set by either a variable density filter or an electronic system. Due to limited precision in most measuring instruments, it is hard to calibrate an individual filter with a density greater than 3.0.

**Varying the Intensity of the Field** The intensity of a channel can be varied by using fixed or variable filters, or by controlling the radiance of the source. For simple, relatively slow (1/min) changes in retinal illuminance, neutral density filters are appropriate. For faster changes, modulators or direct variation of the source radiance are typically required.

**Filters** Neutral density filters either have a uniform attenuation across their spatial extent (fixed filters) or vary in their attenuation characteristics depending on the spatial location (variable filters or “wedges”). Absorbing filters should be used in a collimated beam to avoid different path lengths through the filter. Wedges are used at a pupil conjugate plane to avoid imaging the density gradient on the retina. Fixed filters are typically used to set fixed parameters (such as the average luminance), variable filters to control the brightness continuously (for instance, in an increment threshold test). Table 3 presents some types of neutral density filters.

**Modulators** Passing linearly polarized light through a rotating polarizer produces sinusoidal modulation of the transmitted light. Careful optical design allows generation of an excellent variable modulation, flicker stimulus.<sup>19,20</sup> Acousto-optic modulators (AOM) can be used to vary a light rapidly, but have only a small aperture. Mechanical methods have also been used for temporally modulating a target.<sup>21–23</sup> For the most part, mechanical modulators and moving polarizers have been replaced by direct control of light sources by high-speed shutters and by video systems (see Chap. 22 in this volume of the *Handbook*).

**TABLE 3** Neutral Density Filters

Method	Advantages	Disadvantages	Other Comments
Metal film neutral density filters	Readily available, stable, spectral neutrality is good.	Must be cleaned with care, pinholes can cause a problem. Interreflections between filters can cause deviations from density obtained when calibrated alone.	Silver and inconel are most common. Tilt to keep reflections out of the system.
Multilayer dielectric filters	Easier to clean.	Expensive, chromatic.	
Metal film “wedges”	Same as metal film filters.	Same as metal film filters.	Tilt to aim reflections out of the system. Use in a pupil conjugate plane to avoid spatial variation of image.
Absorbing gel filters	Less expensive, easily available at good photography stores.	They become brittle with age. Somewhat chromatic. Damaged by intense light.	Keep dry and away from a hot lamp. Use in a collimated light path.
Sector disk and mechanical devices (cams, variable slits, etc.)	Spectrally flat.	Must be used in a pupil plane to get desired results.	Use in a pupil conjugate (or in a uniform beam before an integrating sphere). Variable slits can be a simple, inexpensive way to control intensity.
Photographic film or film wedges	Cheap, easily made by flashing a film and developing.	Spectral neutrality will be poor in relation to commercial filters. Must be calibrated. Stability over time will depend on the film base used.	Best used with monochromatic light.

*Varying the Source* Varying the radiance of the light source is fast, and is particularly straightforward with electro-optical sources such as LEDs and diode lasers. However, control of the source is not limited to these devices and can be used with thermal sources (incandescent bulbs), xenon arcs, and fluorescent lights as well. There are four major problems to be overcome when directly varying a light source: (1) nonlinear current versus radiance characteristics, (2) changes in the current-radiance relation over time, (3) temporal hysteresis in the current-radiance relation, and (4) changes in the wavelength distribution of the source with changes in current.

For instance, LEDs have been considered ideal light sources by some investigators, yet they show the same deviations from ideal behavior common to other sources<sup>24–29</sup> (see Chap. 15, “Artificial Sources,” by Anthony LaRocca, Chap. 17, “Light-Emitting Diodes,” by Roland H. Haity, M. George Craford, and Robert M. Weissman, and Chap. 18, “High-Brightness Visible LEDs,” by Winston V. Schoenfeld in Vol. II). The current/radiance relation of LEDs depends on the internal temperature of the device. Thus, the relation changes with both the measurable *temperature* (due to both the environment and to the time average current) and with the immediate current history (for instance, it was just driven at a high current). Finally, the temperature also affects the probability of electronic transitions in the semiconductor, which can change the *wavelength* emitted. For LEDs used in visual work this wavelength dependence has been measured to be on the order of a 1.6-nm change in dominant wavelength with variation of the duty cycle for an ultrabright red LED.<sup>26</sup> Thermal sources, such as a tungsten-halogen bulb, undergo especially large changes in spectral output with changes in current. With these sources, only a narrow spectral band should be used if the source is controlled. In addition, the thermal inertia of an incandescent source precludes rapid modulation, although slow modulation (~1 Hz) can be achieved. Even fluorescent sources which have been widely used in some areas of research show significant changes in spectral output with time.<sup>30</sup>

While heat sinks can help to stabilize the response characteristics of many devices, linear control still requires careful driver design. Drivers can use either an analog or binary design. Analog drivers control the source radiance by varying the current, while digital drivers turn the source either on or off, with the average radiance set by the proportion of time that the source is on (the duty cycle). An advantage of the binary scheme is that transistors dissipate less power in the on and off states than in intermediate states.

There are four major approaches to linearization:

1. Calibrate the current-luminance relation and build a driver with the inverse nonlinearity. For simple applications this technique is adequate. However, for demanding applications the driver becomes complex and the demands for stability in the driver raise new problems.
2. Calibrate the current-luminance relation and use an inverse nonlinearity in computer memory (a lookup table). This technique is quick and easy to implement. A static lookup table will not compensate for dynamic nonlinearities (hysteresis) but a variation of this approach, known as *delta modulation*, will. With delta modulation a linear detector is used to precalibrate the output of a binary driver to produce the desired waveform. By later playing back the binary sequence, the waveform can be re-created exactly. Thus, delta modulation can be used to compensate for all source nonlinearities except for changes in wavelength. The disadvantage is that the waveform must be precalibrated.
3. Vary the ratio of the on and off time periods (the duty cycle) of the source using a binary driver. With a fixed cycle time, the on time can be varied (pulse-width modulation or PWM). PWM works fairly well, but is sensitive to capacitance and nonlinear switching effects that can alter the waveform for either short on or off periods. A similar approach is to use fixed (short) pulses and vary the frequency of the pulses (pulse-frequency modulation or PFM). This approach has the advantage that every pulse is identical, thus capacitance and switching effects are minimized. PFM has been used to control the luminance of LEDs<sup>27</sup> and AOMs<sup>16</sup> linearly over a 1000:1 luminance range.
4. Detect the light with a linear photodetector and use feedback within the driver to linearize the output. By using a PIN photodiode in photovoltaic mode (see Chap. 24, “Photodectors,” by Paul R. Norton in Vol. II), it is possible to construct a very linear circuit.<sup>28,31</sup> Light feedback can be used with either analog or binary drivers.



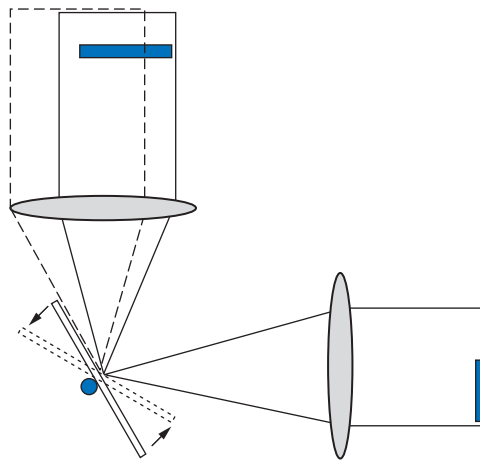
## Generating Complex Temporal and Spatial Patterns

Any type of light source or visual stimulus, from video monitors to street scenes, can be integrated into an optical system, giving improved control of luminance and pupil position. Thus, almost anything can be used as a target. Stimuli can be moved without varying the position of light entry in the eye's pupil by (1) translating a target in the retinal plane, (2) using a cathode ray tube (CRT) or a liquid crystal display (LCD) in a retinal conjugate plane, or (3) rotating a mirror in a pupil conjugate plane.<sup>22</sup>

Rotating a mirror in a pupil conjugate plane changes the angle at which light enters the eye, but not the pupil entry position. This technique has been used for generating motion and counterphase gratings<sup>21</sup> and for decreasing the apparent inhomogeneity of a stimulus by moving the image at a rate above flicker fusion. Galvanometers are commonly used to generate the motion (Fig. 5).

**Video Monitors and CRTs** The combination of a video display with a maxwellian view optical system allows the experimenter the advantages of precise control of the pupil, use of a broad spectral range of backgrounds (provided by the traditional optics), and the fine spatial control available in a video system. The video monitor or CRT is placed at a retinal conjugate plane and the other optical channels are added to it. By passing the combined light path through an aperture at a pupil conjugate plane it is possible to control the pupil size for the video system, insert achromatizing lenses, and the like (see Fig. 8 later in the chapter).

**Liquid Crystal Displays** One type of video display that is of increasing utility is the liquid crystal display (LCD) (Chap. 22). LCDs, unlike video monitors, work by transmission or reflection; they are not self-luminous. Thus, while the wavelength composition of a monitor is determined by the characteristics of its phosphors, the LCD can be used with *any* light source. LCD displays work by changing the polarization state of local regions of the display (pixels). The pattern of pixels is controlled by a video signal or a computer. Like most polarization devices, the extinction ratio is wavelength-dependent and the proportion of light transmitted in the off state may be relatively high. With color versions of the newer active matrix LCDs it is possible to pass three wavelengths of monochromatic light through the color LCD and have a spatially controlled, high-intensity display with the maximum possible gamut. In this case the LCD is placed at a retinal conjugate plane.



**FIGURE 5** A movable mirror located in a pupil conjugate plane allows motion of the retinal image without moving the pupil conjugate images.

## Calibration

There are several approaches to calibrating the retinal illuminance of a maxwellian view system. All depend on either a standard source of known luminance (a reference source) or a calibrated, linear<sup>31,32</sup> photodetector. In the past, calibration typically depended on referring the illuminance of an optical system to a carefully maintained standard light or by measuring the luminance of a diffuser. Well-calibrated photometers and radiometers are now readily available, and we will cover techniques based on using these (the reader is referred to Westheimer<sup>3</sup> for other approaches). We assume that all of the light arriving at the exit pupil of the instrument enters the eye and that the exit pupil can be adequately approximated as circular with a radius of  $r$ .

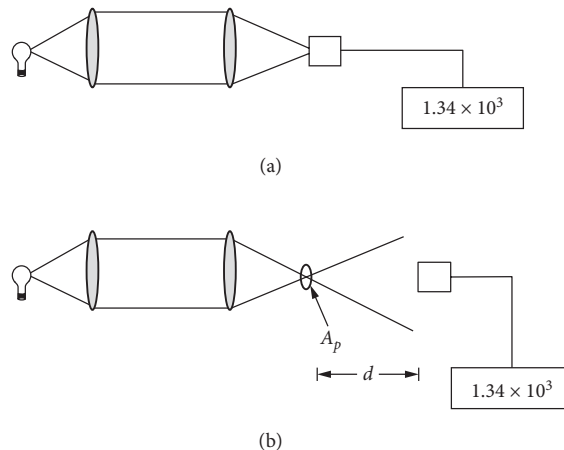
**Measurement of Power at Exit Pupil** With the first technique (Fig. 6a) a calibrated detector is placed at the exit pupil of the maxwellian view device. First, the total luminous power at the exit pupil is measured. Next, the retinal area illuminated is computed from the geometry of the stimulus [Eq. (12)]. From these two quantities the retinal illuminance can be computed. This calculation specifically assumes that all of the power is uniformly distributed across the retinal image. Conversion of retinal illuminance from  $\text{lm}/\text{m}^2$  to trolands is achieved using Eq. (6).

**Measurement of the Illuminance on a Detector** In this technique the illuminance falling on a detector at a fixed distance from the exit pupil is measured (Fig. 6b; also see Ref. 33). A circular source of luminance  $L$  and radius  $r$  produces an illuminance  $E$  on a detector at distance  $d^2$ . If we assume that the source dimensions affect the calibration negligibly ( $r < d/10$ ;  $r < f_e/10$ ), then

$$E = \frac{L\pi r^2}{(d^2)} \quad (18)$$

We want to relate  $E$  as measured by the detector to  $E_R$ , the retinal illuminance, which is

$$E_R = \frac{L\pi r^2}{(f_e)^2} \quad (19)$$



**FIGURE 6** Techniques for calibrating an optical system. (a) Measure the luminous power of the exit pupil and assume that it is uniformly distributed in the retinal image. (b) Measure the illuminance produced by an exit pupil of radius  $r$  at a detector located at distance  $d$  from the exit pupil.

Combining Eqs. (18) and (19) we have

$$E_R = \frac{(d)^2}{(f_e)^2} E \quad (20)$$

If the radius of the exit pupil ( $r_p$ ) cannot be ignored then we need to account for its properties as a light source. If the exit pupil is circular and can be approximated as a lambertian emitter, then the illumination on the detector [Eq. (18)] is

$$E = \frac{L\pi r^2}{(r_p^2 + d^2)} \quad (21)$$

(Ref. 2, Chap. 56, Table I(4.4)). Likewise, the retinal illuminance [Eq. (20)] is

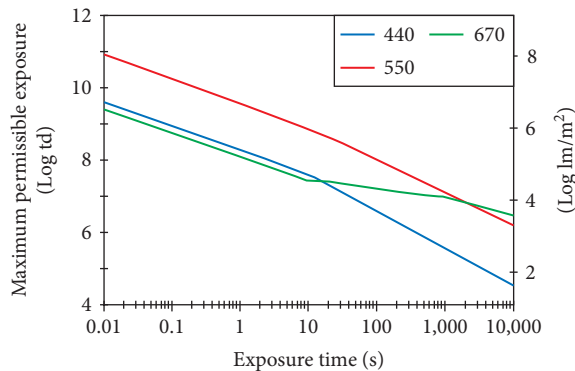
$$E_R = \frac{(r_p^2 + d^2)}{(r_p^2 + (f_e)^2)} E \quad (22)$$

The troland value of  $E_R$  can be computed from Eq. (16).

## 5.7 LIGHT EXPOSURE AND OCULAR SAFETY

There are two main mechanisms by which light can damage the eye. The first is simply by heating: too much radiation burns the retina. The second mechanism is photochemical. Light, especially short-wavelength light, causes photochemical oxidation. The by-products of this oxidation are toxic to the retina (see Chap. 7).

Figure 7 shows the relation of the danger thresholds as defined by the ANSI 136 standard,<sup>34</sup> expressed in tds and lumens/m<sup>2</sup>, for a 440-, a 550-, and a 670-nm light. These thresholds are for field sizes greater than 1 deg. The damage threshold for visible light is quite high, and thus lights capable of producing damage are intensely unpleasant. However, care should be taken to exclude IR and UV light. Most



**FIGURE 7** The maximum permissible retinal illuminance at different exposure times for 440 nm (blue line), 550 nm (red line), and 670 nm (green line) lights and extended fields according to the ANSI standards.

light sources emit considerable radiant energy in the infrared, and sources such as xenon arcs can emit ultraviolet radiation. Failures of blocking filters in these spectral regions may be visually undetectable. For safety it is advisable to have separate blocking filters to exclude unwanted radiation. It should also be noted that when using coherent light sources speckle can cause large focal variations in the retinal illuminance and the safety limits should be lowered by a factor of between 10 and 100 times.<sup>35</sup> Point sources and lasers have different safety standards that take into account the motion of the eye.

## 5.8 LIGHT SOURCES

Common light sources are presented in Table 4 and lasers are presented in Table 5 and in Chap. 15. Wyszecki and Stiles<sup>2</sup> present detailed discussions of light sources. The considerations for deciding on a light source are stability and uniformity of the luminous area, intrinsic luminosity of the source, and the spectral distribution of light from the source. Light sources of various types are discussed in Chaps. 15 to 23 in Vol. II.

## 5.9 COHERENT RADIATION

Coherent light sources can be used for generation of interference fringes,<sup>36</sup> a common use of lasers in vision research.<sup>8-10</sup> Light from the laser is directed into two similar channels. The light from the two channels arrives at the pupil of the eye separated by some distance  $d$ . Light spreading from the two beams then overlaps on the retina and creates an interference pattern in the intersection zone. The spacing of the interference pattern is controlled by the angle at which the beams intersect at the retina. The orientation of the pattern is controlled by the relative orientation of the two pupils. However, HeNe and easily controlled solid-state visible lasers are also excellent general-purpose light sources. The major drawback to their use is speckle.

### Generation of Speckle Fields

Speckle arises when coherent light is reflected from an optically rough surface. The surface roughness causes variations in the path length between the retina and adjacent areas on the surface. This variation causes phase differences from point to point in the retinal image. The spatial frequency of

**TABLE 4** Common Light Sources

Type	Properties	Comments
Tungsten	Broadband, thermal emitter.	Available in a number of filament and power choices, hotter bulbs are short-lived.
Tungsten-halogen	Broadband, thermal emitter.	Inclusion of halogen regeneration cycle allows tungsten-halogen to run hotter (more blue light and a higher luminance) at a longer rated life than tungsten. If running at below the rated wattage the bulbs can be inverted to extend the lifetime.
Xenon arcs	Nearly flat spectrum, high luminance, some emission lines are present.	UV light can create ozone which must be vented. Tendency for the arc to move relative to the electrodes, causing movement of the exit pupil in maxwellian view systems. Luminous area is fairly small.
Fluorescent tube	Broad spectral output with superimposed emission lines. Easily obtainable, efficient.	Can be modulated rapidly. But there may be problems with compound phosphors (see Ref. 28).
Light-emitting diodes (LEDs)	Small, cheap, efficient light sources. Easily controlled.	Can be modulated rapidly, come in a variety of wavelength and power ranges (see Chap. 13).

**TABLE 5** Types of Lasers

Type	Most Common Visible Wavelengths	Comments	Typical Power	Typical Noise %
Argon	514, 488, (other lines are available, depending upon the design).	Power of each wavelength varies with particular laser design and cost. Most common wavelengths are 514 nm and 488 nm.	5 mW–5 W	1
HeCd	444 nm	Fairly expensive, have some high-frequency noise (>300 kHz) that may be important in scanning or short exposure applications.	0.5–50 mW	3
HeNe	543, 594, 632.8, also have orange lines.	Most common and least expensive gas laser. The 632.8-nm designs are easily available, inexpensive, and long-lived. Other wavelengths have many of the same advantages but power and beam options are limited and cost is higher.	0.1–20 mW	0.5
Krypton	588, 677.	Expensive, tend to be unstable, large.	50 mW–5 W	3
Dye lasers	Variable.	Expensive, need to be pumped by another laser. Require costly and toxic dyes to operate, not particularly stable.	50 mW–5 W	Depends on pump laser.
Slab diode lasers	~630 nm, ~670 nm, ~690 nm, some in blue.	Development is very rapid with new wavelengths and power options. The beam geometry is not ideal but they have great potential as light sources. Can be easily controlled using standard electronic techniques as are common with LEDs. Have high impedance and are easily damaged by static electricity. Rapid developments at shorter wavelengths.	0.2–100 mW	1 percent or better if temperature stabilized.
VCSELS vertical cavity lasers (microlasers)	Infrared.	These are under rapid development. They are solid-state lasers but have superior efficiency and beam properties as compared to slab lasers. Low impedance.	1 mW	
Frequency doubled YAG	532 nm.	Readily available, solid-state designs. Expensive.	5–50 mW	<1 percent.
Tunable solid-state and solid dye lasers	Variable, pumped by lasers.	An area of rapid development. By using materials that fluoresce, and tuning the laser cavity, the output wavelength can be varied while maintaining excellent beam qualities.	mW–W	Depends on pump laser.

speckle depends on pupil size, since the size of those adjacent areas is smaller for bigger pupils. Thus, for a very small pupil, speckles will be large. Speckle has been used to develop optometers,<sup>37–39</sup> and to generate a pattern on the retina in the presence of poor optic media.<sup>40,41</sup> However, when using lasers as general-purpose light sources, speckle needs to be minimized or eliminated. There are three ways to despeckle a source: *spatial filtering*, *temporal despeckling*, and *raster scanning*.

In spatial filtering<sup>10</sup> the light remains spatially coherent. A light source is imaged onto a pinhole which diffracts the light so that a spherical wave emerges from the pinhole. A lens is then used to collimate the light, resulting in a plane wave. Filtering should be restricted to those experiments that require the use of coherent light. The introduction of dust or even structure in the anterior segment of the eye can introduce undesirable diffraction effects.

Temporal despeckling<sup>14,15</sup> uses the temporal integration of the visual system to blur the speckle field. Since speckle arises as a result of surface irregularities at a very small scale, a small amount of motion can decrease speckle contrast considerably.

Scanning moves a diffraction-limited spot of light across the retina to create a visual pattern. Typically, the pattern is a raster pattern, like that of a television, and the stimulus is generated by temporal variation in the intensity of the beam.<sup>42–44</sup> Since at any one time only a single, diffraction-limited, retinal region is illuminated, then there is no opportunity for speckle.

## 5.10 DETECTORS

Detectors are used for measuring the light produced by an optical system. They can generally be characterized by their quantum sensitivity, their spectral response curve, and the temporal amplitude spectrum of their noise (see Ref. 45 and Chap. 24, “Photodetectors,” by Paul R. Norton and Chap. 27, “Signal Detection and Analysis,” by John R. Willison in Vol. II). For most purposes detectors can be treated as detecting all light power incident upon their active area, but there are limitations that make it advisable to use them for normally incident light whenever possible.<sup>46</sup> Table 6 presents the most common detectors (also see Chaps. 24 to 28 in Vol. II). At low light levels photomultiplier tubes (PMT)<sup>47</sup> or an avalanche photodiodes (APD)<sup>48</sup> can be used in a photon-counting mode (Chap. 26, “High-Speed Photodetectors,” by John E. Bowers and Yih G. Wey in Vol. II).

## 5.11 PUTTING IT TOGETHER

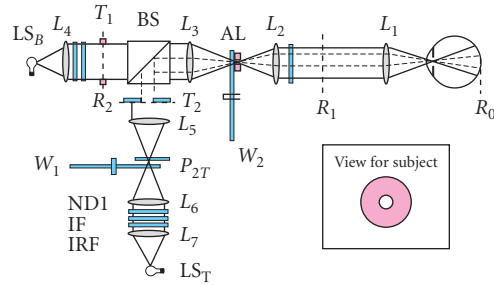
This section briefly describes construction of more complex maxwellian view systems. The goal is simply to help the researcher get started building an optical device by presenting real examples.

### A Two-Channel Maxwellian View Apparatus

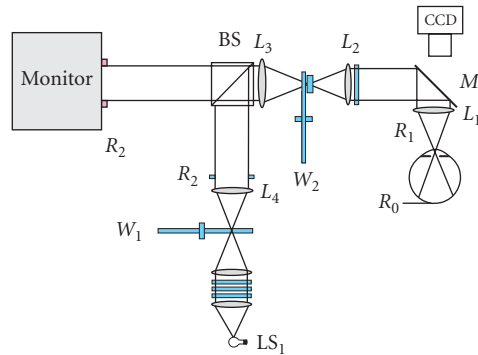
We first consider a simple two-channel maxwellian view device (Fig. 8a) for measuring the detectability of a small circular target on a large circular background. The background is provided via the straight optical channel. It has a light source  $LS_B$  followed by a lens  $L_4$  that collimates the light from the source. A field stop (target  $T_1$ ) controls the size of the background field. Adjacent to the target is a beamsplitter, followed by another lens  $L_3$ .  $L_3$  creates a source image where we place a 2.0-log-unit circular neutral density wedge, a 2-mm (diameter) stop (artificial pupil) and an achromatizing lens AL. An achromatizing lens minimizes the effects of the chromatic aberrations of the eye.<sup>49–51</sup> The artificial pupil will be imaged at 1:1 in the pupil of the eye, thus providing a limiting pupil. Lenses  $L_2$  and  $L_1$  relay the artificial pupil to the pupil of the eye. The test channel is derived from the second light source  $LS_T$ . Light is collimated by  $L_7$ , passed through a heat-rejecting filter IRF, interference filter  $IF_1$ , and a neutral density filter  $ND_1$ . An image of  $LS_T$  is then formed at  $P_{2T}$  by  $L_6$ . In this plane we set an electromechanical shutter and a neutral density wedge. We then collimate

**TABLE 6** Detectors

Detector	Sensitivity	Speed	Problems/Limitations
Photomultipliers	Best in the blue, extended multialkali cathodes extend sensitivity into the red and near IR. Many designs have very low noise.	Fast, MHz, capable of photon counting.	Low sensitivity in the red and infrared, can be damaged by high light levels, though recovery can be aided by leaving in the dark for a long time (months). Fragile. Require high voltage (kV).
Avalanche photodiodes	Have the silicon sensitivity curve, peaking in the near IR. Higher quantum efficiency than PMTs.	Fast MHz, capable of photon counting, better quantum efficiency than a PMT, but higher noise level.	Noise increases rapidly at avalanche voltage, light can affect the breakdown point. Higher noise level than PMTs, but the higher quantum sensitivity makes them better for video rate imaging. Small sensitive area.
CCDs, CIDs	Very linear, easy to use, sensitivity peaks in the near IR. These have a high quantum efficiency.	Variable, depending upon implementation, integrate between readings (kHz).	Blooming and charge spread can affect spatial properties at high radiances. <sup>52</sup> CID's are resistant to blooming.
PIN silicon photodiodes	Easy to use, sensitivity peaks in the near IR.	With care they can be operated at high rates (MHz); however, in most photometers the circuitry is specifically designed for low-noise, lower-frequency operation.	
Thermal (bolometers)	Measures energy, thus flat spectral responsivity.	Very slow <10 Hz.	The only real use for these devices in vision research is for calibration of the spectral sensitivity of another detector. In general, it is better to have the detector calibrated against an NIST standard by a manufacturer.
See also photoresistors, vacuum photodiodes, and phototransistors	Seldom used.		



(a) Two channel increment threshold apparatus



(b) Using a video display in a maxwellian view and monitoring eye position

**FIGURE 8** (a) An example of a two-channel increment threshold maxwellian view device. Two light sources are used ( $LS_B$  and  $LS_T$ ) to form a background and a test channel respectively. The two channels are combined using a beamsplitter (BS). The intensity of the test field is controlled by a circular neutral density wedge  $W_1$ , and the intensity of both channels is controlled using wedge  $W_2$ . Additional details are provided in the text. (b) A similar apparatus, except now a video monitor is superimposed on a background derived from source  $LS_T$ . An infrared sensitive CCD camera is used to monitor the pupil through the back of a dichroic beamsplitter plate ( $M_1$ ) which reflects visible light and transmits infrared light (cold mirror). Infrared illumination of the eye is provided by infrared LEDs (not shown).

the light with lens  $L_5$ , pass it through an aperture  $T_2$  that creates the target, and combine it with the background light at the beamsplitter. In some cases it may be desirable to obtain both the test and background light from a single light source. This is readily achieved by the addition of mirrors and avoids the possibility that slight fluctuations in the sources might produce variability in the ratio of the target and background illuminances. When appropriate, use a single light source for multiple optical channels.



A similar optical design allows full spatial control of the stimulus by incorporating a video monitor (Fig. 8*b*) or an LCD (not shown). In general, the retinal image of the monitor will need to be minified. As shown, the size of the retinal image of the monitor is set by the ratio of the focal length of lenses  $L_1$ ,  $L_2$ , and  $L_3$  [Eq. (14)]. The main advantages of the hybrid system over free-viewing the monitor are the ability to control the pupil precisely, the ability to add spectrally pure backgrounds (and, if only low contrasts are needed, the background can be quite bright), and the ability to align an achromatizing lens precisely and to monitor the eye position. We show a simple monitoring system, where the eye is monitored by a CCD camera, through the back side of a “cold mirror” (M1). A cold mirror is a dichroic beamsplitter that reflects visible light and transmits infrared radiation. IR-emitting LEDs are used to illuminate the eye diffusely. Using the full aperture of the lens  $L_1$  means that the depth of focus will be small, and the eye can be precisely positioned (or monitored) in three dimensions. The resolution of the camera can be relatively low.

To use an LCD in transmission mode rather than a video monitor we can simply introduce the display at the retinal conjugate plane ( $R_2$ ) in Fig. 8*a*. Again, it will be desirable to decrease the magnification of the LCD on the retina by choosing  $L_3$  to have a longer focal length than lenses  $L_1$  and  $L_2$ .

## An Interferometer

We next describe a research interferometer that was built using modern electro-optical components.<sup>10</sup> To study the spatial sampling of the photoreceptor matrix of the eye, this instrument needed to have the following properties: (1) generation of spatial patterns on the retina with a frequency content above that imaged by the optics of the eye, (2) rapid and precise control of the modulation (contrast) and frequency (spacing) of the spatial patterns, (3) minimization of speckle, and (4) control of the two entrance pupils of the interferometer to displace them symmetrically about the center of the eye’s pupil. We outline below how these objectives were achieved. The researcher should refer to the original paper for a more detailed account.

Interference can be used to generate diffraction patterns on the surface of the retina. Two coherently related sources are imaged in the plane of the eye’s pupil (Fig. 9*a*). The spacing of the resulting diffraction pattern depends on the separation of the two sources in the eye’s pupil, and the orientation of the pattern depends on the relative position of the two sources. Williams<sup>10</sup> (see also Refs. 8 and 9 for other examples) constructed a modified Mach-Zender interferometer<sup>52</sup> (Fig. 9*b*) that satisfies the requirements outlined above.

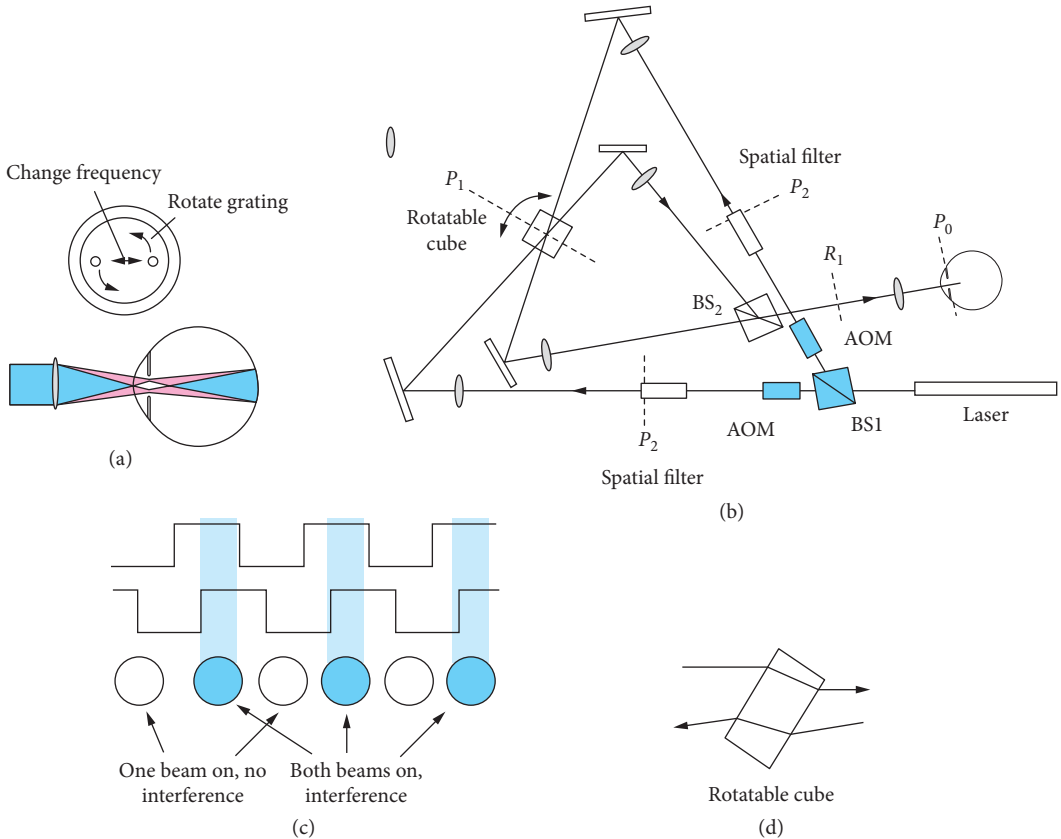
Modulation of the interference pattern is controlled using AOMs. Each of the beams is either on or off, with a 50 percent duty cycle. If both beams are on at the same time (in phase), the interference pattern is seen at full contrast. If one is off when the other is on (counterphase), then there is no interference pattern. Intermediate phases of the two beams will cause intermediate contrasts of the interference pattern (Fig. 9*c*). Thus, both beams are always operated in the same state (avoiding luminance artifacts due to temporal nonlinearities), and all control of modulation is electronic. The use of an all-electronic design allows precise control of contrast without audible cues to the subject. Spatial masking from speckle is minimized by placing spatial filters in a pupil conjugate plane. The spacing of the interference pattern is controlled by a rotatable cube. Each beam traverses the cube in opposite directions (Fig. 9*d*). As the cube is rotated the beams, and thus the images in the eye’s pupil, are displaced symmetrically.

## 5.12 CONCLUSIONS

We have covered only some of the basics of stimulus generation. We urge the interested reader to look elsewhere in this *Handbook*. In addition, the following references are good general sources for this topic.

## 5.13 ACKNOWLEDGMENTS

Supported by NIH EYO4395 (Burns) and DOE DE-FG02-91ER61229 (Webb). We thank Francois Delori for aid in computing the troland values for the ANSI standards and David Williams for valuable editorial help.



**FIGURE 9** An example of a modern interferometer for vision research. (Modified from Ref. 10.) (a) Factors controlling the spacing and orientation of the interference pattern on the retina. Increasing the distance between the entry pupils decreases the spacing of the pattern (increases the spatial frequency in cycles per deg). It is also possible to change the orientation of the interference pattern by rotating the two entry pupil in the plane of the eye's pupil. (b) A schematic of William's interferometer. This is a modified Mach-Zender interferometer and includes acousto-optic modulators (AOM) to control the modulation of the interference pattern, spatial filters to reduce speckle, a rotatable cube to control spacing and beamsplitters to separate the two beams (BS1) and to recombine them (BS2). (For additional details see Ref. 10.) (c) Contrast of the interference pattern is controlled by electronically varying the phase of the signals from the two AOMs. Each AOM is square wave modulated. The interference pattern is produced only when both beams are on at the same time. Thus, varying the relative phase of the two beams varies the proportion of the time when both are on, and thus the contrast. (d) The spacing of the beams in the plane of the eye's pupil is varied by rotating a deviation plate. The two beams propagate through the glass plate in opposite directions. Thus, as the plate is rotated, the beams are displaced equal but opposite amounts, resulting in symmetric deviations of the beams about the center of the eye's pupil. The actual apparatus uses an anti-reflection coated cube.

## 5.14 GENERAL REFERENCES

- J. R. Meyer-Arendt, "Radiometry and Photometry: Units and Conversion Factors," *Applied Optics* 7:2081–2084 (1968).  
 G. Westheimer, "The Maxwellian View," *Vision Res.* 6:669–682 (1966).  
 J. W. T. Walsh, *Photometry*, Constable, London, 1953.

- G. Wyszecki and W. S. Stiles, *Color Science*, Wiley, New York, 1982.
- R. M. Boynton, "Vision," J. B. Sidowski (ed.), *Experimental Methods and Instrumentation in Psychology*, McGraw-Hill, New York, 1966.
- W. J. Smith, *Modern Optical Engineering*, 2d ed., McGraw-Hill, New York, 1990.

## 5.15 REFERENCES

1. Y. LeGrand, *Light Color and Vision*, Wiley, New York, 1957.
2. G. Wyszecki and W. S. Stiles, *Color Science*, Wiley, New York, 1982.
3. G. Westheimer, "The Maxwellian View," *Vis. Res.* **6**:669–682 (1966).
4. F. W. Fitzke, A. L. Holden, and F. H. Sheen, "A Maxwellian-View Optometer Suitable for Electrophysiological and Psychophysical Research," *Vis. Res.* **25**:871–874 (1985).
5. R. J. Jacobs, I. L. Bailey, and M. A. Bullimore, "Artificial Pupils and Maxwellian View," *Appl. Opt.* **31**:3668–3677 (1992).
6. G. M. Byram, "The Physical and Photochemical Basis of Visual Resolving Power Part I: The Distribution of Illumination in Retinal Images," *J. Opt. Soc. Amer.* **34**:571–591 (1944).
7. O. Dupuy, "La perception visuelle," *Vis. Res.* **8**:1507–1520 (1968).
8. F. W. Campbell and D. G. Green, "Optical and Retinal Factors Affecting Visual Resolution," *J. Physiol.* **181**:576–593 (1965).
9. F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *J. Physiol.* **186**:558–578 (1966).
10. D. R. Williams, "Aliasing in Human Foveal Vision," *Vis. Res.* **25**:195–205 (1985).
11. W. D. Wright, *Researches into Normal and Defective Colour Vision*, Kimpton, London, 1946.
12. T. W. Liebman, "Wedge Plate Beam Splitter without Ghosts Reflections," *Appl. Opt.* **31**:5905–5906 (1992).
13. R. M. Boynton and A. L. Nagy, "La Jolla Analytic Colorimeter," *J. Opt. Soc. Amer.* **72**:666–667 (1982).
14. R. H. Webb, "Concentrator for Laser Light," *Appl. Opt.* **31**:5917–5918 (1992).
15. J. R. Krauskopf, D. R. Williams, and D. H. Heeley, "Computer Controlled Color Mixer with Laser Primaries," *Vis. Res.* **21**:951–953 (1981).
16. S. A. Burns, M. R. Kreitz, and A. E. Eisner, "Apparatus Note: A Computer Controlled, Two Color, Laser-Based Optical Stimulator for Vision Research," *Appl. Opt.* **30**:2063–2065 (1991).
17. C. A. Taylor and B. J. Thompson, "Some Improvements in the Operation of the Optical Diffractometer," *J. Sci. Instr.* **34**:439–447 (1957).
18. R. M. Boynton, M. M. Hayhoe, and D. I. A. MacLeod, "The Gap Effect: Chromatic and Achromatic Visual Discrimination as Affected by Field Separation," *Optica Acta* **24**:159–177 (1977).
19. M. C. Rynders and L. N. Thibos, "Single Channel Sinusoidally Modulated Visual Signal Generator, with Variable Temporal Contrast," *J. Opt. Soc. Amer.* **A10**:1642–1650 (1993).
20. V. V. Toi and P. A. Gronauer, "Visual Stimulator," *Rev. Sci. Instr.* **49**:1403–1406 (1978).
21. G. A. Fry, "Square Wave Grating Convolved with a Gaussian Spread Function," *J. Opt. Soc. Amer.* **58**:1415–1416 (1968).
22. T. W. Butler and L. A. Riggs, "Color Differences Scaled by Chromatic Modulation Sensitivity Functions," *Vis. Res.* **18**:1407–1416 (1978).
23. D. Vincent, "Amplitude Modulation with a Mechanical Chopper," *Appl. Opt.* **25**:1035–1036 (1986).
24. T. E. Cohn, "A Fast, Accurate Monochromatic Light Stimulus Generator," *Am. J. Optom. and Arch. Am. Acad. Optom.* **49**:1028–1030 (1972).
25. R. W. Nygaard and T. E. Frumkes, "LEDs: Convenient, Inexpensive Sources for Visual Experimentation," *Vis. Res.* **22**:435–440 (1982).
26. M. Yamashita and S. Takeuchi, "Temperature-Compensated Pulsed Reference Light Source Using a LED," *Review of Sci. Instr.* **54**:1795–1796 (1983).

27. W. H. Swanson, T. Ueno, V. C. Smith, and J. Pokorny, "Temporal Modulation Sensitivity and Pulse-Detection Thresholds for Chromatic and Luminance Perturbations," *J. Opt. Soc. Amer.* **A4**:1992–2005 (1987).
28. G. R. Cole, C. F. Stromeyer, III, and R. E. Kronauer, "Visual Interaction with Luminance and Chromatic Stimuli," *J. Opt. Soc. Amer.* **7**:128–140 (1990).
29. T. U. Watanabe, N. Mori, and F. Nakamura, "Technical Note: A New Suberbright LED Stimulator: Photodiode Feedback Design for Linearizing and Stabilizing Emitted Light," *Vis. Res.*, 1992.
30. J. D. Mollen and P. G. Polden, "On the Time Constant of Tachistoscopes," *Quarterly J. Exp. Psychol.* **30**:555–568 (1978).
31. C. L. Sanders, "Accurate Measurements of and Corrections for Nonlinearities in Radiometers," *J. Res. Natl. Bur. Stand., sec. A*, **76**:437–453 (1972).
32. R. G. Frehlich, "Estimation of the Nonlinearity of a Photodetector," *Appl. Opt.* **31**:5927–5929 (1992).
33. R. W. Nygaard and T. E. Frumkes, "Calibration of the Retinal Illuminance Provided by Maxwellian Views," *Vision Res.* **22**:433–434 (1982).
34. American National Standards Institute, "Safe Use of Lasers," Z-136.1 ANSI, N.Y., 1976.
35. D. L. Fried, "Laser Eye Safety: The Implication of Ordinary Speckle Statistics of Speckle-Speckle Statistics," *J. Opt. Soc. Amer.* **71**:914–916 (1981).
36. M. Francon, *Optical Interferometry*, Academic Press, N.Y., 1966.
37. R. T. Hennessy and H. W. Leibowitz, "Laser Optometer Incorporating the Badal Principle," *Behav. Res. Meth. Instr.* **4**:237–239 (1972).
38. N. W. Charman, "On the Position of the Plane of Stationarity in Laser Refraction," *Am. J. Optom. Physiol. Opt.* **51**:832–838 (1974).
39. A. Morrell and W. N. Charman, "A Bichromatic Laser Optometer," *Am. J. Optom. Physiol. Opt.* **64**:790–795 (1987).
40. W. W. Dawson and M. C. Barris, "Cortical Responses Evoked by Laser Speckle," *Invest. Ophthalmol. Vis. Sci.* **17**:1207–1212 (1978).
41. J. Fukuhara, H. Uozotot, S. Nojima, M. Saishin, and S. Nakao, "Visual-Evoked Potentials Elicited by Laser Speckle Patterns," *Invest. Ophthalmol. Vis. Sci.* **24**:1400–1407 (1983).
42. R. H. Webb, G. W. Hughes, and F. C. Delori, "Confocal Scanning Laser Ophthalmoscope," *Appl. Opt.* **26**:1492–1499 (1987).
43. R. H. Webb, "Optics for Laser Rasters," *Appl. Opt.* **23**:3680–3683 (1984).
44. A. E. Eisner, S. A. Burns, R. W. Webb, and G. H. Hughes, "Reflectometry with a Scanning Laser Ophthalmoscope," *Applied Optics* **31**:3697–3710 (1992).
45. R. H. Webb and G. H. Hughes, "Detectors for Video Rate Scanning Imagers," *Appl. Optics* **32**:6227–6235 (1993).
46. J. Durnin, C. Reece, and L. Mandel, "Does a Photodetector Always Measure the Rate of Arrival of Photons?," *J. Opt. Soc. Amer.* **71**:115–117 (1981).
47. D. van Norren and J. van der Kraats, "A Continuously Recording Retinal Densitometer," *Vision Res.* **21**:897–905 (1981).
48. P. P. Webb, R. J. McIntyre, and J. Conradi, "Properties of Avalanche Photodiodes," *RCA Review* **35**:234–278 (1974).
49. R. E. Bedford and G. Wyszecki, "Axial Chromatic Aberration of the Eye," *J. Opt. Soc. Amer.* **47**:564–565 (1957).
50. A. L. Lewis, M. Katz, and C. Ohrlein, "A Modified Achromatizing Lens," *Am. J. Optom. Physiol. Opt.* **59**:909–911 (1982).
51. I. Powell, "Lenses for Correcting Chromatic Aberration of the Eye," *Appl. Opt.* **20**:4152–4155 (1981).
52. D. Malacara, *Optical Shop Testing*, chap. 4.5, Wiley Interscience, New York, 1991.
53. M. Marchywka and D. G. Socker, "Modulation Transfer Function Measurement Technique for Small-Pixel Detectors," *Appl. Opt.* **31**:7198–7213 (1992).

*This page intentionally left blank*

---

# THE MAXWELLIAN VIEW\*† WITH AN ADDENDUM ON APODIZATION

---

Gerald Westheimer

*Division of Neurobiology  
University of California  
Berkeley, California*

**Abstract**—Optical questions arising in the so-called Maxwellian View method of illuminating the retina have been analyzed theoretically. Problems discussed in detail include those of photometry, of magnification, of focus, and finally of pupil size insofar as it relates to the transmission of spatial frequencies in coherent and incoherent illumination.

---

## 6.1 GLOSSARY

---

**Diffraction theory.** Formulation for calculating the distribution of electromagnetic energy when the propagation of radiation is restricted by apertures.

**Emmetropia.** Refractive state of eye in which the retina is conjugate to infinity. Its absence is called ametropia.

**Fourier theory of optics.** Object-image relationship in which the basic object structure is a sinusoidal grating and all other targets are described in their Fourier terms.

In the Fourier Theory of Optics, the complex amplitude of the electromagnetic vibrations in the plane of focus is the Fourier transform of that in the pupil plane.

**Incoherent light sources.** Sources for which light originating from different elements, when overlapping in forming the image, sums in intensity rather than by interference of the amplitudes of the electromagnetic vibrations.

**Pupil of eye.** Aperture stop, limiting the width of bundle of light entering the eye. The entrance pupil is the virtual aperture stop situated in the eye's object space; it can be directly measured and related to objects without further consideration of the eye's refracting apparatus.

**Transfer function.** Ratio of amplitudes of images to their generating objects for targets whose intensity changes sinusoidally, as a function of spatial frequency.

---

\*This chapter is the full text of an article on the Maxwellian view published in 1966 in the journal *Vision Research* (vol. 6, p. 669–682, 1966). This article is republished with permission from Pergamon Press. Since this is a classic article, it is reprinted in full here. Prof. Westheimer has added a post-script to this article.

†Aided in part by a Grant NB-03154 from the National Institute of Neurological Diseases and Blindness, U.S. P.H.S. and a contract between the Office of Naval Research and the University of California.

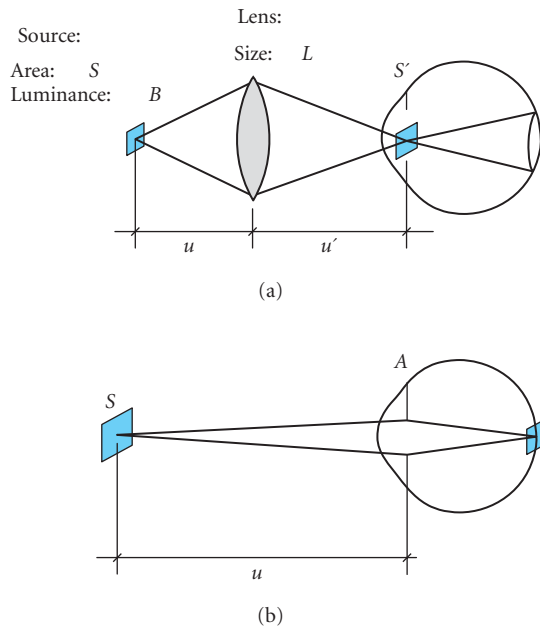
## 6.2 INTRODUCTION

In his experiments on color mixing, James Clerk Maxwell (1968) devised a way by which he could increase the quantity of monochromatic light reaching the retina of his eye. Using the sky as his source, he produced a spectrum by means of a prism and then isolated narrow regions of this spectrum with slits. He thus had monochromatic sources, but the special virtue of his approach is that he then imaged each slit on the pupil of his eye by a lens and saw the lens “uniformly illuminated with light.” The procedure of imaging a light source in the eye’s pupil, instead of looking at it directly, has since been applied widely and is now known as Maxwellian viewing. Its simplest form is illustrated schematically in Fig. 1, and compared there with ordinary viewing.

In the following discussion certain simplifying assumptions are used. (a) The eye is taken to be perfectly transparent; if the transmissivity of the ocular media is known, this can be multiplied into the final results. (b) The Stiles-Crawford effect is not taken into account. (c) No distinction is made between the plane of the eye’s entrance pupil and its anterior principal plane—the error introduced is relatively unimportant here. (d) The eye is assumed to be accommodated for the object plane which is not necessarily at infinity. (e) Angles are regarded as small enough that the approximation  $\cos \theta = 1$  is adequate.

In Fig. 1(a) and Fig. 1(b), the primary source of light is a small self-luminous patch of size  $S(\text{m}^2)$  and luminance  $B(\text{cd}/\text{m}^2)$ ; it is imaged directly on the retina in the case of Fig. 1(b), ordinary viewing, and it is imaged by a lens of aperture  $L(\text{m}^2)$  in the pupil of the eye, in the case of Maxwellian viewing, Fig. 1(a).

Concise discussions of the Maxwellian view can be found in Legrand (Vol. II) and in Walsh (1958). The present paper deals with several aspects of using the Maxwellian view in vision experiments.



**FIGURE 1** (a) Schematic diagram of the principles of Maxwellian viewing. A source of area  $S$  is imaged by a lens in the pupil of the eye. The image of the source has area  $S' = Su^2/u'^2$ . (b) Ordinary viewing of an extended source of area  $S$  placed at a distance from an eye whose entrance pupil has area  $A$ .

## Retinal Illuminance

An immediate concern in any arrangement in which visual stimuli are presented, is the quantity of light reaching the subject's retina. The retinal illuminance should really be measured in terms of the luminous flux per unit area of retina. But there is the difficulty that, while we may be able to specify the light reaching the eye, additional assumptions are necessary to specify the retinal illuminance. The assumptions concern such things as the transmissivity of the ocular media and the size relationship between objects in front of the eye and their images on the retina. To obviate this, an approach is here adopted which retains, for as long as possible, all quantities in realistic units measurable in the eye's object space. For example, the size of the stimulated retinal area is expressed in terms of the object-sided visual angle, rather than square millimeters of retina. Knowledge of the dimension of the structures of the eye permits conversion of one set of units to others, but the advantage claimed for the present approach is that its validity is independent of the constants of a particular eye and it may be applied readily, therefore, to non-standard conditions, and to the eyes of other species.

Take the case of ordinary viewing first (Fig. 1*b*). A luminous surface [area  $S(\text{m}^2)$ , luminance  $B(\text{cd}/\text{m}^2)$ ] is observed by an eye with an entrance pupil [area  $A(\text{m}^2)$ ], directly facing the surface and at a distance  $u(\text{m})$  from it. The total luminous flux entering the eye is equal to the product of the flux emitted by the source per unit solid angle ( $BS$ ) and the solid angle subtended by the eye's entrance pupil at the source ( $A/u^2$ ). This flux ( $B.S.A./u^2$ ) covers a retinal region represented in the eye's object space by the solid angle subtended by the source at the center of the entrance pupil ( $S/u^2$ ). The measure of the retinal illuminance which retains all quantities in the eye's object space is luminous flux per unit solid visual angle and it is given by

$$E (\text{lm/sterad}) = B \cdot A \quad (1a)$$

where  $B$  is the source's luminance in  $\text{cd}/\text{unit area}$  and  $A$  is the eye's pupillary area. It is to be noted that the retinal illuminance is independent of the size of the source and the observation distance.

The most widely used unit of retinal illuminance, the troland, is defined by the condition in which a surface of luminance  $1 \text{ cd}/\text{m}^2$  is observed by an eye with an entrance pupil area of  $1 \text{ mm}^2$ . A troland is thus equal to  $10^{-6} \text{ lm/sterad}$  visual angle. A steradian is not a convenient unit of solid visual angle. We are usually interested in much smaller areas, often of the order of magnitude of single retinal receptors. Thus, it may sometimes be better to give the luminous flux not in each steradian of visual angle, but in each square minute of arc of which there are  $(60 \times 57.3)^2 = 11.8 \times 10^6$  per steradian. A useful unit in which to express the rate at which light is impinging on the eye then is luminous flux per square minute of arc visual angle for each square millimeter of entrance pupil area. When this pupillary area is  $A \text{ mm}^2$  and when the source has luminance  $B \text{ cd}/\text{m}^2$ , we have

$$E (\text{trolands}) = B \cdot A \quad (1b)$$

or

$$E (\text{lm/sterad}) = 10^{-6} B \cdot A \quad (1c)$$

or

$$E (\text{lm/sq. min of arc}) = \frac{B \cdot A 10^{-6}}{11.8 \times 10^{-6}} = 0.85 \times 10^{-13} B \cdot A \quad (1d)$$

In the human eye, cones in the center of the fovea, and rods in an optimal region for scotopic vision (about  $10^\circ$  peripheral) are packed about 4 to a sq. min of arc, and thus the next kind of derived unit would be

$$E (\text{lm/receptor}) = 0.21 \times 10^{-13} B \cdot A \quad (1e)$$



Each scotopic lumen represents a flux of  $1.46 \times 10^{15}$  quanta (507 m $\mu$ ) per second (Aguilar and Stiles, 1954) and hence we have yet another derived unit

$$E \text{ (quanta (507)/receptor} \times \text{sec)} = 30 B \cdot A \quad (1f)$$

Note that in equations (1 b-f),  $B$  is the luminance of the extended source measured in cd/m<sup>2</sup>, and  $A$  is the pupil area in square millimeters.

All the above units are still in the eye's object space, and before the number of quanta that are absorbed in each receptor per second can be computed, it is necessary to know the proportion of externally incident quanta that is transmitted by the ocular media, and the proportion of the latter that is absorbed by the visual pigment. Rushton (1956b) has given an estimate of 0.1 for these two factors combined for human rod vision.

Consider the case of Maxwellian viewing, i.e. in which the source is not imaged on the retina but in the eye's entrance pupil. We may distinguish two important categories, one in which the image of the source (area  $S'' = S \cdot u'^2/u^2$ ) is larger than the pupil area ( $A$ ) and the other in which it is smaller. In the former, it is  $A$  that limits the light entering the eye. The retinal area illuminated in each case is given, in steradians of visual angle, by  $(L/u'^2)$  where  $L$  is the area (m<sup>2</sup>) of the clear aperture of the field lens, and  $u'$  is the distance between this lens and the eye. The total luminous flux reaching the image  $S''$  is  $B \cdot S \cdot L/u'^2$ . If the pupil area  $A$  is smaller than  $S''$ , the proportion of this total flux that can enter the eye is  $A/S''$ .

It is understood that when the pupil size remains constant, the retinal illuminance when viewing a Lambert surface is independent of observation distance. This is so because the angle subtended by an element of area increases when it is brought closer, but so does the angle the pupil subtends at the surface, and hence the fraction of the total radiation emitted by the source. The two effects exactly counterbalance each other.

A similar kind of invariance obtains when using an extended source in a Maxwellian view. So long as the image of the source is larger than the pupil, it is immaterial with what magnification the source is imaged in the pupil plane. The closer the source comes to the lens (i.e. the smaller  $u$ ) the larger the flux per unit area there. Yet at the same time, for a portion of lens surface of given size to subtend a fixed angle at the eye, the distance between the lens and the eye has to be controlled and this in turn controls the lens power and the magnification of the image of the source in the pupil. The result is that the flux per unit angular size of lens surface per unit pupil area always remains constant, so long as all the available pupil area is occupied by the image of the source.

The actual flux entering the eye is either  $S \cdot B \cdot L \cdot u'^2$  for  $S' < A$  and  $S \cdot B \cdot L \cdot A/S' u'^2$  for  $S' > A$ . In the latter case, i.e. when the source image is larger than the pupil, the flux is equal to  $B \cdot L \cdot A/(u')^2$  which is spread over an angular area  $L(u'^2)$ : the flux incident per unit solid visual angle is, therefore,

$$E \text{ (lm/sterad)} = B \cdot A$$

which is the same as that in the case of ordinary viewing.

The principal advantage of the Maxwellian view technique is here brought out; the retinal illuminance attainable is at most equal to that of ordinary viewing of the light source, but a larger area of the retina may be illuminated at that level. This is apparent in Fig. 1. When the pupil area is smaller than the image of the source at the pupil, the retinal illuminance, i.e. flux per unit retinal area, is the same in the two cases, only in the case of Maxwellian viewing the visual angle is equal to  $L/u'^2$ , while otherwise it is  $S/u'^2$ .

When the image of the source is smaller than the pupil, the retinal illuminance will be smaller in proportion to the fraction of the pupil covered by the image of the source, i.e.

$$E \text{ (lm/sterad)} = B \cdot A \quad S'/A = B \cdot S' \quad (2)$$

The determination of the retinal illuminance in an experimental arrangement employing Maxwellian viewing follows from the above consideration. When the effective entrance pupil of the eye (real or artificial) is completely covered by the image of the source, the retina is illuminated at the same flux/unit area as if the eye with that particular pupil were looking directly at the source—allowance having to be made for transmission losses in the optical system between the source and the eye.

When the image of the source is smaller than the pupil, an indirect method of determining the retinal illuminance may be employed, predicated by the fact that all the light coming towards the eye out of the Maxwellian view optical system enters the eye's pupil. The retinal illuminance in this case, by Eq. (2), is  $B \cdot S'$  lm/sterad target area. Figure 1(a) shows how this retinal illuminance is in fact achieved by having the equivalent of a source emitting  $B \cdot S'$  1m/sterad, i.e. a source of  $B \cdot S'$  candelas, imaged in the pupil. A good method\* of measuring the resulting retinal illuminance is to find the number of candelas in the areal image of the source by measuring the luminance it produces when illuminating a perfectly diffusing surface placed at a given distance from it. One candela placed 1 m in front of a perfectly reflecting and diffusing surface gives it a luminance of 0.1 mL. When a small source placed at a distance  $x$  meters in front of a perfectly diffusing surface of reflectance  $r$  produces a luminance of  $\bar{B}$  mL, the source has intensity  $10 \bar{B}x^2/r$  cd, and by equation (1b) such a source image in the plane of the pupil would give a retinal illuminance of  $10 \bar{B}x^2/r$  lm/sterad or  $10^7 \bar{B}x^2/r$  trolands.

The procedure for measuring the retinal illuminance in a Maxwellian view system where all the radiation passing through the instrument enters the pupil is:

1. Place a perfectly diffusing surface of reflectivity  $r$  at a distance  $x$  m beyond the source image that would ordinarily be in the plane of eye's pupil.
2. Measure the luminance of the surface with one of the standard instruments. Let it be  $\bar{B}$  mL.
3. Calculate the retinal illuminance using the formula

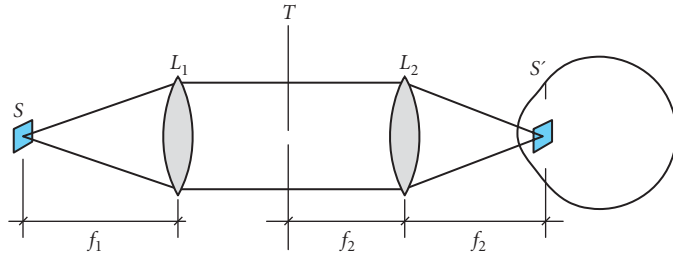
$$E \text{ (trolands)} = 10^7 \bar{B}x^2/r \quad (3)$$

## Focus

Our discussion has so far concerned itself only with the retinal illuminance achievable with a Maxwellian view arrangement. Unless a uniform field of view is desired, a target is placed in the beam. In the typical experiment it is intended to stimulate the subject's retina in a certain way, for example, by illuminating certain areas and not others. The almost universally adopted way of creating a desired pattern of light stimulation on the retina, is to place in the object plane, conjugate to the retina, a target which is then imaged on the retina by the optics of the eye. (Rare exceptions to this approach may occur, for example, when setting up interference fringes on the retina, Westheimer, 1960.) In the arrangement of Fig. 1(a), a reasonable place for the target is just in front of or behind the lens  $L$ . Unfortunately, the resulting retinal image will only be clear when the eye is accommodated on  $L$ . This restriction is no longer necessary, however, if we re-arrange the optical system by splitting the lens  $L$  into two components,  $L_1$  and  $L_2$ . In Fig. 1(a) it is seen that the function of  $L$  is to image the source in the pupil. Suppose this is done in two stages, first by using a lens,  $L_1$ , to produce a parallel beam of the light from the source, and then by having another lens,  $L_2$ , image this parallel beam in the pupil (Fig. 2). Because the beam in its passage between  $L_1$  and  $L_2$  is parallel, these two lenses may be separated by any desired distance without altering their effect of imaging the source in the pupil. But the target position can now be chosen so that the subject's accommodative state may be anything within a wide limit. For example, if the subject is emmetropic and is not to accommodate, the target would be placed at the first focal point of lens  $L_2$ , which then images it at infinity, and this is its situation relative to the subject's eye. If, on the other hand, the subject is 1 diopter myopic (or is emmetropic and is to accommodate 1 diopter) the target's place with respect to lens  $L_2$  must be such that this lens images it at a distance of 1 m in front of the eye. Suppose the lens has focal length  $f_2$ . Let  $x$  be the distance of the object from the first focal point of lens  $L_2$ , and  $x'$  be the distance of the image from the second focal point of lens  $L_2$  (which in our present arrangement is the eye's entrance pupil); then by Newton's lens formula we have the relationship

$$xx' = -f_2^2 \quad \text{or} \quad 1/x' = -x/f_2^2$$

\*The author was introduced to this technique by W. A. H. Rushton, who had originally described in the literature (Rushton, 1956a) a related approach.



**FIGURE 2** More versatile arrangement for Maxwellian view. The source  $S$  is placed at the first principal focal plane of lens  $L_1$  and the target  $T$  is now trans-illuminated by a parallel beam. A laser may be substituted for  $S$  and  $L_1$ . For the image of the source to fall in the plane of the eye's entrance pupil, the latter should be in the second principal focal plane of lens  $L_2$ . If the observer is emmetropic, the target  $T$  should be placed at the first principal focal plane of lens  $L_2$ .

Now  $1/x'$  is the reciprocal of the image distance from the eye, i.e., it is the desired accommodative state of an emmetropic eye (or the far point of an unaccommodated ametropic eye) in diopters. It follows that a linear relationship exists between  $x$  the displacement of the target from the first primary focal point of lens  $L_2$ , and the target focus with respect to the eye in diopters. Each diopter of target focus corresponds to a distance  $x = f_2^2$  ( $x, f_2$  in meters). For example, if the lens  $L_2$  has a focal length of 0.20 m, each diopter of target focus as far as the eye is concerned is represented by 0.04 m displacement of the target from the first focal point of the lens  $L_2$ . The displacement is towards  $L_2$  if the subject is accommodating, or is myopic, and is away from  $L_2$  if he is hyperopic. The device clearly makes a good optometer, and was described in an arrangement for this purpose by Schmidt-Rimpler in 1877 (Emsley, 1939).

## Field of View

Another important variable in setting up a vision experiment, in addition to the level of retinal illuminance and the state of focus, which were discussed above, is the area of retina illuminated. We shall follow the approach adopted throughout this paper of specifying retinal distances in angular measure in the eye's object space. The angle should be measured at the center of the entrance pupil, which is about 3 mm behind the corneal vertex; the specification of retinal distance is then more invariant with such factors as accommodation and out-of-focusness than if another reference point, e.g. nodal point, is used. For a typical human eye, 1 min of arc in the object space corresponds to a retinal distance of just under 5  $\mu$ , so that there arc about  $3.4^\circ$  in each millimeter of retinal distance.

If the target is placed in the plane of lens  $L_2$  in Fig. 2, and if its diameter is  $c$ , its angular subtense is approximately  $c/f_2$  rad. Here,  $c$  and  $f_2$  should be measured in the same units. To convert to minutes of arc, it should be remembered that there are  $3.44 \times 10^3$  min of arc in each radian. In a hypothetical situation where  $f_2 = 20$  cm, and it is desired to present the eye with a stimulus 60 min of arc in diameter, its linear dimension must then be  $c = 20 \times 60/3.44 \times 10^3$  cm = 0.3 cm. Clearly, if small targets have to be presented to the eye, the focal length of the lens  $L_2$  should be long so that the linear size of the targets does not become too small. For example, if a lens of focal length 34.4 cm is chosen, each minute of arc of target size is 0.1 mm.

The above arguments hold identically if the target is not in the plane of  $L_2$  but is placed anywhere in the parallel beam between lenses  $L_1$  and  $L_2$ . The important property of the principle of placing the

second focal point of the lens  $L_2$  in the plane of eye's entrance pupil, the so-called telecentric principle, is now apparent: for purposes of changing focus (Section "Focus") the target may be moved in the space between the two lenses  $L_1$  and  $L_2$ , but such a move does not change either the retinal illuminance or the image size.

The maximum image size that can be achieved in the arrangement of Fig. 2 is one in which the apertures of lenses  $L_1$  and  $L_2$  remain unobscured. If the aperture of the smaller of the two lenses is  $c$ , the maximum field size is  $cf_2$  rad. If, for example, lens  $L_2$  is filled and has an aperture of 50 mm and focal length 20 cm, the largest field is  $5/20 = 1/4$  rad.  $\approx 14^\circ$  in diameter.

## Pupil Size

The effect of pupil size on the quantity of light entering the eye has already been discussed. The Stiles-Crawford effect has an influence, of course; whether it needs to be taken into consideration is a decision that depends on the particular application since, for example, the *S-C* effect does not show up in rod vision.

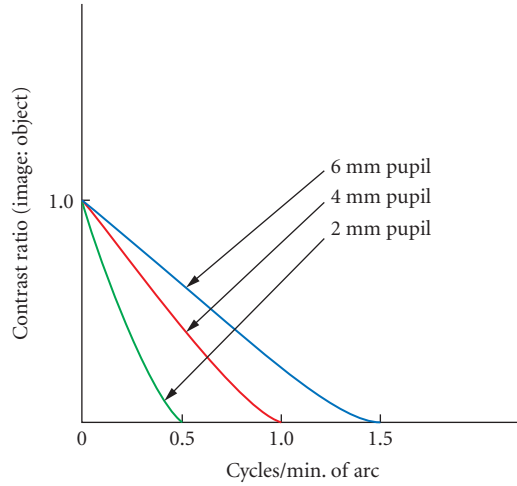
The pupil aperture also affects the quality of the retinal image. While in an important way everything to be said here will be modified by the optical aberrations of the particular eye to which it is applied, there still remains a substantial body of generally valid facts based on the diffraction theory of perfect optical instruments. Only under exceptional circumstances can an eye be expected to be as good as this, but our considerations allow us to set a bound to the best performance of any eye. In the present study no attempt will be made to go beyond this approach.

**Incoherent Target\*** It used to be taught that the resolution limit of an eye with an entrance pupil diameter of  $a$  mm for light of wave length  $\lambda$  is  $1.22 \lambda/a$  rad of visual angle in object space. This is at best an incomplete description of the situation, based on the fact that the diffraction-limited point-spread function, the Fraunhofer diffraction pattern, then has a radius (to its first zero) of  $1.22 \lambda/a$  rad of visual angle. A fuller analysis of the performance limitation of the eye is afforded by Fourier theory. The fundamental description of the eye's imagery, which the above statement expresses in the space domain (a point object is imaged in the Fraunhofer diffraction spread pattern), may also be phrased in terms of the spatial frequency domain. A grating with sinusoidal intensity pattern with a period  $\alpha$  min of arc visual angle may also be said to have repetition rate or spatial frequency of  $1/\alpha$  c/min of arc. In the spatial frequency domain, the way of describing the performance limitation of an optical system is to state the demodulation experienced by a sinusoidal grating target in the process of being imaged. When this quantity is plotted against spatial frequency we have what is now called the modulation transfer function. Such curves for a hypothetical aberration-free eye with various entrance pupils are shown in Fig. 3. Their main feature is a progressive reduction in the modulation of the image (when the object grating retains constant contrast) as the spatial frequency is increased up to the cut-off spatial frequency, beyond which there is no indication in the image that the object contains any features of those repetition frequencies.

The above statement is based on the acceptable premise that optical imagery obeys the principles of linearity. When the target is self-luminous, as it is assumed to be the case with ordinary viewing, this is indeed the case. The following is a summary in mathematical language of the basic concepts of image formation for an incoherent target.

Let the two dimensions in the eye's object space be  $\alpha, \beta$ ; distances involving  $\alpha$  refer to the horizontal direction, and  $\beta$  the vertical direction. Let distances in the two corresponding directions of the eye's entrance pupil be  $y_\alpha, y_\beta$ . Let spatial frequencies in these two directions be denoted by  $\omega_\alpha, \omega_\beta$ . For example a grating made up of vertical bars with period 10 min of arc has spatial frequency  $\omega_\alpha = 0.1$  c/min of arc.

\* An excellent introduction to the following discussions can be found in O'Neill (1963) or in Hopkins (1953).



**FIGURE 3** Contrast transfer functions (ratios of contrast in images and corresponding objects for sinusoidal light distributions) of optical system like that of the human eye but free from aberrations with round pupil of various diameters in monochromatic light of wavelength  $\lambda = 555 \text{ m}\mu$ .

Let the pupil aperture be described by the function  $g(y_\alpha, y_\beta)$ . For example, if the pupil is rectangular with a horizontal diameter of  $a$ , and vertical diameter  $b$ , and is quite unobstructed,

$$g(y_\alpha, y_\beta) = 1 \quad \begin{aligned} -\frac{a}{2} \leq y_\alpha \leq \frac{+a}{2} \\ -\frac{b}{2} \leq y_\beta \leq \frac{+b}{2} \end{aligned}$$

$$= 0 \quad \text{elsewhere.}$$

If it is round,

$$g(y_\alpha, y_\beta) = 1 \quad y_\alpha^2 + y_\beta^2 \leq \left(\frac{a}{2}\right)^2$$

$$= 0 \quad \text{elsewhere.}$$

The following theorems from diffraction theory are applicable:

1. Let the pupil function be expressed in terms of the variables  $\omega_\alpha = y_\alpha/\lambda$  and  $\omega_\beta = y_\beta/\lambda$  where  $\lambda$  is the wavelength of light. In this way we define the function  $S(\omega_\alpha, \omega_\beta) = g(y_\alpha, y_\beta)$ . It is the Fourier transform of  $\hat{s}(\alpha, \beta)$ , the amplitude distribution of the disturbance in the image of a point object, that is to say

$$\hat{s}(\alpha, \beta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(y_\alpha, y_\beta) e^{-2\pi i(y_\alpha \alpha + y_\beta \beta)} dy_\alpha dy_\beta \quad (4)$$

2. The intensity distribution in the image of a point object,  $s(\alpha, \beta)$  is the product of the amplitude distribution  $\hat{s}(\alpha, \beta)$  and its complex conjugate  $\hat{s}^*(\alpha, \beta)$ . Thus

$$s(\alpha, \beta) = \hat{s}(\alpha, \beta) \cdot \hat{s}^*(\alpha, \beta) \quad (5)$$

3. The Fourier transform of the intensity distribution in the image of a point source  $s(\alpha, \beta)$ , is the modulation transfer function,  $\tau(\omega_\alpha, \omega_\beta)$ . Thus

$$\tau(\omega_\alpha, \omega_\beta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} s(\alpha, \beta) e^{-2\pi i(\alpha\omega_\alpha + \beta\omega_\beta)} d\alpha d\beta \quad (6)$$

4. The modulation transfer function is the self-convolution of the pupil function

$$\tau(\omega_\alpha, \omega_\beta) = g(\gamma_\alpha, \gamma_\beta) \otimes g(\gamma_\alpha, \gamma_\beta) \quad (7)$$

In practice, the following steps can lead from the knowledge of the aperture function of an optical system to the detailed knowledge of the image intensity distribution:

1. The modulation transfer function is given by the self-convolution (auto-correlation) of the pupil aperture function, equation (7).
2. The product of the Fourier transform of the object intensity distribution with the modulation transfer function gives the Fourier transform of the image intensity distribution. In the space domain, the equivalent statement is that the image intensity distribution is the convolution of the object intensity distribution, with the intensity distribution in the image of a point source. These relationships, which are derived and discussed in considerable detail in the relevant texts (O'Neil, Born and Wolf) are summarized in Table 1.

The fundamental imaging property of an instrument like the eye is, therefore, contained in the modulation transfer function  $\tau(\omega_\alpha, \omega_\beta)$  in the Fourier domain, and in the image intensity distribution of a point object  $s(\alpha, \beta)$  in the space domain. This holds not only for purely diffraction-limited instruments but also for those afflicted with aberrations. In such cases the modulation transfer function will be lower (less transmission of spatial frequencies) and the point-spread function larger.

The application of this theory to imaging in the eye of incoherently illuminated objects has been discussed in a previous study (Westheimer, 1964).

**Coherent Illumination** The whole situation is radically altered if the target is transilluminated by light which has its origin in a single source, so that the radiation emitted from different portions of the target is capable of mutually interfering. This is realized in Maxwellian viewing when the source S

**TABLE 1** Incoherent Imagery

	Object Space $\alpha, \beta$ Visual Angle	Pupil Plane $\gamma_\alpha, \gamma_\beta$ Distance	Fourier Domain $\omega_\alpha, \omega_\beta$ (c/min of arc)	Image Space $\alpha, \beta$ (min of arc)
Object	$o(\alpha, \beta)$		$O(\omega_\alpha, \omega_\beta)$	
Pupil aperture function		$g(\gamma_\alpha, \gamma_\beta)$		
Amplitude function in image of point object			$\hat{S}(\omega_\alpha, \omega_\beta)$	$\hat{s}(\alpha, \beta)$
Transfer function			$\tau(\omega_\alpha, \omega_\beta) = g(\gamma_\alpha, \gamma_\beta) \otimes g(\gamma_\alpha, \gamma_\beta)$	
Intensity function in image of point object				$s(\alpha, \beta) = \hat{s}(\alpha, \beta) \cdot \hat{s}^*(\alpha, \beta)$
Image intensity function			$I(\omega_\alpha, \omega_\beta) = \tau(\omega_\alpha, \omega_\beta) \cdot O(\omega_\alpha, \omega_\beta)$	$i(\alpha, \beta) = o(\alpha, \beta) \hat{s}^*(\alpha, \beta)$

See footnote to Table 2.

in Fig. 1(a) is very small, and monochromatic. Then the light after being collimated by lens  $L_1$  may be regarded as perfectly coherent, and the following theory holds rigidly. (It also holds for targets illuminated by lasers which emit perfectly coherent light.<sup>†</sup>)

The following is a convenient way of visualizing the situation. Since the wavefront impinging on the target is plane, there will be formed in the plane of the eye's pupil the Fraunhofer diffraction pattern produced at the passage of the light through the target. The amplitude distribution in this diffraction image is the Fourier transform of the target amplitude function.

If  $\delta(\alpha, \beta)$  is the distribution of amplitude in the object, the distribution of amplitude in the disturbance in the plane of the pupil is given by

$$O(\omega_a, \omega_\beta) = \iint_{-\infty}^{+\infty} \delta(\omega_a, \omega_\beta) e^{-2\pi i(\omega_a \alpha + \omega_\beta \beta)} d\alpha d\beta \quad (8)$$

Naturally, only that portion of  $O(\omega_a, \omega_\beta)$  that is transmitted by the pupillary aperture can be effective in the formation of the image amplitude distribution  $i(\alpha, \beta)$  which is again the Fourier transform of the pupil amplitude function. Thus

$$i(\alpha, \beta) = \iint_{\text{pupil}} O(\omega_a, \omega_\beta) e^{-2\pi i(\omega_a \alpha + \omega_\beta \beta)} d\omega_a d\omega_\beta \quad (9)$$

Finally, the image intensity distribution  $i(\alpha, \beta)$  is the product of the image amplitude distribution with its complex conjugate:

$$i(\alpha, \beta) = i(\alpha, \beta) \cdot i^*(\alpha, \beta)$$

It is seen that the limitation to the fidelity of reproduction of the original target on the retina (aside from any aberrations that may be present) is the fact that the pupil aperture is necessarily finite, while for any finite target the diffraction image, or Fourier transform, that is formed in the pupillary plane will be infinite in extent. All portions of the diffraction image of the target that fall outside the pupil will be cut off. The diffraction limitation during Maxwellian viewing lies therefore in equation (9) which states that the range of integration in the Fourier re-synthesis of the image extends only over the pupil instead of going to infinity. Now there is a very simple relationship between the target's amplitude spatial frequencies  $(\omega_a, \omega_\beta)$ , and the corresponding positions in the diffraction image in the pupil. Each spatial frequency  $\omega$  is situated at a distance  $y = \lambda \omega$  from the geometrical image of the source in the pupil. The pupil aperture function  $g(y_\alpha, y_\beta)$ , written in terms of the variables  $\omega_\alpha = y_\alpha / \lambda$ ,  $\omega_\beta = y_\beta / \lambda$ , thus, gives us the amplitude transmission function  $\tau(\omega_a, \omega_\beta)$ . It is this function which, when multiplied with  $O(\omega_a, \omega_\beta)$ , the Fourier transform of the object amplitude function, gives  $\hat{I}(\omega_a, \omega_\beta)$  the Fourier transform of the image amplitude function. All this is also summarized in Table 2.

Two examples follow. Suppose that the disturbance issuing from the target plane has an amplitude distribution that is sinusoidal with spatial frequency  $\omega$ , i.e.

$$O(\alpha, \beta) = 1 + \cos(2\pi\alpha\omega)$$

The associated amplitude distribution in the plane of the pupil is three delta functions, i.e. there is a concentration of disturbance in 3 spots in the pupillary plane and nowhere else. The three are:

$$(1) \quad \omega_a = 0, \text{ i.e. } y_a = \omega_a \lambda = 0, \text{ i.e. at the center of the pupil and}$$

$$(2 \text{ and } 3) \quad \omega_a = \pm\omega, \text{ i.e. } y_a = \pm\omega\lambda$$

<sup>†</sup> A laser may be substituted for the source and lens  $L_1$ . It is to be noted that all the light from the laser is then focused in the eye's pupil. Obviously, precautions are necessary here, just as much as in using a laser and letting the eye's optical system focus its beam on the retina.

**TABLE 2** Incoherent Imagery

	Object Space $\alpha, \beta$ Visual Angle	Pupil Plane $\gamma_\alpha, \gamma_\beta$ Distance	Spatial Frequency Domain $\omega$ (c/min of arc)	Image Space
Object amplitude function	$o(\alpha, \beta)$		$O(\omega_\alpha, \omega_\beta)$	
Aperture function		$g(\gamma_\alpha, \gamma_\beta)$		
Complex amplitude transmission function			$\tau(\omega_\alpha, \omega_\beta)$	
Image amplitude function			$I(\omega_\alpha, \omega_\beta) = \tau(\omega_\alpha, \omega_\beta) \cdot O(\omega_\alpha, \omega_\beta)$	$\hat{i}(\alpha, \beta)$
Image intensity function				$i(\alpha, \beta) = i(\alpha, \beta) i^*(\alpha, \beta)$

The relationship between variables is as follows:

- Where quantities in 1st and 3rd columns are entered between a pair of horizontal lines, they are Fourier transforms:

$$o(\omega_\alpha, \omega_\beta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} O(\alpha, \beta) e^{-2\pi i(\alpha\omega_\alpha + \beta\omega_\beta)} d\alpha d\beta$$

where  $\alpha$ , and  $\beta$  are visual angles in minutes of arc,  $\omega_\alpha, \omega_\beta$  are spatial frequency in c/min of arc.

- Distances  $\gamma_\alpha, \gamma_\beta$  in plane of entrance pupil measured from its center are related to the spatial frequencies in the same orientation by the equations

$$\gamma_\alpha = \omega_\alpha \lambda \quad \gamma_\beta = \omega_\beta \lambda$$

$\lambda$  being the wavelength of light used. Functions in the second column (pupil space) are transferred directly to the third column, spatial frequency, by this change in variables.

The above example shows that the representation of a sinusoidal grating in the frequency domain consists of a d.c. term and a pair of a.c. terms which are physically placed in the pupillary plane in positions according to their spatial frequency. Frequencies higher than the one present at the edge of the pupil can clearly not be available to make up the image. If the pupil aperture has diameter  $a$ , this limit of sinusoidal components of amplitude is given by

$$\omega_\alpha = a/2\lambda$$

Another example: suppose the target is a trans-illuminated circular disk of diameter  $\alpha_0$  rad, centered on the optical system. This case has been dealt with in detail in a previous study (Westheimer, 1959). The amplitude distribution in the diffraction image in the pupil is here the familiar Airy disk, with the first zero at a distance from the center given by

$$y = 1.22 \lambda / \alpha_0$$

$y$  and  $\lambda$  being in the same units. For a disk 2.0 min of arc in diameter (0.00058 rad) and light of a representative wavelength,  $y = 1.2$  mm. Thus, if the pupil diameter is smaller than 2.4 mm (or if an artificial pupil of smaller diameter than 2.4 mm is placed in front of the eye) only the center portion of the diffraction image of the disk is admitted into the eye. This constitutes a severe restriction of the “bandwidth” of spatial frequencies making up the original disk. The apparent anomaly of the situation is that most of the amplitude of the disturbance (>90 per cent) from the point source passing through the target disk, is contained in the central part of the diffraction image. Yet much of the “information” about the target—whether or not it is a sharply bounded disk and the position and exact nature of its edge—is contained in the higher spatial frequencies. They are imaged in the periphery of the eye’s entrance pupil and could thus be screened out needlessly by an experimenter who, unaware of their information content, might stop them from entering the eye by placing too small an artificial pupil in front of an eye seeing the target in Maxwellian view.

It should be emphasized that the process of coherent imaging is linear in *amplitude* of the electromagnetic radiation, not in *intensity* as it is with incoherent targets. When a target is lit by coherent light, it is helpful to consider the amplitude of the disturbance emerging from the target plane as being focused in the pupil plane. The amplitude of the disturbance there, which is the Fourier



transform of that in the target plane, represents the spatial frequency content of the latter displayed at distances from the geometrical image of the source proportional to the spatial frequencies. Since the pupil necessarily cuts out all portions of the diffraction pattern further out than its radius, the disturbance entering the eye will now make up an image amplitude distribution on the retina that is restricted in bandwidth of its spatial frequencies. In addition, intraocular effects and aberrations may also superimpose inhomogeneities and phase shifts. Ultimately the amplitude distribution of the disturbance in the plane of the retina is subjected to a non-linear transformation, viz. squaring, and the resulting intensity distribution constitutes the retinal image. Once its origin is properly understood, the Stiles-Crawford effect can be allowed for in this series of transformations. The place of the non-linearity in coherent imagery—it occurs as the last step in the transformation—prevents the formulation of any general statements about the retinal image. All particular cases of target configurations, pupil openings (not to mention aberrations and defocusing) will give rise to their own kinds of imaging characteristics. For incoherent target illumination, the non-linearity occurs at a more convenient juncture in the sequence, for once the point-spread function, or in the Fourier domain, the modulation transfer function, has been fixed by the particulars of the pupil size, aberrations and focus errors, the image for any target can be readily determined.

When it is coherent, the amplitude of the electromagnetic disturbance emerging from the target plane is being spatially Fourier analyzed (displayed as a diffraction image in the plane of the pupil), filtered, and then resynthesized as a distribution of amplitude of electromagnetic disturbance in the retinal plane. Ultimately the amplitude is squared at each point in the image and this constitutes the image intensity pattern. It is seen that filtering the spatial frequency in the domain of the amplitude of the disturbance will lead to results which, when looked at in the realm of the intensity pattern, may be non-linear. As an example, consider the case of an amplitude distribution that is sinusoidal. When squared this becomes an intensity distribution that is also sinusoidal but has twice the frequency. Suppose that now a constant (d.c.) level is added in the realm of amplitude. The laws of linearity are observed, so that the new amplitude distribution is the sum of the a.c. and d.c. terms. But when this is squared to yield the intensity distribution, a result is obtained which is quite different from that obtained when the amplitude distribution due to the a.c. term alone is squared.

## Partial Coherence

The theory of coherent imagery (Beran and Parrant, 1964; Born and Wolf, 1959) developed above is applicable when the source  $S$  is essentially a monochromatic point source. (A laser in place of source  $S$  and lens  $L_1$  realizes this situation in practice.) It is then possible for light coming from a point on the target to interfere with all the light coming from all the other parts of the target. When the source is larger than that, a point on the target is transilluminated by light from a range of points on the source, and the capacity to subsequently interfere is limited. In the limit, when the source is quite extensive, the target may be regarded as incoherently illuminated. These effects are described by the theory of partial coherence, the chief theorem of which, the van Cittert–Zernike theorem, permits us to define the target region beyond which no significant degree of coherence exists.

Applied to the situation illustrated in Fig. 2, the van Cittert–Zernike theorem states that the degree of coherence in the target plane is distributed in the manner of the Fraunhofer pattern that would be produced in the hypothetical case due to diffraction at the aperture of the source  $S$ . Let the source be circular with linear diameter  $y_0$ , let the ratio of the focal lengths of the lens  $L_2$  to that of lens  $L_1$  (the magnification of the source in the plane of the pupil) be  $M$ . Then the separation of two points in the source for the first reduction of their mutual coherence to zero (i.e. the radius of the Fraunhofer diffraction pattern) expressed in terms of  $\alpha$ , the angular distance at the eye, is

$$\alpha = 1.22\lambda/My_0$$

It is seen that for  $\alpha$  to be, say, 1 min of arc,  $My_0$  should be about 2.2 mm, i.e. the geometrical image of the source in the plane of the pupil should be 2.2 mm. This is an outer limit, since two points on the source separated by, say, half the radius of the Airy disk due to the source, would have a mutual coherence that is already small.

It is thus possible to lay down the following guidelines on whether a particular experimental situation calls for the application of the theory of coherent or incoherent imagery.

When the source has an image in the plane of the pupil larger than, say, 2.5 mm in diameter, the target may be regarded as incoherently illuminated. An artificial pupil can with advantage be employed and the approach to the light distribution in the retinal image would be by way of the modulation transfer function of the eye.

On the other extreme, one may wish to utilize the characteristics of coherent imagery for a particular purpose, such as spatial filtering. How small does the geometrical image of the source have to be for a satisfactory degree of coherence to be achieved? We have mentioned that the laser will do this, but all the essential ingredients are also present in a tungsten filament or mercury or xenon arc, provided that the angular subtense of the source at lens  $L_1$  is small enough and the light has been restricted to a narrow enough wavelength band. If one wishes to ensure that the light coming from a 10 min of arc region of the visual field is at least, say, 90 per cent coherent, the geometrical image of the source in the pupil should be no larger than

$$My_0 = 0.32\lambda/2.9 \times 10^{-3} \quad \text{i.e.} \quad 0.06 \text{ mm}$$

When an interference filter is used to make the beam monochromatic, the resultant retinal illuminance cannot be very high. Nevertheless, a tungsten filament with a luminance of  $10^3$  cd/cm<sup>2</sup> can still give a few hundred trolands of quasi-monochromatic retinal illuminance with good coherence properties. With the use of a high-pressure mercury arc this can be increased by at least two orders of magnitude.

## References

- Aguilar, M. and Stiles, W. S. (1954). Saturation of the rod mechanism of the retina at high levels of stimulation. *Optica Acta* 1, 59–65.
- Beran, M. J. and Parrent, G. B. (1964). *Theory of Partial Coherence*. Prentice Hall, Englewood Cliffs, N.J.
- Born M. and Wolf, E. (1959). *Principles of Optics*. Pergamon Press, Oxford.
- Emsley, H. H. (1939). *Visual Optics*. 2nd Edition, Hatten Press, London.
- Hopkins, H. H. (1953). On the diffraction theory of optical images. *Proc. R. Soc.* A217, 408–432.
- Le Grand, Y. (1948). *Optrique Physiologique*. Tome 2ieme, Ed. Rev, d'Optique.
- Maxwell, J. C. (1860)—On the theory of compound colours, and the relations of the colours of the spectrum. *Phil. Trans. R. Soc.* 150, 57.
- O'Neill, E. L. (1963). *Introduction to Statistical Optics*. Addison-Wesley, Reading, Mass.
- Rushton, W. A. H. (1956a). The difference spectrum and the photo-sensitivity of the rhodopsin in the living eye. *J. Physiol.* 134, 11–29.
- Rushton, W. A. H. (1956b). The rhodopsin density in the human rods. *J. Physiol.* 134, 30–46.
- Walsh, J. W. T. (1958). *Photometry*. 3rd Edition, Constable, London.
- Westheimer, G. (1959). Retinal light distributions for circular apertures in Maxwellian View, *J. Opt. Soc. Am.* 49, 41.
- Westheimer, G. (1960). Modulation thresholds for sinusoidal light distribution on the retina. *J. Physiol.* 152, 67–74.
- Westheimer, G. (1964). Pupil size and visual resolution. *Vision Res.* 4, 39–45.

## 6.3 POSTSCRIPT (2008)

These supplementary remarks are informed by developments in the two intersecting topics which this article, written 45 years ago, covers: physical optics and dioptrics of the eye.

The Maxwellian view is a nontraditional way of generating the retinal image. In ordinary viewing, the energy over the whole bundle of rays—or the wavefront—reaching the entrance pupil is uniform.

The recent great advances in adaptive optics, that is, the modifying of the wavefront for purposes of neutralizing deviations caused by the eye's dioptric defects, deal only with the phase, not amplitude (Chap. 15 in this volume, and Chap. 5, "Adaptive Optics," by Robert Q. Fugate in Vol. V).

But it is precisely the amplitude of the electromagnetic disturbance across the entering wavefront that is in play in Maxwellian viewing. A simple way of formulating the eye's imaging is the statement that Fourier transformation relates the complex amplitude distributions of the electromagnetic disturbance in the planes of the exit pupil and the retinal image. The image intensity distribution is then the product of the amplitude with its complex conjugate—in practice squaring, once phase effects have been factored in. Hence in principle one can configure a desired retinal image by manipulating the incoming wavefront as it reaches the eye. In the most elementary case of Maxwellian viewing, the disturbance in the pupil plane is merely the diffraction pattern generated when a coherent beam passes through a transparent target comes to a focus there. But spatial amplitude and phase distributions can be created by other means. Limiting it to two points by a double prisms, for example, is a time-honored manner of producing interference fringes on the retina. If, a technology were available that allows both amplitude and phase modulation of a uniform plane wavefront incident on the eye—instead of only the phase as currently implemented in "adaptive optics"—an unprecedented control of the retinal image would be achieved in the absence of any physical objects. A start of this has already been made in attempts to measure the performance of the retinal mosaic by not just interference fringes but also in two-point separation acuity.<sup>1</sup>

The original paper discussed the always difficult problem of partial coherence. This is only appropriate, if for no other reason than that resolution measurements as a function of the degree of coherence had been performed in the first instance by visual observation<sup>2</sup>. In the early days of this research, generating a highly coherent light signal was quite difficult. It involved strictly monochromatic light, usually produced by mercury arcs limited to a single spectral line by narrow-band interference filters, focused on a pin-hole aperture only a fraction of a millimeter in diameter. The possibility of coherence being only partial needed to be kept in mind, particularly since its mathematical handling involves delicate problems.

All this changed radically with the advent of the laser. This advance in technology took place in the interval between two papers devoted to measuring retinal resolution bypassing the eye's optics and illustrates how it allowed researchers to concentrate on the biology rather than the physics of the process.<sup>3,4</sup> As the result of the easy access to fully coherent beams, partial coherence is now of more theoretical than practical interest.

Utilizing lasers and sources other than incandescent ones has opened up a question that at one time occupied center stage in vision research: photon statistics at the absolute visual threshold. Hecht, Schlaer and Pirenne's epochal work,<sup>5</sup> done while the quantum theory of light was still novel, pointed to the physical basis of threshold fluctuation through the contention that a Poisson distribution was the governing rule when the number of absorbed photons was only a handful. Since then the subject of photon statistics has been opened more widely with the utilization of photon emission both from lasers<sup>6,7</sup> and from oscilloscope screens,<sup>8</sup> each of which are used in vision research.

In the process of estimating the retinal image by way of Fourier transformation of the complex amplitude distribution of the electromagnetic disturbance in the plane of the entrance pupil, the question cannot be skirted of how to factor in the Stiles-Crawford effect<sup>9</sup> (see Chaps. 8 and 9). The original measurement involved the reduction in luminous efficiency for a narrow bundle as it was moved from the center to the edge of the pupil. The most trivial explanation of absorption in the eye media was quickly eliminated by the fact that the phenomenon differs between rod and cone vision.

The problem arises when computing the light distribution in the image formed by the eye's optics as it is in turn is funneled into the receptors to form the immediate stimulus for the visual process leading to photoisomerizations of photopigment molecules in rods and cones. Most calculations make no particular assumptions about the uptake of radiation in retinal receptors, yet their properties in channeling and accepting radiation cannot be ignored, especially if that involves conditioning the impinging wavefront. There are two possible interpretations of the role of the Stiles-Crawford effect at the level just prior to the phototransduction:

- a. The effect is an apodization phenomenon, where the entering wavefront near the edge of the pupil somehow is diminished in amplitude, perhaps as a result of waveguide properties of receptors (Chap. 9) or acceptance-lobe limitations at the molecular absorption levels. In such a case, the

calculation from pupil aperture to image distributions would have to fold in amplitude variations across the wavefront; in other words, the amplitude of the wavefront is in effect diminished toward the pupil periphery as the square root of the measured Stiles-Crawford effect, the latter measurements having been obtained in the intensity, not the amplitude domain of the electromagnetic disturbance.

- b. The effect is a diminution of the Poynting vector of the electromagnetic disturbance as a function of its direction with respect to the orientation of the receptor cells. This would be implemented by physical light screening at the retina, for example, inter-receptor absorbing pigments.

The case for (a) has been made well enough by arguments in favor of retinal receptors acting as waveguides (Chap. 8) and resultant improvement in image quality through reducing the weight of the wavefront in the pupil periphery more prone to be afflicted by ocular aberrations. It is strengthened by the fact that the Stiles-Crawford effect is much more prominent in cone vision where light level is not an issue but image quality is, than in rod vision.<sup>10</sup>

But the alternative (b) cannot yet be dismissed. There is evidence that the photopigment pools reached by light entering at opposite edges of the pupil are not identical.<sup>11</sup> The situation has been laid out in greater detail elsewhere.<sup>12,13</sup>

An effective retinal image, that is, a pattern of intra-receptor light absorption, based on proposition (a) will differ from one based on (b) and would lead to a rod-cone difference in spatial visual stimulus patterns.

## References

1. J. Liang and G. Westheimer, "Method for Measuring Visual Resolution at the Retinal Level," *J. Opt. Soc. Am. A* **10**(8):1691–1696 (1993).
2. A. Arnulf, O. Dupuy, and F. Flamant, "Étude expérimentale de la variation de la limite de résolution en fonction de la cohérence," *Revue d'Optique* **32**:529–552 (1953).
3. G. Westheimer, "Modulation Thresholds for Sinusoidal Light Distributions on the Retina," *J. Physiol. (London)* **152**(1):67–74 (1960).
4. F. W. Campbell and D. G. Green, "Optical and Retinal Factors Affecting Visual Resolution," *J. Physiol.* **181**:576–593 (1965).
5. S. Hecht, S. Shlaer, and M. H. Pirenne, "Energy, Quanta and Vision," *J. Physiol.* **25**:819–840 (1942).
6. E. Wolf and C. L. Mehta, "Determination of the Statistical Properties of Light from Photoelectric Measurements," *Physical Rev. Letters* **13**:1–3 (1964).
7. R. J. Glauber, "Optical Coherence and Photon Statistics," *Quantum Optics and Electronics—Les Houches Lectures 1964*, C. DeWitt, (ed.), Gordon & Breach, New York, 1965.
8. M. C. Teich, P. R. Prucnal, G. Vannucci, M. E. Breton, and W. J. McGill, "Multiplication Noise in the Human Visual System at Threshold. 3. The role of non-Poisson quantum fluctuations," *Biol. Cybernetics* **44**:157–165 (1982).
9. W. S. Stiles and B. H. Crawford, "The Luminous Efficiency of Rays Entering the Eye Pupil at Different Points," *Proc. Roy. Soc. (London)* **B 112**:428–450 (1933).
10. W. S. Stiles, "The Directional Sensitivity of the Retina and the Spectral Sensitivities of the Rods and Cones," *Proc. Roy. Soc. (London)* **B 127**:64–105 (1939).
11. W. Makous, "A Transient Stiles-Crawford Effect," *Vis. Res.* **8**:1271–1284 (1968).
12. G. Westheimer, "Specifying and Controlling the Optical Image on the Retina," *Progress in Retinal and Eye Research* **25**:19–42 (2006).
13. G. Westheimer, "Directional Sensitivity of the Retina: Seventy-Five Years of Stiles-Crawford Effect," *Proc. Roy. Soc. (London)* **B 275**:2777–2789 (2008).

*This page intentionally left blank*

---

# OCULAR RADIATION HAZARDS\*

---

David H. Sliney

*Consulting Medical Physicist*

*Fallston, Maryland*

*Retired, U.S. Army Center for Health Promotion and Preventive Medicine*

*Laser/Optical Radiation Program*

*Aberdeen Proving Ground, Maryland*

---

## 7.1 GLOSSARY

---

**Bunsen-Roscoe law of photochemistry.** The reciprocal relation between irradiance (dose-rate) and exposure duration to produce a constant radiant exposure (dose) to produce an effect.

**Coroneo effect.** See *Ophthalmoheliosis*.

**Erythema.** Skin reddening (acute damage) such as that caused by UV radiation; sunburn.

**Macula lutea.** Yellowed-pigment central region (macular area) of the retina of the human eye.

**Maxwellian view.** A method for directing a convergent beam into the eye's pupil to illuminate a large retinal area.

**Ophthalmoheliosis.** The concentration of UV radiation upon the nasal region of the cornea (the limbus) and nasal equatorial region of the lens, thus contributing to the development of pterygium on the cornea and cataract in the lens beginning in these regions.

**Photokeratitis.** A painful inflammation of the cornea of the eye.

**Photoretinitis.** Retinal injury due to viewing the sun or other high-intensity light source arising from photochemical damage due to short, visible wavelengths (also termed *photic maculopathy*).

**Pterygium.** A fleshy growth that invades the cornea (the clear front “window” of the eye). It is an abnormal process in which the conjunctiva (a membrane that covers the white of the eye) grows into the cornea. A pterygium may be small or grow large enough to interfere with vision and commonly occurs on the inner (nasal) corner of the eye.

**Solar retinopathy.** Retinal injury caused from staring at the sun—photochemical damage produced by short, visible (blue) light (also referred to as *solar maculopathy*, since it normally occurs in the macula lutea).

---

\*Note: The opinions or assertions herein are those of the author and should not be construed as official policies of the U.S. Department of the Army or Department of Defense.

## 7.2 INTRODUCTION

Optical radiation hazards to the eye vary greatly with wavelength and also depend upon the ocular exposure duration. Depending upon the spectral region the cornea, conjunctiva, lens, and/or retina may be at risk when exposed to intense light sources such as lasers and arc sources and even the sun. Photochemical damage mechanisms dominate in the ultraviolet (UV) end of the spectrum, and thermal damage mechanisms dominate at longer wavelengths in the visible and infrared (IR) spectral regions. Natural aversion responses to very bright light sources limit ocular exposure, as when one glances at the sun, but artificial sources may be brighter than the solar disk (e.g., lasers) or have different spectral or geometrical characteristics that overcome natural avoidance mechanisms, and the eye can suffer injury. Guidelines for the safe exposure of the eye from conventional light sources as well as lasers have evolved in recent decades. Because of special concerns about laser hazards, safety standards and regulations for lasers exist worldwide. However, the guidelines for exposure to conventional light sources are more complex because of the likelihood of encountering more than one type of ocular hazard from the broader emission spectrum.

The adverse effects associated with viewing lasers or other bright light sources such as the sun, arc lamps, and welding arcs have been studied for decades.<sup>1-4</sup> During the last three decades, guidelines for limiting exposure to protect the eye have evolved.<sup>4-12</sup> The guidelines were fostered to a large extent by the growing use of lasers and the quickly recognized hazard posed by viewing laser sources. Injury thresholds for acute injury in experimental animals for corneal, lenticular, and retinal effects have been corroborated for the human eye from accident data. Exposure limits are based upon this knowledge.<sup>3</sup> These safe exposure criteria were based not only upon studies aimed at determining thresholds in different ocular structures, but also upon studies of injury mechanisms and how the different injury mechanisms scaled with wavelength, exposure duration, and the area of irradiated tissue. The exposure guidelines also had to deal with competing damage to different ocular structures such as the cornea, lens, and retina, and how these varied with wavelength.<sup>3-11</sup> Although laser guidelines could be simplified to consider only the type of damage dominating for a single wavelength and exposure duration, the guidelines for incoherent, broadband sources were necessarily more complex.<sup>11</sup> To understand the guidelines for safe exposure, it is first necessary to consider the principal mechanisms of injury and which ocular structures may be injured under different exposure conditions.

## 7.3 INJURY MECHANISMS

For most exposure durations ( $t > 1$  s), optical radiation injury to ocular structures is dominated by either photochemically or thermally initiated events taking place during or immediately following the absorption of radiant energy. At shorter durations, nonlinear optical interaction mechanisms may play a role.<sup>12</sup> Following the initial insult; biological repair responses may play a significant role in determining the final consequences of the event. While inflammatory, repair responses are intended to reduce the sequelae; in some instances, this response could result in events such as scarring, which could have an adverse impact upon biological function.<sup>3</sup>

### Action Spectra

Photochemical and thermal effects scale differently with wavelength, exposure duration, and irradiated spot size. Indeed, experimental biological studies in humans and animals make use of these different scaling relationships to distinguish which mechanisms are playing a role in observed injury. Photochemical injury is highly wavelength dependent; thermal injury depends upon exposure duration.

### Exposure Duration and Reciprocity

The Bunsen-Roscoe law of photochemistry describes the *reciprocity* of exposure rate and duration of exposure, which applies to any photochemical event. The product of the dose-rate (in watts per square centimeter) and the exposure duration (in seconds) is the exposure dose (in joules

per square centimeter at the site of absorption) that may be used to express an injury threshold. Radiometrically, the irradiance  $E$  in watts per square centimeter, multiplied by the exposure duration  $t$ , is the radiant exposure  $H$  in joules per square centimeter, as shown in Eq. (1).

$$H = E \cdot t \quad (1)$$

This reciprocity helps to distinguish photochemical injury mechanisms from thermal injury (burns) where heat conduction requires a very intense exposure within seconds to cause photocoagulation; otherwise, surrounding tissue conducts the heat away from the absorption site (e.g., from a retinal image). Biological repair mechanisms and biochemical changes over long periods and photon saturation for extremely short periods will lead to reciprocity failure.

Thermal injury is a *rate process* that is dependent upon the time-temperature history resulting from the volumic absorption of energy across the spectrum. The thermochemical reactions that produce coagulation of proteins and cell death require critical temperatures for detectable biological injury. The critical temperature for an injury gradually decreases with the lengthening of exposure duration. This approximate decrease in temperature varies over many orders of magnitude in time and decreases approximately as a function of the exposure duration  $t$  raised to the  $-0.25$  power [i.e.,  $\{f(t^{-0.25})\}$ ].

As with any photochemical reaction, the *action spectrum* should be known.<sup>13</sup> The action spectrum describes the relative effectiveness of different wavelengths in causing a photobiological effect. For most photobiological effects (whether beneficial or adverse), the full width at half-maximum is less than 100 nm and a long-wavelength cutoff exists where photon energy is insufficient to produce the effect. This is not at all characteristic of thermal effects; in which the effect occurs over a wide range of wavelengths where optical penetration and tissue absorption occur. For example, significant radiant energy can penetrate the ocular media and be absorbed in the retina in the spectral region between 400 and nearly 1400 nm; the absorbed energy can produce retinal thermal injury.

Although the action spectra for acute photochemical effects upon the cornea,<sup>3,14-16</sup> upon the lens<sup>3,16</sup> and upon the retina<sup>3,17</sup> have been published, the action spectra of some other effects appear to be quite imprecise or even unknown.<sup>14</sup> The action spectrum for neuroendocrine effects mediated by the eye is still only approximately known. This point to the need to specify the spectrum of the light source of interest as well as the irradiance levels if one is to compare experimental results from different experimental studies.

Although both  $E$  and  $H$  may be defined over the entire optical spectrum, it is necessary to employ an action spectrum for photochemical effects. The International Commission on Illumination (CIE) photopic  $V(\lambda)$ , UV hazard  $S(\lambda)$ , and blue-light hazard  $B(\lambda)$  curves of Fig. 1 are examples of action spectra that may be used to spectrally weight the incident light. With modern computer spreadsheet programs, one can readily spectrally weight a lamp's spectrum by a large variety of photochemical action spectra. These computations all take the following form:

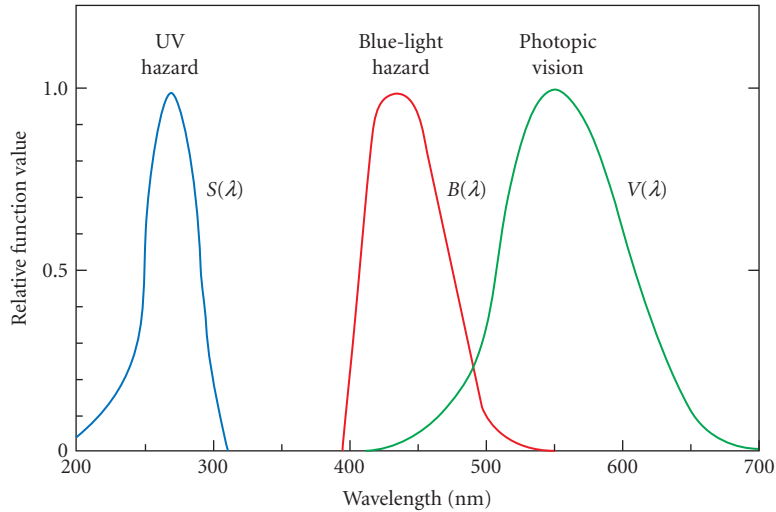
$$E_{\text{eff}} = \Sigma E_{\lambda} \cdot A(\lambda) \cdot \Delta(\lambda) \quad (2)$$

where  $A(\lambda)$  may be any action spectrum (unitless) of interest. One can then compare different sources to determine the relative effectiveness of the same irradiance from several lamps for a given action spectrum.

## 7.4 TYPES OF INJURY

The human eye is actually quite well-adapted to protect itself against the potential hazards from optical radiation (UV, visible, and IR radiant energy) from most environmental exposures encountered from sunlight. However, if ground reflections are unusually high, as when snow is on the ground, reflected UV radiation may produce "snow blindness," that is, UV photokeratitis. Another example occurs during a solar eclipse, if one stares at the solar disc for more than 1 to 2 min without eye protection, the result may be an "eclipse burn" of the retina.<sup>3,17</sup> Lasers may injure the eye, and under unusual situations other artificial light sources, such as mercury-quartz-discharge lamps or arc lamps, may pose a potential ocular hazard. This is particularly true when the normal defense





**FIGURE 1** Action spectra. The ACGIH UV hazard function  $S(\lambda)$  describes approximately the relative spectral risk for photokeratitis, and the blue-light hazard function  $B(\lambda)$  describes the spectral risk for photoreinitis. For comparison, the CIE spectral sensitivity (standard observer) function  $V(\lambda)$  for the human eye is also provided.

mechanisms of squinting, blinking, and aversion to bright light are overcome. Ophthalmic, clinical exposures may increase this risk if the pupil is dilated or the head is stabilized.

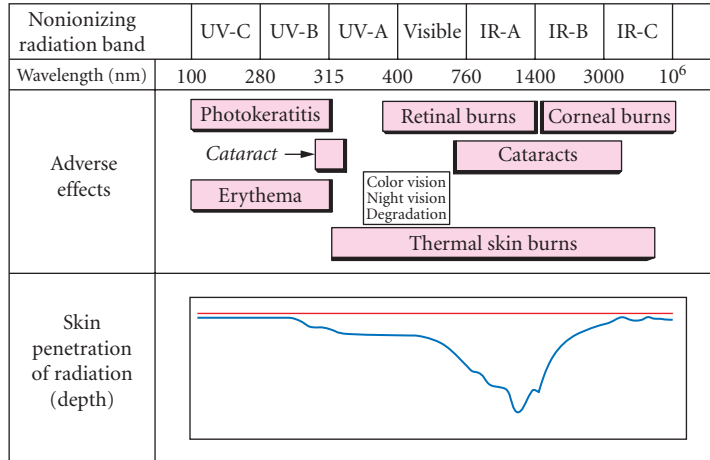
There are at least six separate types of hazards to the eye from lasers and other intense optical sources, and these are shown in Fig. 2 (along with the spectral range of responsible radiation):<sup>3</sup>

1. Ultraviolet photochemical injury to the cornea (photokeratitis); also known as “welder’s flash” or “snow blindness” (180 to 400 nm)<sup>3,15,16</sup>
2. Ultraviolet photochemical injury to the crystalline lens (cataract) of the eye (295 to 325 nm, and perhaps to 400 nm)<sup>16</sup>
3. Blue-light photochemical injury to the retina of the eye (principally, 400 to 550 nm; unless aphakic, 310–550 nm)<sup>3,4,17</sup>
4. Thermal injury (photocoagulation) to the retina of the eye (400 to nearly 1400 nm) and thermoacoustic injury at very short laser exposure durations<sup>3,4,12,18</sup>
5. Near-IR thermal hazards to the lens (~800–3000 nm)<sup>3,19</sup>
6. Thermal injury (burns) of the cornea of the eye (~1300 nm to 1 mm)<sup>3</sup>

The potential hazards to the eye and skin are illustrated in Fig. 2, which shows how effects are generally dominant in specific CIE photobiological spectral bands, UV-A, -B, and -C and IR-A, -B, and -C.<sup>20</sup> Although these photobiological bands are useful shorthand notations, they do not define fine lines between no effect and an effect in accordance with changing wavelengths.

## Photokeratitis

This painful but transient (1 to 2 d) photochemical effect normally occurs only over snow or open arc. The relative position of the light source and the degree of lid closure can greatly affect the proper calculation of this UV exposure dose. For assessing risk of photochemical injury, the spectral distribution of the light source is of great importance.



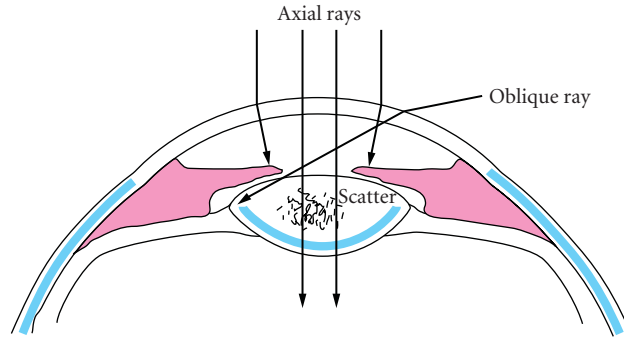
**FIGURE 2** The separate types of hazards to the eye from lasers and other intense optical sources, along with the spectral range of responsible radiation.<sup>3</sup>

### Cataract

Ultraviolet cataract can be produced from exposure to UV radiation (UVR) of wavelengths between 295 and 325 nm (and even to 400 nm). Action spectra can only be obtained from animal studies, which have shown that anterior, cortical, and posterior subcapsular cataract can be produced by intense exposure delivered over a period of days. Human cortical cataract has been linked to chronic, lifelong UV-B radiation exposure. Although the animal studies and some epidemiological studies suggest that it is primarily UV-B radiation in sunlight, and not UV-A, that is most injurious to the lens, biochemical studies suggest that UV-A radiation may also contribute to accelerated aging of the lens. It should be noted that the wavelengths between 295 and 325 nm are also in the wavelength region where solar UVR increases significantly with ozone depletion. Because there is an earlier age of onset of cataract in equatorial zones, UVR exposure has frequently been one of the most appealing of a number of theories to explain this latitudinal dependence. The UVR guideline, established by the American Conference of Governmental Industrial Hygienists (ACGIH) and the International Commission on Non-Ionizing Radiation Protection (ICNIRP) is also intended to protect against cataract.

Despite the collection of animal and epidemiological evidence that exposure of the human eye to UVR plays an etiological role in the development of cataract, this role in cataractogenesis continues to be questioned by others. Indeed, more recent studies would support a stronger role for ambient temperature as the primary environmental etiological factor in one type of cataract—nuclear cataract.<sup>19</sup>

Because the only direct pathway of UVR to the inferior germinative area of the lens is from the extreme temporal direction, it has been speculated that side exposure is particularly hazardous. For any greatly delayed health effect, such as cataract or retinal degeneration, it is critical to determine the actual dose distribution at critical tissue locations. A factor of great practical importance is the actual UVR that reaches the germinative layers of any tissue structure. In the case of the lens, the germinative layer where lens fiber cell nuclei are located is of great importance. The DNA in these cells is normally well-shielded by the parasol effect of the irises. However, Coroneo<sup>21</sup> has suggested that the focusing of very peripheral rays by the temporal edge of the cornea, those which do not even reach the retina, can enter the pupil and reach the equatorial region as shown in Fig. 3. He terms this effect, which can also produce a concentration of UVR at the nasal side of the limbus and lens, “ophthalmoheliosis.” He also noted the more frequent onset of cataract in the nasal quadrant of the lens and the formation of pterygium in the nasal region of the cornea. Figure 4 shows the percentage of cortical cataract that

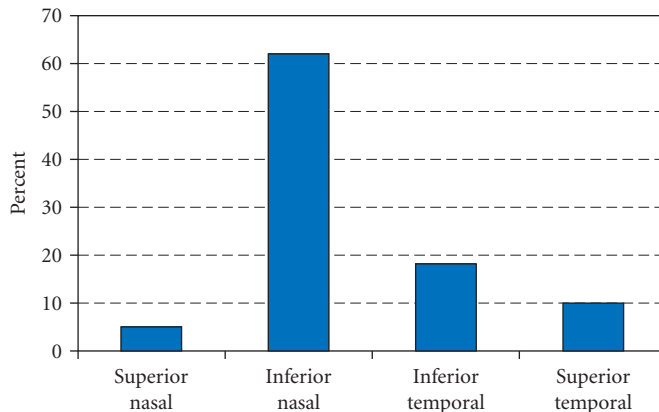


**FIGURE 3** The Coroneo effect. Very oblique, temporal rays can be focused near the equator of the lens in the inferior nasal quadrant.

actually first appears in each quadrant of the lens from the data of Barbara Klein and her colleagues in their study of a population in the U.S. Midwest, in Beaver Dam, Wisconsin.<sup>22</sup> This relationship is highly consistent with the Coroneo hypothesis.

### Pterygium and Droplet Keratopathies

The possible role of UVR in the etiology of a number of age-related ocular diseases has been the subject of many medical and scientific papers. However, there is still a debate as to the validity of these arguments. Although photokeratitis is unquestionably caused by UVR reflection from the snow,<sup>3,23,24</sup> pterygium and droplet keratopathies are less clearly related to UVR exposure.<sup>24</sup> Pterygium, a fatty growth over the conjunctiva that may extend over the cornea, is most common in ocean island residents (where both UVR and wind exposure is prevalent). Ultraviolet radiation



**FIGURE 4** Distribution of cortical cataract by segment. The percentage of cortical cataract that actually first appears in each quadrant of the lens.<sup>22</sup>

is a likely etiological factor,<sup>24,25</sup> and the Coroneo effect may also play a role.<sup>21,25</sup> To better answer these epidemiological questions, far better ocular dosimetry is required. Epidemiological studies can arrive at erroneous conclusions if assignments of exposure are seriously in error, and assumptions regarding relative exposures have been argued to be incorrect.<sup>25</sup> Before one can improve on current epidemiological studies of cataract (or determine the most effective UVR protective measures), it is necessary to characterize the actual solar UVR exposure to the eye.

## Photoretinitis

The principal retinal hazard resulting from viewing bright continuous-wave (CW) light sources is photoretinitis [e.g., *solar retinitis* with an accompanying scotoma (“blind spot”), which results from staring at the sun]. Solar retinitis was once referred to as “eclipse blindness” and associated “retinal bum.” At one time, solar retinitis was thought to be a thermal injury, but it has been since shown conclusively (1976) to result from a photochemical injury related to exposure of the retina to shorter wavelengths in the visible spectrum (i.e., violet and blue light).<sup>3,17</sup> For this reason, it has frequently been referred to as the “blue-light” hazard. The action spectrum for photoretinitis peaks at about 445 nm in the normal phakic eye; however, if the crystalline lens has been removed (as in cataract surgery), the action spectrum continues to increase for shorter wavelengths into the UV spectrum, until the cornea blocks shorter wavelengths, and the new peak shifts down to nearly 305 nm.

As a consequence of the Bunsen-Roscoe law of photochemistry, blue-light retinal injury (photoretinitis) can result from viewing either an extremely bright light for a short time or a less bright light for longer periods. The approximate retinal threshold for a nearly monochromatic source at 440 nm is approximately 22 J/cm<sup>2</sup>, hence a retinal irradiance of 2.2 W/cm<sup>2</sup> delivered in 10 s, or 0.022 W/cm<sup>2</sup> delivered in 1000 s will result in the same threshold retinal lesion. Eye movements will reduce the hazard, and this is particularly important for sources subtending an angle of less than 11 milliradians (mrad), because saccadic motion even during fixation is of the order of 11 mrad.<sup>27</sup>

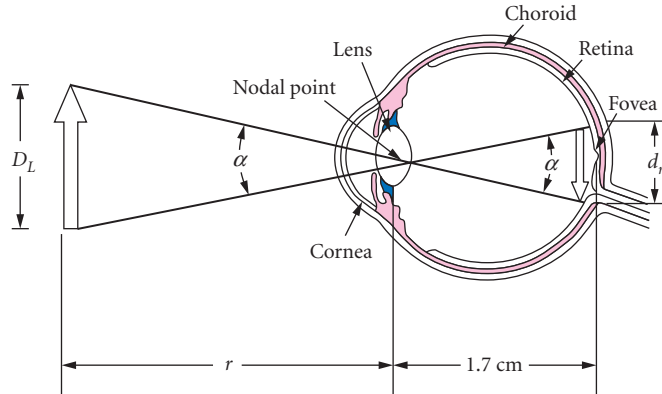
## Infrared Cataract

Infrared cataract in industrial workers appears only after lifetime exposures of the order of 80 to 150 mW/cm<sup>2</sup>. Although thermal cataracts were observed in glassblowers, steelworkers, and others in metal industries at the turn of the century, they are rarely seen today. These aforementioned irradiances are almost never exceeded in modern industry, where workers will be limited in exposure duration to brief periods above 10 mW/cm<sup>2</sup>, or they will be (should be) wearing eye protectors. Good compliance in wearing eye protection occurs at higher irradiances if only because the worker desires comfort of the face.<sup>3</sup>

## 7.5 RETINAL IRRADIANCE CALCULATIONS

Retinal irradiance (exposure rate) is directly related to source radiance (brightness). It is not readily related to corneal irradiance.<sup>3</sup> Equation (3) gives the general relation, where  $E_r$  is the retinal irradiance in watts per square centimeter,  $L_s$  is the source radiance in watts per square centimeter per steradian (sr),  $f$  is the effective focal length of the eye in centimeters,  $d_e$  is the pupil diameter in centimeters, and  $t$  is the transmittance of the ocular media:

$$E_r = \frac{(\pi \cdot L_s \cdot \tau \cdot d_e^2)}{(4f^2)} \quad (3)$$



**FIGURE 5** Retinal image size related to source size. The retinal image size can be calculated based upon the equal angular subtense of the source and the retinal image at the eye's nodal point (17 mm in front of the retina).

Equation (3) is derived by considering the equal angular subtense of the source and the retinal image at the eye's nodal point (Fig. 5). The detailed derivation of this equation is given elsewhere<sup>3</sup> and in Chap. 37 "Radiometry and Photometry for Vision Optics," by Yoshi Ohno in Vol. II. The transmittance  $t$  of the ocular media in the visible spectrum for younger humans (and most animals) is as high as 0.9 (i.e., 90 percent).<sup>3</sup> If one uses the effective focal length  $f$  of the adult human eye (Gullstrand eye), where  $f = 1.7$  cm, one has

$$E_r = 0.27 \cdot L_s \cdot \tau \cdot d_e^2 \quad (4)$$

All of the preceding equations assume that the iris is pigmented, and the pupil acts as a true aperture. In albino individuals, the iris is not very effective, and some scattered light reaches the retina. Nevertheless, imaging of a light source still occurs, and Eq. (4) is still valid if the contribution of scattered light (which falls over the entire retina) is added.

## 7.6 EXAMPLES

As an example, a typical cool-white fluorescent lamp has an illumination of 200 lx, and the total irradiance is 0.7 mW/cm<sup>2</sup>. By spectral weighting, the effective blue-light irradiance is found to be 0.15 mW/cm<sup>2</sup>. From measurement of the solid angle subtended by the source at the measurement distance,<sup>3,11</sup> the blue-light radiance is 0.6 mW/(cm<sup>2</sup>·sr) and the radiance is 2.5 mW/(cm<sup>2</sup> · sr). When fluorescent lamps are viewed through a plastic diffuser, the luminance and blue-light radiance are reduced.

As another example, a 1000-watt tungsten halogen bulb has a greater radiance, but the percentage of blue light is far less. A typical blue-light radiance is 0.95 W/(cm<sup>2</sup> · sr) compared with a total radiance of 58 W/(cm<sup>2</sup> · sr). The luminance is 2600 candelas (cd)/cm<sup>2</sup>—a factor of 3000 times brighter than the cool-white fluorescent. However, because the retinal image is small, eye movements spread the exposure over a much larger area.

Once the source radiance,  $L$  or  $L_{bp}$ , is known, the retinal irradiance  $E_r$  is calculated by Eq. (4). The preceding examples illustrate the importance of considering the size of a light source and the impact of eye movements in any calculation of retinal exposure dose. If one were exposed to a focal beam

of light (e.g., from a laser or LED) that was brought to focus in the anterior chamber (the aqueous humor), the pupil plane, or the lens, the light beam would diverge past this focal point and could be incident upon the retina as a relatively large image. This type of retinal illumination is frequently referred to as *Maxwellian view* and does not occur in nature. The retinal irradiance calculation in this case would be determined by the depth of the focal spot in the eye; the closer to the retina, the smaller the retinal image and the greater the irradiance.

Because the iris alters its diameter (and pupil center) dynamically, one must be alert to vignetting. The pupil aperture also diminishes with age. Near-sighted, or myopic, individuals tend to have larger pupils, and far-sighted, or hyperopic, individuals tend toward smaller pupils. Accommodation also results in a decrease in pupil diameter.

## 7.7 EXPOSURE LIMITS

A number of national and international groups have recommended occupational or public exposure limits (ELs) for optical radiation (i.e., UV, light, and IR radiant energy). Although most such groups have recommended ELs for UV and laser radiation, only one group has recommended for some time ELs for visible radiation (i.e., light). This one group is well-known in the field of occupational health—the American Conference of Governmental Industrial Hygienists (ACGIH).<sup>5,6</sup> The ACGIH refers to its ELs as *threshold-limit values* (TLVs); these are issued yearly, so there is an opportunity for a yearly revision. The current ACGIH TLVs for light (400 to 760 nm) have been largely unchanged for the last decade—aside from an increase in the retinal thermal limits in 2008. In 1997, with some revisions, the International Commission on Non-Ionizing Radiation Protection (ICNIRP) recommended these as international guidelines.<sup>11</sup> The ICNIRP guidelines are developed through collaboration with the World Health Organization (WHO) by jointly publishing criteria documents that provide the scientific database for the exposure limits.<sup>4</sup> The ACGIH TLVs and ICNIRP ELs are generally applied in product safety standards of the International Commission on Illumination (CIE), the International Electrotechnical Commission (IEC), the International Standardization Organization (ISO), and consensus standards from the American National Standards Institute and other groups. They are based in large part on ocular injury data from animal studies and data from human retinal injuries resulting from viewing the sun and welding arcs. All of the guidelines have an underlying assumption that outdoor environmental exposures to visible radiant energy is normally not hazardous to the eye except in very unusual environments such as snow fields, deserts, or out on the open water.

### Applying the UV Limits

To apply the UV guideline, one must obtain the average spectrally weighted UV irradiance  $E_{\text{eff}}$  at the location of exposure. The spectrally weighted irradiance is

$$E_{\text{eff}} = \sum E_{\lambda} \cdot S(\lambda) \cdot \Delta(\lambda) \quad (5)$$

where the summation covers the full spectral range of  $S(\lambda)$ , which is the normalized “Ultraviolet Hazard Function” (a normalized action spectrum for photokeratitis). The maximum duration of exposure to stay within the limit  $t_{\text{max}}$  is determined by dividing the daily EL of  $3 \text{ mJ} \cdot \text{cm}^{-2}$  by the measured effective irradiance to obtain the duration in seconds, as noted in Eq. (6):

$$t_{\text{max}} = \frac{3 \times 10^{-3} \text{ J} \cdot \text{cm}^{-2}}{E_{\text{eff}}} \quad (6)$$

In addition to the  $S(\lambda)$  envelope action spectrum-based EL, there has always been one additional criterion to protect the lens and to limit the dose rate to both lens and the skin from very high irradiances.

Initially, this was based only upon a consideration to conservatively protect against thermal effects. This was later thought essential not only to protect against thermal damage, but to also hedge against a possible unknown photochemical damage in the UV-A to the lens.

Sharply defined photobiological action spectra apply to the ultraviolet hazard function  $[S(\lambda)]$ , the blue-light hazard function  $[B(\lambda)]$ , and the retinal thermal hazard function  $[R(\lambda)]$ . These are shown in Fig. 1.

## Guidelines for the Visible

The ocular exposure limits for intense visible and IR radiation exposure of the eye (from incoherent radiation) are several, because they protect against either photochemical or thermal effects to the lens or retina.

The two primary hazards that must be assessed in evaluating an intense visible light source are (1) the photoretinitis (blue-light) hazard and (2) the retinal thermal hazard. Additionally, lenticular exposure in the near IR may be of concern. It is almost always true that light sources with a luminance less than  $1 \text{ cd/cm}^2$  ( $10^4 \text{ cd/m}^2$ ) will not exceed the limits, and this is generally a maximal luminance for comfortable viewing. Although this luminance value is not considered a safety limit, and may not be sufficiently conservative for a violet LED, it is frequently provided as a quick check to determine the need for further hazard assessment.<sup>3,5</sup>

The retinal thermal criteria, based upon the action spectrum  $R(\lambda)$ , applies to pulsed light sources and to intense sources. The longest viewing duration of potential concern is 10 s, because pupillary constriction and eye movements limit the added risk from greater exposure durations. The retinal thermal hazard EL is therefore not specified for longer durations.<sup>12</sup> Indeed, in 2008 ACGIH limited the changing value to 0.25 s. These retinal thermal limits are

$$\sum L_{\lambda} \cdot R(\lambda) \cdot \Delta\lambda \leq 5/\alpha \cdot t^{0.25} \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1}, \quad 1 \mu\text{s} < t < 10 \text{ s} \quad (\text{ICNIRP, 1997}) \quad (7)$$

$$\sum L_{\lambda} \cdot R(\lambda) \cdot \Delta\lambda \leq 640/t^{0.25} \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1}, \quad 1 \mu\text{s} < t < 0.625 \text{ ms} \quad (\text{ACGIH, 2008}) \quad (8a)$$

$$\sum L_{\lambda} \cdot R(\lambda) \cdot \Delta\lambda \leq 16/t^{0.75} \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1}, \quad 0.625 \text{ ms} < t < 0.25 \text{ s} \quad (\text{ACGIH, 2008}) \quad (8b)$$

$$\sum L_{\lambda} \cdot R(\lambda) \cdot \Delta\lambda \leq 45 \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1}, \quad t > 0.25 \text{ s} \quad (\text{ACGIH, 2008}) \quad (8c)$$

The blue-light photoretinitis hazard criteria were based upon the work of Ham et al.<sup>6,11,17</sup> The limit for time  $t$  is expressed as a  $B(\lambda)$  spectrally weighted radiance:

$$\sum L_{\lambda} \cdot B(\lambda) \cdot t \cdot \Delta\lambda \leq 100 \text{ J} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1} \quad (8)$$

## Applying IR Limits

There are two criteria in the near-IR region. To protect the lens against IR cataract, the EL that is applicable to IR-A and IR-B radiant energy (i.e., 780 to 3000 nm) specifies a maximal irradiance for continued exposure of  $10 \text{ mW/cm}^2$  (average) over any 1000-s period, but not to exceed an irradiance of  $1.8t^{-0.75} \text{ W/cm}^2$  (for times of exposure less than 1000 s). This is based upon the fact that IR cataract in workers appears only after lifetime exposure of 80 to  $150 \text{ mW/cm}^2$ .<sup>2,3,6</sup>

The second IR EL is to protect against retinal thermal injury from low-luminance IR illumination sources. This EL is for very special applications, where near-IR illuminators are used for night surveillance applications or IR LEDs are used for illumination or signaling. These illuminators have a very low visual stimulus and therefore would permit lengthy ocular exposure with dilated pupils. Although the retinal thermal limit [based upon the  $R(\lambda)$  function] for intense, visible, broadband sources is not provided for times greater than 0.25 to 10 s because of pupillary constriction and the like, the

retinal thermal hazard—for other than momentary viewing—will only realistically occur when the source can be comfortably viewed, and this is the intended application of this special-purpose EL. The IR illuminators are used for area illumination for nighttime security where it is desirable to limit light trespass to adjacent housing. If directly viewed, the typical illuminator source may be totally invisible, or it may appear as a deep cherry red source that can be comfortably viewed. The EL is proportional to  $l/\alpha$  and is simply limited to:

$$L_{\text{NIR}} = \sum L_{\lambda} \cdot R(\lambda) \cdot \Delta\lambda \leq 3.2/\alpha \cdot t^{0.25} \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1} \quad t < 810 \text{ s} \quad (\text{ACGIH, 2008}) \quad (10a)$$

$$L_{\text{NIR}} = \sum L_{\lambda} \cdot R(\lambda) \cdot \Delta\lambda \leq 0.6/\alpha \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1} \quad t < 810 \text{ s} \quad (\text{ACGIH, 2008}) \quad (10b)$$

It must be emphasized that this criterion is not applied to white-light sources, because bright light produces an aversion response and Eq. (8) or (9) would apply instead.

## 7.8 DISCUSSION

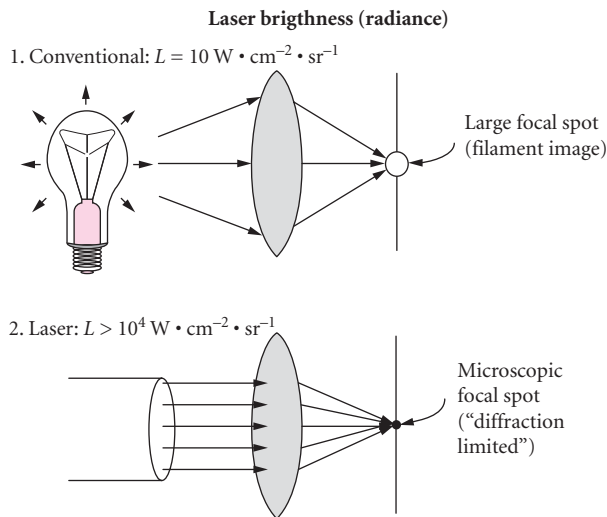
### Exceeding the Exposure Limits

When deriving the ELs for minimal-image-size visible and infrared laser radiations, the ICNIRP generally includes a factor of 10 to reduce the 50-percent probability of retinal injury by a factor of 10. This is not a true “safety factor,” because there is a statistical distribution of damage, and this factor was based upon several considerations. These included the difficulties in performing accurate measurements of source radiance or corneal irradiance, the measurement of the source angular subtense, as well as histological studies showing retinal changes occurring at the microscopic level at levels of approximately 2 below the ED-50 value.<sup>3</sup> In actual practice, this means that an exposure at two to three times the EL would not be expected to actually cause a physical retinal injury. At five times the EL, one would expect to find some injuries in a population of exposed subjects. The ELs are guidelines for controlling human exposure and should not be considered as fine lines between safe and hazardous exposure. By employing benefit-versus-risk considerations, it would be appropriate to have some relaxed guidelines; however, to date, no standards group has seen the need to do this.

### Laser Hazards

The very high radiance (brightness) of a laser ( $\text{MW}$  and  $\text{TW cm}^{-2} \text{sr}^{-1}$ ) is responsible for the laser’s great value in material processing and laser surgery, but it also accounts for its significant hazard to the eye (Fig. 6). When compared with a xenon arc or the sun, even a small He–Ne alignment laser is typically 10 times brighter (Fig. 7). A collimated beam entering the relaxed human eye will experience an increased irradiance of about  $10^5$  (i.e.,  $1 \text{ W} \cdot \text{cm}^{-2}$  at the cornea becomes  $100 \text{ kW/cm}^2$  at the retina). Of course, the retinal image size is only about 10 to 20  $\mu\text{m}$ , considerably smaller than the diameter of a human hair. So you may wonder: “So what if I have such a small lesion in my retina? I have millions of cone cells in my retina.” The retinal injury is always larger because of heat flow and acoustic transients, and even a small disturbance of the retina can be significant. This is particularly important in the region of central vision, referred to by eye specialists as the *macula lutea* (yellow spot), or simply the *macula*. The central region of the macula, the fovea centralis, is responsible for your detailed 20/20 vision. Damage to this extremely small (about 150- $\mu\text{m}$  diameter) central region can result in severe vision loss even though 98 percent of the retina is unscathed. The surrounding retina is useful for movement detection and other tasks but possesses limited visual acuity (after all, this is why your eye moves across a line of print, because your retinal area responsible for detailed vision has a very small angular subtense). Outside the retinal hazard region (400 to 1400 nm), the cornea—and even the lens—can be damaged by laser beam exposure.



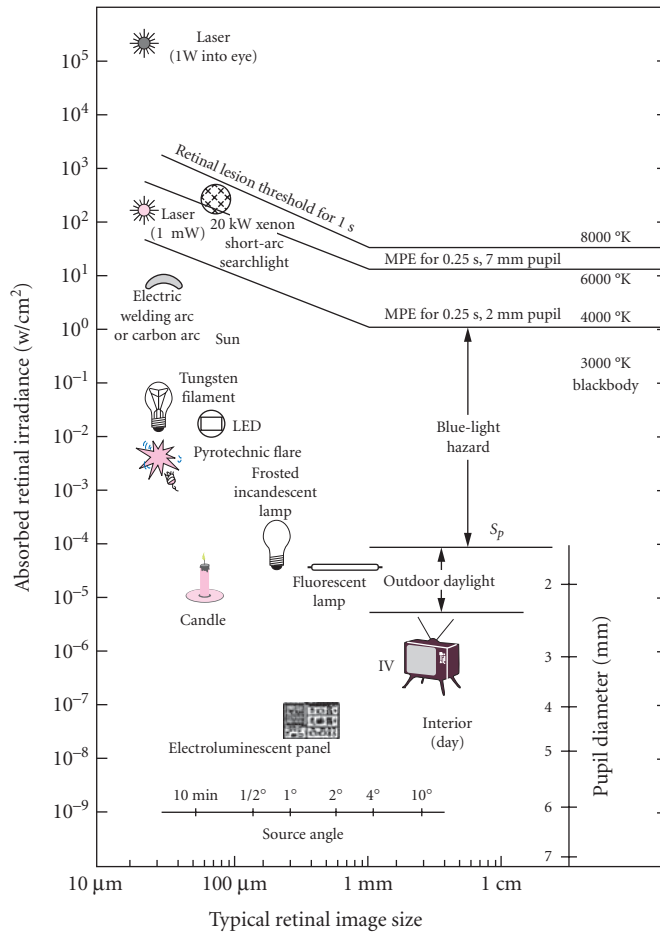


**FIGURE 6** The radiance of a light source determines the irradiance in the focal spot. Hence, the very high radiance of a laser permits one to focus laser radiation to a very small image of very high irradiance.

**Laser Safety Standards** In the United States, the American National Standard, ANSI Z136.1-2000, *The Safe Use of Lasers*, is the national consensus standard for laser safety in the user environment. It evolved through several editions since 1973. Maximum permissible exposure (MPE) limits are provided as sliding scales, with wavelength and duration for all wavelengths from 180 nm to 1 mm and for exposure durations of 100 fs to 30 ks (8-h workday). Health and safety specialists want the simplest expression of the limits, but some more mathematically inclined scientists and engineers on standards committees argue for sophisticated formulas to express the limits (which belie the real level of biological uncertainty). These exposure limits (which are identical to those of ACGIH) formed the basis for the U.S. Federal Product Performance Standard (21 CFR 1040).<sup>3</sup> The latter standard regulates only laser manufacturers. On the international scene, the International Electrotechnical Commission Standard IEC 60825-1 Ed. 2 (2007) originally grew out of an amalgam of the ANSI standard for user control measures and exposure limits and the U.S. Federal product classification regulation. The ICNIRP, which has a special relationship with the WHO in developing criteria documents on laser radiation, now recommends exposure limits for laser radiation.<sup>10</sup> All of the aforementioned standards are basically in agreement.

Laser safety standards existing worldwide group all laser products into four general hazard classes and provide safe measures for each hazard class (classes 1 to 4).<sup>4,5</sup> The U.S. Federal Product Performance Standard (21 CFR 1040) requires all commercial laser products to have a label indicating the hazard class. Once one understands and recognizes the associated ocular hazards, the safety measures recommended in these standards are quite obvious (e.g., beam blocks, shields, baffles, eye protectors). The ocular hazards are generally of primary concern. Many laser products sold in the USA and worldwide, will also be certified to meet the corresponding standard of IEC 60825-1:2007.

**Laser Accidents** A graduate student in a physical chemistry laboratory is aligning a Nd:YAG-pumped optical parametric oscillator (OPO) laser beam to direct it into a gas cell to study photodissociation parameters for a particular molecule. Leaning over a beam director, he glances down over an upward, secondary beam and approximately 80  $\mu\text{J}$  enters his left eye. The impact produces a microscopic hole in his retina, a small hemorrhage is produced over his central vision, and he sees only red in his left eye. Within an hour, he is rushed to an eye clinic where an ophthalmologist tells him



**FIGURE 7** Relative retinal irradiances for staring directly at various light sources. Two retinal exposure risks are depicted: (1) retinal thermal injury, which is image-size dependent and (2) photochemical injury, which depends on the degree of blue light in the source’s spectrum. The horizontal scale indicates typical image sizes. Most intense sources are so small that eye movements and heat flow will spread the incident energy over a larger retinal area. The range of photoretinitis (the “blue-light hazard”) is shown to be extending from the normal outdoor light levels and above. The xenon arc lamp is clearly the most dangerous of nonlaser hazards.

he has only 20/400 vision. In another university, a physics graduate student attempts to realign the internal optics in a Q-switched Nd:YAG laser system—a procedure normally performed by a service representative that the student had witnessed several times before. A weak secondary beam reflected upward from a Brewster window enters the man’s eye, producing a similar hemorrhagic retinal lesion with a severe loss of vision. Similar accidents occur each year and frequently do not receive publicity because of litigation or for administrative reasons.<sup>3,26</sup> Scientists and engineers who work with open-beam lasers really need to realize that almost all such lasers pose a very severe hazard to the eye if eye protection is not worn or if other safety measures are not observed.<sup>12-4</sup>

A common element in most laser laboratory accidents is an attitude that “I know where the laser beams are; I do not place my eye near to a beam; safety goggles are uncomfortable; therefore, I do not need to wear the goggles.” In virtually all accidents, eye protectors were available, but not worn. The probability that a small beam will intersect a 3- to 5-mm pupil of a person’s eye is small to begin with, so injuries do not always happen when eye protectors are not worn. However, it is worthwhile to consider the following analogy. If these individuals were given an air rifle with 100 BBs to fire, were placed in a cubical room that measures 4 m on each side and is constructed of stainless-steel walls, and were told to fire all of the BBs in any directions they wished, how many would be willing to do this without heavy clothing and eye protectors?!! Yet, the probability of an eye injury is similar. In all fairness, there are laser goggles that are comfortable to wear. The common complaints that one cannot see the beam to align it are mere excuses that are readily solved with some ingenuity once you accept the hazard. For example, image converters and various fluorescent cards (such as are used to align the Nd: YAG 1064-nm beam) can be used for visible lasers as well.

**Laser Eye Protectors** In the ANSI Z136.1 standard, there is broad general guidance for the user to consider factors such as comfort and fit, filter damage threshold, and periodic inspection, as well as the critical specification of wavelength and optical density (OD). However, there has never been any detailed U.S. specifications for standard marking and laboratory proofing (testing) of filters. By contrast, the approach in Germany (with DIN standards) and more recently in Europe (with CEN standards) has been to minimize the decision making by the user and place heavy responsibility upon the eyewear manufacturer to design, test, and follow standardized marking codes to label the eye protector. Indeed, the manufacturer had been required to use third-party test houses at some considerable expense to have type eyewear tested. Each approach has merit, and a new standard in the United States is now being developed for testing for standardized marking that is understandable to the wearer. The most important test of a protective filter is to measure the OD under CW, Q-switched, and mode-locked pulse irradiation conditions to detect saturable absorption-reversible bleaching.<sup>28,29</sup>

The marking of laser eye protection in an intelligible fashion to ensure that the user will not misunderstand and select the wrong goggle has been a serious issue, and there appears to be no ideal solution. Several eye injuries appear to have been caused by a person choosing the wrong protector. This is particularly likely in an environment where multiple and different laser wavelengths are in use, as in a research laboratory or a dermatological laser setting. A marking of OD may be clearly intelligible to optical physicists, but this is seldom understandable to physicians and industrial workers. Probably the best assurance against misuse of eyewear has been the application of customized labeling by the user. For example, “Use only with the model 12A,” or “Use only for alignment of mode-locked YAG,” or “Use only for port-wine stain laser.” Such cautions supplement the more technical marking. The terms ruby, neodymium, carbon dioxide, and similar labels in addition to the wavelength can reduce potential confusion with a multi-wavelength laser. Marking of broadband eye protection, such as welding goggles, does not pose a problem to the wearer. The welding shade number (e.g., WG-8), even if misunderstood by the welder, poses little risk. The simple suggestion to select the filter based upon visual comfort prevents the welder from being injured; wearing too light a shade such that the BL for blue light is exceeded is virtually impossible because of the severe disability glare that would result. The same is not true for laser eye protectors: One could wear a goggle that protects against a visible wavelength but that doesn’t protect against the invisible 1064-nm wavelength. Improved, standardized marking is certainly needed! Recently, a new standard in the ANSI Z136 series of standards was published: ANSI Z136.7:2008 (American National Standard for Testing and Labeling of Laser Protective Equipment), which provides more reasonable guidance.<sup>28</sup>

## Lamp Safety Standards

Lamp safety is dealt with in ANSI/IESNA RP27.3-2005 [*Recommended Practice for Photobiological Safety for Lamps—Risk Group Classification & Labeling, and in CIE Standard S-009:2006(IEC62471:2006) Photobiological Safety for Lamps and Lamp Systems*]. In addition, LEDs are also included in one laser safety standard: *American National Standard for the Safe Use of Lasers in Optical Fiber Communications Systems (OFCS)*], ANSI Z136.2 (1997), since LEDs used in this special application may be viewed by

optical aids such as a microscope or eye loupe, and such viewing actually happens in realistic splicing conditions and the instrumentation to distinguish the narrow bandwidth of the laser from the wider, nominal 50-nm bandwidth of the LED is a needless burden. All current LEDs have a maximum radiance of about  $12 \text{ W} \cdot \text{cm}^{-2} \cdot \text{sr}^{-1}$  and can in no way be considered hazardous, but for a period an IEC laser standard 60825-1 included LEDs (an inclusion dropped in 2007). Some observers of the IEC group accused the group of having an ulterior motive of providing test houses and safety delegates with greater business!

## 7.9 REFERENCES

1. L. R. Solon, R. Aronson, and G. Gould, "Physiological Implications of Laser Beams," *Science* **134**:1506–1508 (1961).
2. D. H. Sliney and B. C. Freasier, "The Evaluation of Optical Radiation Hazards," *Applied Optics* **12**(1):1–24 (1973).
3. D. H. Sliney and M. L. Wolbarsht, *Safety with Lasers and Other Optical Sources*. Plenum Publishing, New York, 1980.
4. World Health Organization (WHO), *Environmental Health Criteria No. 23, Lasers and Optical Radiation*, joint publication of the United Nations Environmental Program, the International Radiation Protection Association, and the World Health Organization, Geneva, 1982.
5. American Conference of Governmental Industrial Hygienists (ACGIH), *TLVs and BEIs Based on the Documentation of the Threshold Limit Values for Chemical Substances and Physical Agents and Biological Exposure Indices*, American Conference of Governmental Industrial Hygienists, Cincinnati, OH, 2008.
6. ACGIH, *Documentation for the Threshold Limit Values*, American Conference of Governmental Industrial Hygienists, Cincinnati, OH, 2007.
7. A. S. Duchene, J. R. A. Lakey, and Michael H. Repacholi (eds.), *IRPA Guidelines on Protection Against Non-Ionizing Radiation*, MacMillan, New York, 1991.
8. D. H. Sliney, J. Mellerio, V. P. Gabel, and K. Schulmeister, "What is the Meaning of Threshold in Laser Injury Experiments? Implications for Human Exposure Limits," *Health Phys.* **82**(3):335–347, 2002.
9. American National Standards Institute (ANSI), *Safe Use of Lasers*, ANSI Z136.1-2007, Laser Institute of America, Orlando, FL, 2007.
10. International Commission on Non-Ionizing Radiation Protection (ICNIRP), "Guidelines on Limits for Laser Radiation of Wavelengths between 180 nm and 1,000  $\mu\text{m}$ ," *Health Phys.* **71**(5):804–819 (1996); update (in press).
11. ICNIRP, "Guidelines on limits of Exposure to Broad-Band Incoherent Optical Radiation (0.38 to 3  $\mu\text{m}$ )," *Health Phys.* **73**(3):539–554 (1997).
12. W. P. Roach et al., "Proposed Maximum Permissible Exposure Limits for Ultrashort Laser Pulses," *Health Phys.* **77**:61–68 (1999).
13. T. P. Coohill, "Photobiological Action Spectra—What Do They Mean?" *Measurements of Optical Radiation Hazards*, CIE/ICNIRP, Munich, pp. 27–39 (1998).
14. D. H. Sliney, "Radiometric Quantities and Units Used in Photobiology and Photochemistry: Recommendations of the Commission Internationale de l'Eclairage (International Commission on Illumination)," *Photochem. Photobiol.* **83**:425–432 (2007).
15. J. A. Zuclich, "Ultraviolet-Induced Photochemical Damage in Ocular Tissues," *Health Phys.* **56**(5):671–682 (1989).
16. D. G. Pitts et al., "Ocular Effects of Ultraviolet Radiation from 295 to 365 nm," *Invest. Ophthalmol. Vis. Sci.* **16**(10):932–939 (1977).
17. W. T. Ham, Jr., "The Photopathology and Nature of the Blue-Light and Near-UV Retinal Lesions Produced by Lasers and Other Optical Sources," *Laser Applications in Medicine and Biology*, M. L. Wolbarsht (ed.), Plenum Publishing, New York, 1989.
18. W. T. Ham et al., "Evaluation of Retinal Exposures from Repetitively Pulsed and Scanning Lasers," *Health Phys.* **54**(3):337–344 (1988).
19. D. H. Sliney, "Physical Factors in Cataractogenesis—Ambient Ultraviolet Radiation and Temperature," *Invest. Ophthalmol. Vis. Sci.* **27**(5):781–789 (1986).

20. Commission Internationale de l'éclairage (International Commission on Illumination), *International Lighting Vocabulary*, CIE Publication No. 17.4, CIE, Geneva, Switzerland (1987).
21. M. T. Coroneo et al., "Peripheral Light Focussing by the Anterior Eye and the Ophthalmohelioses," *Ophthalmic Surg.* **22**:705–711 (1991).
22. B. E. K. Klein, R. Klein, and K. L. Linton, "Prevalence of Age-Related Lens Opacities in a Population, the Beaver Dam Eye Study," *Am. J. Pub. Hlth.* **82**(12):1658–1662 (1992).
23. D. H. Sliney, "Eye Protective Techniques for Bright Light," *Ophthalmology* **90**(8):937–944 (1983).
24. P. J. Dolin, "Assessment of the Epidemiological Evidence that Exposure to Solar Ultraviolet Radiation Causes Cataract," *Doc. Ophthalmol.* **88**:327–337 (1995).
25. D. H. Sliney, "Geometrical Gradients in the Distribution of Temperature and Absorbed Ultraviolet Radiation in Ocular Tissues," *Dev. Ophthalmol.* **35**:40–59 (2002).
26. D. H. Sliney, "Ocular Injuries from Laser Accidents," *SPIE Proceedings of Laser-Inflicted Eye Injuries: Epidemiology, Prevention, and Treatment*, San Jose, CA, 1996, pp. 25–33.
27. J. W. Ness et al., "Retinal Image Motion during Deliberate Fixation: Implications to Laser Safety for Long Duration Viewing," *Health Phys.* **72**(2):131–142 (2000).
28. ANSI, "American National Standard for Testing and Labeling of Laser Protective Equipment," ANSI Z136.7-2008 (2008).
29. D. H. Sliney and W. J. Marshall (eds.), "LIA Guide for the Selection of Laser Eye Protection," *Laser Institute of America*, 2008.

---

# BIOLOGICAL WAVEGUIDES

---

## Vasudevan Lakshminarayanan

*School of Optometry and Departments of Physics and Electrical Engineering  
University of Waterloo  
Waterloo, Ontario, Canada*

## Jay M. Enoch

*School of Optometry  
University of California at Berkeley  
Berkeley, California*

---

### 8.1 GLOSSARY

---

**Apodization.** Modification of the pupil function. In the present context, the Stiles-Crawford directional effect can be considered as an apodization, that is, variable transmittance across the pupil.

**Bessel functions.** These special functions, first defined by the mathematician Daniel Bernoulli and generalized by Friedrich Bessel, are canonical solutions of a particular differential equation called the Bessel differential equation. Bessel's equation arises when finding separable solutions to Laplace's equation and the Helmholtz equation in cylindrical or spherical coordinates systems. Bessel functions are therefore especially important for many problems of wave propagation, static potentials, such as propagation of electromagnetic waves in cylindrical waveguides, heat conduction in cylindrical objects, and the like. A specific form of the Bessel functions are called Hankel functions. These functions are very useful to describe waveguiding.

**Born approximation.** If an electromagnetic wave is expressed as the sum of an incident wave and the diffracted secondary wave, the scattering of the secondary wave is neglected. This neglect represents what is known as Born's first-order approximation.

**Bruch's membrane.** Inner layer of the choroid.

**Cilia.** Hair.

**Cochlea.** The spiral half of the labyrinth of the inner ear.

**Coleoptile.** The first true leaf of a monocotyledon.

**Copepod.** A subclass of crustacean arthropods.

**Cotyledons.** The seed leaf—the leaf or leaves that first appear when the seed germinates.

**Ellipsoid.** The outer portion of the inner segment of a photoreceptor (cone or rod). It is located between the myoid and the outer segment. It often contains oriented mitochondria for metabolic activity.

**Fluence.** A measure of the time-integrated energy flux usually given in units of Joules/cm<sup>2</sup>. In biology, it is used as a measure of light exposure.

**Helmholtz's reciprocity theorem.** Also known as the reversion theorem. It is a basic statement of the reversibility of optical path.

**Henle's fibers.** Slender, relatively uniform fibers surrounded by more lucent Müller cell cytoplasm in the retina. They arise from cone photoreceptor bodies and expand into cone pedicles.

**Hypocotyl.** The part of the axis of the plant embryo that lies below the cotyledons.

**Interstitial matrix.** Space separating photoreceptors. It is of a lower refractive index than the photoreceptors and, hence, can be considered as "cladding."

**Mesocotyl.** The node between the sheath and the cotyledons of seedling grasses.

**Mode.** The mode of a dielectric waveguide is an electromagnetic field that propagates along the waveguide axis with a well-defined phase velocity and that has the same shape in any arbitrary plane transverse along that axis. Modes are obtained by solving the source-free Maxwell's equations with suitable boundary conditions.

**Myoid.** It is the inner portion of the inner segment of a photoreceptor (cone or rod), located between the ellipsoid and the external limiting membrane of the retina.

**Organ of Corti.** The structure in the cochlea of the mammals which is the principal receiver of sound; contains hair cells.

**Pedicle.** Foot of the cone photoreceptor; it is attached to a narrow support structure.

**Phototropism.** Movement stimulated by light.

**Spherule.** End bulb of a rod photoreceptor.

**V-parameter.** Waveguide parameter [Eq. (3)]. This parameter completely describes the optical properties of a waveguide and depends upon the diameter of the guide, the wavelength of light used, and the indices of refraction of the inside (core) and outside (cladding) of the guide.

**Equation (1).** This is a statement of Snell's law.

- $n_1$  refractive index of medium 1; light is incident from this medium
- $n_2$  refractive index of medium 2; light is refracted into this medium
- $\theta_1$  angle of incidence of light measured with respect to the surface normal in medium 1
- $\theta_2$  angle of refraction of light in medium 2 measured with respect to the surface normal

**Equation (2).** This is an expression for the numerical aperture.

- $n_1$  refractive index of the inside of the dielectric cylinder (the "core")
- $n_2$  refractive index of the surround (the "cladding")
- $n_3$  refractive index of medium light is incident from before entering the cylinder
- $\theta_L$  limiting angle of incidence

**Equation (3).** Expression for the waveguide V-parameter.

- $d$  diameter of the waveguide cylinder
- $\lambda$  wavelength of incident light
- $n_1$  refractive index of core
- $n_2$  refractive index of cladding

## 8.2 INTRODUCTION

In this chapter, we will be studying two different types of biological waveguide model systems in order to relate the models' structures to the waveguide properties initiated within the light-guide construct. In these models, we will be discussing how these biological waveguides follow the Stiles-Crawford effect of the first kind, wavelength sensitivity of transmission, and Helmholtz's reciprocity theorem of optics. Furthermore, we will be investigating the different sources of biological waveguides such as vertebrate photoreceptors, cochlear hair cells (similar to cilia), and fiber-optic plant tissues, as well as considering their applications and comparing their light-guiding effects to the theoretical models. The emphasis will be on waveguiding in vertebrate retinal photoreceptors, though similar considerations apply to other systems.

### 8.3 WAVEGUIDING IN RETINAL PHOTORECEPTORS AND THE STILES-CRAWFORD EFFECT

*Photoreceptor optics* is defined as the science that investigates the effects of the optical properties of the retinal photoreceptors—namely, their size, shape, refractive index, orientation, and arrangement—on the absorption of light by the photopigment, as well as the techniques and instrumentation necessary to carry out such studies. It also explores the physiological consequences of the propagation of light within photoreceptor cells. The *Stiles-Crawford effect of the first kind*<sup>1</sup> (*SCE I*), discovered in 1933, represents a major breakthrough in our understanding of retinal physiology and the modern origin of the science of photoreceptor optics. The SCE I refers to the fact that visual sensitivity in normal eyes is greatest for light entering near the center of the eye pupil, and the response falls off roughly symmetrically from this peak. The SCE I underscores the fact that the retina is a complex optical processing system whose properties play a fundamental role in visual processing. The individual photoreceptors (rods and cones) behave as light collectors which capture the incident light and channel the electromagnetic energy to sites of visual absorption, the photolabile pigments, where the transduction of light energy to physicochemical energy takes place. The photoreceptors act as classic fiber-optic elements and the retina can be thought of as an enormous fiber bundle (a typical human retina has about  $130 \times 10^6$  receptors). Some aspects of the behavior of this fiber optic bundle can be studied using the SCE I function, a psychophysical measurement of the directional sensitivity of the retina. The SCE I is an important index reflecting waveguide properties of the retina and is a psychophysical measurement of the directional sensitivity of the retina. The SCE I can be used to elucidate various properties, and the photodynamic nature, of the photoreceptors. The SCE I can also be used to indicate the stage and degree of various retinal abnormalities. The reader is referred to the literature for detailed discussion of various aspects of photoreceptor optics.<sup>2–5</sup> A collection of classic reprints on the Stiles-Crawford effect and photoreceptor optics has been published by the Optical Society of America.<sup>6</sup> The books by Dowling<sup>7</sup> and Rodieck,<sup>8</sup> for example, give good basic information on the physiology of the retina as well as discussions of the physical and chemical events involved in the early stages of visual processing. Additionally, there is a color effect, known as the Stiles-Crawford effect of the second kind (*SCE II*). Here, if the locus of entry of a beam in the entrance pupil of the eye is changed, it altered the perceived hue and saturation of the retinal image. The SCE II will not be dealt with in this chapter and the reader is referred to the literature.<sup>9,10</sup> Added features of the SCE are addressed in the following chapter.

### 8.4 WAVEGUIDES AND PHOTORECEPTORS

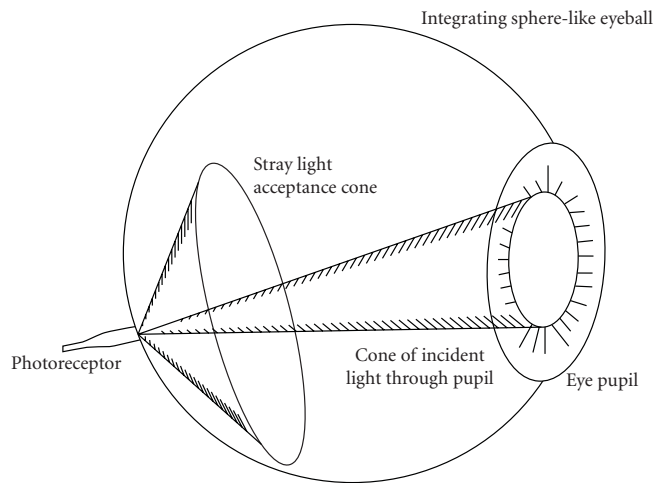
It appears as though there are two major evolutionary lines associated with the development of visual receptors: the ciliary type (vertebrates) and the rhabdomeric type (invertebrates). Besides being different in structure, the primary excitation processes also differ between these two types of detectors. Ciliary type (modified hair) detectors are found in all vertebrate eyes. They have certain common features with other sensory systems, for example, the hair cells of the auditory and vestibular systems. It is possible to say that all vertebrate visual detectors have waveguides associated with the receptor element. The incident light is literally guided or channeled by the receptor into the outer segment where the photosensitive pigments are located. In other words, the fiber transmits radiant energy from one point (the entrance pupil or effective aperture of the receptor) to other points (the photolabile pigment transduction sites). Given the myriad characteristics found in photoreceptors in both vertebrate and invertebrate eyes (including numerous superposition and apposition eyes), what is common in all species is the presence of the fiber-optic element. It appears as though fiber-optic properties evolved twice in the vertebrate and invertebrate receptors. This fact emphasizes the importance of these fiber-optic properties.

Like any optical device, the photoreceptor as a waveguide can accept and contain light incident within a solid angle about its axis. Therefore, all waveguides are directional in their acceptance of light and selection of only a certain part of the incident electromagnetic wavefront for transmittance to the transducing pigment.



The detailed analysis of optics and image formation in various forms of vertebrate and invertebrate eyes is beyond the scope of this chapter (see Refs. 11 and 12 as well as Chap. 1 in this volume for excellent reviews). However, all such eyes must share a common characteristic. That is, each detecting element must effectively collect light falling within its bound and each must “view” a slightly different aspect of the external world, the source of visual signal. Because of limitations of ocular imagery, there must be a degree of overlap between excitation falling on neighboring receptors; there is also the possibility of optical cross talk. It has been postulated that there is an isomorphic correspondence between points on the retina and a direction in (or points in) real (the outside world) space. A number of primary and secondary sources of light stimulus for the organism are located in the outside world. The organism has to locate those stimuli falling within its field of view and to relate those sensed objects to its own egocentric (localization) frame of reference. Let us assume that there is a fixed relationship between sense of direction and retinal locus. The organism not only needs to differentiate between different directions in an orderly manner, but it must also localize objects in space. This implies that a pertinent incident visual signal is identifiable and used for later visual processing. Further, directionality-sensitive rods and cones function most efficiently only if they are axially aligned with the aperture of the eye, that is, the source of the visual signal. Thus, it becomes necessary to relate the aperture of the receptor as a fiber-optic element relative to the iris aperture. The retinal receptor-waveguide-pupillary aperture system should be thought of as an integrated unit that is designed for optimization of light capture of quanta from the visual signal in external space and for rejection of stray light noise in the eye/image. A most interesting lens-aperture-photoreceptor system in the invertebrate world is that of a copepod, *Copilia*. The *Copilia* has a pair of image-forming eyes containing two lenses, a corneal lens and another lens cylinder (with an associated photoreceptor), separated by a relatively enormous distance. Otherwise, the photoreceptor system is similar to the insect eye. *Copilia* scans the environment by moving the lens cylinder and the attached photoreceptor. The rate of scan varies from about five scans per second to one scan per two seconds. This is like a mechanical television camera. There are other examples of scanning eyes in nature.<sup>12</sup>

The retina in the vertebrate eye is contained within the white, translucent, diffusing, integrating-sphere-like eye. It is important to realize that the retinal receptor “field of view” is not at all limited by the pupillary aperture. The nonimaged light may approach the receptor from virtually any direction. In Fig. 1 the larger of the two converging cones of electromagnetic energy is meant to portray (in a



**FIGURE 1** Schematic drawing showing light incidence at a single photoreceptor. Light incident on the retina has a maximum angle of incidence of about  $\pm 10^\circ$  for a receptor at the posterior pole of the eye. Stray light may enter the receptor over a much larger acceptance angle (shown by the larger cone).

limited sense) the extent of the solid angle over which energy may impinge on the receptor. While stray light from any small part of this vast solid angle may be small, the total, when integrated over the sphere surrounding the receptor, may be quite a significant sum. To prevent such noise from destroying image quality, the organism has evolved specific mechanisms, including screening dark pigments in the choroid and pigment epithelium, morphological designs, photomechanical changes, aligned photolabile pigment and fiber-optic properties, and in a number of species, a tapetum.

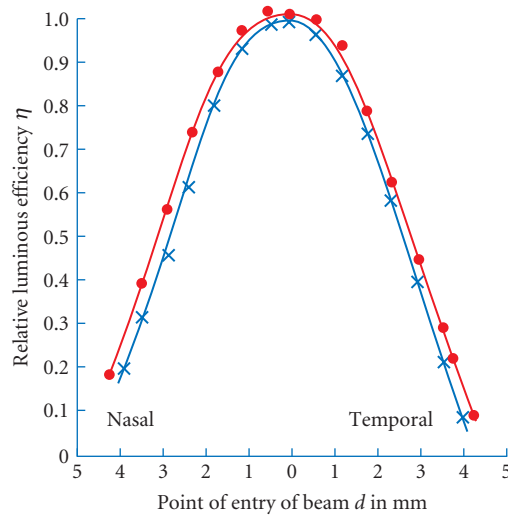
Since the pupillary aperture is the source of the pertinent visual signal, the primary purpose of the retinal fiber-optic element, the receptor, is to be a directionally selective mechanism, which accepts incident light from only within a rather narrow solid angle. Given that the receptor has a rather narrow, limiting aperture, exhibits directionality, and favors absorption of light passing along its axis, this optical system will be most effective if it is aligned with the source of the signal, the pupillary aperture. This argument applies equally for cones and rods.

In order for a material to serve as a fiber-optic element, the material must be reasonably transparent and have a refractive index that is higher than the immediately surrounding medium. To reduce cross talk in an array of such fiber-optic elements, they must be separated by a distance approximately equal to or greater than a wavelength of light by a medium having a lower index of refraction than the fiber. The inner and outer segments of retinal photoreceptors and the separating interstitial matrix satisfy such condition. The index of refraction of biological materials is generally dependent on the concentration of large molecules of proteins, lipids, and lipoproteins. The receptor has a high concentration of such solids. These materials are in low concentrations in the surrounding interstitial matrix space. As examples, Sidman<sup>13</sup> has reported refractive index values of 1.4 for rod outer segments, 1.419 for foveal cone outer segments, and a value between 1.334 and 1.347 for the interstitial matrix in the primate retina. Enoch and Tobey<sup>14</sup> found a refractive index difference of 0.06 between the rod outer segment and the interstitial matrix in frogs. Photoreceptors are generally separated by about 0.51  $\mu\text{m}$ . A minimum of about 0.5  $\mu\text{m}$  (1 wavelength in the green part of the spectrum) is needed. Separation may be disturbed in the presence of pathological processes. This can alter waveguiding effects, transmission as a function of wavelength, directionality, cross talk, stray light, resolution capability, and the like.

## 8.5 PHOTORECEPTOR ORIENTATION AND ALIGNMENT

As noted above, Stiles and Crawford<sup>1</sup> found that radiant energy entering the periphery of the dilated pupil was a less effective stimulus than a physically equal beam entering the pupil center. Under photopic foveal viewing conditions, it was found that a beam entering the periphery of a pupil had to have a radiance of from 5 to 10 times that of a beam entering the pupil center to have the same subjective brightness. Stiles and Crawford plotted a parameter  $\eta$ , defined as the ratio of the luminance of the standard beam (at pupil center) to that of the displaced beam at photometric match point as a function of the entry portion of the displaced beam. The resultant curve has essentially a symmetric falloff in sensitivity (Fig. 2). This cannot be explained as being due to preretinal factors and this directional sensitivity of the retina is known as the Stiles-Crawford Effect of the first kind (SCE I). Later a chromatic effect (alteration of perceived hue and saturation of displaced beam) was also discovered that is called the *Stiles-Crawford Effect of the second kind*.<sup>9,10</sup> SCE I is unaffected by polarized light. Normative values of the parameters of this function describing the effect (as well as various mathematical descriptions) can be found in Applegate and Lakshminarayanan.<sup>15</sup>

The directionality of photoreceptors depends upon the acceptance angles of individual photoreceptors and on the variability of their orientation within the tested photoreceptor population. MacLeod<sup>16</sup> developed a selective adaptation technique to study the variability of photoreceptor orientations and concluded that foveal cones were aligned with great precision (similar results were shown, for example, Burns et al.<sup>17</sup> using reflectometry methods). Roorda and Williams<sup>18</sup> measured the orientations of individual cones and found the average disarray to be 0.17 mm in the pupil plane and point out that this disarray accounts for less than 1 percent of the breadth of the overall tuning function.



**FIGURE 2** Psychophysically measured Stiles-Crawford function for left and right eyes of an observer. Nasal and temporal refer to direction with respect to pupil center. The observer's task is to match the brightness of a beam whose point of entry is varied in the pupil to a fixed beam coming in through the center of the pupil.

The directionality of photoreceptors is due to their structure that makes them act as optical fibers. Therefore, measurement of photoreceptor directionality is an important tool for testing the physical properties of photoreceptors *in vivo*. Much valuable information has been obtained from studies of patients with pathologies or conditions that affect the photoreceptor (and adjoining structures) layers. However clinical studies are limited by the fact that psychophysical methods are time consuming and require excellent cooperation from the subject. Therefore, different reflectometric methods have been developed.

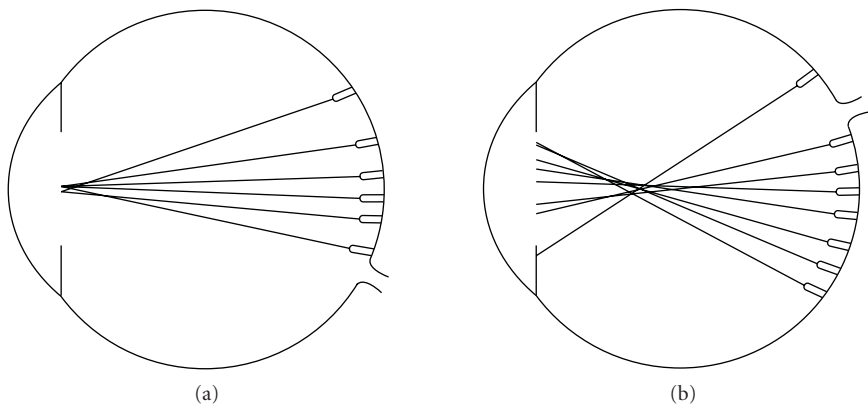
Krauskopf<sup>19</sup> was the first to show a direct correspondence between the psychophysical SCE and changes in reflectivity when the entry and exit pupils are moved. There are three major methods to measure the directional properties of photoreceptors by reflectometry. They are: (a) Moving entry and exit pupils method—here the angles of incidence and reflection at the retina are changed in tandem. This method has been applied to design the photoreceptor alignment reflectometer<sup>20</sup> and a custom built Scanning Laser Ophthalmoscope.<sup>21</sup> (b) Moving exit pupil method—in this method the entrance pupil remains fixed and this design was applied to the reflectometers of van Bloklant<sup>22,23</sup> and Burns et al.<sup>24</sup> (c) Moving entrance pupil method<sup>25</sup>—Roorda and Williams<sup>18</sup> applied this method to sample fields as small as a single cone by means of the Rochester adaptive optics ophthalmoscope. (See Chap. 15 in this volume for a discussion of adaptive optics in vision science).

Histological and x-ray diffraction studies show that in many species, including human, photoreceptors tend to be oriented not toward the center of the eye, but toward an anterior point. This pointing tendency is present at birth. The directional characteristics of an individual receptor are difficult to assess, and cannot, as yet, be directly related to psychophysical SCE I data. If one assumes that there is a retinal mechanism controlling alignment and that the peak of the SCE I represents the central alignment tendencies of the receptors contained within the sampling area under consideration, it is possible to investigate where the peak of the receptor distribution is pointed at different loci tested across the retina and to draw conclusions relative to the properties of an alignment system or mechanism. Results of the SCE I functions studied over a wide area of the retina show that the overall pattern of receptor orientation is toward the approximate center of the exit pupil of the eye.

This alignment is maintained even up to  $35^\circ$  in the periphery. There is a growing evidence that receptor orientation is extremely stable over time in normal observers.<sup>26</sup>

Photoreceptor orientation disturbance and realignment of receptors following retinal disorders have also been extensively studied. A key result of these studies is that with remission of the pathological process, the system is often capable of recovering orientation throughout life. Another crucial conclusion from these studies is that alignment is *locally controlled* in the retina (e.g., Ref. 27). The stability of the system, as well as corrections in receptor orientation after pathology, implies the presence of at least one mechanism (possibly phototropic) for alignment. Even though the center-of-the-exit-pupil-of-the-eye receptor alignment characteristic is predominant, rare exceptions to this have been found. These individuals exhibit an approximate center-of-the-retinal-sphere-pointing receptor alignment tendency (Fig. 3).<sup>28</sup> Though certain working hypotheses have been made,<sup>29</sup> what maintains the receptor is an open unanswered question and is the subject of current research that is beyond the scope of this chapter. Eckmiller,<sup>30</sup> for example, has proposed that the alignment of photoreceptors is accomplished by a feedback controlled bending of the cell at the myoid. Some unspecified local changes are thought to occur within the myoid that activate molecular motors and induce local movements by cytoskeletal elements, and bending can be accomplished by a differential change in the myoid at different positions around the cell perimeter. To summarize, the resultant directionality must reflect the properties of the media, the waveguide properties of receptors, the alignment properties of photolabile pigments, and so forth. Mechanical tractional effects also affect receptor alignment (e.g., Refs. 31 and 32). Other factors could include stress due to microscars, traction or shear forces due to accommodation, inertial forces during saccades, local growth for example, in high myopia, and the like. The SCE I function obtained experimentally depends upon all these factors as well as the sum of photoreceptor alignment properties sampled in the retinal test area. The response is dominated by the receptor units most capable of responding in the test area and is affected by the level of light adaptation.

What are the functional visual consequences of SCE I? Recent studies have dealt with this question (e.g., Ref. 33). SCE is important in photometry and retinal image quality (see Chap. 9). Calculations for change in effective retinal illuminance were first given by Martin<sup>34</sup> and expanded on by Atchison et al.<sup>35</sup> Baron and Enoch,<sup>36</sup> used a half-sensitivity, half-width measurement of retinal directional sensitivity as the basis for integrating a parabolic approximation of the SCE over the pupillary area. It is possible to define an “effective” troland which takes into account the SCE compensated entrance pupil area. These studies show that increasing the pupil size from 2- to 8-mm diameter provides an increase in effective retinal illuminance of about 9 times rather than the 16 times increase in pupil area.



**FIGURE 3** Inferred photoreceptor alignment from SCE measurements. (a) The more usual center-of-the-exit-pupil pointing tendency. (b) Center-of-the-retinal-sphere pointing tendency (rare).

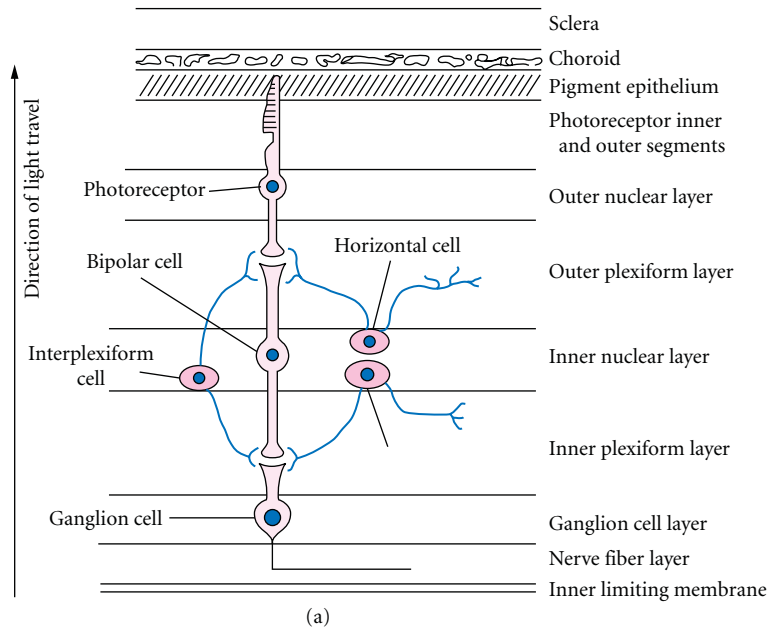
The influence of the SCE on visual functions can be investigated by using filters based on the apodization model of the SCE.<sup>37,38</sup> Rynders et al.<sup>39</sup> developed SCE neutralizing filters, but provided few details. Following up, Scott et al.<sup>40</sup> have described construction of practical filters to achieve neutralization of the SCE. When light from the edge of a large pupil plays a significant role in degrading retinal image quality, attenuating this light by the effective SCE I apodization will, of course, improve image quality. In general, the magnitude of this improvement will depend on the magnitude of the SCE and the level of blur caused by marginal rays. The presence of aberrations influences the impact of the SCE apodization on defocused image quality. Since defocus of the opposite sign to that of the aberration can lead to well-focused marginal rays, attenuating these well-focused rays with apodization will not improve image quality.<sup>41</sup> In addition to generally attenuating the impact of defocus on image contrast, this apodization will effectively remove some of the phase reversals (due to zero crossings of the modulation transfer function) created by positive defocus. The displacement of phase reversals to higher spatial frequencies will have a direct impact on defocused visual resolution. Spatial tasks in which veridical phase perception is critical will be significantly more resistant to positive defocus because of the SCE. However, Atchison et al.<sup>41,42</sup> suggest that SCE has little effect on image quality when dealing with well centered pupils. Zhang et al.<sup>33</sup> have used a wave optics model to examine the effects of SCE and aberrations and have compared the results with psychophysical measurements of the effect of defocus on contrast sensitivity and perceived phase reversals. They found that SCE apodization had the biggest effect on defocused image quality only when defocus and spherical aberration have the same sign and conclude that SCE can significantly improve defocused image quality and defocused vision particularly for tasks that require veridical phase perception. It has been argued that measurements of the *transverse chromatic aberration (TCA)* can be affected due to SCE. Various studies show that the effect of the SCE on the amount of TCA varies strongly across individuals, and between eyes in the same individual. In conclusion, it is thought that aberrations and SCE modify the amount and direction of TCA, and SCE does not necessarily reduce the impact of TCA.<sup>43</sup> Longitudinal chromatic aberration, on the other hand, is only slightly affected by SCE I. It has also been shown that decentering the SCE produces an appreciable shift in subjective TCA for large pupil sizes.<sup>42</sup>

The wavelength dependence of the SCE has been explained by using a model of fundus reflectance developed by van de Kraats et al.<sup>44</sup> Berendschot et al.<sup>45</sup> showed that a there is a good fit between the model (based on geometrical optics) and experimental data, if self-screening and backscattered choroidal light are included in the model.

Is the retina an optimal instrument? To answer this question, Marcos and Burns<sup>46</sup> studied both wavefront aberration and cone directionality. They concluded that cone directionality apodization does not always occur at the optically best pupillary region and that in general ocular optics and cone alignment do not develop toward an optimal optical design.

## 8.6 INTRODUCTION TO THE MODELS AND THEORETICAL IMPLICATIONS

First, we need to explore the theoretical constructs or models of the biological waveguide systems based on the photoreceptors to get a full appreciation of how these models transmit light through their structure. There are two models of biological waveguide systems that will be discussed here. The first model, which is the retinal layer of rods and cones, is defined as beginning approximately at the *external limiting membrane (ELM)* (Fig. 4), and it is from that boundary that the photoreceptors are believed to take their orientation relative to the pupillary aperture. In any local area within that layer, the photoreceptors are roughly parallel cylindrical structures of varying degrees of taper, which have diameters of the order of magnitude of the wavelength of light, and which are separated laterally by an organized lower refractive index substance, called the *interphotoreceptor matrix (IPM)*. The photoreceptors thus form a highly organized array of optical waveguides whose longitudinal axes are, beginning at the ELM, aligned with the central direction of the pupillary illumination. The IPM serves as a cladding for those waveguides. We will assume that the ELM approximates the inner bound of the receptor fiber bundle for light that is incident upon the retina in the physiological



**FIGURE 4** Schematic diagram of the retina and the photoreceptor. (a) Retinal layers. Note: the direction of light travel from the pupil is shown by the arrow.

direction.<sup>47</sup> Portions of the rod and cone cells that lie anterior to the ELM are not necessarily aligned with the pupillary aperture, and might be struck on the side, rather than on the axis, by the incident illumination. Some waveguiding is observed for a short distance into this region. However, these portions of the rod and cone cells anterior to the ELM would only be weakly excited by the incoming light.

Figure 5 shows the three-segment model that is generally used, denoting the idealized photoreceptor. All three segments are assumed to have a circular cross section, with uniform myoid and outer segment and a smoothly tapered ellipsoid. For receptors with equal myoid and outer segment radii, the ellipsoid is untapered. In this model, the sections are taken to be homogeneous, isotropic, and of higher refractive index than the surrounding near homogeneous and isotropic medium, which is the IPM. Both the myoid and the ellipsoid are assumed to be lossless, or nonabsorbing, since the cytochrome pigments in the mitochondria of the ellipsoids are not considered in this model. Absorption in the outer segments by the photolabile pigment molecules aligned within the closely packed disk membranes is described by the *Naperian absorption coefficient* ( $\alpha$ ). The aperture imposed at the ELM facilitates calculation of the equivalent illumination incident upon the receptor.

A number of assumptions and approximations have been made for these two models described above:

1. The receptor cross section is approximately circular and its axis is a straight line. This appears to be a reasonable assumption for the freshly excised tissue that is free of ocular diseases or pathology.
2. Ellipsoid taper is assumed to be smooth and gradual. Tapers that deviate from these conditions may introduce strong mode coupling and radiation loss.
3. Individual segments are assumed to be homogeneous. This is a first-order approximation, as each of the photoreceptor sections has different inclusions, which produce local inhomogeneities.

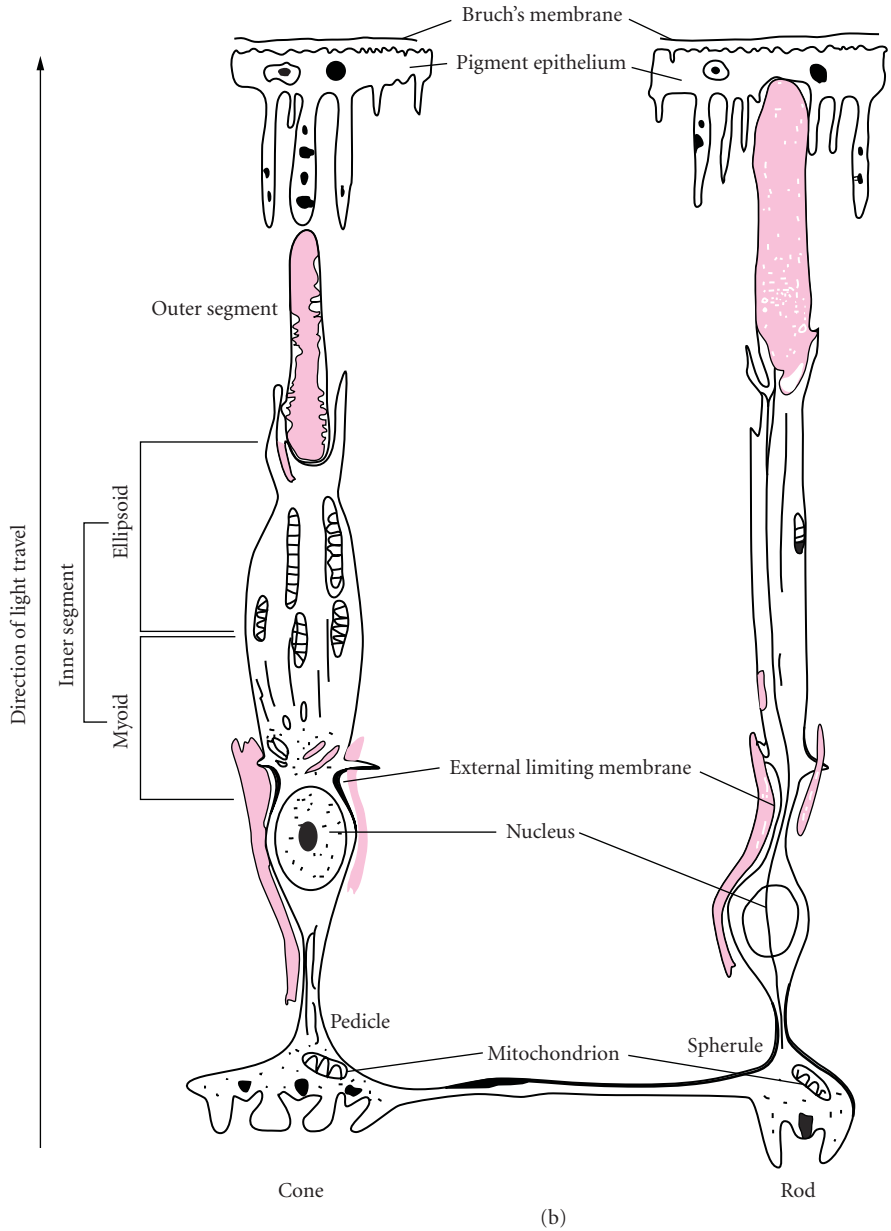
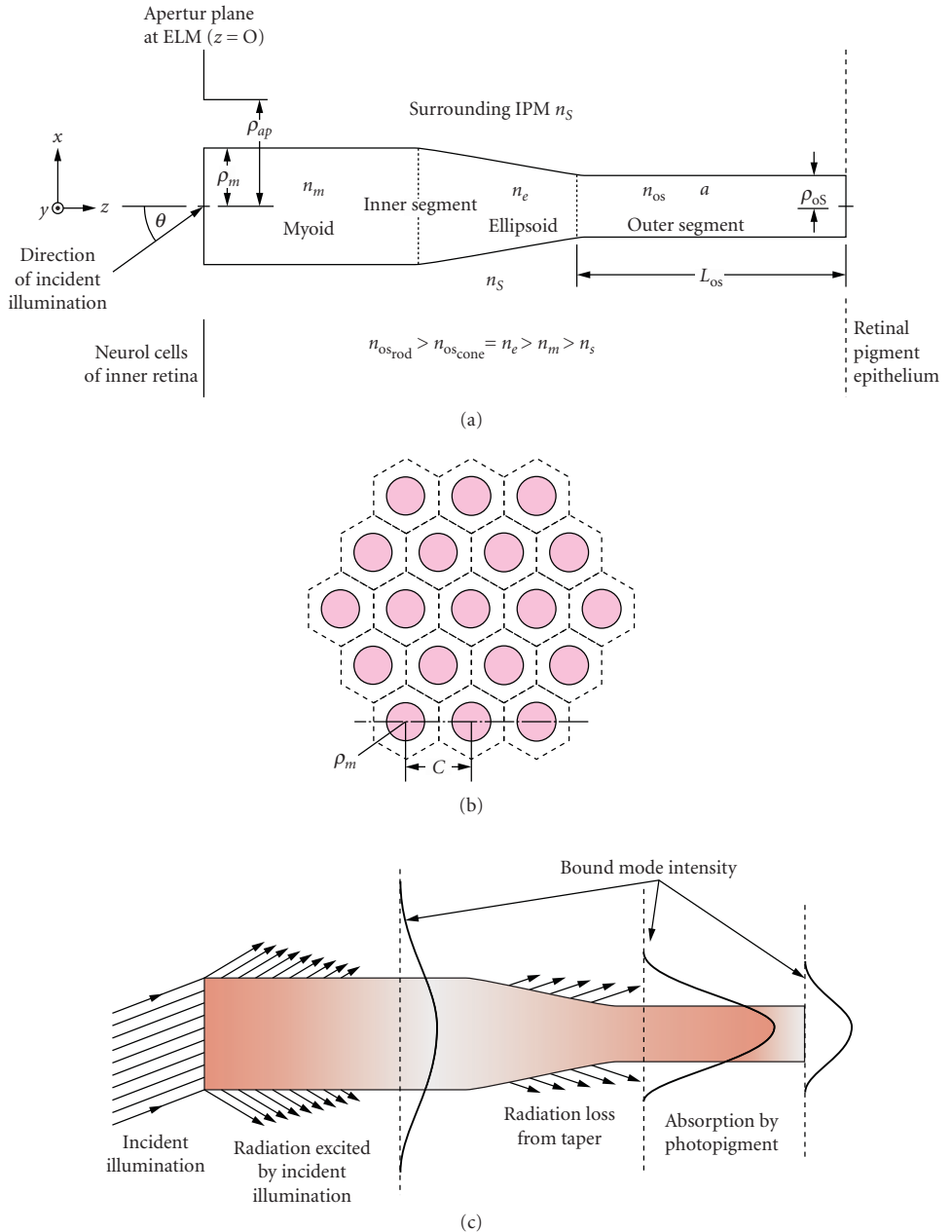


FIGURE 4 (Continued) (b) Structure of the cone and rod photoreceptor cells.



**FIGURE 5** (a) Generalized three-segment model of an idealized photoreceptor. (b) View of retinal mosaic from the ELM (external limiting membrane), which is the aperture plane of the biological waveguide model for photoreceptors. The solid circles represent myoid cross sections. (c) Illustration of optical excitation of retinal photoreceptor showing radiation loss, funneling by ellipsoid taper, and absorption in the outer segment. Regions of darker shading represent greater light intensity.



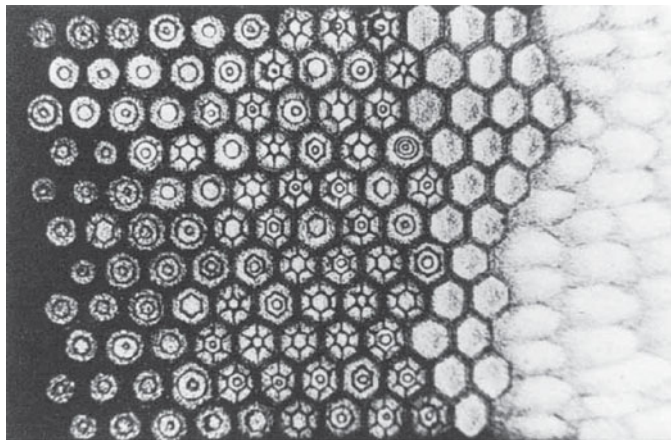
4. Individual segments are assumed to be isotropic as a first-order approximation. However, in reality, the ellipsoid is packed with highly membranous mitochondria whose dimensions approach the wavelength of visible light, and which are shaped differently in different species. The mitochondrion-rich ellipsoid introduces light scattering and local inhomogeneity. The outer segment is composed of transversely stacked disks, having a repetition period an order of magnitude smaller than the wavelength of light; in addition, these outer segments contain photolabile pigment.
5. Linear media are assumed. Nonlinearities may arise from a number of sources. These could include the following:
  - Absorption of light is accompanied by bleaching of the photopigments and the production of absorbing photoproducts that have different absorption spectra from those of the pigments. Bleaching is energy and wavelength dependent and the bleached pigment is removed from the pool of available pigment, thus affecting the absorption spectrum. Under physiological conditions in the intact organism where pigment is constantly renewed, the relatively low illumination levels that result in a small percent bleach can be expected to be consistent with the assumed linearity. However, with very high illumination levels where renewal is virtually nonexistent, researchers must utilize special procedures to avoid obtaining distorted spectra (e.g., albino retina).
  - Nonlinearity may arise in the ellipsoid owing to illumination-dependent activity of the mitochondria, since it has been shown that mitochondria of ellipsoid exhibit metabolically linked mechanical effects. If such effects occurred in the retina, then illumination-driven activity could alter inner segment scattering.
  - Some species exhibit illumination-dependent photomechanical responses that result in gross movement of retinal photoreceptors. That movement is accompanied by an elongation or contraction of the myoid in the receptor inner segment and has been analyzed in terms of its action as an optical switch.<sup>48</sup> In affected species, rod myoids elongate and cone myoids shorten in the light, while rod myoids contract and the cone myoids elongate in the dark. In the light, then, the cones lie closer to the outer limiting membrane and away from the shielding effect of the pigment epithelial cell processes, and contain absorbed pigment, while the pigment granules migrate within the pigment epithelial processes.<sup>49</sup>
6. The refractive indices of the segment are assumed to be independent of wavelength, which completely contradicts the definition of the *refractive index*, which is  $n = c/v$  where  $v$  (velocity) =  $f$  (frequency)  $\cdot \lambda$  (wavelength), and  $v$  is proportional to wavelength because  $f$  is constant for all media. Therefore, a change in the velocity dependent on wavelength also changes the refractive index in real life.
7. The medium to the left of the external limiting membrane is assumed to be homogeneous. However, in reality, although the ocular media and inner (neural) retina are largely transparent, scattering does occur within the cornea, aqueous, lens, vitreous, and inner retina. Preparations of excised retina will be devoid of the ocular media, but the inner retina will be present. In turn, the preparations would account for light scattering upon light incidence.
8. It is assumed that the medium surrounding the photoreceptors is homogeneous and nonabsorbing. Again, this is an approximation. Microvilli originating in the Müller cells extend into the space between the inner segments, and microfibrils of the *retina pigment epithelium* (RPE) surround portions of the outer segments. Also, the assumption of homogeneity of the surround neglects the reflection and backscatter produced by the RPE, choroid, and sclera. Note that preparations of excised retinae that are used for observing waveguiding are usually (largely) devoid of interdigitating RPE.
9. In some species the outer segment is tapered. Gradual taper can be accommodated in the model. Sharp tapers, however, must be dealt with separately.
10. The exact location of the effective input plane to the retinal fiber optics bundle is not known; however, in early studies it was assumed that the ELM was the effective inner bound of the retinal fiber bundle. To date, though, universal agreement does not exist regarding the location of the effective input plane; some investigators assume it to be located at the outer segment level.

11. In species with double cones (e.g., goldfish), the two components may often not be of equal length or contain identical photolabile pigments.

Next, we will evaluate the electromagnetic validity of the preceding models based on the assumptions discussed above. An additional assumption is that meaningful results relating to the in situ properties of a retinal photoreceptor can be obtained by means of calculations performed upon a single photoreceptor model. Involved here are considerations regarding optical excitation, optical coupling between neighboring receptors, and the role of scattered light within the retina. These assumptions may be summarized as follows:

1. The illumination incident upon the single photoreceptor through the aperture in Fig. 5 is taken to be representative of that illumination available to an individual receptor located in the retinal mosaic. Figure 5*b* illustrates the cross section of a uniform receptor array as seen from the ELM. For a uniformly illuminated, infinitely extended array, the total illumination available to each receptor would be that falling on a single hexagonal area.
2. The open, or unbounded, transverse geometry of the dielectric waveguide supports two classes of modes, referred to as bound, guided, or trapped modes, and unbound, unguided, or radiation modes. Different behavior is exhibited by the two mode species. As their names suggest, the bound modes carry power along the waveguide, whereas the unbound or radiation modes carry power away from the guide into the radiation field (Fig. 5*c*). From a ray viewpoint, the bound modes are associated with rays that undergo total internal reflection and are trapped within the cylinder; unbound modes are associated with rays that refract or tunnel out of the cylinder. These modes are obtained as source-free solutions of Maxwell's equations and are represented in terms of Bessel and Hankel functions (for a cylindrical geometry). These mode shapes show a high degree of symmetry. A full discussion of waveguide theory can be found in Snyder and Love.<sup>50</sup>
3. Coherent illumination of the photoreceptors is assumed. Here, the equivalent aperture of the photoreceptor is defined to be the circle whose area is equal to that of the hexagon seen in Fig. 5*b* specified by the appropriate intercellular spacing and inner segment diameter.

The earliest analyses of light propagation within the retinal photoreceptors employed geometric optics. As early as 1843, Brucke discussed the photoreceptor's trapping of light via total internal reflection (as quoted by von Helmholtz),<sup>51</sup> but stopped just short of attributing a directionality to vision. In fact, Hannover had observed waveguide modes in photoreceptors,<sup>52</sup> but interpreted them as being cellular fine structure (Fig. 6). Interest resumed after the discovery of the Stiles-Crawford effect of the



**FIGURE 6** Drawing of frog outer segment viewed end-on. (From Ref. 21.)

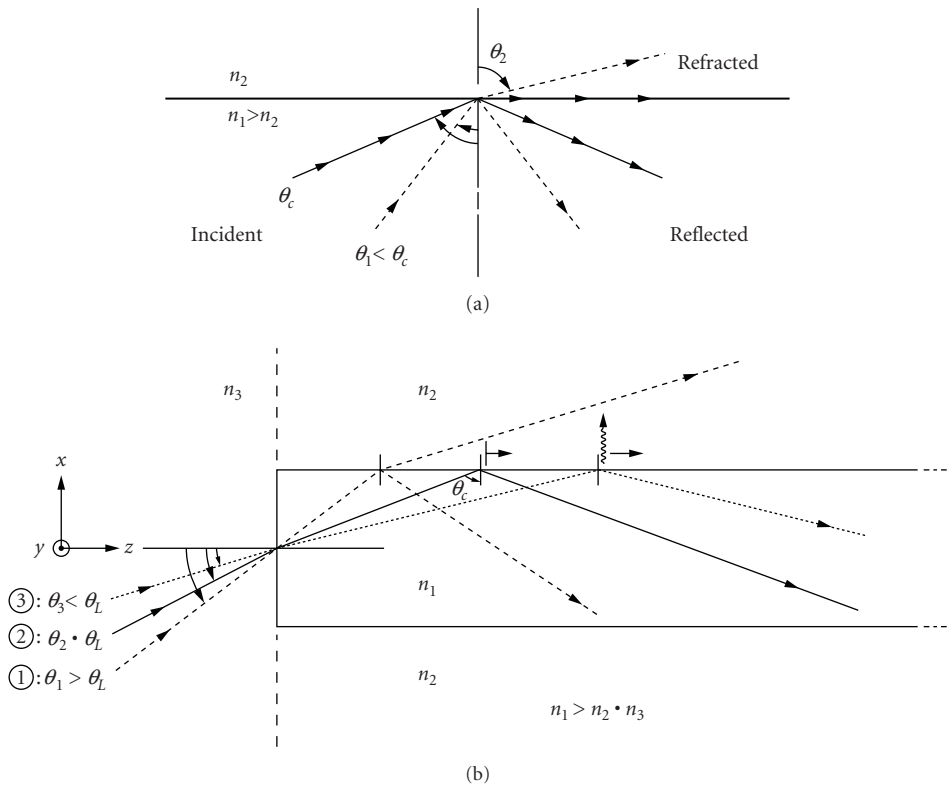
first kind (SCE I). O'Brien<sup>53</sup> suggested a light-funneling effect produced by the tapered ellipsoids, and Winston and Enoch<sup>54</sup> employed geometric optics to investigate the light-collecting properties of the retinal photoreceptor.

Recall the behavior of rays incident upon the interface separating two dielectric media (see Fig. 7a). Rays incident from medium 1 at an angle  $\theta_1$  to the interface normal, will be refracted at an angle  $\theta_2$  in medium 2; the angles are related by Snell's law where:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 \tag{1}$$

where  $n_1$  and  $n_2$  are the refractive indices of media 1 and 2, respectively. As shown in Fig. 6a, the refracted rays are bent toward the interface, since the incidence is from a medium of greater refractive index toward a medium of lesser refractive index. In this situation, the limiting angle of incidence angle  $\theta_L$  equals the critical angle of the total internal reflection, for which the refracted ray travels along the interface (angle  $\theta_2 = 90^\circ$ ); from Snell's law, the critical angle =  $\arcsin(n_2/n_1)$ . Rays incident at angles smaller than critical angle will undergo partial reflection and refraction; rays incident at angles equal to or larger than critical angle will undergo total internal reflection.

Let us now consider the pertinent geometry involved with a dielectric cylinder. A circular dielectric cylinder of refractive index  $n_1$  is embedded in a lower-refractive-index surround  $n_2$ . Light is incident upon the fiber end face at an angle  $\theta$  to the  $z$  or cylinder axis from a third medium of refractive index  $n_3 < n_1$  (Fig. 7b). We consider the meridional rays, that is, rays which intersect the fiber axis. There are



**FIGURE 7** (a) Reflection and refraction of light rays incident upon an interface separating two different media. (b) Ray optics analysis of light incident on an idealized cylindrical photoreceptor (see text for details).

three possibilities for meridional rays, shown by rays 1 through 3. Once inside the cylinder, ray 1 is incident upon the side wall at an angle smaller than the critical angle, is only partially reflected, and produces, at each reflection, a refracted ray that carries power away from the fiber. Ray 2 is critically incident and is thus totally reflected; this ray forms the boundary separating meridional rays that are refracted out of the fiber from meridional rays that trapped within the fiber. Ray 3 is incident upon the side wall at an angle greater than the critical angle, and is trapped within the fiber by total internal reflection.

Ray optics thus predicts a limiting incidence angle  $\theta_L$ , related to the critical angle via Snell's law, given by

$$n_3 \sin \theta_L = [(n_1)^2 - (n_2)^2]^{1/2} \quad (2)$$

for which incident meridional rays may be trapped within the fiber. Only those rays incident upon the fiber end face at angles  $\theta < \text{limiting angle } \theta_L$  are accepted; those incident at greater angles are quickly attenuated owing to refraction. The numerical aperture (NA) of the fiber is defined to be  $n_3 \sin \theta_L$ .













Application of meridional ray optics thus yields an acceptance property wherein only rays incident at  $\theta < \theta_L$  are trapped within the fiber. One thing to note is that the greater the angle of incidence is, the faster the attenuation of light will be through the fiber. This result is consistent with results obtained with electromagnetic waveguide modal analysis, which show that the farther away from a cutoff a mode is, the more tightly bound to the cylinder it is, and the closer its central ray is to the cylinder axis.

Geometric optics predicts that an optical fiber can trap and guide light; however, it does not provide an accurate quantitative description of phenomena on fibers having diameters of the order of the wavelength of visible light, where diffraction effects are important. Also, taking a simplified electromagnetic analysis of the photoreceptor waveguide model, illumination is assumed to be monochromatic, and the first Born approximation<sup>55</sup>—which ignores the contribution of waves that are scattered by the apertures—is used to match the incident illumination in the equivalent truncating aperture to the modes of the lossless guide that represents the myoid. Modes excited on the myoid are then guided to the tapered ellipsoid. Some of the initially excited radiation field will have leaked away before the ellipsoid is reached. Each segment is dealt with separately; thus, the ellipsoid will guide or funnel the power received from the myoid to the outer segment. Since each segment is dealt with separately, and because the modes of those segments are identical to those of an infinitely extended cylinder of the same radius and refractive index, the dielectric structure is considered to have an infinite uniform dielectric cylinder of radius  $p$  and refractive index  $n_1$  embedded in a lower-refractive-index surround  $n_2$ .

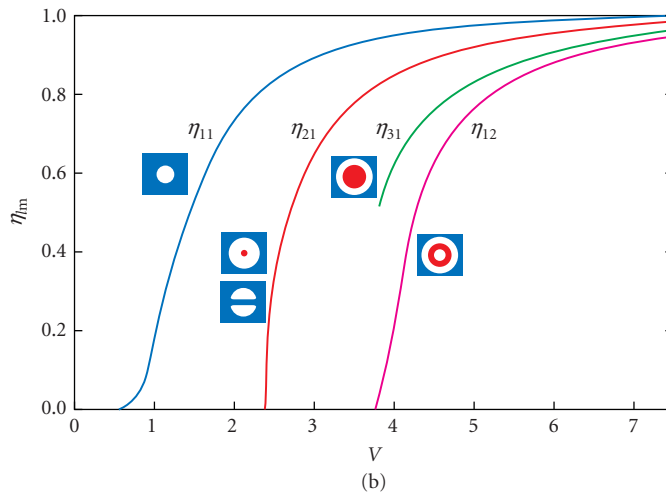
## 8.7 QUANTITATIVE OBSERVATIONS OF SINGLE RECEPTORS

The first observations of waveguide modal patterns in photoreceptors were made by Enoch.<sup>47,56–58</sup> Given that vertebrate photoreceptors act as waveguides, in order to treat the optics of a single receptor theoretically, it is modeled as a cylindrical waveguide of circular symmetry, with two fixed indices of refraction inside and outside the guide. Note that light is propagated along the guide in one or more of a set of modes, each associated with a characteristic three-dimensional distribution of energy density in the guide. In the ideal guide their distributions may be very nonuniform but will show a high degree of symmetry. Modes are grouped into orders, related to the number of distinct modal surfaces on which the electromagnetic fields go to zero. Thus, in Fig. 8, the lowest (first) order mode (designated  $HE_{11}$ ) goes to zero only at an infinite distance from the guide axis. Note that for all modes the fields extend to infinity in principle. At the terminus of the guide the fiber effectively becomes an antenna, radiating light energy into space. If one observes the end of an optical waveguide with a microscope, one sees a more or less complex pattern of light which roughly approximates a cross section through the modes being transmitted.

Modal patterns

Designations	Negative form	Cutoff
HE <sub>11</sub>		0.0
TE <sub>01</sub> , TM <sub>01</sub>		2.405
HE <sub>21</sub>		2.4+
(TE <sub>01</sub> or TM <sub>01</sub> ) + HE <sub>21</sub>		2.4+
HE <sub>12</sub>		3.812
EH <sub>11</sub>		3.812
HE <sub>31</sub>		3.8+
HE <sub>12</sub> + (HE <sub>31</sub> or EH <sub>11</sub> )		3.8+
HE <sub>13</sub> + EH <sub>11</sub>		3.8+
EH <sub>21</sub> or HE <sub>41</sub>		5.2
TE <sub>02</sub> , TM <sub>02</sub> , HE <sub>22</sub>		5.52
(TE <sub>02</sub> or TM <sub>02</sub> ) + HE <sub>22</sub>		5.52

(a)



(b)

**FIGURE 8** (a) Some commonly observed modal patterns found in human photoreceptors; their designations and V values. (b) The fraction of transmitted energy as a function of the V parameter for some commonly observed modal patterns.

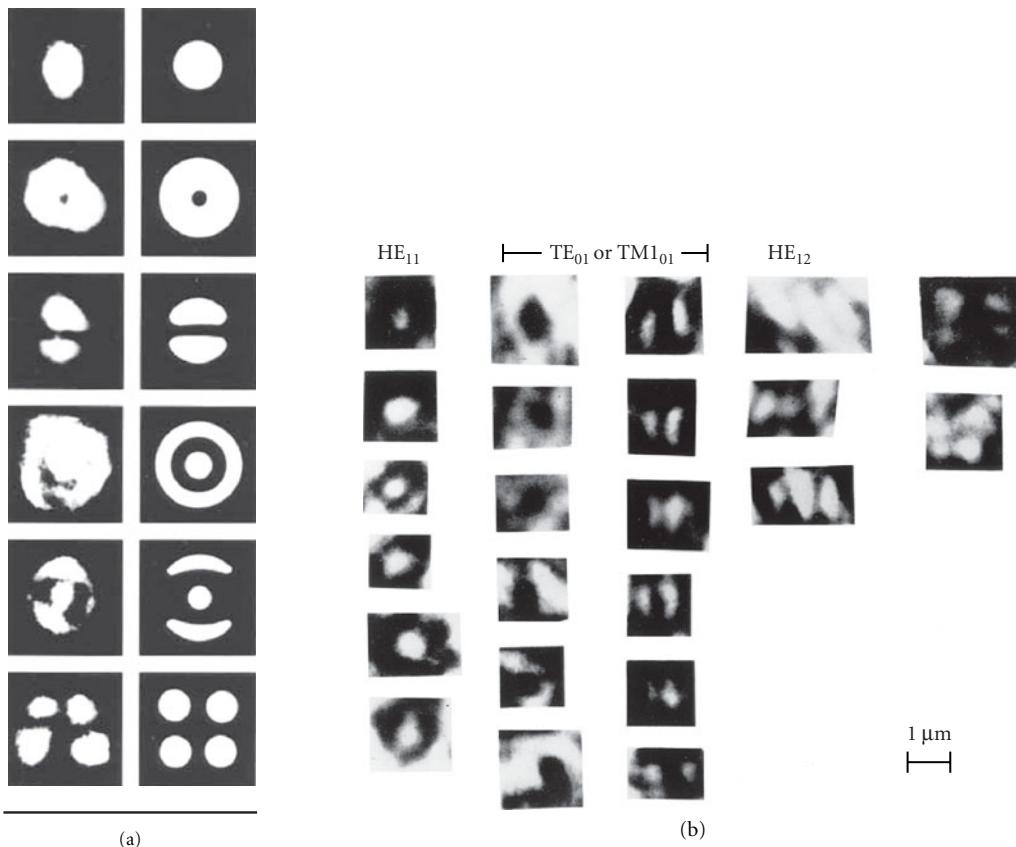
The optical properties of the ideal guide are completely determined by a parameter called the *waveguide parameter*  $V$ , defined by

$$V = (\pi d / \lambda) [(n_1)^2 - (n_0)^2]^{1/2} \quad (3)$$

where  $d$  is the diameter of the guide,  $\lambda$  the wavelength in vacuum, and  $n_1$  and  $n_0$  are the indices of refraction of the inside (core) and outside (cladding) of the fiber, respectively. For a given guide,  $V$  is readily changed by changing wavelength as shown above. With one exception, each of the various modes is associated with a cutoff value,  $V_c$ , below which light cannot propagate indefinitely along the fiber in that mode. Again, the lowest order,  $HE_{11}$ , mode does not have a cutoff value.  $V$  may be regarded as setting the radial scale of the energy distribution; for  $V \gg V_c$ , the energy is concentrated near the axis of the guide with relatively little transmitted outside the core. As  $V$  approaches  $V_c$ , the energy spreads out with proportionally more being transmitted outside the guide wall. A quantity  $\eta$  (not to be confused with the relative luminous efficiency defined for SCE I) has been defined that represents the fraction of transmitted energy which propagates inside the guide wall. For a given mode or combination of modes and given  $V$ ,  $\eta$  can be evaluated from waveguide theory. Figure 8a shows schematic representations of the mode patterns commonly observed in vertebrate receptors, cochlear hair cells, cilia (various), and the like, together with  $V_c$  for the corresponding modes. Figure 8b shows values of  $\eta$  as a function of  $V$  for some of the lower order modes.<sup>59</sup>

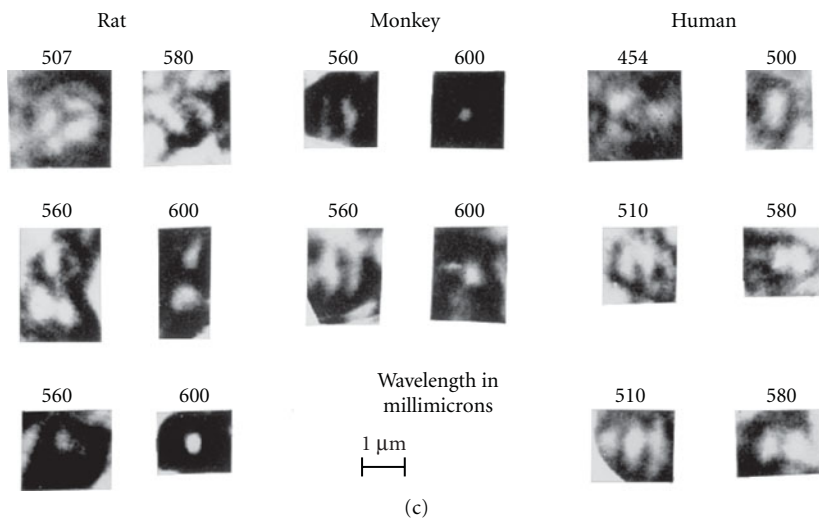
It is evident that an understanding of receptor waveguide behavior requires a determination of  $V$  for a range of typical receptors. An obvious straightforward approach is suggested, where if one could determine  $V$  at any one wavelength, it could be found for other wavelengths by simple calculation. How is this done? If one observes a flat retinal preparation illuminated by a monochromator so that the mode patterns are sharp, then by rapidly changing the wavelength throughout the visible spectrum, some of the receptors will show an abrupt change of the pattern from one mode to another. This behavior suggests that for these receptors  $V$  is moving through the value of  $V_c$  for the more complex mode resulting in a change in the dominant mode being excited. To be valid, this approach requires that the fraction of energy within the guide ( $\eta$ ) should fall abruptly to zero at cutoff. In experimental studies, it has been shown that  $\eta$  does go to zero at cutoff for the second-order modes characterized by bilobed or thick annular patterns as shown in Fig. 8. These modes include  $HE_{12}$ ,  $TE_{01}$ ,  $TM_{01}$ ,  $TM_{01}(TE_{01} + HE_{21})$ , and  $(TM_{01} + HE_{21})$ . Figure 9 displays some waveguide modes obtained in humans and monkeys, as well as variation with wavelength. In turn, it has been discovered that these second-order modal patterns are by far the most frequently observed in small-diameter mammalian receptors. Also, a detailed understanding of receptor-waveguide behavior requires accurate values for guide diameters. Another fact to note concerning parameter  $V$  is that the indices of refraction of the entire dielectric cylinder are fairly constant with no pronounced change in waveguide behavior as determined from past experimental studies. For reinforcement, considerable power is carried outside the fiber for low (above cutoff) values of  $V$ . In the limit as  $V \rightarrow$  infinity, the mode's total power is carried within the fiber.

The most complete study of the electrodynamic of visible-light interaction with the outer segment of the vertebrate rod based on detailed, first-principles computational electromagnetics modeling using a direct time integration of Maxwell's equation in a two-dimensional space grid for both transverse-magnetic and transverse-electric vector field modes was presented by Picket-May, Taflove, and Troy.<sup>60</sup> Detailed maps of the standing wave within the rod were generated (Fig. 10). The standing-wave data were Fourier analyzed to obtain spatial frequency spectra. Except for isolated peaks, the spatial frequency spectra were found to be essentially independent of the illumination wavelength. This finding seems to support the hypothesis that the electrodynamic properties of the rod contribute little if at all to the wavelength specificity of optical absorption.<sup>61</sup> Frequency-independent structures have found major applications in broad band transmission and reception of radio-frequency and microwave signals. As Picket-May et al.<sup>60</sup> point out, it is conceivable that some engineering application of frequency-independent, retinal-rod-like structures may eventually result for optical signal processing.



(a)

(b)



(c)

**FIGURE 9** (a) Human waveguide modal patterns and their theoretical, idealized equivalents. The bar represents  $10 \mu\text{m}$ . (b) Monkey waveguide modal patterns. The bar represents  $1 \mu\text{m}$ . (c) Changes in modal patterns induced by wavelength in three species.





**FIGURE 10** Computed figure showing the magnitude of the optical standing wave within a retinal rod at three different wavelengths (475, 505, and 714 nm). White areas are standing wave peaks, dark areas standing wave nulls. (From Ref. 26.)

## 8.8 WAVEGUIDE MODAL PATTERNS FOUND IN MONKEY/HUMAN RETINAL RECEPTORS

When a given modal pattern is excited, the wavelength composition of the energy associated with the propagation of that pattern in the retinal receptor (acting as a waveguide) may differ from the total distribution of energy at the point of incidence; for it is important to realize that the energy which is not transmitted is not necessarily absorbed by the receptor. It may be reflected and refracted out of the receptor. A distinct wavelength separation mechanism is seen to exist in those retinal



receptors observed. In some instances, one type of modal pattern was seen having a dominant hue. Where multiple colors were seen in a given receptor unit, those colors in some instances were due to changes in modal pattern with wavelength. The appearance of multiple colors, observed when viewing the outer segments, occurred in some receptors where mode coupling or interactive effects were present. There are differences in the wavelength distribution of the energy transmitted to the terminal end of the receptor. In addition, there are differences in the spatial or regional distribution of that energy within the outer segments as a function of wavelength due to the presence of different modal patterns and interactions governed by parameter  $V$ .<sup>62</sup>

If the same well-oriented central foveal receptors of humans are illuminated at three angles of obliquity of incident light, other changes are observed. It is observed subjectively that increased obliquity results in more red being transmitted, or less yellow and green being seen. Why does this occur? This phenomenon is dependent on the angle of incidence of light upon the receptor and the spectral sensitivity of the outer segment. First, the angle of incidence of light upon the receptor determines the amount of light transmitted through the receptor, as shown. Therefore, different values of  $V$  result in different amounts of light being collected for a fixed angle of incidence, as well as in different shapes, hence, in different widths, of the respective selectivity curves. Also, as the angle of incidence approaches the limiting angle, there is a decrease in the amount of bound-mode power of light transmitted through the receptor. In turn, absorption is greater for the larger-diameter receptors which, having higher  $V$  values, are more efficient at collecting light. So for higher amounts of light transmitted through the receptor with less contrast of different hues, the receptor has a high  $V$  value (large myoid diameter) and the light incident upon it is closer to normal to the myoid's surface versus its limiting angle. Another fact to note is that the absorption peak occurs at decreasing wavelengths for decreasing receptor radius, and the absorption curves broaden for decreasing radius of outer segment.

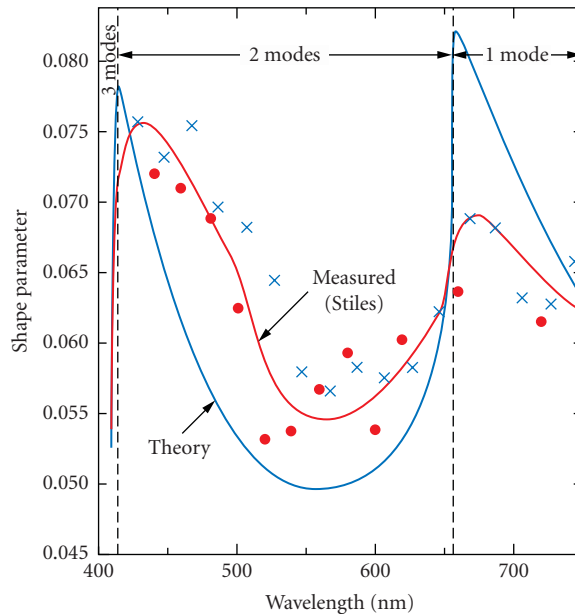
Also, since wavelength transmissivity, as well as total transmissivity, changes with obliquity of incidence, it may be further assumed that receptors that are not oriented properly in the retina will not respond to stimuli in the same manner as receptors that are oriented properly. In other words, disturbance in the orientation of receptors should result in some degree of anomalous color vision as well as visual acuity.<sup>55,63-65</sup>

It has been suggested that Henle fibers found in the retina may act as optical fibers directing light toward the center of the fovea.<sup>66</sup> Henle fibers arise from cone cell bodies, and turn horizontally to run radially outward from the center of the fovea, parallel to the retinal surface. At the foveal periphery, the Henle fibers turn again, perpendicular to the retinal surface, and expand into cone pedicles. Pedicles of foveal cones form an annulus around the fovea. Cone pedicles are located between 100 to 1000  $\mu\text{m}$  from the center of the fovea and the Henle fibers connecting them to cone photoreceptor nuclear regions can be up to 300  $\mu\text{m}$  long. If this hypothesis is correct, then Henle fiber-optic transmission could increase foveal irradiance, channeling short wavelength light from perifoveal pedicles to the central fovea and increasing the fovea's risk of photochemical damage from intense light exposure. This property has also been suggested as a reason for foveomacular retinitis burn without direct solar observation and could also be considered as a previously unsuspected risk in perifoveal laser photocoagulation therapy.

More recently, Franze et al.<sup>67</sup> investigated intact retinal tissue and individual Muller cells, which are radial glial cells spanning the entire thickness of the retina. Transmission and reflection confocal microscopy of retinal tissue *in vivo* and *in vitro* showed that these cells provide a low scattering passage for light from the retinal surface to the photoreceptors. In addition, using a modified dual beam laser trap, these researchers were able to demonstrate that these individual cells act as optical fibers. Their parallel array in the retina is analogous to fiberoptic plates used for low-distortion image transfer. It should be noted that these cells have an extended funnel shape, a higher refractive index than their surrounding tissue and are oriented along the direction of light propagation. Calculation of the  $V$  parameter showed that it varied from 2.6 to 2.9 (at 700 nm) for the different parts along the Muller cell, which is sufficiently high for low loss propagation of a few modes in the structure at this long wavelength. In the more intermediate wavelengths, the  $V$  varies from 3.6 to 4.0. Even though the refractive index and diameter of the cells change along their length, the  $V$  parameter and thus, their light guiding capability is very nearly constant. Unlike photoreceptors, muller cells do not have

the smooth cylindrical shape and have complex side branching processes. Their inclusion, using an “effective index” actually increases the  $V$  parameter. Therefore, even though Muller cells have a complex morphology, they can function as optical waveguides for visible light.

Modern theories of the Stiles-Crawford Effect I (SCE I) have used waveguide analysis to describe the effect. The most noteworthy is that of Snyder and Pask.<sup>68</sup> Snyder and Pask hypothesized an ideal “average” foveal cone with uniform dimensions and refractive indices. The cone is thought to be made up of two cylinders representing inner and outer segments. The response of a single cone is assumed to be a monotonically increasing function of absorbed light. The psychophysical response is assumed to be a linear combination of single-cone responses. Values of various parameters in the associated equations were adjusted to fit the wavelength variation of the SCE I reported by Stiles.<sup>69</sup> The model predicts the magnitude and trends of the experimental data (Fig. 11). The predictions, however, are highly dependent on the choice of refractive indices. Even small variations ( $\sim 0.5\%$ ) produce enormous changes in the results. The model is based not upon the absorption of light, but upon transmitted modal power. Other waveguide models include those of, for example, Alpern,<sup>70</sup> Starr,<sup>71</sup> and Wijngaard et al.<sup>72</sup> These papers deal with SCE II, the color effect. Additionally, Goyal et al.<sup>73</sup> have proposed an inhomogeneous waveguide model for cones. Vohnsen and his colleagues<sup>74,75</sup> have modified the existing models by making an important assumption, namely that instead of assuming that the light that propagates as rays in the eye to/from photoreceptors are propagated as waves. This implies that diffraction of both the incoming beam and the light emanated by the photoreceptors has to be included in the analysis. The mode field is approximated as a Gaussian. This is an important modification, since diffraction effects in relation to waveguide coupling have not been considered in previous models. Using this model it is possible to derive analytical expressions for the directionality parameters. The model also approximates the overall wavelength variation of directionality.



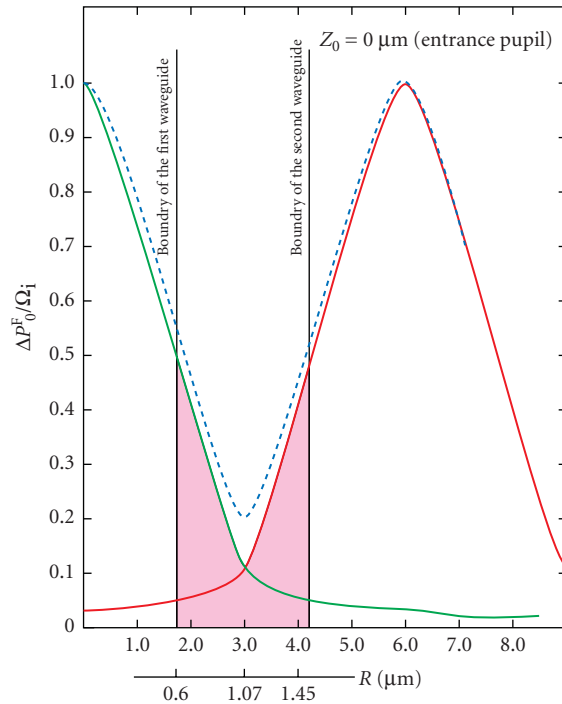
**FIGURE 11** The idealized cone model of Snyder and Pask. (From Ref. 31) Here the magnitude of the shape parameter (the curvature of the SCE function) is plotted as a function of the wavelength. The blue curve is the theoretical prediction and the red curve is plotted through Stiles’s experimental points. Note the qualitative agreement between theory and experiment.

When phase variations (say, due to aberrations) of the field incident at the photoreceptor aperture are considered, the model predicts that the finite width of photoreceptors not only leads to a slight broadening of the point spread function, but also reduces the impact of aberrations on the visual sensation that is produced. This may play a role in accommodation, since a defocused image couples less light.

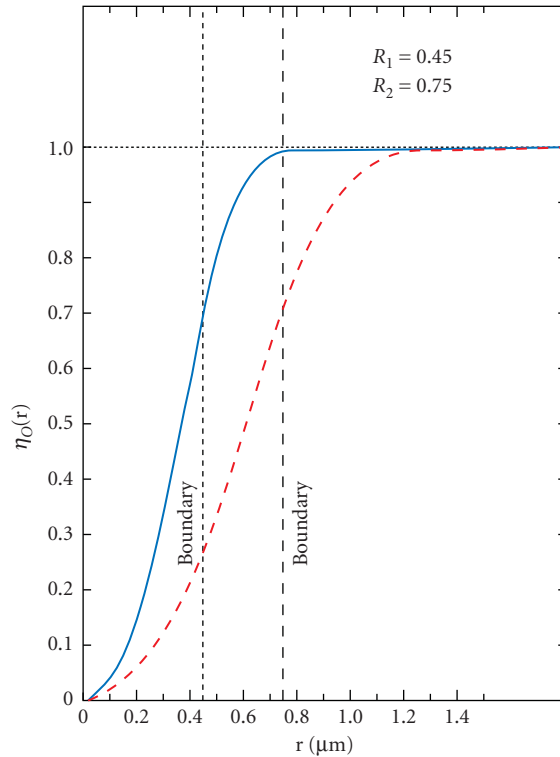
A caveat when considering waveguide models is that attention must be given to the physical assumptions made. These, along with uncertainties in the values of various physical parameters (e.g., refractive index), can seriously affect the results and hence the conclusions drawn from these models.

A formalism for modal analysis in absorbing photoreceptor waveguides has been presented by Lakshminarayanan and Calvo.<sup>76-78</sup> Here, an exact expression for the fraction of total energy confined in a waveguide was derived and applied to the case of a cylindrically symmetric waveguide supporting both the zeroth- and first-order sets of excited modes. The fraction of energy flow confined within the cylinder can be decomposed into two factors, one containing a  $Z$ -dependence ( $Z$  is direction of cylinder axis) as a damping term and the second as a  $Z$ -dependent oscillatory term. It should be noted that because of non-negligible absorption, Bessel and Hankel functions with complex arguments have to be dealt with. Using this formalism, if it is found that, using physiological data, the attenuation of the fraction of confined power increased with distance, the energy is completely absorbed for a penetration distance  $Z > 60 \mu\text{m}$ .

More interesting results are obtained when two neighboring waveguides separated by a distance of the order of the wavelength (a situation found in the foveal photoreceptor mosaic) are considered. An incoherent superposition of the intensity functions in the two waveguides is assumed. The results are shown in Fig. 12. The radial distribution of transmitted energy for two values of core radius ( $0.45 \mu\text{m}$



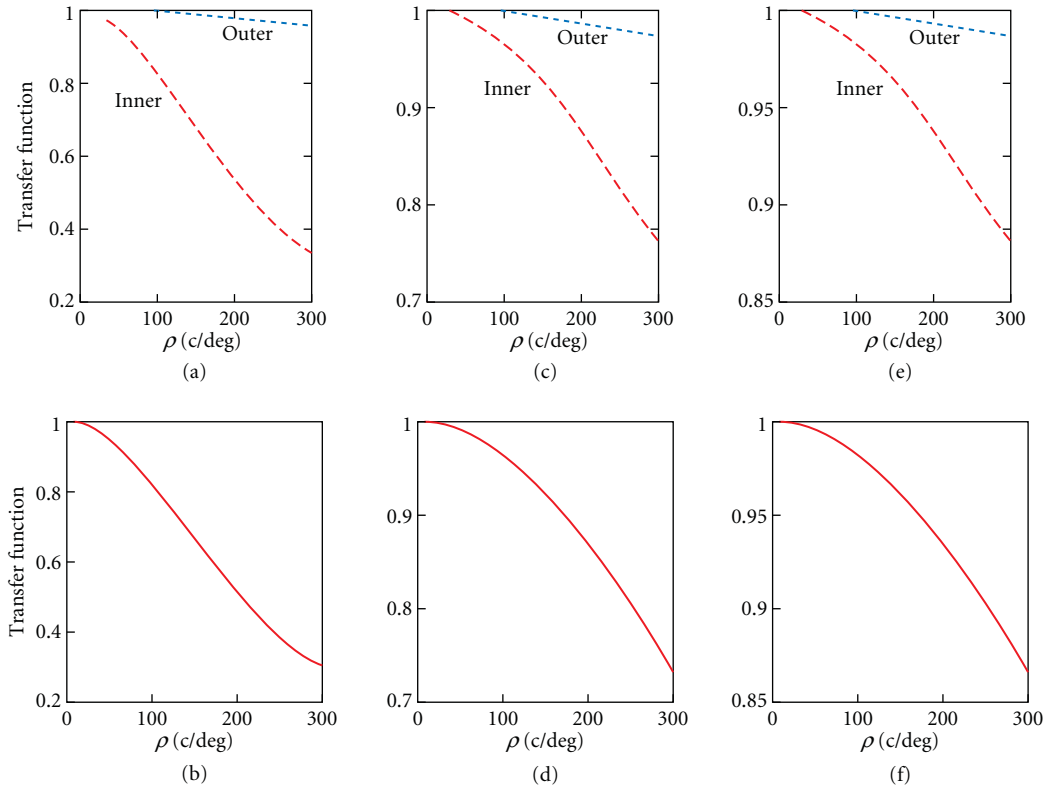
**FIGURE 12** Effect of two neighboring absorbing waveguide photoreceptors on the fraction of confined power per unit area. The pink area denotes the region in which the energy of both waveguides is “cooperative.”



**FIGURE 13** Flow of transmitted energy in a monomode absorbing waveguide for two value of the core radius (blue lines—0.45  $\mu\text{m}$ ; red line—0.75  $\mu\text{m}$ ) (see text for details).

and 0.75  $\mu\text{m}$ ) is shown in Fig. 13. It is seen that 70 percent of the energy is still confined within the waveguide boundary, with some percentage flowing into the cladding, producing an “efficient” waveguide with radius of the order of 0.75  $\mu\text{m}$  and 0.95  $\mu\text{m}$  in the two cases. This effect together with the incoherent superposition implies a mechanism wherein as light propagates along the waveguide, there is a two-way transverse distribution of energy from both sides of the waveguide boundary, providing a continuous energy flow while avoiding complete attenuation of the transmitted signal. Analysis also showed that for specific values of the modal parameters, some energy can escape in a direction transverse to the waveguide axis at specific points of the boundary. This mechanism, a result of the  $Z$ -dependent oscillatory term, also helps to avoid complete attenuation of the transmitted signal.

The spatial impulse response of a single-variable cross-section photoreceptor has been analyzed by Lakshminarayanan and Calvo.<sup>79</sup> The modal field propagating under strict confinement conditions can be shown to be written in terms of a superposition integral which describes the behavior of a spatially invariant system. Using standard techniques of linear systems analysis, it is possible to obtain the modulation transfer functions of the inner and outer segments. This analysis neglects scattering and reflection at the inner-outer segment interface. The transfer function depends on both the modal parameter defining the waveguide regime and spatial frequency (Fig. 14). It is seen that both inner and outer segments behave as low-pass filters, although for the outer segment a wide frequency range is obtained, especially for the longer wavelengths. Using a different analysis, Stacey and Pask<sup>80,81</sup> have shown similar results and conclude that photoreceptors themselves contribute to visual acuity



**FIGURE 14** Transfer function of the inner and outer segments (top) and the total transfer function of the photoreceptor for three wavelengths: 680, 550, and 460 nm. The values of the modal parameters are fixed. Horizontal axis is in units of cycles/degree.

through a wavelength-dependent response at each spatial frequency and that the response is dependent on the coherence of the source. They have used this model to study the Campbell effect. (When measuring human visual acuity with an incoherent source, Campbell,<sup>82</sup> noticed that the response decreased markedly when an artificial pupil placed in front of the eye was displaced perpendicular to the test fringes. There was no observed reduction when the pupil was displaced parallel to the fringes.) They conclude<sup>83</sup> that the waveguiding properties of photoreceptors make them sensitive to obliquely incident exciting waves and this provides some support for the hypothesis that both the SCE and the Campbell effect are manifestations of the same underlying waveguide nature of the photoreceptors.<sup>83,84</sup>

## 8.9 LIGHT GUIDE EFFECT IN COCHLEAR HAIR CELLS AND HUMAN HAIR

The vibration of the organ of Corti in response to sound is becoming evident at the cellular level using the optical waveguide property of cochlear hair cells to study the motion of single hair cells. It has been reported that conspicuous, well-defined bright spots can be seen on the transilluminated organ of Corti. The spots are arranged in the mosaic fashion of hair cells known from surface views of the organ of Corti.<sup>85</sup> Thus, the hair cells appear to act as light guides. Also, a definite property of

biological waveguide requires the extracellular matrix or fluid to be lower in refractive index as compared to the fiber (in this case, cochlear hair cell), and this property has been elucidated by phase-contrast micrographs of cochlear hair cells.<sup>86</sup>

Cochlear hair cells are transparent, nearly cylindrical bodies with a refractive index higher than the surrounding medium. Inner hair cells are optically denser than the surrounding supporting cells, and outer hair cells are optically denser than the surrounding cochlear fluid. Therefore, in optical studies, hair cells may be considered as optical fibers and the organ of Corti as a fiber-optics array. Even though the significance of the function of the organ of Corti in audition is understood, the role of waveguiding in hair cells is not clearly established.

To check for light-guide action, a broad light is usually used so that many hair cells show up simultaneously as light guides. For broad beams the conditions of light entrance into individual cells are not specified. Light can enter hair cells along the cylindrical wall. Part of such light is scattered at optical discontinuities inside the cells. To a large extent scattered light is trapped in cells by total internal reflection. Thus, to maximize contrast for transillumination with broad beams, no direct light is allowed to enter the microscope. Areas with hair cells appear dark. However, the heads of hair cells appear brightly illuminated owing to the light-collecting effect of hair cells on scattered light. Also, the diameter of cochlear hair cells is from 5 to 9  $\mu\text{m}$ , that is, 10 to 20 times the wavelength of light. In fiber optics, fibers with diameters larger than 10 to 20 times the wavelength are considered multimodal because they support a large number of characteristic modes of light-energy propagation. Fibers with diameters smaller than 10 to 20 times the wavelength support fewer modes, and discrete modes of light-energy propagation can be launched more readily. Because of this division, cochlear hair cells fall in between typically single-mode and multimode optical fibers.<sup>87</sup> Also, an important parameter for the evaluation of fibers with small diameters is the parameter  $V$ . Recall that the smaller parameter  $V$ , the smaller the number of discrete modes a given fiber is capable of supporting. In addition, parameter  $V$  is greater for cochlear hair cells than for retinal receptor cells mainly because they have larger diameters than retinal rods and cones (see above for parameter  $V$ ).

Gross features of surface preparations of the organ of Corti can be recognized with the binocular dissecting microscope. When viewed, the rows of hair cells can be seen as strings of bright pearls. Bright spots in the region of hair cells are due to light transmission through hair cells. Bright spots disappear when hair cells are removed.

Helmholtz's reciprocity theorem of optics<sup>55</sup> (reversibility of light path) also applies to hair cells where exit radiation pattern is equivalent to entrance radiation pattern. There is no basic difference in light-guide action when the directions of illumination and observation are reversed compared to the normal directions. Obviously, Helmholtz's theorem applies to visual photoreceptors too, since they are just modified cilia.

It is also seen that the detection of light guidance is shown to strongly correlate with the viewing angle, where incident light hits the numerical aperture of the hair cell, which is 30 to 40°. When the incident light is normal to the numerical aperture of the hair cell, higher transmittance of light and contrast is observed; however, at more oblique angles to the numerical aperture of the hair cell, less contrast is observed because the angle of incident light is approaching the limiting angle, which is a limit to total internal reflection and more refraction or scattering of rays occurs through the fiber.

A useful tool for studying the quality and preservation of preparations of organ of Corti utilizes the concept that hair cells show little change in appearance under the light microscope for up to 20 to 40 min after the death of an animal. Consequently, poor fixation and preservation of the organ of Corti can be recognized immediately at low levels of magnification.

To study the modal light intensity distributions on single hair cells, researchers use a monocular microscope. Hair cells exercise relatively little mode selection, which is to be expected because of their relatively large diameters (large parameter  $V$ ). As the monocular microscope is focused up and down individual hair cells, it is seen that the intensity distributions in different cross sections of hair cells may differ. Focusing above the exit of hair cells shows that the near-field radiation pattern may differ from the intensity distribution on the cell exit. Furthermore, other studies have shown that the modal distribution of light intensity can be changed drastically for individual cells by changing the angle of light incidence relative to the axis of hair cells. Also, several low-order patterns on the exit end of single hair cells are easily identifiable and are designated according to the customary notation.

As a note, since cochlear hair cells show relatively little mode selection, it is not always possible to launch a desired mode pattern on a given cell with ordinary light. In order to restrict the spectrum of possible modes, it is advantageous to work with monochromatic, coherent light, that is, laser light. However, the identification of patterns relative to the cells was found to be difficult. This is due to the fact that laser stray light is highly structured (speckled), which makes it difficult to focus the microscope and to recognize cell boundaries. Light guidance in cochlear hair cells may be considered to be a general research tool for cochlear investigations in situations where the organ of Corti can be transilluminated and where relatively low levels of magnification are desirable.

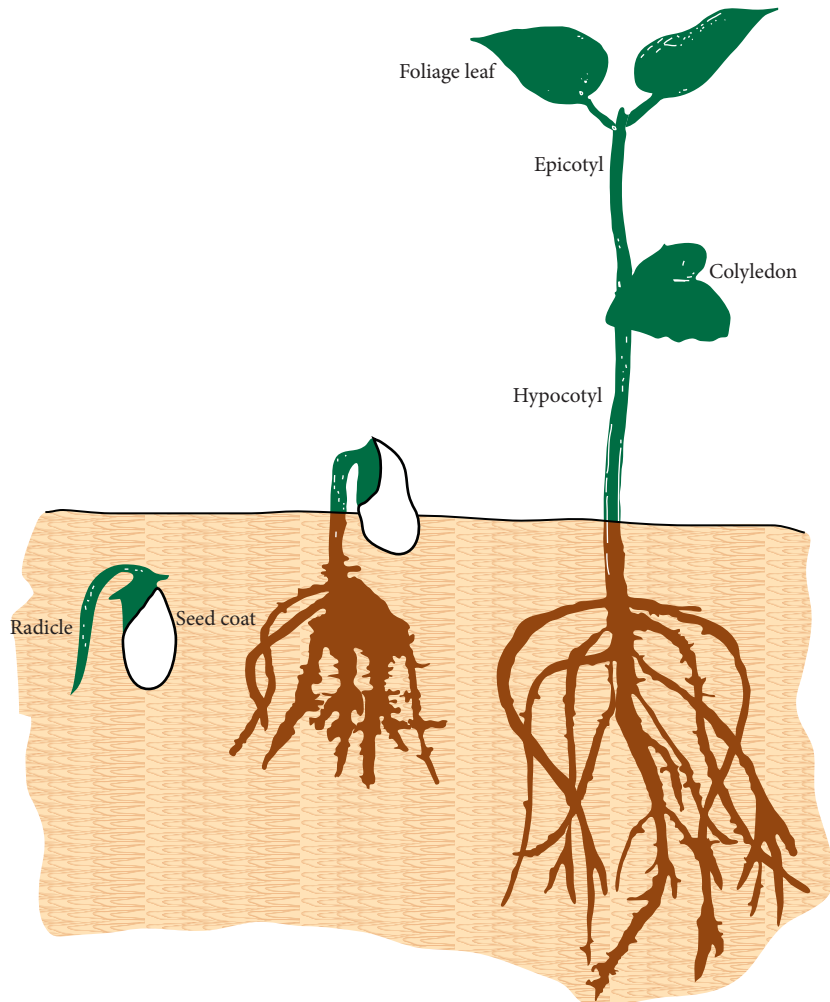
Furthermore, biological waveguide effects have also been shown to occur in human hair,<sup>88</sup> which are similar to cochlear hair cell waveguide effects. There is a reduction in light intensity with increased hair length, but by far the most important factor was hair color. Little or no light was transmitted by brown hair, while gray hair acts as a natural fiber optic which can transmit light to its matrix, the follicular epithelium, and to the dermis. Interestingly, whether light transmission down hairs affects skin and hair needs investigation. As pointed out by Enoch and Lakshminarayanan,<sup>89</sup> many major sensory systems incorporate cilia in their structure. In the eye, the invagination of the neural groove of the surface ectoderm of the embryo leads to the development and elaboration of the nervous system. The infolding and incorporation of surface tissue within the embryo accounts for the presence of ciliated cells within nervous tissue neurons, which after maturation form photoreceptors containing cilia, lining the outer wall of the collapsed brain ventricle that will form the retina (see Oyster<sup>90</sup> for a description of the embryology and growth of the retina). A study of the evolution of the relationships between cilia and specific sensory systems would be of great interest.

## 8.10 FIBER-OPTIC PLANT TISSUES

Dark-grown (etiolated) plant tissues are analogous to multiple fiber-optic bundles capable of coherent transfer of light over at least 20 mm; hence, they can transmit a simple pattern faithfully along their length. Each tissue examined can accept incident light at or above certain angles relative to normal as is expected of any optical waveguide. The peak of the curve that describes this angular dependence, the acceptance angle, and the overall shape of the curve appear to be characteristic of a given plant species and not of its individual organs.<sup>91</sup> Knowledge of this optical phenomenon has permitted localization of the site of photoreception for certain photomorphogenic responses in etiolated oats. Before the discussion of the fiber-optic properties of etiolated plant tissues, it is important to have a general idea of what plant structures are presently being discussed to get a full understanding of the structure to function relationship (Fig. 15). To define, the *epicotyl* is a region of shoot above the point of attachment of the cotyledons, often bearing the first foliage leaves. The *hypocotyl* is the part of the embryo proper below the point of attachment of the cotyledons; it will form the first part of the stem of the young plant. The *coleoptile* is a cylindrical sheath that encloses the first leaves of seedlings of grasses and their relatives.

Total internal reflection of light should be independent of fluence rate since this phenomenon is simply a special case of light scattering. The amount of light emerging from the tissue segment was measured first at the highest fluence rate available and then again when the incident light beam had been attenuated with one or more calibrated neutral density filters over a range of 4.5 absorbance units. The amount of light axially transmitted at a given fluence rate (expressed as a percentage of the amount transmitted at the highest fluence rate for individual tissue segments) decreased log-linearly as filters of increasing absorbance were used for oat, mung bean, and corn tissues. Therefore, light guiding in etiolated tissues is clearly fluence rate-independent over at least 4.5 orders of magnitude.<sup>92</sup>

Light guiding through segments of etiolated root, stem, and coleoptiles, plus primary leaves from several species, all display the spectral dependence expected of any light scattering agent, with low transmission in the blue relative to that in the far-red regions of the spectrum. As tissue length increases, the difference between transmissions in the blue and far-red regions of the spectrum increases.



**FIGURE 15** General structure of plant tissues.

It has been found that the spectra obtained when the incident beam of white light is applied to the side of the intact plant are indistinguishable from those obtained when the cut end of a tissue is irradiated. Spectra taken on tissue segments excised from different regions of the plant may show traces of certain pigments present only in those regions. Whereas there are no differences between spectra of the apical and basal portions of oat mesocotyls, the hook regions of mung bean hypocotyls contain small amounts of carotenoids, which is indicated as a small dip in the transmittance curve that is absent in spectra from the lower hypocotyl or hypocotyl-root transition zone. Oat coleoptilar nodes, which join the mesocotyl to the coleoptile, decrease light transmission from one organ to another when light is passed from the coleoptile to the mesocotyl or from the mesocotyl to the coleoptile, respectively, but show a spectral dependence which is independent of the direction of light transmission based on Helmholtz's reciprocity theorem of light. Older etiolated leaves do not transmit red light axially relative to the cylindrical coleoptile which sheaths them. Consequently, the presence or absence of the primary leaves does not alter the spectrum of the etiolated coleoptile.



Plant tissues mainly consist of vertical columns of cells. Light internally reflected through them passes predominantly along these columns of cells rather than across columns when the incident beam is at or above the tissue acceptance angle.<sup>91</sup> As optical fibers, the tissues are about 10 percent as effective as glass rods of comparable diameter, and about 1 percent as effective as commercial fiber optics.<sup>92</sup> Within a given cell, light could be traveling predominantly in the cell wall, the cytoplasm, the vacuole, or any combination of these three compartments. Observations of the cut ends of tissue segments that are light-guiding indicate that the cell wall does not transmit light axially but that the cell interior does. Also, plant cells contain pigments that are naturally restricted to the vacuole or to the cytoplasmic organelles. Mung bean hypocotyls contain carotenoids (absorption maxima near 450 nm), which are located in the etioplasts, organelles found only in the cytoplasm, whereas etiolated beet and rhubarb stems are composed of cells with clear cytoplasm and vacuoles containing betacyanin and/or anthocyanin (absorption maxima near 525 to 530 nm). Spectra of the light guided through these tissues show that light guiding occurs both through the cytoplasm and the vacuole. The relative size of each compartment, the pigment concentration, and the refractive indices involved will largely determine the amount of light that a given compartment will transmit. As the seedlings are exposed to light and become green, troughs due to absorption by chlorophylls (near 670 nm) appear. Completely greened tissue has very little light-guiding capacity except in the far-red region of the spectrum.

Also, the red to far-red (R/FR) ratio determines the fraction of phytochrome in the active state in the plant cell.<sup>93</sup> This ratio plays an important role in the regulation of growth. Similarly, the blue to red (B/R) ratio may be important for responses mediated by phytochrome alone when transmission is low in the red relative to the blue, or for those mediated by both phytochrome and a pigment absorbing blue light. The R/FR ratio of the light emitted from several curved, etiolated plant tissues decreases as the tissue length increases. However, the patterns of change in the B/R ratio differ with each tissue examined. All B/R ratios presumably decrease along the first 20 mm of tissues but in mung hypocotyl (stem) this ratio then increases, whereas in oat mesocotyl it increases and then decreases further. Changes in the R/FR and B/R ratios also occur in a given length of tissue as it becomes green. The R/FR ratio does not change in tissues that remain white (i.e., oat mesocotyl) but decreases dramatically in those that synthesize pigments (i.e., oat coleoptile or mung bean, beet, and rhubarb stems). The B/R ratios, unlike the R/FR ratios, tend to increase as the tissues become green, reflecting a smaller decrease in transmission at 450 nm than 660 nm as pigments accumulate.

All of these results demonstrate certain properties of light-guiding effects in plants: the fluence-rate independence of light guiding in plants; an increase in tissue length showing a more pronounced difference between transmission in the blue and far-red regions of the spectra; and the effect of the greening status on the pertinent R/FR and B/R ratios for several species of plants grown under a variety of conditions. Certainly, both the B/R and the R/FR ratios change dramatically in light that is internally reflected through plants as etiolated tissue length is increased and as greening occurs. Finally, the sensitivity of etiolated plant tissues to small fluences of light and the inherent capacity of plants to detect and respond to small variations in the R/FR ratio indicate that light guiding occurring in etiolated plants beneath the soil and the concomitant spectral changes that occur along the length of the plant may be of importance in determining and/or fine-tuning the plant's photomorphogenic response.<sup>94,95</sup>

## 8.11 SPONGES

Recently, researchers at Bell Laboratories studied the fiber optical features of a deep sea organism, namely, the glass sponge (*Euplectella*, "venus flower basket").<sup>96</sup> The spicules of these sponges have some remarkable fiber optic properties. The skeleton of the hexactinellid class of sponges is constructed from amorphous hydrated silica. The sponge has a lattice of fused spicules that provide extended structural support. The spicules are 5 to 15 cm long and 40 to 70  $\mu\text{m}$  in diameter. The spicules have a characteristic layered morphology and a cross-sectional variation in composition: a pure silica core of about 2  $\mu\text{m}$  in diameter that enclosed an organic filament, a central cylinder with maximum organic material and a striated shell that has a gradually decreasing organic content.

The researchers conducted interferometric index profiling and found that corresponding to the three regions, there was a core with a high refractive index that is comparable to or higher than that of vitreous silica, a cylinder of lower refractive index that surrounds the core and an oscillating pattern with progressively increasing refractive index at the outer part of the spicule. These embedded spicules act as single or a few-mode waveguides. When light was coupled to free standing spicules, they functioned as multimode fibers.

These fibers are similar to commercial telecommunication fibers in that they are made of the same material and have comparable dimensions. They also function as efficient as single-mode, few-mode, or multi-mode, fibers depending upon optical launch conditions.

In addition, it is possible that these spicules in addition to providing structural anchorage support, could also act as a fiber optic network for distributing light in the deep sea environment. The fascinating aspect of these fibers is that they are made under ambient conditions, as opposed to commercial manufactured fibers which require very high temperatures (about 1700°C). Organic ligands at the exterior of the fiber seem to protect it and provide an effective crack-arresting mechanism and the fibers seem to be doped with specialized impurities that improve the refractive index profile and hence the waveguiding properties.

## 8.12 SUMMARY

In short, we have pointed out some similar relationships between the theoretical waveguide models and the actual biological waveguides such as vertebrate photoreceptors, cochlear hair cells of the inner ear, human hair (cilia), and fiber-optic plant tissues. The similar relationships between the theoretical model of waveguides and actual biological waveguides were demonstrated as (a) almost entirely following the Stiles-Crawford effect of the first kind (SCE-I) for vertebrate photoreceptors, (b) exhibiting wavelength sensitivity of transmission, and (c) conforming to Helmholtz's reciprocity theorem of optics. Of special interest is the fact that visual photoreceptors exhibit self-orientation. Similar phototropic self-orientation is exhibited by growing and fully-grown plants (e.g., sunflowers). For theoretical or possible applications, the light-guiding properties of biological waveguides can be used not only to determine the structure of such guides, but also to provide the basis for determining the functioning of other structures related to these guides. Lastly, these studies which give us greater insight into low temperature, biologically inspired processes could possibly result in better fiber optic materials and networks for commercial applications such as telecommunications.

## 8.13 REFERENCES

1. W. S. Stiles and B. H. Crawford, "The Luminous Efficiency of Rays Entering the Eye Pupil at Different Points," *Proc. R. Soc. Lond. B.* **112**:428–450 (1933).
2. A. W. Snyder and R. Menzel, *Photoreceptor Optics*, Springer-Verlag, New York, 1975.
3. J. M. Enoch and F. L. Tobey Jr., *Vertebrate Photoreceptor Optics*, vol. 23, *Springer Series in Optical Science*. Springer-Verlag, Berlin, 1981.
4. J. M. Enoch and V. Lakshminarayanan, "Retinal Fiber Optics," in W. N. Charman (ed.), *Vision and Visual Dysfunction: Visual Optics and Instrumentation*, vol. 1, MacMillan Press, London, 1991, pp. 280–309.
5. V. Lakshminarayanan, "Waveguiding in Retinal Photoreceptors: An Overview," *Proc. SPIE* **3211**:182–192 (1998).
6. V. Lakshminarayanan, J. M. Enoch, and A. Raghuram, *The Stiles-Crawford Effects*, Classic Reprints on CD-ROM, vol. 4, Optical Society of America, Washington, D.C. (2003).
7. J. Dowling, *The Retina: An Approachable Part of the Brain*, Harvard University Press, Cambridge, MA, 1987.
8. R. W. Rodieck, *The First Steps in Seeing*, Sinauer, Sunderland, MA, 1998.

9. W. S. Stiles, "The Luminous Efficiency of Monochromatic Rays Entering the Eye Pupil at Different Points and a New Color Effect," *Proc. R. Soc. Lond. B* **123**:90–118 (1937).
10. J. M. Enoch and W. S. Stiles, "The Color Change of Monochromatic Light with Retinal Angle of Incidence," *Optica Acta* **8**:329–358 (1961).
11. M. F. Land, "The Optics of Animal Eyes," *Contemp. Physiol.* **29**:435–455 (1988).
12. M. F. Land, "Optics and Vision in Invertebrates," in H. Autrum (ed.), *Handbook of Sensory Physiology*, vol. VII/6B, Springer-Verlag, Berlin, 1981, pp. 471–592.
13. R. Sidman, "The Structure and Concentration of Solids in Photoreceptor Cells Studied by Refractometry and Interference Microscopy," *J. Biophys. Biochem. Cytol.* **3**:15–30 (1957).
14. J. M. Enoch and F. L. Tobey Jr., "Use of the Waveguide Parameter  $V$  to Determine the Differences in the Index of Refraction between the Rat Rod Outer Segment and Interstitial Matrix," *J. Opt. Soc. Am.* **68**(8):1130–1134 (1978).
15. R. A. Applegate and V. Lakshminarayanan, "Parametric Representation of Stiles-Crawford Function: Normal Variation of Peak Location and Directionality," *J. Opt. Soc. Am. A* **10**(7):1611–1623 (1993).
16. D. I. A. MacLeod, "Directionally Selective Light Adaptation: A Visual Consequence of Receptor Disarray?" *Vis. Res.* **14**:369–378 (1974).
17. S. A. Burns, S. Wu, F. C. DeLori, and A. E. Elsner, "Variations in Photoreceptor Directionality across the Central Retina," *J. Opt. Soc. Am. A* **14**:2033–2040 (1997).
18. A. Roorda and D. Williams, "Optical Fiber Properties of Human Cones," *J. Vision* **2**:404–412 (2002).
19. J. Krauskopf, "Some Experiments with a Photoelectric Ophthalmoscope," *Excerpta Medica International Congress Series* **125**:171–181 (1965).
20. J. M. Gorrand and F. C. DeLori, "A Reflectometric Technique for Assessing Photoreceptor Alignment," *Vis. Res.* **35**:999–1010 (1995).
21. P. J. deLint, T. T. J. M. Berendschott, and D. van Norren, "Local Photoreceptor Alignment Measured with a Scanning Laser Ophthalmoscope," *Vis. Res.* **37**:243–248 (1997).
22. G. J. van Blokland and D. van Norren, "Intensity and Polarization of Light Scattered at Small Angles from the Human Fovea," *Vis. Res.* **26**:485–494 (1986).
23. G. J. van Blokland, "Directionality and Alignment of the Foveal Receptors Assessed with Light Scattering from the Fundus in Vivo," *Vis. Res.* **26**:495–500 (1986).
24. S. A. Burns, S. Wu, J. C. He, and A. Elsner, "Variations in Photoreceptor Directionality Across the Central Retina," *J. Opt. Soc. Am. A* **14**:2033–2040 (1997).
25. S. Marcos and S. A. Burns, "Cone Spacing and Waveguide Properties from Cone Directionality Measurements," *J. Opt. Soc. Am. A* **16**:2437–2447 (1999).
26. J. M. Enoch, J. S. Werner, G. Haegerstrom-Portnoy, V. Lakshminarayanan, and M. Rynders, "Forever Young: Visual Functions Not Affected or Minimally Affected by Aging: A Review," *J. Gerontol. Biol. Sci.* **54**(8):B336–351 (1999).
27. E. Campos, J. M. Enoch, and C. R. Fitzgerald, "Retinal Receptive Field Like Properties and the Stiles-Crawford Effect in a Patient with a Traumatic Choroidal Rupture," *Doc. Ophthalmol.* **45**:381–395 (1978).
28. V. Lakshminarayanan, J. M. Enoch, and S. Yamade, "Human Photoreceptor Orientation: Normals and Exceptions," in *Advances in Diagnostic Visual Optics*, A. Fiorentini, D. L. Guyton, and I. M. Siegel, (eds.), Springer Verlag, Heidelberg, Germany, 1987, pp. 28–32.
29. V. Lakshminarayanan, *The Stiles Crawford effect in Aniridia*, Ph.D. Dissertation, University of California, Berkeley (1985).
30. M. S. Eckmiller, "Defective Cone Photoreceptor Cytoskeleton, Alignment, Feedback and Energetics can Lead to Energy Depletion in Macular Degeneration," *Prog. Retin Eye Res.* **23**:495–522 (2004).
31. V. Lakshminarayanan, J. E. Bailey, and J. M. Enoch, "Photoreceptor Orientation and Alignment in Nasal Fundus Ectasia," *Optom. Vis. Sci.* **74**(12):1011–1018 (1997).
32. S. S. Choi, J. M. Enoch, and M. Kono, "Evidence for Transient Forces/Strains at the Optic Nervehead in Myopia: Repeated Measurements of the Stiles Crawford Effect of the First Kind (SCE-1) over Time," *Ophthalm. Physiol. Optics* **24**:194–206 (2004).
33. X. Zhang, Ye. Ming, A. Bradley, and L. Thibos, "Apodization of the Stiles Crawford Effect Moderates the Visual Impact of Retinal Image Defocus," *J. Opt. Soc. Am. A* **16**(4):812–820 (1999).

34. L. Martin, *Technical Optics*, vol. 2, Pitman, London, UK, 1953.
35. D. Atchison, D. Scott, and G. Smith, "Pupil Photometric Efficiency and Effective Center," *Ophthalm. Physiol. Optics* **20**:501–503 (2001).
36. W. S. Baron and J. M. Enoch, "Calculating Photopic Illuminance," *Amer. J. Optometry Physiol. Optics* **59**(4): 338–341 (1982).
37. H. Metcalf, "Stiles-Crawford apodization," *J. Opt. Soc. Am.* **55**:72–74 (1965).
38. J. P. Carrol, "Apodization Model of the Stiles-Crawford Effect," *J. Opt. Soc. Am.* **70**:1155–1156 (1980).
39. M. Rynders, L. Thibos, A. Bradley, and N. Lopez-Gil, "Apodization Neutralization: A New Technique for Investigating the Impact of the Stiles-Crawford Effect on Visual Function," in: *Basic and Clinical Applications of Vision Science*, V. Lakshminarayanan, (ed.), Kluwer, Dordrecht, The Netherlands, 1997, pp. 57–61.
40. D. H. Scott, D. A. Atchison, and P. A. Pejeski, "Description of a Method for Neutralizing the Stiles-Crawford Effect," *Ophthalm. Physiol. Optics* **21**(2):161–172 (2001).
41. D. A. Atchison, A. Joblin, and G. Smith, "Influence of Stiles-Crawford Apodization on Spatial Visual Performance," *J. Opt. Soc. Am. A* **15**(9):2545–2551 (1998).
42. D. A. Atchison, D. H. Scott, A. Joblin, and F. Smith, "Influence of Stiles-Crawford Apodization on Spatial Visual Performance with Decentered Pupils," *J. Opt. Soc. Am. A* **18**(6):1201–1211 (2001).
43. S. Marcos, S. Burns, E. Moreno-Barriuso, and R. Navarro, "A New Approach to the Study of Ocular Chromatic Aberrations," *Vis. Res.* **39**:4309–4323 (2000).
44. J. van de Kraats, T. J. M. Berendschot, and D. van Norren, "The Pathways of Light Measured in Fundus Reflectometry," *Vis. Res.* **36**:2229–2247 (1996).
45. T. T. J. Berendschot, J. van de Kraats, and D. van Norren, "Wavelength Dependence of the Stiles-Crawford Effect Explained by Perception of Backscattered Light from the Choroid," *J. Opt. Soc. Am. A* **18**(7):1445–1451 (2001).
46. S. Marcos and S. A. Burns, "On the Symmetry between Eyes of Wavefront Aberration and Cone Directionality," *Vis. Res.* **40**:2437–4580 (2000).
47. J. M. Enoch, "Optical Properties of Retinal Receptors," *J. Opt. Soc. Am.* **53**:71–85 (1963).
48. W. H. Miller and A. W. Snyder, "Optical Function of Myoids," *Vis. Res.* **12**(11):1841–1848 (1972).
49. B. Burnside and C. King-Smith, "Retinomotor Movements," in L. Squire, *New Encyclopedia of Neuroscience*, Elsevier, in press 2008.
50. A. W. Snyder and J. D. Love, *Optical Waveguide Theory*, Chapman and Hall, London, 1983.
51. H. von Helmholtz, *Treatise in Physiological Optics*, Dover, New York, 1962, p. 229.
52. A. Hannover, *Vid. Sel. Naturv. Og Math. Sk.* **X** (1843).
53. B. O'Brien, "A Theory of the Stiles Crawford Effect," *J. Opt. Soc. Am.* **36**(9):506–509 (1946).
54. R. Winston and J. M. Enoch, "Retinal Cone Receptors as an Ideal Light Collector," *J. Opt. Soc. Am.* **61**(8):1120–1122 (1971).
55. M. Born and E. Wolf, *Principles of Optics*, 6th ed., Pergamon Press, Oxford, 1984.
56. J. M. Enoch, "Waveguide Modes in Retinal Receptors," *Science* **133**(3461):1353–1354 (1961).
57. J. M. Enoch, "Nature of the Transmission of Energy in the Retinal Photoreceptor," *J. Opt. Soc. Am.* **51**:1122–1126 (1961).
58. J. M. Enoch, "Optical Properties of the Retinal Receptor," *J. Opt. Soc. Am.* **53**:71–85 (1963).
59. K. Kirschfeld and A. W. Snyder, "Measurement of Photoreceptors Characteristic Waveguide Parameter," *Vis. Res.* **16**(7):775–778 (1976).
60. M. J. May Pickett, A. Taflove, and J. B. Troy, "Electrodynamics of Visible Light Interaction with the Vertebrate Retinal Rod," *Opt. Letters* **18**:568–570 (1993).
61. A. Fein and E. Z. Szuts, *Photoreceptors: Their Role in Vision*, Cambridge University Press, Cambridge, UK, 1982.
62. J. M. Enoch, "Receptor Amblyopia," *Am. J. Ophthalmol.* **48**:262–273 (1959).
63. V. C. Smith, J. Pokorny, and K. R. Diddie, "Color Matching and the Stiles-Crawford Effect in Observers with Early Age-Related Macular Changes," *J. Opt. Soc. Am. A.* **5**:2113–2121 (1988).
64. V. C. Smith, J. Pokorny, J. T. Ernest, and S. J. Starr, "Visual Function in Acute Posterior Multifocal Placoid Pigment Epitheliopathy," *Am. J. Ophthalmol.* **85**:192–199 (1978).

65. F. W. Campbell and A. H. Gregory, "The Spatial Resolving Power of the Human Retina with Oblique Incidence," *J. Opt. Soc. Am.* **50**:831 (1960).
66. M. A. Mainster, "Henle Fibers May Direct Light toward the Center of the Fovea," *Lasers and Light in Ophthalmol.* **2**:79–86 (1988).
67. K. Franze, J. Grosche, S. N. Skatchkov, S. Schinkinger, C. Foja, D. Schiold, O. Uckermann, K. Travis, A. Reichenbach, and J. Guck, "Muller Cells are Living Optical Fibers in the Vertebrate Retina," *Proc. Natl. Acad. Sciences (USA)* **104**(20):8287–8292 (2007).
68. A. W. Snyder and C. Pask, "The Stiles Crawford Effect—Explanation and Consequences," *Vis. Res.* **13**(6):1115–1137 (1973).
69. W. S. Stiles, "The Directional Sensitivity of the Retina and the Spectral Sensitivities of the Rods and Cones," *Proc. R. Soc. Lond. B* **127**:64–105 (1939).
70. M. Alpern, "A Note on Theory of the Stiles Crawford Effects," in J. D. Mollon and L. T. Sharpe (eds.), *Color Vision: Physiology and Psychophysics*, Academic Press, London, 1983, pp. 117–129.
71. S. J. Starr, *Effects of Luminance and Wavelength on the Stiles Crawford Effect in Dichromats*, Ph.D. dissertation, Univ. of Chicago, Chicago, 1977.
72. W. Wijngaard, M. A. Bouman, and F. Budding, "The Stiles Crawford Color Change," *Vis. Res.* **14**(10):951–957 (1974).
73. I. C. Goyal, A. Kumar, and A. K. Ghatak, "Stiles Crawford Effect: An Inhomogeneous Model for Human Cone Receptor," *Optik* **49**:39–49 (1977).
74. B. Vohnsen, I. Iglesias, and P. Artal, "Guided Light and Diffraction Model of Human-Eye Photoreceptors," *J. Opt. Soc. Am. A* **22**(11):2318–2328 (2005).
75. B. Vohnsen, "Photoreceptor Waveguides and Effective Retinal Image Quality," *J. Opt. Soc. Am. A* **24**:597–607 (2007).
76. V. Lakshminarayanan and M. L. Calvo, "Initial Field and Energy Flux in Absorbing Optical Waveguides II. Implications," *J. Opt. Soc. Am.* **A4**(11):2133–2140 (1987).
77. M. L. Calvo and V. Lakshminarayanan, "Initial Field and Energy Flux in Absorbing Optical Waveguides I. Theoretical formalism," *J. Opt. Soc. Am.* **A4**(6):1037–1042 (1987).
78. M. L. Calvo and V. Lakshminarayanan, "An Analysis of the Modal Field in Absorbing Optical Waveguides and Some Useful Approximations," *J. Phys. D: Appl. Phys.* **22**:603–610 (1989).
79. V. Lakshminarayanan and M. L. Calvo, "Incoherent Spatial Impulse Response in Variable-Cross Section Photoreceptors and Frequency Domain Analysis," *J. Opt. Soc. Am.* **A12**(10):2339–2347 (1995).
80. A. Stacey and C. Pask, "Spatial Frequency Response of a Photoreceptor and Its Wavelength Dependence I. Coherent Sources," *J. Opt. Soc. Am.* **A11**(4):1193–1198 (1994).
81. A. Stacey and C. Pask, "Spatial Frequency Response of a Photoreceptor and Its Wavelength Dependence II. Partial Coherent Sources," *J. Opt. Soc. Am.* **A14**:2893–2900 (1997).
82. F. W. Campbell, "A Retinal Acuity Directional Effect," *J. Physiol. (London)* **144**:25P–26P (1958).
83. C. Pask and A. Stacey, "Optical Properties of Retinal Photoreceptors and the Campbell Effect," *Vis. Res.* **38**(7):953–961 (1998).
84. J. M. Enoch, "Retinal Directional Resolution" in J. Pierce and J. Levine (eds.), *Visual Science*, Indiana University Press, Bloomington, IN, 1971, pp. 40–57.
85. H. Engstrom, H. W. Ades, and A. Anderson, *Structural Pattern of the Organ of Corti*, Almqvist and Wiskell, Stockholm, 1966.
86. R. Thalmann, L. Thalmann, and T. H. Comegys, "Quantitative Cytochemistry of the Organ of Corti. Dissection, Weight Determination, and Analysis of Single Outer Hair Cells," *Laryngoscope* **82**(11):2059–2078 (1972).
87. N. S. Kapany and J. J. Burke, *Optical Waveguides*, New York, Academic Press, 1972.
88. J. Wells, "Hair Light Guide" (letter), *Nature* **338**(6210):23 (1989).
89. J. M. Enoch and V. Lakshminarayanan, "Biological Light Guides" (letter), *Nature* **340**(6230):194 (1989).
90. C. Oyster, *The Human Eye: Structure and Function*, Sinauer, Sunderland, MA, 1999.
91. D. F. Mandoli and W. R. Briggs, "Optical Properties of Etiolated Plant Tissues," *Proc. Natl. Acad. Sci. USA* **79**:2902–2906 (1982).

92. D. F. Mandoli and W. R. Briggs, "The Photoreceptive Sites and the Function of Tissue Light-Piping in Photomorphogenesis of Etiolated Oat Seedlings," *Plant Cell Environ.* 5(2):137–145 (1982).
93. W. L. Butler, H. C. Lane, and H. W. Siegelman, "Nonphotochemical Transformations of Phytochrome, In Vivo," *Plant Physiol.* 38(5):514–519 (1963).
94. D. C. Morgan, T. O'Brien, and H. Smith, "Rapid Photomodulation of Stem Extension in Light Grown *Sinapis Alba*: Studies on Kinetics, Site of Perception and Photoreceptor," *Planta* 150(2):95–101 (1980).
95. W. S. Hillman, "Phytochrome Conversion by Brief Illumination and the Subsequent Elongation of Etiolated *Pisum* Stem Segments," *Physiol. Plant* 18(2):346–358 (1965).
96. V. C. Sundar, A. D. Yablon, J. L. Grazul, M. Ilan, and J. Aizenberg, "Fiber Optical Features of a Glass Sponge," *Nature*, 424:899–900 (2003).

*This page intentionally left blank*

---

# THE PROBLEM OF CORRECTION FOR THE STILES-CRAWFORD EFFECT OF THE FIRST KIND IN RADIOMETRY AND PHOTOMETRY, A SOLUTION

---

Jay M. Enoch

*School of Optometry  
University of California at Berkeley  
Berkeley, California*

Vasudevan Lakshminarayanan

*School of Optometry and Departments of Physics and Electrical Engineering  
University of Waterloo  
Waterloo, Ontario, Canada*

---

## 9.1 GLOSSARY\*

---

**Aniridia.** (Greek “an” = without; “iridia” = iris.) One or both eyes of a patient who is either born without an iris (usually this is a bilateral anomaly), or, for whatever reason the individual has no iris (e.g., as a result of disease, or trauma, or surgery, etc.). Note: In some cases, on examination, a very tiny remnant or very small iris nubbin may be present in congenital cases.

**Aniseikonia.** (Greek “anis” = unequal; “eikon or ikon” = image.) The result of aniseikonia, or the presence of unequal image sizes in the two eyes, is the formation of characteristic spatial distortions in binocular vision. If this anomaly is large enough, there is a loss of fusion of the two images.

**Aphakia.** (Greek “a” = not or without; “phakia” = lens.) The state of an eye after surgical removal of an eye lens. In some cases, the eye lens may become displaced rather than removed (this used to be seen commonly with an older form of cataract surgery), or for example, as a complication of certain diseases such as Marfan’s syndrome. Today, the excised eye lens is commonly replaced with a plastic “intraocular lens” (see Chap. 21).

---

\*Added useful glossaries: please also refer to glossaries of Chap. 8, “Biological Waveguides” by Vasudevan Lakshminarayanan and Jay M. Enoch and Chap. 6, “The Maxwellian View with an Addendum on Apodization” by Gerald Westheimer in this volume and to Chap. 37, “Radiometry and Photometry for Vision Optics” by Yoshi Ohno in Vol. II.



**Keratoconus.** This is an anomaly of the cornea which results in a cone-like forward-projection of the central or near-central cornea (usually slightly decentered). It is also often accompanied by progressive thinning of the area within and surrounding the cone.

**Stiles-Crawford Effect of the first kind (SCE-1).** The directional sensitivity of the retina. SCE-1 is the result largely of photoreceptor waveguide properties (see Chap. 8). We test vision for beams of light entering the eye at different points in the entrance pupil of the eye, and then refracted to the retina by the optical system of the eye. *The SCE-1 affects perceived brightness of luminous objects.*

**Stiles-Crawford Effect of the second kind (SCE-2).** In addition to perceived brightness effects, there occur alterations in perceived hue and saturation of objects viewed. This is associated with the angle of incidence of the viewed beam at the retinal receptor (e.g., see Refs. 45 and 46).

**Optical Stiles-Crawford effect.** Distribution of reverse path (reflected) luminous energy distribution assessed in the entrance pupil of the eye by reverse path. It is radiant energy which had been transmitted to and through the retina and has been reflected back through the ocular media after having been reradiated after reflection by the photoreceptor waveguides. It is assumed that these retinal receptors are preferentially aligned with the exit pupil of the eye. Also present are back-reflected scattered and nonguided light, for example, see Refs. 25–27.

## 9.2 INTRODUCTION

### The Troland

In the vision science literature, the stimulus to vision is defined as the product of (1) the luminance of the object of regard assessed in the plane of the entrance pupil of the eye, expressed in candelas/m<sup>2</sup>, times (2) the area of the entrance pupil of the eye expressed in mm<sup>2</sup>. These values are expressed in either photopic or scotopic “trolands.” The unit is named after the late Dr. Leonard Troland.<sup>1–3</sup>

### It Is Time for a Reasoned Reassessment of “the Troland,” the Unit of Retinal Illuminance

Leonard Troland, 1889 to 1932, died in a tragic accident as a relatively young man. He left behind a respectable body of research which led to the modern approach to assessment of retinal illuminance, the accepted unit used to specify the stimulus to vision.<sup>1–3</sup> Please also refer to Dr. Yoshi Ohno’s discussion in Chap. 37 in the section on “Radiometry and Photometry for Vision Optics,” in Vol. II. He discusses current practices. We also call attention to the recent paper by Cheung et al. which addresses the standard units of optical radiation.<sup>4</sup>

There remains, of course, the larger issue of considering whether the currently employed means of estimating/assessing the stimulus to vision (often expressed today as “the photopic or scotopic troland”) might be improved and/or modified in order to bring this form of analysis, and measurement units employed, more effectively into parallel with other modern measures of radiometry and photometry.

Please note the spectral response of the eye, when exposed to brighter/more intense stimuli exciting the cone-dominant-*photopic*-visual-response-system ( $V_\lambda$ ), differs from the rod-receptor-dominant-*scotopic*-visual-response-system ( $V'_\lambda$ ). The latter determination(s) pertain to less intense visual stimuli. Readers are referred to the book by Wyszecki and Stiles for further discussion.<sup>5</sup>

An “equivalent troland” also has been defined. This unit makes use of specification of the SCE-1 in the entrance pupil where the stimulus has 50 percent effectiveness relative to the level of response determined at the SCE-1 peak location.<sup>6–9</sup> The more complete SCE-1 function, etc., can be inferred using data obtained with this approach.

## The Photometric Efficiency Factor

For completeness, we call attention to another approach to integration of and utilization of SCE-1 data for photometric and other purposes. This is based on the photometric efficiency (PE) factor.<sup>10,11</sup> The PE is defined by the following ratio:

$$\frac{\text{Effective light collected by pupil}}{\text{Actual light collected through the pupil}} \quad (1)$$

This is an interesting concept, but it is not easy to compute.<sup>10</sup> What is sought here is determination of “an equivalent entrance pupil size” where all points in the entrance pupil of the eye would theoretically contribute equally to the visual stimulus. Particularly for photopic vision, this equivalent pupil, would be smaller than the measured entrance pupil of the eye.<sup>10,11</sup> By using this approach, an integrated “equivalent entrance pupil” of given diameter and area may be determined. In addition, an adjusted SCE-1 centrum<sup>10,11</sup> is provided by developing appropriate weighting factors for the asymmetries encountered. Please remember, for different size pupils, the center of the entrance pupil may change by as much as 0.5 mm with dilation of the pupil.<sup>12</sup> For the same dilation of the eye pupil, usually this centrum is quite stable.

## Current Practice

The troland,<sup>1-3</sup> as it is ordinarily computed (also see Chap. 37, “Radiometry and Photometry for Vision Optics,” by Yoshi Ohno in Vol. II), does not take into consideration the Stiles-Crawford effects, nor factors associated with existing blur of the retinal image, nor uncorrected blur encountered when large entrance pupil areas are included in the experimental design (using a large natural, or a dilated pupil of the eye, or a decentered eye pupil). It is in our collective interest to enhance the quality of determinations, to improve the precision of measurements, to reconsider units used, and to correct these deficiencies.

## The Argument Presented in This Chapter

Here we address a number of aspects of these issues, and offer an approach to the solution of resultant photometric problems. We suggest incorporation of the integrated Stiles-Crawford effect of the first kind (SCE-1)<sup>13-16</sup> in photometric analyses *when degradation of images caused by peripheral portions of the optical elements of the eye (the cornea and the eye lens) are/have been effectively neutralized*. Even though this is still not a perfect solution to the specification of the visual stimulus, it does represent an advance relative to techniques employed today. This will allow us to move effectively toward a finer estimate of the magnitude of the visual stimulus seen either by the experimental subject or patient tested.

## 9.3 THE PROBLEM AND AN APPROACH TO ITS SOLUTION

### Statement of the Problem

When one seeks to determine effective “visual stimulation induced by radiant energy” within the visible spectrum (i.e., *light*, the stimulus to vision),<sup>9</sup> and integrates all energy entering within the entrance pupil of the eye,<sup>13</sup> for example, one needs to consider the Stiles-Crawford effect of the first kind (SCE-1).<sup>13-16</sup> SCE-1 is also known as the directional sensitivity of the retina (see Chap. 8). The issues considered here have rarely been treated adequately in experimental procedures. *If one integrates data derived from point-by-point SCE-1 determinations (obtained by measuring responses obtained when using tiny apertures imaged at a number of different locations in the entrance pupil of the eye)*

*the resultant integrated function may not predict properly the visual result obtained when different diameter areas in the entrance pupil of the eye are sampled* (please see further arguments presented in this chapter).

Most modern assessments of the integrated SCE-1 are based on point-by-point measurements of SCE-1 determined experimentally at a number of discrete points across the entrance pupil of the eye in a number of meridians (commonly tested in two meridians, i.e., horizontal and vertical). These are then fitted by a parabolic function, proposed in 1937, by *both* Stiles (using a log base 10 relationship)<sup>14,15</sup> and Crawford (employing a  $\log_e$  or  $\ln$  base  $e$  equation),<sup>16</sup> and that function is then integrated around the entrance pupil. Other functions have been proposed, for example, a Gaussian function by Safir and Hyams,<sup>17</sup> Burns et al.<sup>18</sup> as well as the relationship used in figures in this chapter by Enoch.<sup>19,20</sup> A review of various analytical expressions used to fit SCE-1 data is discussed in Applegate and Lakshminarayanan.<sup>21</sup>

In his dissertation, Lakshminarayanan recorded SCE-1 functions in a unique patient. This was a young lady born aniridic (without an iris), and (importantly) she had developed only marginal nystagmus (i.e., approximately  $\frac{1}{4}$  mm amplitude nystagmus—in others, it is almost always of greater amplitude). She had quite normal visual acuity. In this eye, not only was the SCE-1 recorded, but, in her case, it was possible to assess better the associated side lobes of SCE-1.<sup>22–24</sup> Assuming (1) her data correspond to characteristic population (central lobe) SCE-1 patterns in normal observers, (2) then none of the above cited equations adequately conforms to the SCE-1 coupled with *the side lobes recorded*.

Notes: (1) In most experiments measuring the SCE-1 performed in recent years, very small-size projections of the aperture stop of the test apparatus have been imaged in the plane of the observer's entrance pupil (i.e., they are usually less than 1 mm in diameter). (2) This discussion is limited to monocular testing. (3) Measured SCE-1 functions obtained vary with wavelength (see Fig. 8). (4) We do not consider here recent discussions of reflected, re-emitted, and projected light gathered from the retinal photoreceptor waveguides which can be assessed by reverse path irradiation/illumination. This relationship is commonly termed “the optical SCE-1” (see Glossary and Refs. 25–27).

## Confounds

Fundamentally, when an observer views a display or a scene in an eye with a relatively large natural eye pupil, or with a dilated iris resulting in an enlarged pupil (for purposes considered here, let us say this occurs when the entrance pupil *diameter* is greater than 3 mm), one must consider *both the SCE-1 and additional blur effects* present in the peripheral portions of the cornea and the eye lens. These two factors affect perceived brightness of a stimulus!

For simplicity, assume we are testing at photopic levels, the refraction of the individual has been appropriately corrected for spherical and astigmatic errors (these are first- and second-order aberrations), and the observer is properly accommodated (or corrected visually) for the object viewed. When point-to-point assessments for SCE-1 in the entrance pupil of the eye are properly determined and analyzed, *the resultant uncertainty occurring is largely assignable to the remaining/residual blur of the retinal image* (see data and figures). And as will be seen, this uncertainty is associated with blur induced by peripheral corneal and eye lens aberrations. In discussions of related topics, the senior author has referred to this blur factor as *a dirty variable*, that is, this is an uncontrolled or poorly controlled factor both in vision research and in clinical practice. To further simplify this discussion, we do not address chromatic aberrations (these factors can be addressed/considered separately), nor the use of binocular stimuli. Similarly, here, we do not consider a variety of anomalous optical conditions encountered in clinical settings (e.g., keratoconus, cataracts, or aphakia, etc.), or decentered and/or misshapen irises/pupils, or intraocular lenses centered or decentered, refractive surgery and its complications, movements/decentration(s) of contact lenses, binocular anomalies, etc. (see Chaps. 13, 14, 16, 20, and 21).

Both the magnitude of blur encountered and the nature of a blurred image vary with accommodation of the eye, and, of course, with mismatches between the retinal plane and the plane of focus of the retinal image formed within the eye. We restate the point that the center of the entrance pupil of the eye varies naturally with pupillary dilation, and is virtually always decentered to some extent relative to the optical axis of the eye. Further, the optical axis of the cornea as best determined is not the

same as that of the eye lens (e.g., Refs. 28–30). Simply, the optical system of the eye is not a perfectly centric system. We also know that dilation does not affect the SCE-1 function, *per se* (e.g., Ref. 31). For radiometric and/or photometric purposes, we seek the very best result possible depending on the level of accuracy required for the task.

## A Bit of History

When Walter “Stanley” Stiles and Brian “Hewson” Crawford first reported the SCE effect 75 years ago in 1933,<sup>13</sup> their results were based on data obtained from their own eyes. And it is apparent that, by chance, their eyes manifested little blur of the retinal image, and hence, the complications one often encounters during integration of SCE-1 in the entrance pupil of the eye were largely not encountered in their data.<sup>13</sup> In such relatively rare individuals, the point-by-point estimates obtained through SCE-1 studies predict well the results of integrating such data across the entrance pupil of the eye.<sup>19,20</sup> (For example, see the example given under “Sample Data from Enoch” in Sec. 9.4 for experimental subject B.W.)<sup>19,20</sup>

On the other hand, if there is meaningful degradation of the retinal image caused by peripheral corneal and lenticular aberrations of the eye resulting in image blur, this factor, on its own, can alter the effective integrated perceived luminance of the visual stimulus (e.g., Refs. 18, 19, 32–37). The results found need not be the same between eyes or individuals, nor as predictable as SCE-1.

Collectively, these points have been demonstrated in number of studies, particularly in those of Enoch and coworkers<sup>18,19,32,33</sup> and Drum.<sup>34,35</sup>

## Today, A Solution of the Problem Is Possible by Using (in Addition) Adaptive Optics (AO) Techniques

What is needed is a joint test of the Stiles-Crawford function using a suitable SCE-1 apparatus (e.g., Ref. 38) *and* utilization of one of a number of modern adaptive optics (AO) devices (see Chap. 15) capable of determining and correcting the incident wavefront in order to deliver a corrected or a nearly corrected image at the retinal test locus sampled. By such an approach, we have the ability to determine more accurately the magnitude of the visual stimulus for the designated entrance pupil size. This is achieved by correcting blur (and higher-order aberrations), and adjusting or correcting for the SCE-1, and properly integrating the resultant data. Earlier, except for use of small entrance pupil diameters or imaged apertures in experiments (see Chap. 6), such corrections were not always possible. Note, devices utilizing adaptive optics are moving toward incorporation of means of correction for chromatic aberrations in their designs or overall experimental techniques.

## If a Nonmonochromatic Stimulus to Vision Is Employed

For a more complete treatment of stimuli which are not monochromatic, one might use an achromatic approach/technique proposed by Powell<sup>39</sup> (see Chap. 15), or other special lens(es) such as designed and used by the late Gunter Wyszecki to correct the chromatic aberrations of the human eye (e.g., Ref. 5). Using the Wyszecki lens, JME made such adjustments in some of his earlier research, and found this approach quite useful. Please note, the Wyszecki lens must be aligned carefully with the optic axis.

## A Reasonable Estimate Is Possible without Full Application of Adaptive Optics

By correcting refraction, *and* spherical aberration in annular zones within the eye pupil, and by using monochromatic stimuli, in 1956, Enoch was able to approximate a more complete correction of blur in the eyes of his experimental subjects.<sup>19,20</sup>

A rather similar research project to that of Enoch was conducted by Drum just a few years later.<sup>34,35</sup>

These added issues relate to image formation in the retinal plane, and the assumption is made (often implied, but rarely stated) that energy is transferred from a single plane into the rod and cone photoreceptor waveguides. Please realize that the living cell is a highly complex entity (e.g., Ref. 43). Cell morphology and dimensions (particularly cones) often alter quite a bit across small/modest distances on the retina. Clearly, diffraction and interference effects, as well as rapid changes occurring in the spatial retinal image itself, are occurring across small distances about the entrance to the photoreceptor waveguides. For example, it is useful to consider image alterations occurring over small distances about a plane of focus (both longitudinally and laterally) of the retinal image. As but one example, consider the contents of a paper by Bachinski and Bekkefi,<sup>40</sup> and data recorded by Enoch and Fry<sup>41</sup> in greatly enlarged and simplified rod and cone receptor models assessed in the microwave spectrum. Without addressing such issues further, here we suggest that a somewhat better estimate of the retinal stimulus to vision *per se* is achievable by incorporating the integrated SCE-1 function in a blur-corrected or blur-minimized retinal image. We argue that this provides a better assessment of the visual stimulus than that provided in what has become now a “classical” approach defined by the late Leonard Troland many years ago, and which is still used for assessing retinal illuminance.<sup>2,3</sup> See Chap. 37, “Radiometry and Photometry for Vision Optics,” by Yoshi Ohno in Vol. II. So saying, there is more that needs consideration relative to this topic in future.

## 9.4 SAMPLE POINT-BY-POINT ESTIMATES OF SCE-1 AND INTEGRATED SCE-1 DATA

### Sample Data from Stiles and Crawford, 1933

In this section, data are presented which demonstrate reasonably the magnitude and nature of the errors induced in calibration and measurement without correction of the defined errors encountered. While these are not extreme errors, *per se*, these are real effects, they are readily correctable, and should be considered in studies of vision.

Data here are taken from the dissertation of Enoch<sup>19,20</sup> and the function utilized for integrations of “relative directional sensitivity” here indicated as “ $\eta$ ” (plotted as a  $\log_{10}$  scale) are defined as

$$\eta = 0.25(1 + \cos 9.5 \theta)^2 \quad (2)$$

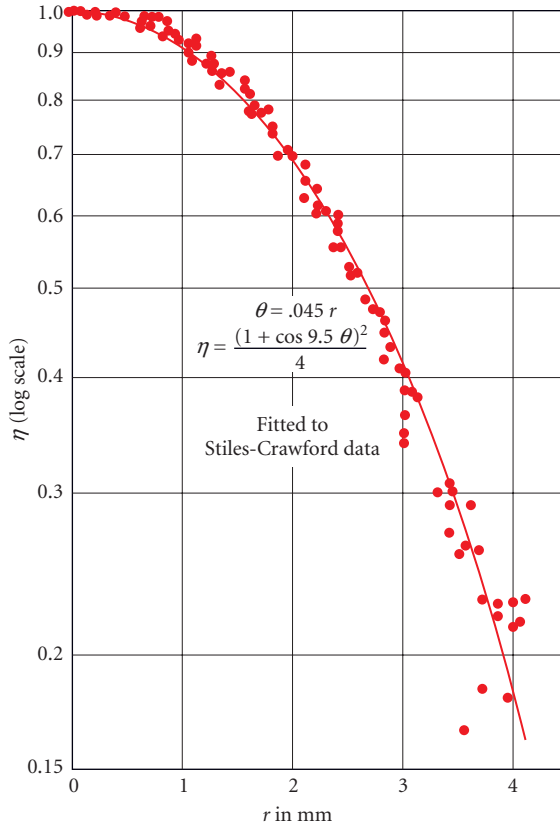
$\eta$  equals 1.0 at the peak of the SCE-1 curve;  $\theta$  is the angle of incidence of the ray of light at the retinal plane. A 1-mm beam displacement at the entrance pupil of the eye alters the angle of incidence of the incident ray of light by 2.5° of oblique incidence at the retinal plane (based on computations using the constants of the Gullstrand schematic eye).<sup>23</sup>

This particular equation provides a bit better fit of SCE-1 data<sup>13–16</sup> than the Stiles, 1937 (log base 10) relationship<sup>14</sup> or Crawford’s 1937 (ln base e) equation<sup>16</sup> for a parabola. It is used here to simplify transfer of the figures employed below. Because the improvements made when fitting SCE-1 data using this relationship were modest, the authors of this chapter have not used this format for some years, and, rather, have used the more broadly employed Stiles 1937 formulation<sup>14</sup> in its stead

$$\text{Log}_{10} \eta = -\rho r^2 \quad (3)$$

In this equation “ $\eta$ ” is the measured relative visual sensitivity of the stimulus to vision (in trolands,  $\log_{10}$  scale); “ $\rho$ ” is a constant defining the curvature of this parabolic function; and “ $r$ ,” in mm, is the distance of the point being sampled from the peak of the measured SCE-1 function in the plane of the entrance pupil of the eye.

SCE-1 data shown in Fig. 1 were taken from Stiles and Crawford<sup>13</sup> and were fitted by Eq. (2).



**FIGURE 1** The 1933 SCE-1 data of W. S. Stiles and B. H. Crawford were adjusted such that the maxima of measured relative visual sensitivity functions were located at  $r = 0.0$  mm in their entrance pupils, and were assigned a relative sensitivity value of 1.0 (at/as the peak of these data functions). Radial distance settings, “ $r$ ,” of the test beam in the entrance pupils of their eyes from the peaks of SCE-1 are displayed on the abscissa. The ordinate displays  $\eta$  ( $\log_{10}$  scale). Data used in this illustration originated in the paper of W. S. Stiles and B. H. Crawford, 1933.<sup>13</sup> (This illustration was copied from Fig. 1, Enoch, 1956, Dissertation; and Fig. 1, Enoch, 1958, J.O.S.A.<sup>19,20</sup> These figures in this paper are reproduced with permission of the author and the publisher of J.O.S.A.)

In their initial study, Stiles and Crawford, 1933,<sup>13</sup> were seeking to measure precisely the area of the eye pupil (i.e., their instrument had been designed as a straightforward pupillometer). They assumed, at the outset, that energy passing through each part of the eye pupil contributed equally and proportionately to the retinal image and visual sensitivity. They discovered that the device gave results which were not consistent with their *à priori* assumption, and they properly interpreted their results to indicate that there was evidence for the presence of directional sensitivity of the retina, which became known as the Stiles-Crawford effect of the first kind (SCE-1).

Their instrument sought to assess the integrated visual response resulting from irradiating the entire entrance pupil, and for different size entrance pupils of the eye. *Since they had little image blur in their own eyes*, the integrated result gave little evidence that when the full pupil was measured, the resultant additivity (of the contributions of different parts of the pupil) might also be different in eyes where there were meaningful peripheral corneal and eye lens aberrations resulting in additional degradation (blur) of the retinal image. That is, they did not encounter other than near perfect additivity from integrating point-by-point SCE-1 determinations and comparing them to full eye-pupil assessments. Thus, we need to differentiate between (1) simple addition of the contributions of the entire eye pupil, versus (2) that sum adjusted for SCE-1, and (3) *the result obtained in eyes where there are meaningful aberrations in the periphery of the optical elements of the eye.*

### Sample Data from Enoch

Data were obtained from three well-trained graduate student subjects.<sup>19,20</sup> Subject B.W. had an eye with fine optical properties. Both subjects R. V. and A. M. had greater peripheral ocular aberrations in their measured eyes than B. W.; with subject A. M. exhibiting somewhat greater image degradation than observer R. V. Note these are all photopic determinations. By avoiding scotopic measurements here, we simplify the argument.<sup>42</sup> Except for overall refractive correction, the single aberration measured independently as part of this dissertation (1955 to 1956) was spherical aberration. The latter was achieved by using a technique similar to one employed by Ivanoff.<sup>43</sup> That is, the pupil was divided into discrete nonoverlapping annular zones, and the required refractive correction was altered as needed within each annular zone. Table 1 provides values useful if one is performing the described integrations.

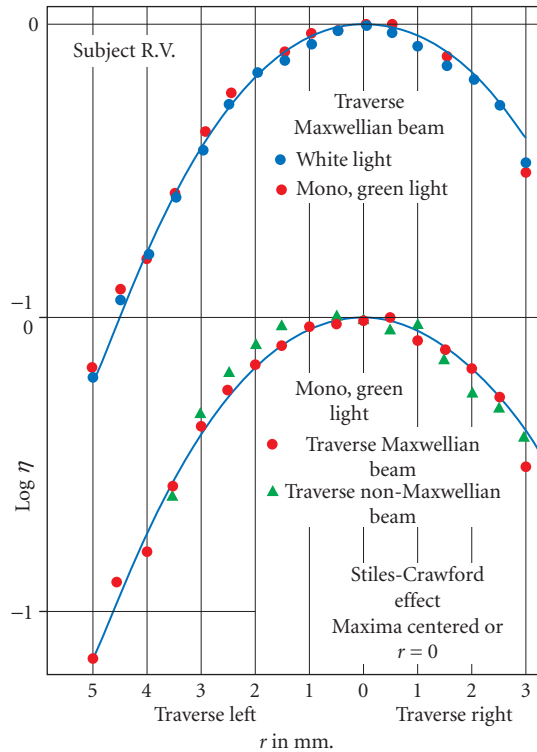
The SCE-1 was measured separately for both (1a) white light (obtained by using a ribbon filament lamp run at 16 amperes, with the beam passing through a heat filter), and paired neutral density filters, or (1b) by appropriately adding an interference filter having a near monochromatic green light band of light (peak  $\lambda = 552 \text{ m}\mu$ ) into the same optical system.

A second variable was introduced in this experiment. There was the traverse of the test beam across the entrance pupil of the eye employing (2a) an instrument design where the ribbon-lamp-filament was imaged in the plane of the entrance pupil of the eye (a classic Maxwellian view instrument design) (see Chap. 6), and, alternatively (2b) the light source was imaged directly on the retina of the observer having passed through the same number of surfaces and by using the same dimension beam image in the entrance pupil of the eye, and the same viewed stimulus target. This was regarded as a non-Maxwellian-view illumination system. Each of these cases utilized a projected aperture which subtended slightly less than 1-mm diameter image in the entrance pupil of the eye. Differences in measured SCE-1 results encountered for these different test conditions were very modest for all three subjects. Direct comparisons were made using both (a) bipartite fields, and, separately, (b) flicker photometry.

**TABLE 1** Area in the Entrance Pupil of the Human Eye Based on Gullstrand Schematic Eye

Ent. Pupil Area (Abscissa)*	Solve for Radius and Diameter of Entrance Pupil	
	Radius	Diameter
10 mm <sup>2</sup>	1.78 mm	3.57 mm
20	2.52	5.05
30	3.09	6.18
40	3.57	7.14
50	3.99	7.98
60	4.37	8.74
70	4.72	9.44

\* The Abscissas in Figs. 3 to 6 are plotted as  $\log_{10}$  entrance pupil area.

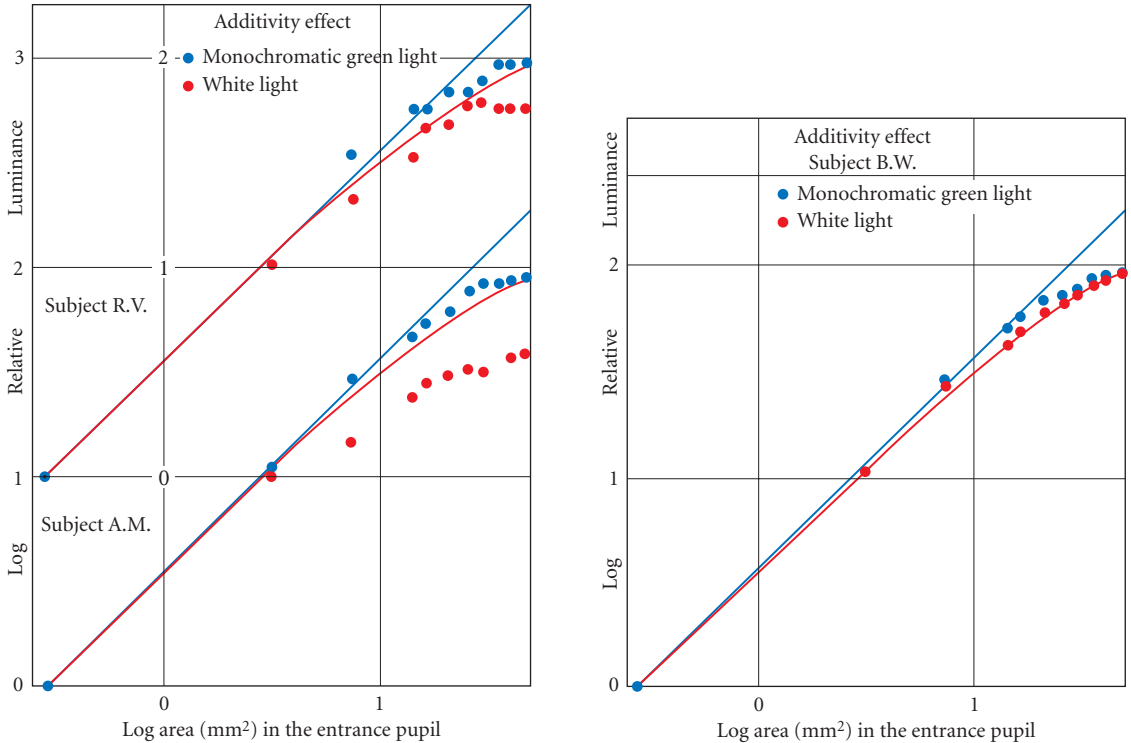


**FIGURE 2** SCE-1 data for subject R. V. are presented as an example of the many data sets acquired. His SCE-1 maxima were set at  $r = 0.0$  mm on the abscissa. *Top*: SCE-1 data obtained for both white light and monochromatic green light (552 m $\mu$  max.) are shown. *Bottom*: Both data sets were obtained using monochromatic light. Two different optical conditions were employed. That is, in one experiment, a Maxwellian beam traversed in the eye pupil, and in a second set of sessions, a non-Maxwellian beam was used to traverse the entrance pupil of the eye.<sup>19,20</sup> Also please refer to Chap. 6. These data are presented with Eq. (2) fitted to these data. (This illustration was taken from Fig. 14, Enoch, 1956, Dissertation;<sup>19</sup> and Fig. 9, Enoch, 1958, J.O.S.A.<sup>19,20</sup> These figures are reproduced with permission of the author and the publisher of J.O.S.A.)

Figure 2 shows sample SCE-1 test data for subject R. V.<sup>19,20</sup> As is obvious, there were only small differences between measured SCE-1 functions for the stated conditions. These data are presented as an example of the many data sets acquired. The peaks of his SCE-1 maxima were set at  $r = 0.0$  mm on the abscissa.

In Figs. 3 and 4, two different predictive additivity (integrated) plots are presented. The straight line on these double logarithmic plots assumes each point within the entrance pupil of the observer's eye contributes equally to the perceived visual stimulus. The modestly curved line represents the integrated SCE-1 function [here based on Eq. (2) above]. In Fig. 3, data for subjects A. M. and R. V. are presented, and in Fig. 4, data for subject B. W. appears. In Fig. 3 data for one observer were displaced by 1 log unit for clarity. For monochromatic green light [wavelength 522 m $\mu$ , (blue circles)] data obtained from all subjects show quite good agreement with the integrated SCE-1 function plot in





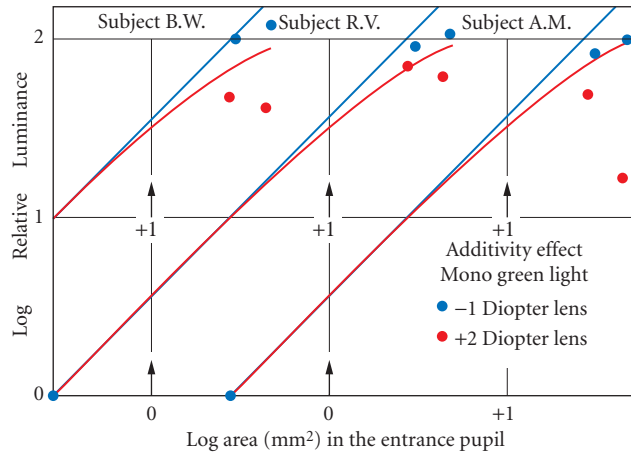
**FIGURES 3 AND 4** These figures demonstrate the equivalent of integrated or “additivity” diagrams for excitation entering defined entrance pupil areas. The straight line would predict perfect additivity of luminous stimuli entering the entrance pupil of the eye tested of the designated subject. The curved line incorporates a correction for the integration of SCE-1 data about the entrance pupil of the eye. In Fig. 3 data for Subjects R. V. and A. M. are presented; in Fig. 4 data are presented for Subject B. W. (This illustration was taken from Figs. 18 and 19, Enoch, 1956, *Dissertation*19; and Figs. 12,13, Enoch, 1958, *J.O.S.A.*20 These figures are reproduced with permission of the author and the publisher of *J.O.S.A.*)

both Figs. 3 and 4 (blue circles). The performance of subjects R. V. and A. M. clearly was less good for white light (red circles). On the other hand, data obtained from the eye of B. W. (Fig. 4) shows remarkably good “additivity” for both monochromatic and white light stimuli. We can infer that ocular aberrations in B. W.’s eye exhibited lower amplitudes than those encountered in the eyes of A. M. and R. V. That is, the optical quality of the periphery of B. W.’s cornea and eye lens was better than that found in the peripheral portions of the ocular components of the other two observers. This difference is the result of factors contributing to both monochromatic and chromatic aberrations.

In each case, near monochromatic green stimuli more closely matched the integrated SCE-1 curve than did white light, and in the case of B. W. both data sets were in close approximation to the predicted result.<sup>19,20</sup>

The added chromatic aberrations apparently made little difference to B. W. (or he retained his focus about mid-spectrum?), while the chromatic aberrations noticeably affected performance of the two subjects, R. V. and A. M.

In Fig. 5, the additivity effect obtained when placing an added  $-1.00$  diopter spherical (D.S.) lens (before the test eye of each subject) is shown for a limited number of different test aperture sizes projected into the entrance pupil of the eye, and this result is contrasted with the addition of a  $+2.00$  D.S. lens placed in front of the same eye (with the  $-1.00$  D.S. lens removed). The former lens (the  $-1.00$  D.S.) makes that eye more hyperopic “far-sighted,” while the  $+2.00$  D.S. makes the same eye more myopic “near-sighted.” The test eye could at least partially self-correct for the  $-1.00$  D.S.



**FIGURE 5** A comparison was made between the SCE additivity effect obtained when a  $-1.00$  D.S. lens was placed in the apparatus before the eye of each observer's test eye, and (separately) for addition of a  $+2.00$  D.S. lens. Two differing relatively large entrance pupil areas were selected for this test. These lenses slightly altered the area of the imaged aperture stop in the observer's entrance pupil. Hence, the  $+2$  D.S. lens is shown as subtending a slightly smaller area in the entrance pupil (on the abscissa) than the  $-1.00$  D.S. lens. (This illustration was taken from Fig. 22, Enoch, 1956, *Dissertation*;<sup>19</sup> and Fig. 16, Enoch, 1958, *J.O.S.A.*<sup>20</sup> These figures are reproduced with permission of the author and the publisher of *J.O.S.A.*)

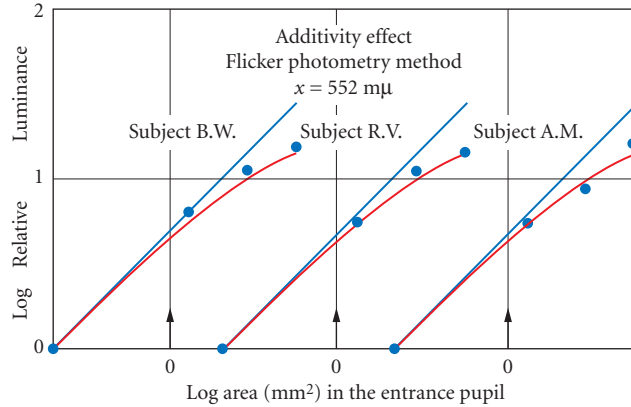
lens by accommodating on the test target. That is, these were nonpresbyopic test subjects. That is, the pupil-dilating drops employed had only slight/modest effect on accommodation in these observers. All tests were monocular.

When a flicker photometric test method was employed,<sup>19,20</sup> *blur of the retinal image proved to have much less of an influence upon brightness matches made by the subjects* (note, observers were instructed to compare the apparent brightness of the central parts of the flickering fields). When this same technique was combined with use of monochromatic  $\lambda = 552$  m $\mu$  stimuli, the matches of all three subjects proved to be similar and closely matched the integrated SCE-1 estimated curve. That is, all three subjects exhibited near perfect additivity approximately matching the SCE-1 predicted values (Fig. 6).<sup>19,20</sup>

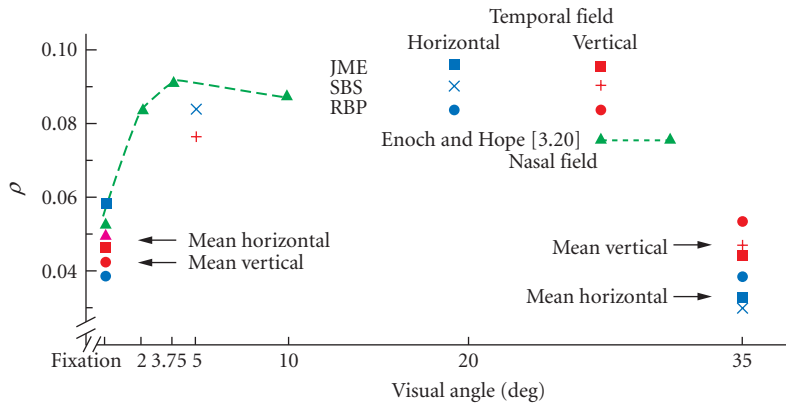
Thus, by eliminating and/or minimizing blur effects on perceived brightness due to aberrations (i.e., both monochromatic and chromatic aberrations!), *the integrated SCE function* becomes a good estimate of visual performance for any given pupil aperture utilized. Use of such methods, now (today) are made more readily achievable by using modern adaptive optics techniques for nonchromatic image formation, and, thus, for visual stimulus control. Taking the defined steps allows a superior estimate to be made of perceived brightness, and for the specification of the visual stimulus.

Finally, please note, the factor rho (or  $\rho$ ) is not a constant across the retina (i.e., it varies with eccentricity from the point of fixation (Fig. 7)).<sup>12,14,15,44,45</sup> It also varies within the visual spectrum with test wavelength utilized (as implied in Fig. 8).<sup>12</sup> Here we do not consider the Stiles-Crawford effect of the second kind (SCE-2), that is, the alteration of perceived hue and saturation associated with variation of the angle of incidence at the retina.<sup>14-16,44-46</sup> Separately, the SCE-1 is also affected to some degree by the distribution of and density of yellow (blue-absorbing) pigment in the eye lens.<sup>47</sup> The absorbance of this pigment also varies with wavelength. This factor is addressed in Fig. 8. See Chap. 3 in Ref. 12.

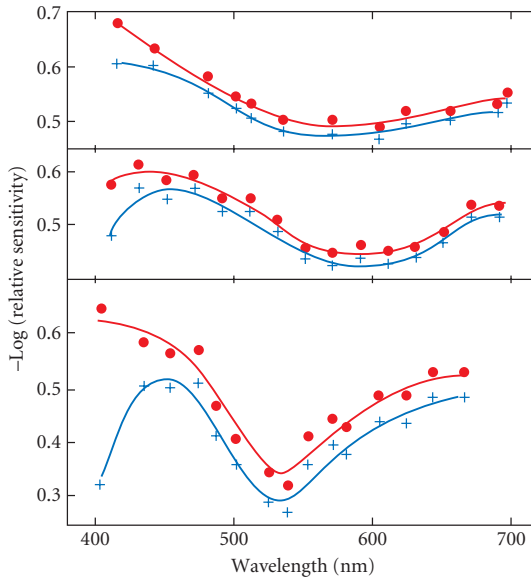
As noted above, Drum addressed some similar issues from a somewhat different point of view.<sup>34,35</sup> He also clearly demonstrated additivity of the SCE-1 to exist for all assessed tests and conditions. The simple fact is that SCE-1 and associated integrated results are rather robust under quite a variety of test conditions.



**FIGURE 6** These are integrated additivity data for each of the three trained subjects. The flicker photometry method was used. Effects of peripheral corneal and eye lens blur of the observers were minimized by testing with (1) monochromatic light, and by employing (2) a photometric matching technique less influenced by blurred imagery. (This illustration was taken from Fig. 25, Enoch, 1956, *Dissertation*;<sup>19</sup> and Fig. 19, Enoch, 1958, *J.O.S.A.*<sup>20</sup> These figures are reproduced with permission of the author and the publisher of J.O.S.A.)



**FIGURE 7** This figure addresses the variation of rho,  $\rho$ , with eccentricity from the point of fixation (assume this locus corresponds to the center of the fovea). This seemingly complex figure combines data from more than one paper on Stiles'  $\rho$  factor [Eq. (3)]. Tests were performed at fixation, in the para-foveal area, and separately at a test locus located  $35^\circ$  from the point of fixation. Horizontal and vertical in this figure refer to SCE-1 tests conducted in these two meridians within the entrance pupils of human eyes. (This figure is reproduced from Fig. 3.8 in Enoch and Tobey, 1980, (see Ref. 12, p. 99), and Enoch and Hope, 1973.<sup>12,45</sup> It is printed with the permission of the authors and publishers.)



**FIGURE 8** These data relate to values obtained from measured human foveal SCE-1 functions (see Ref. 12, Fig. 3.13 located on p. 109).<sup>15,47</sup> The upper curves present values of (–) log relative sensitivity for different subjects measured at a number of different wavelengths. The upper data curves presented were affected by the presence of yellow eye lens pigment(s), and the lower curves have been corrected for these pigment effects. Please be aware that the density of the yellow lens pigments increases with age. (*This illustration has origin in data of Stiles, 1939,<sup>15</sup> with additions obtained from the paper by Vos and Van Os, 1975.<sup>47</sup> These figures are reproduced with permission of the authors and the publishers.*)

## 9.5 DISCUSSION

### When Do Matters Considered Here Warrant Consideration?

Like in all psychophysical studies, the demand for precision is greater the more sensitive the determination, and the more complex the problem encountered in the observer or patient. We think we would all agree that the approach argued here is indicated whenever large or dilated pupils are used, particularly for tests of photopic vision! If a pupil is not well centered, or even reasonably asymmetric, or relatively nonresponsive, or otherwise not normal, one needs to rule out any unnecessary variable. Similarly, if the SCE-1 is abnormal, or displaced from its natural centrum, then it is clearly advisable to pursue the matter further. Note, given retinal photoreceptor packing properties, that it is almost unheard of for rods to be normally aligned if cones are found to be disturbed in their orientations in a given retinal area, and vice versa.

Zwicky and his coworkers (e.g., Ref. 48) have studied the effects of laser exposures on retinal survival, recovery, and/or lack thereof mainly in snakes. In their experiments on animals, Zwicky et al. used (1) subthreshold laser exposures, (2) threshold burns, (3) above threshold burns, and (4) more severe exposures. These studies were backed up by assessments of individual humans who have been exposed to comparable laser burns. Such laser burns result in immediate and substantial disturbances in photoreceptor alignments (often extending for some distance about the burn site), resultant early and late sensitivity losses. Snake eyes have large photoreceptors visible directly through the animal's natural pupil *in vivo*, and are available for study. Snakes are relatively nonreactive when exposed

to high intensity laser light when sedated, and show relatively rapid recovery *if* there is a potential for recovery. A meaningful area of the retina is affected. These authors<sup>48</sup> have carefully documented recovery and failure to recover in their studies.

One must remember that alignment of photoreceptors with the center of the exit pupil of the eye is maintained normally throughout life, except in cases where there has been some retinal disorder or disease.<sup>49,50</sup> In a number of cases, the retina can recover if the inducing anomaly is corrected or self corrects (remits).<sup>51,52</sup> So saying, residual scar tissue, local adhesions, or other impediments to recovery can alter the result. If ever doubts exist, suitable tests are indicated.

As has been inferred above, and can be seen in Fig. 3 to 6, in the normal observer the smooth, integrated SCE-1 function curve differs only modestly from the linear additivity curve until there is about a 4- to 5-mm-diameter entrance pupil.

Once again, looking at Figs. 3 and 4, uncertainty enters with peripheral blur of the optical components of the eye. The effects are greater for white light than monochromatic light. Subject A. M. apparently manifests less aberrations in his peripheral eye lens and cornea than observer V. R., and subject B. W. exhibits virtually no effect due to aberrations. At the outset, it was pointed out that image blur or degradation enters as a “dirty variable.” Without correction of these degrading peripheral lens blur factors, one introduces uncertainty. If one adds refractive error to the mix, the situation is made worse (see Fig. 5). If refractive and peripheral blur are corrected or at least meaningfully minimized, then a SCE correction applied to such data predict quite well the visual stimulus presented to the observer (see Fig. 6)!

Among other very interesting results, Makous and his coworkers,<sup>53-58</sup> and Applegate and Lakshminarayanan,<sup>21</sup> correctly point out, as can be inferred from comments made above, for small pupil apertures, certainly for 3-mm-diameter entrance pupils (1.5-mm radial distances) and perhaps a bit more, the two functions differ only slightly if refraction is well corrected, and the system is centered on the subject’s entrance pupil. And Makous et al. have also addressed and thoughtfully considered issues regarding effects of coherence and noncoherence of imagery, loci of adaptation, quantum activation rates in photoreceptors, effects of cell obliquity on excitation processes, etc.

## 9.6 TELEOLOGICAL AND DEVELOPMENTAL FACTORS

In a sense, it is also useful to consider these visual functions from a teleological point of view. It is apparent that the ultimate purpose(s) of the SCE-1 and associated waveguide properties of retinal photoreceptors are to enhance the detection of critical visual signal, both photopic and scotopic, and to help suppress intraocular stray light “noise” present in the “integrating-sphere-like” eye (e.g., Ref. 12). The feedback mechanisms controlling receptor alignment (under normal circumstances) and their directional properties *together* serve to favor the detection of directed visual signal content passing through the center of the exit pupil in eyes of vertebrates. In many invertebrate species the eyes serve similarly, for example, octopus, etc.; or in a different (but effectively comparable) fashion in the vast number of other invertebrate species. These optical and anatomical features serve to enhance greatly visual processes, and, as such, play critical roles in vision and survival of species. This argument emphasizes the great importance exerted by such mechanisms in evolutionary processes.

Related to such matters, the attention of the reader is called to a *remarkable recent paper* by Detlev Arendt et al. in *Science* in 2004.<sup>59</sup> These authors located a primitive and ancient form of invertebrate aquatic worm which had the usual paired invertebrate eye structures *as well as vertebrate-type, cylindrically shaped photoreceptors* containing cilia located in its brain in that area which controlled circadian rhythms. And the latter cells were also shown to contain a cone-type of opsin!<sup>59</sup>

## 9.7 CONCLUSIONS

There is a need to optimally correct and control the quality of the images formed in the eye, and/or the eye plus associated optical apparatuses. This is needed in order to define better the observer’s stimulus to vision, and to understand, in a superior way, the observer’s visual responses. This will include (1) a satisfactory correction of refraction, that is, lower-order aberrations, (2) correction

of higher-order monochromatic aberrations, and also (3) correction of chromatic aberrations. The totality can be aided by utilization of modern adaptive optics (AO) techniques. In radiometric and photometric studies, there is also need to include a factor which corrects for the Stiles-Crawford effect of the first kind (SCE-1), and that need increases with pupil diameter, particularly for photopic vision.

Note, because we do not wholly understand the nature of the stimulus to accommodation, we must be careful not to be too overly aggressive in seeking to eliminate *all* image blur.

And, as pointed out in the introductory remarks, it is time to reconsider the definition and units used to describe/define the troland.

Here, we only have considered monocular visual corrections. That is, in this discussion, we have not considered issues associated with maintenance of sustained and comfortable binocular vision (see Chap. 13) both in instrument design as well as in assessment and corrections of vision. For effective binocular results, it is clear that issues associated with maintaining an observer's comfort while performing extended binocular visual tasks need to be carefully addressed in both design and assessment of vision roles/function, including careful attention being paid to equating the sizes of the two retinal images (i.e., countering the effects of aniseikonia), and fusion of those images. That is, in design, we always need to address those additional factors affecting ocular motility and binocular fusion (Chap. 13).

The list of references appended to this chapter includes a number of citations which address additional aspects of issues considered here.<sup>60–68</sup>

## 9.8 REFERENCES

1. J. Howard, "The History of OSA; Profile in Optics, Leonard Thompson Troland, Optics and Photonic News (O.P.N.)," *Optical Society of America* **19**(6):20–21 (2008).
2. L. T. Troland, "Theory and Practice of the Artificial Pupil," *Psych. Review* **22**(3):167–176 (1915).
3. L. T. Troland, "On the Measurement of Visual Sensation Intensities," *J. Exp. Psych.* **2**(1):1–34 (1917).
4. J. Y. Cheung, C. J. Chunnillall, E. R. Woolliams, N. P. Fox, J. R. Mountford, J. Wang, and P. J. Thomas, "The Quantum Candela: A Redefinition of the Standard Units for Optical Radiation," *J. Mod. Opt.* **54**(2–3):373–396 (2007).
5. G. Wyszecki and W. S. Stiles, *Color Science: Concepts, and Methods, Quantitative Data and Formulae*, 2nd (ed.), Wiley, New York, 1982.
6. W. S. Baron and J. M. Enoch, "Calculating Photopic Illuminance," *Am. J. Optom. Physiol. Optics* **59**(4):338–341 (1982).
7. J. M. Enoch, "Vision; Physiology," *Modern Ophthalmology*, vol. 1, Arnold Sorsby, (ed.), 1st (ed.), Butterworths, Washington DC, 1963, sec. 1, Chap. 3, pp. 202–289.
8. J. M. Enoch and Harold E. Bedell, "Specification of the Directionality of the Stiles-Crawford Function," *Am. J. Optom. Physiol. Optics* **56**:341–344 (1979).
9. *Handbook of the Illumination Engineering Society* (any issue), Appendix, Conversion Factors, p. A-1.
10. D. Atchison, D. H. Scott, and G. Smith, "Pupil Photometric Efficiency and Effective Centre," *Ophthalmol. Physiol. Optics* **20**(6):501–503 (2000).
11. L. C. Martin, *Technical Optics*, 1st (ed.), vol. 2, Pitman, London, 1954.
12. J. M. Enoch, F. L. Tobey, Jr., (eds.), *Vertebrate Photoreceptor Optics. Springer Series in Optical Sciences*, vol. 23, Springer-Verlag, Berlin, Heidelberg, New York, 1981, ISBN 3-540-10515-8 Berlin, etc., and ISBN 0-387-10515-8 New York, etc.
13. W. S. Stiles and B. H. Crawford, "The Luminous Efficiency of Light Entering the Eye Pupil at Different Points," *Proc. Roy. London, Ser. B* **112**:428–450 (1933).
14. W. S. Stiles, "The Luminous Efficiency of Monochromatic Rays Entering the Eye Pupil at Different Points and a New Color Effect," *Proc. Roy. Soc. London, Ser. B* **123**:(830) 90–118 (1937).
15. W. S. Stiles, "The Directional Sensitivity of the Retina and the Spectral Sensitivities of the Rods and Cones," *Proc. Roy. Soc. London Ser. B* **127**:64–105 (1939).

16. B. H. Crawford, "The Luminous Efficiency of Light Entering the Eye Pupil at Different Points and Its Relation to Brightness Threshold Measurements," *Proc. Roy. Soc. London, Ser. B* **124**:(834) 81–96 (1937).
17. A. Safir, L. Hyams, and J. Philpott, "The Retinal Directional Effect: A Model Based on the Gaussian Distribution of Cone Orientations," *Vis. Res.* **11**:819–831 (1971).
18. S. Marcos and S. Burns, "Cone Spacing and Waveguide Properties from Cone Directionality Measurements," *J. Opt. Soc. Am. A* **16**:995–1004 (1999).
19. J. M. Enoch, *Summated Response of the Retina to Light Entering Different Parts of the Pupil*, Dissertation, Professor Glenn A. Fry, Advisor, Ohio State University, 1956.
20. J. M. Enoch, "Summated Response of the Retina to Light Entering Different Parts of the Pupil," *J. Opt. Soc. Am.* **48**:392–405 (1958).
21. R. A. Applegate and V. Lakshminarayanan, "Parametric Representation of the Stiles-Crawford Functions: Normal Variation of Peak Location and Directionality," *J. Opt. Soc. Am. A* **10**:1611–1623 (1993).
22. Vasudevan Lakshminarayanan, *The Stiles-Crawford Effect in Aniridia*, Ph.D. Dissertation, Jay M. Enoch, Advisor, University of California, Berkeley, 1985.
23. J. M. Enoch, V. Lakshminarayanan, and S. Yamade, "The Stiles-Crawford Effect (SCE) of the First Kind: Studies of the SCE in an Aniridic Observer," *Perception* **15**:777–784 (1986) (the W.S. Stiles Memorial issue).
24. V. Lakshminarayanan, J. M. Enoch, and S. Yamade, "Human Photoreceptor Orientation: Normals and Exceptions," *Advances in Diagnostic Visual Optics*, A. Fiorentini, D. L. Guyton, and I. M. Siegel. (eds.), Heidelberg, Springer-Verlag, 1987, pp. 28–32.
25. J. C. He, S. Marcos, and S. A. Burns, "Comparison of Cone Directionality Determined by Psychophysical and Reflectometric Techniques," *J. Opt. Soc. Am. A* **16**:2363–2369 (1999).
26. M. J. Kanis, *Foveal Reflection Analysis in a Clinical Setting*, Dissertation, Utrecht University, Faculty of Medicine, the Netherlands, p. 155, 2008. ISBN: 978-90-39348536, Dr. Dirk van Nooren, Advisor.
27. W. Gao, B. Cense, Y. Zhang, R. S. Jonnal, and D. T. Miller, "Measuring Retinal Contributions to the Optical Stiles-Crawford Effect with Optical Coherence Tomography," *Opt. Express* **16**:6486–6501 (2008).
28. H. von Helmholtz, *Helmholtz's Treatise on Physiological Optics*, 3d German (ed.), vol. 1, English Translation by J.P.C. Southall, Rochester, NY, Optical Society of America, 1924.
29. C. Cui and V. Lakshminarayanan, "The Choice of Reference Axis in Ocular Wavefront Aberration Measurement," *J. Opt. Soc. Am. A* **15**:2488–2496 (1998).
30. C. Cui and V. Lakshminarayanan, "The Reference Axis in Corneal Refractive Surgeries—Visual Axis or the Line of Sight?" *J. Mod. Opt.* **50**:1743–1749 (2003).
31. L. Ronchi, "Influence d'un mydriatique sur l'effet Stiles-Crawford," *Optica Acta* **2**(1):47–49 (1955).
32. A. M. Laties and J. M. Enoch, "An Analysis of Retinal Receptor Orientation: I. Angular Relationship of Neighboring Photoreceptors," *Invest. Ophthalmol.* **10**(1):69–77 (1971).
33. J. M. Enoch and A. M. Laties, "An Analysis of Retinal Receptor Orientation: II. Predictions of Psychophysical Tests," *Invest. Ophthalmol.* **10**(12):959–970 (1971).
34. B. Drum, *Additivity of the Stiles-Crawford Effect for a Fraunhofer Image*, Dissertation, Ohio State University, 1973, Carl R. Ingling, Advisor.
35. B. Drum, "Additivity of the Stiles-Crawford Effect for a Fraunhofer Image," *Vis. Res.* **15**:291–298 (1975).
36. G. Bocchino, "Studio della variazione della luminosità di un cannocchiale a variare della pupilla d'uscita," *Ottica* **1**:136–142 (1936).
37. G. T. di Francia and W. Sbroli, "Sulla legge integrale dell'effetto Stiles-Crawford," *Atti della Fond. G. Ronchi* **2**:100–104 (1947).
38. J. M. Enoch and G. M. Hope, "An Analysis of Retinal Receptor Orientation: III Results of Initial Psychophysical Tests," *Invest. Ophthalmol.* **11**(9):765–782 (1972).
39. I. Powell, "Lenses for Correcting Chromatic Aberration of the Eye," *Appl. Opt.* **20**:4152–4155 (1981).
40. M. P. Bachynski and G. Bekefi, "Study of Optical Diffraction Images at Microwave Frequencies," *J. Opt. Soc. Am.* **47**:428–438 (1957).
41. J. M. Enoch and G. A. Fry, "Characteristics of a Model Retinal Receptor Studied at Microwave Frequencies," *J. Opt. Soc. Am.* **48**(12):899–911 (1958).
42. J. M. Enoch, H. E. Bedell, and E. C. Campos, "Local Variations in Rod Receptor Orientation," *Vis. Res.* **18**(1): 123–124 (1978).



43. A. Ivanoff, *Les Aberrations de l'Œil, Leur Role dans l'Accommodation*, (The relation between pupil efficiencies for small and extended pupils of entry.) Editions de la Revue d'Optique, Paris, 1953.
44. J. M. Enoch and G. M. Hope, "Directional Sensitivity of the Foveal and Parafoveal Retina," *Invest. Ophthalmol.* **12**:497–503 (1973).
45. J. M. Enoch and W. S. Stiles, "The Colour Change of Monochromatic Light with Retinal Angle of Incidence," *Optica Acta* **8**:329–358 (1961).
46. P. L. Walraven and M. A. Bouman, "Relation Between Directional Sensitivity and Spectral Response Curves in Human Cone Vision," *J. Opt. Soc. Am.* **60**:780–784 (1960).
47. J. J. Vos and F. L. Van Os, "The Effect of Lens Density on the Stiles-Crawford Effect," *Vis. Res.* **15**:749–751 (1975).
48. H. Zwick, P. Edsall, B. E. Stuck, E. Wood, R. Elliott, R. Cheramie, and H. Hacker, "Laser Induced Photoreceptor Damage and Recovery in the High Numerical Aperture Eye of the Garter Snake," *Vis. Res.* **48**:486–493 (2008).
49. M. Rynders, T. Grosvenor, and J. M. Enoch, "Stability of the Stiles-Crawford Function in a Unilateral Amblyopic Subject over a 38 year Period: A Case Study," *Optom. Vis. Sci.* **72**(3):177–185 (1995).
50. J. M. Enoch, J.S. Werner, G. Haegerstrom-Portnoy, V. Lakshminarayanan, and M. Rynders, "Forever Young: Visual Functions Not Affected or Minimally Affected By Aging," *J. Gerontology: Biological Sciences* **55A**(8): B336–B351 August (1999).
51. E. C. Campos, H. E. Bedell, J. M. Enoch, and C. R. Fitzgerald, "Retinal Receptive Field-like Properties and Stiles-Crawford Effect in a Patient with a Traumatic Choroidal Rupture," *Doc. Ophthalmol.* **45**:381–395 (1978).
52. J. M. Enoch, C. R. Fitzgerald, and E. C. Campos, *Quantitative Layer-by-Layer Perimetry: An Extended Analysis*, Grune and Stratton, New York, 1981.
53. W. Makous and J. Schnapf, "Two Components of the Stiles-Crawford Effect: Cone Aperture and Disarray," (Abstract) *Program A.R.V.O.* 1973, p. 88.
54. M. J. McMahan and D. I. A. MacLeod, "Retinal Contrast Losses and Visual Resolution with Obliquely Incident Light," *J. Opt. Soc. Am. A* **18**(11):2692–2703 (2001).
55. J. Schnapf and W. Makous, "Individually Adaptable Optical Channels in Human Retina," (Abstract) *Program A.R.V.O.* 1974, p. 26.
56. B. Chen and W. Makous, "Light Capture in Human Cones," *J. Physiol. (London.)* **414**:89–108 (1989).
57. W. Makous, "Fourier Models and the Loci of Adaptation," *J. Opt. Soc. Am. A* **14**:2323–2345 (see p. 2332 pertinent to this discussion), (1997).
58. W. Makous, "Scotopic vision," in John Werner and L. M. Chapula (eds.), *The Visual Neurosciences*, Boston, MIT Press, 2004, (this paper corrects an erroneous table in the prior reference) pp. 838–850.
59. D. Arendt, K. Tessmar-Raible, H. Snyman, A. Dorresteijn, and J. Wittbrodt, "Ciliary Photoreceptors with a Vertebrate-type Opsin in an Invertebrate-brain," *Science* 306(29 October):869–871 (2004). See the interesting discussion of this paper by Elizabeth Pennisi, pp. 796–797.
60. L. Lundström and P. Unsbo, "Transformation of Zernike Coefficients: Scaled, Translated, and Rotated Wavefronts with Circular and Elliptical Pupils," *J. Opt. Soc. Am. A* **24**(3):569–577 (2007).
61. R. A. Applegate, W. J. Donnelly III, J. D. Marsack, D. E. Koenig, and K. Pesudovs, "Three-Dimensional Relationship between High-order Root-mean-square Wavefront Error, Pupil Diameter, and Aging," *J. Opt. Soc. Am.* **23**(3):578–587 (2007).
62. X. Zhang, M. Ye, A. Bradley, and L. Thibos, "Apodization by the Stiles-Crawford Effect Moderates the Visual Impact of Retinal Image Defocus," *J. Opt. Soc. Am. A* **16**:812–820 (1999).
63. J. M. Enoch, "Retinal Directional Resolution," International Conference on Visual Science, Bloomington, Indiana (April 1968), in *Visual Science* by J. Pierce and J. Levene, (eds.), Indiana University Press, Bloomington, Indiana, pp. 40–57 (1971).
64. H. Metcalf, "Stiles-Crawford Apodization," *J. Opt. Soc. Am.* **55**:72–74 (1965).
65. J. P. Carroll, "Apodization of the Stiles-Crawford Effect," *J. Opt. Soc. Am.* **70**:1155–1156 (1980).
66. D. A. Palmer, "Stiles-Crawford Apodization and the Stiles-Crawford Effect," *J. Opt. Soc. Am. A* **2**:1371–1374 (1985).



67. L. L. Sloan, "Size of Pupil as a Variable Factor in Measurements of the Threshold: An Experimental Study of the Stiles-Crawford Phenomenon," (Abstract) *J. Opt. Soc. Am.* **30**:271, (June, 1940), Paper: *Arch. Ophthalmol.* **24** (New Series, N.S.) July-December, pp. 258–275 (1940).
68. E. J. Fernández, A. Unterhuber, B. Považay, B. Hermann, P. Artal, and W. Drexler, "Chromatic Aberration Correction of the Human Eye for Retinal Imaging in the Near Infrared," *Optics Express* **14**(13):6213–6225, (June 26, 2006).

---

# COLORIMETRY

---

## David H. Brainard

*Department of Psychology  
University of Pennsylvania  
Philadelphia, Pennsylvania*

## Andrew Stockman

*Department of Visual Neuroscience  
UCL Institute of Ophthalmology  
London, United Kingdom*

---

### 10.1 GLOSSARY

---

**Chromaticity coordinates.** Tristimulus values normalized to sum to unity.

**CIE.** Commission Internationale de l'Éclairage or International Commission on Illumination. Organization that develops standards for color and lighting.

**Color-matching functions (CMFs).** Tristimulus values of the equal-energy spectrum locus.

**Color space transformation matrix.** Multiply a vector of tristimulus values for one color space by such a matrix to obtain tristimulus values in another color space.

**Cone coordinates.** Tristimulus values of a light with respect to the cone fundamentals.

**Cone fundamentals.** Estimates of the cone spectral sensitivities at the cornea. Equivalently, the CMFs that would result if primaries that uniquely stimulated the three cones could be and were used.

**Linear model.** Set of spectral functions that may be scaled and added to approximate other spectral functions. For example, the spectral power distributions of three monitor primaries are a linear model for the set of lights that can be emitted by the monitor.

**Metamers.** Two physically different lights that match in appearance to an observer.

**Photopic luminosity function.** Measure of luminous efficiency as a function of wavelength under photopic (i.e., rod-free) conditions.

**Primary lights.** Three independent lights (real or imaginary) to whose scaled mixture a test light is matched (actually or hypothetically). They must be independent in the sense that no combination of any two can match the third.

**Standard observer.** The standard observer is the hypothetical individual whose color-matching behavior is represented by a particular set of CMFs.

**Tristimulus values.** The tristimulus values of a light are the intensities of the three primary lights required to match it.

**Visual angle.** The angle subtended by an object in the external field at the effective optical center of the eye. Colorimetric data are typically specified for centrally fixated 2° or 10° fields of view.

## 10.2 INTRODUCTION

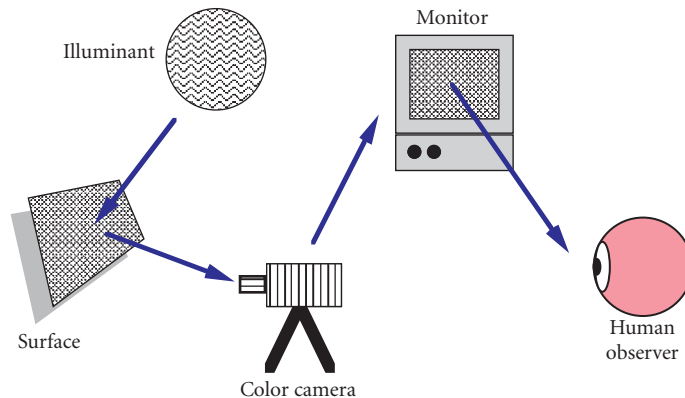
### Scope

The goal of colorimetry is to incorporate properties of the human color vision system into the measurement and numerical specification of visible light. Thanks in part to the inherent simplicity of the initial stages of visual coding, this branch of color science has been quite successful. We now have effective quantitative representations that predict when two lights will appear identical to a human observer and a good understanding of how these matches are related to the spectral sensitivities of the underlying cone photoreceptors. Although colorimetric representations do not directly predict color sensation,<sup>1-3</sup> they do provide the foundation for the scientific study of color appearance. Moreover, colorimetry can be applied successfully in practical applications. Foremost among these is perhaps color reproduction.<sup>4-6</sup>

As an illustrative example, Fig. 1 shows an image processing chain. Light from an illuminant reflects from a collection of surfaces. This light is recorded by a color camera and stored in digital form. The digital image is processed by a computer and rendered on a color monitor. The reproduced image is viewed by a human observer. The goal of the image processing is to render an image with the same color appearance at each image location as the original. Although exact reproduction is not always possible with this type of system, the concepts and formulas of colorimetry do provide a reasonable solution.<sup>4,7</sup> To develop this solution, we will need to consider how to represent the spectral properties of light, the relation between these properties and color camera responses, the representation of the restricted set of lights that may be produced with a color monitor, and the way in which the human visual system encodes the spectral properties of light. We will treat each of these topics in this chapter, with particular emphasis on the role played by the human visual system.

### Reference Sources

A number of excellent references are available that provide detailed treatments of colorimetry and its applications. Wyszecki and Stiles' comprehensive book<sup>8</sup> is an authoritative reference and



**FIGURE 1** A typical image processing chain. Light reflects from a surface or collection of surfaces. This light is recorded by a color camera and stored in digital form. The digital image is processed by a computer and rendered on a color monitor. The reproduced image is viewed by a human observer.

provides numerous tables of standard colorimetric data. Smith and Pokorny<sup>9</sup> provide a treatment complementary to the one developed here. Several publications of the Commission Internationale de l'Éclairage (International Commission on Illumination, commonly referred to as the CIE) describe current international technical standards for colorimetric measurements and calculations.<sup>10</sup> The most recent CIE proposal is for a set of physiologically relevant color-matching functions or cone fundamentals based mainly on the results of human psychophysical measurements.<sup>11</sup> Other sources cover colorimetry's mathematical foundations,<sup>12,13</sup> its history,<sup>14–16</sup> its applications,<sup>2,5,17,18</sup> and its relation to neural mechanisms.<sup>19–21</sup> Chapters 3, 5, 11, and 22 in this volume, and Chap. 37, “Radiometry and Photometry for Vision Optics,” by Yoshi Ohno in Vol. II of this *Handbook* are also relevant.

## Chapter Overview

The rest of this chapter is organized into three main sections. Section 10.3, “Fundamentals of Colorimetry,” reviews the empirical foundation of colorimetry and introduces basic colorimetric methods. In this section, we adhere to notation and development that is now fairly standard in the field.

Section 10.4, “Color Coordinate Systems,” discusses practicalities of using basic colorimetric ideas and reviews standard coordinate systems for representing color data.

Desktop computers can easily handle all standard colorimetric calculations. In Sec. 10.5 we introduce vector and matrix representations of colorimetric data and formulas. This development enables direct translation between colorimetric concepts and computer calculations. Matrix algebra is now being used increasingly in the colorimetric literature.<sup>4,22,23</sup>

Section 10.6 uses the vector and matrix formulation developed in Sec. 10.5 to treat some advanced topics.

The appendix (Sec. 10.7) reviews the elementary facts of matrix algebra required for this chapter. Numerous texts treat the subject in detail.<sup>24–27</sup> Many software packages (e.g., MATLAB, S-Plus, R) provide extensive support for numerical matrix algebra.

## 10.3 FUNDAMENTALS OF COLORIMETRY

### Introduction

We describe the light reaching the eye from an image location by its spectral power distribution. The spectral power distribution generally specifies the radiant power density at each wavelength in the visible spectrum. For human vision, the visible spectrum extends roughly between 400 and 700 nm (but see subsection “Sampling the Visible Spectrum” in Sec. 10.5). Depending on the viewing geometry, measures of radiation transfer other than radiant power may be used. These measures include radiance, irradiance, exitance, and intensity. The distinctions between these measures and their associated units as well as equivalent photometric measures are treated in Chaps. 34, 36, and 37 of Vol. II on this *Handbook* and are not considered here.

Color and color perception are limited at the first stage of vision by the spectral properties of the layer of light-sensitive *photoreceptors* that cover the rear surface of the eye (upon which an inverted image of the world is projected by the eye's optics). These photoreceptors transduce arriving photons to produce the patterns of electrical signals that eventually lead to perception. Daytime (photopic) color vision depends mainly upon the three classes of cone photoreceptor, each with different spectral sensitivity. These are referred to as long-, middle-, and short-wavelength-sensitive cones (L, M, and S cones), according to the part of the visible spectrum to which they are most sensitive (see Fig. 6). Night-time (scotopic) vision, by contrast, depends on a single class of photoreceptor, the rod.

**TABLE 1** Glossary of Conventional Colorimetric Terms and Notation

Chromaticity coordinates	$x, y$ , or in terms of the tristimulus values $X/(X+Y+Z)$ and $Y/(X+Y+Z)$ , respectively (or $r, g$ for <i>RGB</i> space, or $l, m$ for <i>LMS</i> space).
Color-matching functions or CMFs	$\bar{x}(\lambda)$ , $\bar{y}(\lambda)$ , and $\bar{z}(\lambda)$ . Tristimulus values of the equal-energy spectrum locus.
Cone fundamentals	$\bar{l}(\lambda)$ , $\bar{m}(\lambda)$ , and $\bar{s}(\lambda)$ in CMF notation, or often $L(\lambda)$ , $M(\lambda)$ , and $S(\lambda)$ . These are the CMFs that would result if primaries that uniquely stimulated the three cones could be used.
Photopic luminosity function	Photometric measure of luminous efficiency as a function of wavelength under photopic ( <i>i.e.</i> , rod-free) conditions: $V(\lambda)$ or $\bar{y}(\lambda)$ .
Primary lights	<b>R, G, B</b> , the three independent primaries (real or imaginary) to which the test light is matched (actually or hypothetically). They must be independent in the sense that no combination of two can match the third.
Standard observer	The standard observer is the hypothetical individual whose color-matching behavior is represented by a particular set of mean CMFs.
Tristimulus values	$R, G, B$ , the amounts of the three primaries required to match a given stimulus.
Visual angle	The angle subtended by an object in the external field of view at the effective optical center of the eye. Colorimetric data are typically for centrally fixated 2 or 10° fields of view.

## Conventional Colorimetric Terms and Notation

Table 1 provides a glossary of conventional colorimetric terms and notation. We adhere to these conventions in our initial development, in this section and in Sec.10.4. See also Table 3.1 of Ref. 28, and compare with the matrix algebra glossary in Table 2.

## Trichromacy and Univariance

Normal human vision is *trichromatic*. With some important provisos (see subsection “Conditions for Trichromatic Color Matching” in Sec. 10.3), observers can match a test light of any spectral composition to an appropriately adjusted mixture of just three other lights. Consequently, colors can be defined by three variables: the intensities of the three primary lights with which they match. These are called *tristimulus values*.

The range of colors that can be produced by the additive combination of three lights is simulated in Fig. 2. Overlapping red, green, and blue lights produce regions that appear cyan, purple, yellow, and white. Other, intermediate, colors can be produced by varying the relative intensities of the three lights.

Human vision is *trichromatic* because there are only three classes of cone photoreceptor in the eye, each of which responds univariantly to the rate of photon absorption.<sup>29,30</sup> *Univariance* refers to the fact that the effect of a photon, once absorbed, is independent of wavelength. What varies with wavelength is the probability that a photon is in fact absorbed, and this variation is described by the photoreceptor’s spectral sensitivity. Photoreceptors are, in effect, sophisticated photon counters the outputs of which vary according to the rate of absorbed photons. Changes in the absorption rate can result from a change in photon wavelength or from a change in the number of incident photons. This confound means that individual photoreceptors are effectively color blind. Normal observers are able to see color by comparing the outputs of the three, individually color-blind, cone types.

**TABLE 2** Glossary of Notation Used in Matrix Algebra Development

		<a href="#">Link to Conventional Notation</a>
$\lambda$	Wavelength	
$N_\lambda$	Number of wavelength samples	
$\mathbf{b}$	Spectral power distribution; basis vector	
$\mathbf{B}$	Linear model basis vectors	
$\mathbf{a}$	Linear model weights	
$N_b$	Linear model dimension	
$\mathbf{p}$	Primary spectral power distribution	
$\mathbf{P}$	Linear model for primaries	
$\mathbf{t}$	Tristimulus coordinates	$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$
$\mathbf{T}$	Color-matching functions ( $\bar{x}$ , $\bar{y}$ , and $\bar{z}$ are rows of $\mathbf{T}$ )	$\begin{bmatrix} \bullet & \dots & \bar{x} & \bullet & \dots & \bullet \\ \bullet & \dots & \bar{y} & \bullet & \dots & \bullet \\ \bullet & \dots & \bar{z} & \bullet & \dots & \bullet \end{bmatrix}$
$\mathbf{r}$	Cone (or sensor) coordinates	$\begin{bmatrix} L \\ M \\ S \end{bmatrix}$
$\mathbf{R}$	Cone (or sensor) sensitivities ( $\bar{l}$ , $\bar{m}$ , and $\bar{s}$ are rows of $\mathbf{R}$ )	$\begin{bmatrix} \bullet & \dots & \bar{l} & \bullet & \dots & \bullet \\ \bullet & \dots & \bar{m} & \bullet & \dots & \bullet \\ \bullet & \dots & \bar{s} & \bullet & \dots & \bullet \end{bmatrix}$
$\nu$	Luminance	$[Y]$
$\mathbf{V}$	Luminous efficiency function ( $V_\lambda$ is the single row of vector $\mathbf{v}$ )	$[\bullet \dots \bullet V_\lambda \bullet \dots \bullet]$
$\mathbf{M}$	Color space transformation matrix	

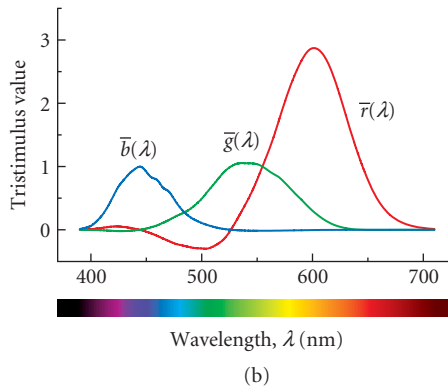
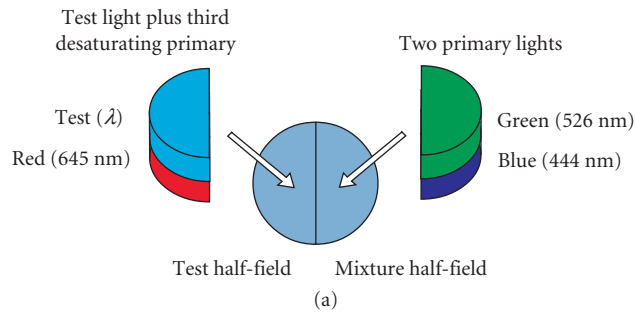


**FIGURE 2** Additive color mixing. Simulated overlap of projected red, green, and blue lights. The additive combination of red and green is seen as yellow, red and blue as purple, green and blue as cyan, and red, green, and blue as white.

## Color Matching

Trichromacy, together with other critical properties of color matching described in subsection “Critical Properties of Color Matching” in Sec. 10.3 mean that the color-matching behavior of an individual can be characterized as the intensities of three independent *primary lights* that are required to match a series of monochromatic spectral lights spanning the visible spectrum. Two experimental methods have been used to measure color matches: the maximum saturation method and Maxwell’s method. Most standard color-matching functions have been obtained using the maximum saturation method, though it is arguably inferior.

**Maximum Saturation Method** The maximum saturation method was used by Wright<sup>31</sup> and Guild<sup>32</sup> to obtain the matches that form the basis of the CIE 1931 color-matching functions (see subsection “CIE 1931 2° Color Matching Functions” in Sec. 10.4). In this method, the observer is presented with a half field illuminated by a monochromatic test light of variable wavelength  $\lambda$  as illustrated in Fig. 3a and an abutting half field illuminated by a mixture of red (R), green (G), and blue (B) primary lights.



**FIGURE 3** (a) Maximum saturation method of color matching. A monochromatic test field of wavelength  $\lambda$  can be matched using a mixture of red (645 nm), green (526 nm), and blue (444 nm) primary lights, one of which must usually be added to the test field to complete the match. (b) Color-matching functions. The amounts of each of the three primaries required to match equal energy monochromatic lights spanning the visible spectrum are known as the red  $\bar{r}(\lambda)$ , green  $\bar{g}(\lambda)$ , and blue  $\bar{b}(\lambda)$ , CMFs. These are shown as the red, green, and blue lines respectively. A negative sign means that primary must be added to the target to complete the match. (Based on Fig. 2.6 of Stockman and Sharpe.<sup>21</sup> The data are from Stiles and Burch.<sup>33</sup>)

(Note that in this section of the chapter, bold uppercase symbols denote primary lights, not matrices.) Often the primary lights are chosen to be monochromatic, although this is not necessary. For each test wavelength  $\lambda$ , the observer adjusts the intensities and arrangement of the three primary lights to make a match between the half field containing the test light and the adjacent half field. Generally, one of the primary lights is admixed with the test, while the other two are mixed together in the adjacent half field. Figure 3*b* shows the mean  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  color-matching functions (hereafter abbreviated as CMFs) obtained by Stiles and Burch<sup>33</sup> for primary lights of 645, 526, and 444 nm. Notice that one of the CMFs is usually negative. There is no “negative light.” Negative values mean that the primary in question has been added to the test light in order to make a match. Matches using real primaries result in negative values because the primaries do not uniquely stimulate single cone photoreceptors, the spectral sensitivities of which overlap throughout the visible spectrum (see Fig. 6). Although color-matching functions are generally plotted as functions of wavelength, it is helpful to keep in mind that they represent matches, not light spectral power distributions.

The maximum saturation match between  $E_\lambda$ , a monochromatic constituent of the equal unit energy stimulus of wavelength  $\lambda$ , and the three primary lights (**R**, **G**, and **B**) is denoted by

$$E_\lambda \sim \bar{r}(\lambda)\mathbf{R} + \bar{g}(\lambda)\mathbf{G} + \bar{b}(\lambda)\mathbf{B} \quad (1)$$

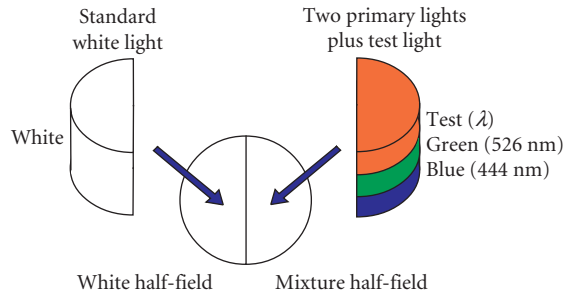
where  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  are the three CMFs, and where negative CMF values indicate that the corresponding primary was mixed with the test to make the perceptual match. CMFs are usually defined for a stimulus, **E**, which has equal unit energy throughout the spectrum. However, in practice the spectral power of the test light used in most matching experiments is varied with wavelength. In particular, longer-wavelength test lights are typically chosen to be intense enough to saturate the rods so that rods do not participate in the matches (see, e.g., Ref. 34). CMFs and the spectral power distributions of lights are always measured and tabulated as discrete functions of wavelength, typically defined in steps of 1, 5, or 10 nm.

We use the symbol  $\sim$  in Eq. (1) to indicate that two lights are a perceptual match. Perceptual matches are to be carefully distinguished from physical matches, which are denoted by the  $=$  symbol. Of course, when two lights are a physical match, they must also be a perceptual match. Two lights that are a perceptual match but not a physical match are referred to as metameric color stimuli or metamers. The term metamerism is often used to refer to the fact that two physically different lights can appear identical.

The color-matching functions are defined for equal energy monochromatic test lights. More generally any test light, whether monochromatic or not, may be matched in the color-matching experiment. As noted above, we refer to the primary weights *R*, *G*, and *B* required to match any light as its tristimulus values. As with CMFs, tristimulus values may be negative, indicating that the corresponding primary is mixed with the test to make the match. Once the matching primaries are specified, the tristimulus values of a light provide a complete description of its effect on the human cone-mediated visual system, subject to the caveats discussed below. In addition, knowledge of the color-matching functions is sufficient to compute the tristimulus values of any light (see subsection “Tristimulus Values for Arbitrary Lights” in Sec. 10.3).

**Conditions for Trichromatic Color Matching** There are a number of qualifications to the empirical generalization that it is possible for observers to match any test light by adjusting the intensities of just three primaries. Some of these qualifications have to do with ancillary restrictions on the experimental conditions (e.g., the size of the bipartite field and the overall intensity of the test and matching lights). The other qualifications have to do with the choice of primaries and certain conventions about the matching procedure. First the primaries must be chosen so that it is not possible to match any one of them with a weighted superposition of the other two. Second, the observer sometimes wishes to increase the intensity of one or more of the primaries above its maximum value. In this case, we must allow him to scale the intensity of the test light down. We follow the convention of saying that the match was possible and scale up the reported primary weights by the same factor. Third, as discussed in more detail above, the observer sometimes wishes to decrease the intensity of one or more of the primaries below zero. This is always the case when the test light is a spectral light unless its wavelength is equivalent to one of the primaries. In this case, we must allow the observer





**FIGURE 4** Maxwell's method of color matching. A monochromatic test field of wavelength  $\lambda$  replaces the primary light to which it is most similar, and a match is made to the white standard by adjusting the intensities of the two remaining primaries and the test field. (Based on Fig. 3 of Stockman.<sup>206</sup>)

to superimpose each such primary on the test light rather than on the other primaries. We follow the convention of saying that the match was possible but report with negative sign the intensity of each transposed primary.

With these qualifications, matching with three primaries is always possible for small fields. For larger fields, spatial inhomogeneities may make it impossible to produce a match simultaneously across the entire field (see subsections “Specificity of CMFs” and “Tristimulus Values for Arbitrary Lights” in Sec. 10.3).

**Maxwell's Matching Method** It is of methodological interest to note that the maximum saturation method is not the only way to instrument the color-matching experiment. Indeed the first careful quantitative measurements of color matching and trichromacy were made by Maxwell.<sup>35</sup> In Maxwell's method, which is illustrated in Fig. 4, the matched fields always appear white, so that at the match point the eye is always in the same state of adaptation whatever the test wavelength (in contrast to the maximum saturation method in which the chromaticity of the match varies with wavelength). In the experiment, the subject is first presented with a white standard half-field, and is asked to match it with the three primary lights. The test light then replaces the primary light to which it is most similar and the match is repeated. Grassmann's laws are invoked to convert the two empirical matches to the form of Eq. (1).

**Critical Properties of Color Matching** Color-matching data are usually obtained for monochromatic test lights. Such data are useful in general only if they can be used to predict matches for other lights with arbitrary spectral power distributions, and by extension the matches that would be made for other sets of primary lights. For this to be possible, the color-matching experiment must exhibit a number of critical properties. We review these properties briefly below. Given that they hold, it is possible to show that tristimulus values provide a complete representation for the spectral properties of light as these affect human vision. Krantz provides a detailed formal treatment.<sup>12</sup>

**Grassmann's laws** Grassmann's laws describe several of the key properties of color matching. They are:<sup>8,12</sup>

1. *Symmetry*: If light **X** matches light **Y**, then **Y** matches **X**.
2. *Transitivity*: If light **X** matches light **Y** and **Y** matches light **Z**, then **X** matches **Z**.
3. *Proportionality*: If light **X** matches light **Y**, then  $n\mathbf{X}$  matches  $n\mathbf{Y}$  (where  $n$  is a constant of proportionality).
4. *Additivity*: If **W** matches **X** and **Y** matches **Z**, then the combination of **W** and **Y** matches the combination of **X** and **Z** (and similarly the combination of **X** and **Y** matches **W** and **Z**).

These laws have been tested extensively and hold well.<sup>8,19</sup> To a first approximation, color matching can be considered to be linear and additive.<sup>12,36</sup>

**Uniqueness of color matches** The tristimulus values of a light should be unique. This is equivalent to the requirement that only one weighted combination of the apparatus primaries produces a match to any given test light. The uniqueness of color matches ensures that tristimulus values are well-defined. In conjunction with transitivity, uniqueness also guarantees that two lights that match each other will have identical tristimulus values. It is generally accepted that, apart from variability, trichromatic color matches are unique for color normal observers.

**Persistence of color matches** The above properties concern color matching under a single set of viewing conditions. By viewing conditions, we refer to the properties of the image surrounding the bipartite field and the sequence of images viewed by the observer before he made the match. An important property of color matching is that lights that match under one set of viewing conditions continue to match when the viewing conditions are changed. This property is referred to as the persistence or stability of color matches.<sup>8,19</sup> It holds to good approximation (but see subsection “Limits of Color Matching Data” in Sec. 10.4). The importance of the persistence law is that it allows a single set of tristimulus values to be used across viewing conditions.

**Consistency across observers** Finally, for the use of tristimulus values to have general validity, it is important that there should be agreement about matches across observers. For the majority of the population, there is good agreement about which lights match. We discuss individual differences in color matching in section “Limits of Color-Matching Data.”

**Specificity of CMFs** Color-matching data are specific to the conditions under which they were measured, and strictly to the individual observers in whom they were measured. By applying the data to other conditions and using them to predict other observer’s matches, some errors will inevitably be introduced.

An important consideration is the area of the retina within which the color matches were made. Standard color matching data (see section “Color-Matching Functions” in Sec. 10.4) have been obtained for centrally viewed fields with diameters of either 2° or 10° of visual angle. The visual angle refers to the angle subtended by an object in the external field at the effective optical center of the eye. The size of a circular matching field used in colorimetry is defined as the angular difference subtended at the eye between two diametrically opposite points on the circumference of the field. Thus, matches are defined according to the retinal size of the matching field *not* by its physical size. A 2° diameter field is known as a *small field*, whereas a 10° one as a *large field*. (One degree of visual angle is roughly equivalent to the width of the fingernail of the index finger held at arm’s length.) Color matches vary with retinal size and position because of changes in macular pigment density and photopigment optical density with visual angle (see section “Limits of Color-Matching Data”).

Standardized CMFs are mean data that are also known as *standard observer* data, in the sense that they are assumed to represent the color-matching behavior of a hypothetical typical human observer. The color matches of individual observers, however, can vary substantially from the mean matches represented by standard observer CMFs. Individual differences in lens pigment density, macular pigment density, photopigment optical density, and in the photopigments themselves can all influence color matches (see section “Limits of Color-Matching Data”).

**Tristimulus Values for Arbitrary Lights** Given that additivity holds for color matches, the tristimulus values,  $R$ ,  $G$ , and  $B$  for an arbitrarily complex spectral radiant power distribution  $P(\lambda)$  can be obtained from the  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs by:

$$R = \int P(\lambda)\bar{r}(\lambda)d\lambda, \quad G = \int P(\lambda)\bar{g}(\lambda)d\lambda, \quad \text{and} \quad B = \int P(\lambda)\bar{b}(\lambda)d\lambda \quad (2)$$

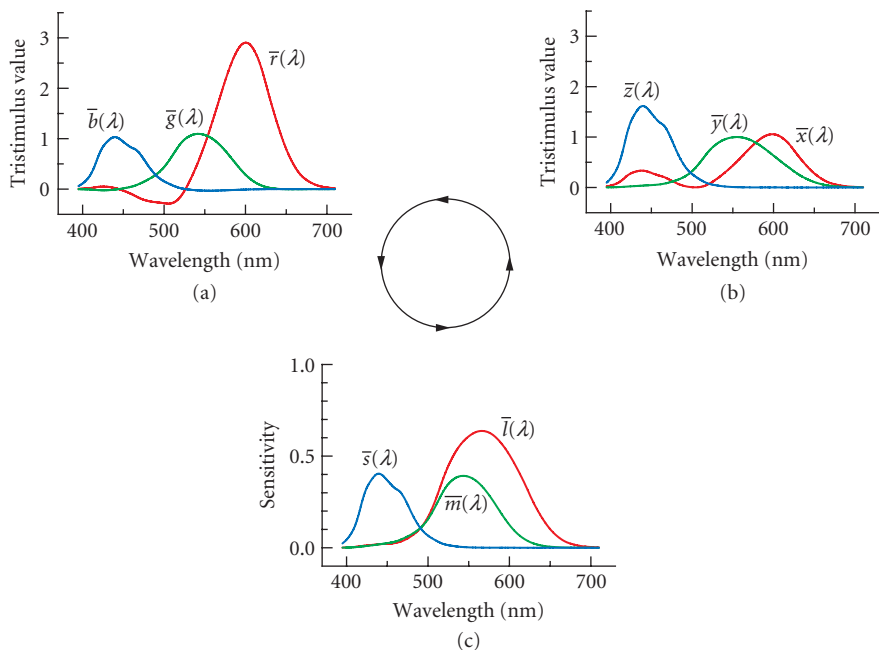
Since spectral power distributions and CMFs are usually discrete functions, the integration in Eq. (2) is usually replaced by a sum.

**Transformability of CMFs** The  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  CMFs shown in Fig. 3 are for monochromatic RGB (red-green-blue) primaries of 645, 526, and 444 nm. These CMFs can be transformed to other sets of real primary lights, and to CMFs for imaginary primary lights, such as the CIE X, Y, and Z primaries, or to CMFs representing the LMS cone spectral sensitivities (*cone fundamentals*). These transformations are illustrated in Fig. 5.

Each transformation of CMFs is accomplished by multiplying the CMFs, viewed as a column vector at each wavelength, by a  $3 \times 3$  matrix. For now we simply assert this result, as our key point here is to note that such transformation is possible to enable a discussion of commonly used tristimulus representations. See Sec. 10.5 or Sec. 3.2.5 of Ref. 8 for more details about transformations between primaries.

The primaries selected by the CIE produced  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  CMFs that are always positive. The  $\bar{y}(\lambda)$  CMF is also the luminosity function (see section “Brightness Matching and Photometry” and also Chap. 11) thus incorporating luminosity information into the CMFs, and linking colorimetry and photometry.

The primaries that yield the cone fundamentals  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  as CMFs are three imaginary primary lights that would uniquely stimulate each of the three classes of cones. Although  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  cannot be obtained directly from color matches, they are strongly constrained by color-matching data since they should be a linear transformation of any other set of CMFs. Derivation of cone fundamentals is discussed in the section “Cone Fundamentals.”



**FIGURE 5** CMFs can be linearly transformed from one set of primaries to another. Illustrated here are CMFs for **R**, **G**, and **B** primaries (a), for the imaginary **X**, **Y**, and **Z** primaries (b), and the cone fundamental **L**, **M**, and **S** primaries (c). The CMFs shown in (a) and (b) are Judd-Vos modified CIE 1931 RGB and XYZ functions, respectively (see subsection “Judd-Vos Modified 2° Color-Matching Functions” in Sec. 10.4) and those shown in (c) are the Smith-Pokorny cone fundamentals (see section “Cone Fundamentals”). (Based on Fig. 4 of Stockman.<sup>206</sup>)

## 10.4 COLOR COORDINATE SYSTEMS

### Overview

For the range of conditions where the color-matching experiment obeys the properties described in the previous sections, tristimulus values (or cone coordinates) provide a complete and efficient representation of human color vision. When two lights have identical tristimulus values, they are indistinguishable to the visual system and may be substituted for one another. When two lights have tristimulus values that differ substantially, they can be distinguished by an observer with normal color vision.

The relation between spectral power distributions and tristimulus values depends on the choice of primaries used in the color-matching experiment. In this sense, the choice of primaries in colorimetry is analogous to the choice of unit (e.g., foot versus meter) in the measurement of length. We use the terms *color coordinate system* and *color space* to refer to a representation derived with respect to a particular choice of primaries. We will also use the term *color coordinates* as synonym for tristimulus values.

Although the choice of primaries determines a color space, specifying primaries alone is not sufficient to compute tristimulus values. Rather, it is the color-matching functions that characterize the properties of the human observer with respect to a particular set of primaries. As noted in section “Fundamentals of Colorimetry” above and developed in detail in Sec. 10.5, “Matrix Representations and Calculations,” knowledge of the color-matching functions allows us to compute tristimulus values for arbitrary lights, as well as to derive color-matching functions with respect to other sets of primaries. Thus in practice we can specify a color space either by its primaries or by its color-matching functions.

A large number of different color spaces are in common use. The choice of which color space to use in a given application is governed by a number of considerations. If all that is of interest is to use a three-dimensional representation that accurately predicts the results of the color-matching experiment, the choice revolves around the question of finding a set of color-matching functions that accurately capture color-matching performance for the set of observers and viewing conditions under consideration. From this point of view, color spaces that differ only by an invertible linear transformation are equivalent. But there are other possible uses for color representation. For example, one might wish to choose a space that makes explicit the responses of the physiological mechanisms that mediate color vision. We discuss a number of commonly used color spaces based on CMFs, cone fundamentals, and transformations of the cone fundamentals guided by assumptions about color vision after the photoreceptors.

Many of the CMFs and cone fundamentals are available online in tabulated form at URL <http://www.cvrl.org/>.

### Stimulus Spaces

A stimulus space is the color space determined by the primaries of a particular apparatus. For example, stimuli are often specified in terms of the excitation of three monitor phosphors. Stimulus color spaces have the advantage that they provide a direct description of the physical stimulus. On the other hand, they are nonstandard and their use hampers comparison of data collected in different laboratories. A useful compromise is to transform the data to a standard color space, but to provide enough side information to allow exact reconstruction of the stimulus. Often this side information can be specification of the apparatus primaries.

### Color-Matching Functions

Several sets of standard CMFs are available for the central 2° or the central 10° of vision. For the central 2° (the small-field matching conditions), they are the CIE 1931 CMFs,<sup>37</sup> the Judd-Vos modified 1931 CMFs,<sup>38,39</sup> and the Stiles and Burch CMFs.<sup>33</sup> For the central 10° (the large-field matching conditions), they are the 10° CMFs of Stiles and Burch,<sup>34</sup> and the related 10° CIE 1964 CMFs. CIE functions are available as  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  for the real primaries **R**, **G**, and **B**, or as  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  for the imaginary primaries **X**, **Y**, and **Z**. The latter are more commonly used in applied colorimetry.

**CIE 1931 2° Color-Matching Functions** In 1931, the CIE integrated a body of empirical data to determine a standard set of CMFs.<sup>37,40</sup> The notion was that the CIE 1931 color-matching functions would characterize the results of a color-matching experiment performed on an “average” or “standard” color-normal human observer known as the CIE 1931 standard observer. They are available in both  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  and  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$  form.

The empirical color-matching data used to construct the 1931 standard observer were those of Wright<sup>41</sup> and Guild,<sup>32</sup> which provided only the ratios of the three primaries required to match spectral test lights. Knowledge of the absolute radiances of the matching primaries is required to generate CMFs, but this was unavailable. The CIE reconstructed this information by assuming that a linear combination of the three unknown CMFs was equal to the 1924 CIE  $V(\lambda)$  function.<sup>37,42</sup> In addition to uncertainties about the validity of this assumption,<sup>43</sup> the  $V(\lambda)$  curve that was used as the standard is now known not to provide an accurate description of typical human performance; it is far too insensitive at short wavelengths (see Fig. 2.13 of Ref. 44).

More generally, there is now considerable evidence that the color-matching functions standardized by the CIE in 1931 differ from those of the average human observer<sup>21,33,34,38,39</sup> and the CIE has recently recommended<sup>11</sup> a new set of color-matching functions based on estimates of the cone photoreceptor spectral sensitivities and the Stiles and Burch 10° CMFs.<sup>34</sup> A large body of extant data is available only in terms of the CIE 1931 system, however, and many colorimetric instruments are designed around it. Therefore it seems likely that the CIE 1931 system will continue to be of practical importance for some time. Its inadequacy at short-wavelengths is well-known, and is often taken into account in colorimetric and photometric applications.

**Judd-Vos Modified 2° Color-Matching Functions** In 1951, Judd reconsidered the 1931 CMFs and came to the conclusion that they could be improved.<sup>38</sup> He increased the sensitivity of  $V(\lambda)$  used to reconstruct the CIE CMFs below 460 nm, and derived a new set of CMFs [see Table 1 (5.5.2) of Ref. 8, which were later slightly modified by Vos,<sup>39</sup> see his Table 1].

The modifications to the  $V(\lambda)$  function introduced by Judd had the unwanted effect of producing CMFs that are relatively insensitive near 460 nm (where they were unchanged). Although this insensitivity can be roughly characterized as being consistent with a high macular pigment density,<sup>33,45,46</sup> the CMFs are somewhat artificial and thus removed from real color matches. Nevertheless, in practice the Judd-Vos modifications lead to a set of CMFs that are probably more typical of the average human observer than the original CIE 1931 color-matching functions. These functions were never officially standardized. However, they are widely used in practice, especially in vision science, because they are the basis of a number of estimates of the human cone spectral sensitivities, including the recent versions of the Smith-Pokorny cone fundamentals.<sup>47</sup>

**Stiles and Burch (1955) 2° CMFs** The assumption used to construct the CIE 1931 standard observer, namely that  $V(\lambda)$  is a linear combination of the CMFs is now unnecessary, since current instrumentation allows CMFs to be measured in conjunction with absolute radiometry. The Stiles and Burch 2° CMFs<sup>33</sup> are an example of directly measured functions. Though referred to by Stiles as “pilot” data, these CMFs are the most extensive set of directly measured color-matching data for 2° vision available, being averaged from matches made by 10 observers. Even compared in relative terms, there are real differences between the CIE 1931 and the Stiles and Burch<sup>33</sup> 2° color-matching data in the range between 430 and 490 nm. These CMFs are seldom used.

**Stiles and Burch (1959) 10° CMFs** The most comprehensive set of color-matching data are the large-field, centrally viewed 10° CMFs of Stiles and Burch.<sup>34</sup> Measured in 49 subjects from approximately 390 to 730 nm (and in nine subjects from 730 to 830 nm), these data are probably the most secure set of existing CMFs. Like the Stiles and Burch 2° functions,<sup>33</sup> the 10° functions represent directly measured CMFs, and so do not depend on measures of  $V(\lambda)$ . These CMFs are the basis of the Stockman and Sharpe<sup>46</sup> cone fundamentals (see section “Cone Fundamentals”) and thus the recent CIE proposal for a set of physiologically relevant CMFs.<sup>11</sup>

**1964 10° Color-Matching Functions** In 1964, the CIE standardized a second set of CMFs appropriate for larger field sizes. These CMFs take into account the fact that human color matches depend on

the size of the matching fields. The CIE 1964 10° color-matching functions are an attempt to provide a standard observer for these larger fields. The use of 10° color-matching functions is recommended by the CIE when the sizes of the regions under consideration are larger than 4°.<sup>10</sup> The large field CIE 1964 CMFs are based mainly on the 10° CMFs of Stiles and Burch<sup>34</sup> and to a lesser extent on the arguably inferior and possibly rod-contaminated 10° CMFs of Speranskaya.<sup>48</sup> These functions are available as  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  and  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ , and  $\bar{z}(\lambda)$ .

While the CIE 1964 CMFs are similar to the 10° CMFs of Stiles and Burch functions, they differ in several ways that compromise their use as the basis for cone fundamentals.<sup>46</sup> The CIE<sup>11</sup> has now recommended a new set of 10° color-matching functions that are more tightly coupled to estimates of the cone spectral sensitivities and are based on the original Stiles and Burch 10° data.

## Cone Fundamentals

An important goal in color science since the establishment of trichromatic color theory,<sup>49–52</sup> has been the determination of the linear transformation between  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ , and  $\bar{b}(\lambda)$  and the three cone spectral sensitivities,  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ .

A match between the test and mixture fields in a color-matching experiment is a match at the level of the cone photoreceptors. The response of each cone class to the mixture of primaries equals the response of that cone class to the test light. Put more formally, the following equations must hold for each unit energy test light:

$$\begin{aligned}\bar{l}_R \bar{r}(\lambda) + \bar{l}_G \bar{g}(\lambda) + \bar{l}_B \bar{b}(\lambda) &= \bar{l}(\lambda) \\ \bar{m}_R \bar{r}(\lambda) + \bar{m}_G \bar{g}(\lambda) + \bar{m}_B \bar{b}(\lambda) &= \bar{m}(\lambda) \\ \bar{s}_R \bar{r}(\lambda) + \bar{s}_G \bar{g}(\lambda) + \bar{s}_B \bar{b}(\lambda) &= \bar{s}(\lambda)\end{aligned}\quad (3)$$

where  $\bar{l}_R$ ,  $\bar{l}_G$ , and  $\bar{l}_B$  are, respectively, the L-cone sensitivities to the **R**, **G**, and **B** primary lights,  $\bar{m}_R$ ,  $\bar{m}_G$ , and  $\bar{m}_B$  are the M-cone sensitivities to the primary lights, and  $\bar{s}_R$ ,  $\bar{s}_G$ , and  $\bar{s}_B$  are the S-cone sensitivities.

Since the S cones are now known to be insensitive to long wavelengths, it can be assumed that  $\bar{s}_R$  is effectively zero for a long-wavelength **R** primary. There are therefore eight unknowns required, and we can rewrite Eq. (3) as a linear transformation:

$$\begin{pmatrix} \bar{l}_R & \bar{l}_G & \bar{l}_B \\ \bar{m}_R & \bar{m}_G & \bar{m}_B \\ 0 & \bar{s}_G & \bar{s}_B \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} \bar{l}(\lambda) \\ \bar{m}(\lambda) \\ \bar{s}(\lambda) \end{pmatrix}\quad (4)$$

Moreover, since we are often more concerned about the relative  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$  cone spectral sensitivities, rather than their absolute values, the eight unknowns become five:

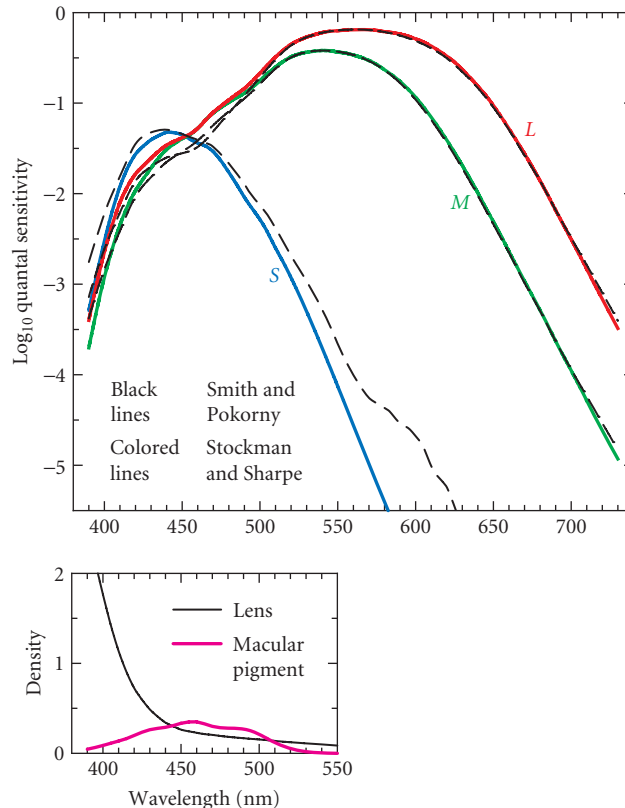
$$\begin{pmatrix} \bar{l}_R/\bar{l}_B & \bar{l}_G/\bar{l}_B & 1 \\ \bar{m}_R/\bar{m}_B & \bar{m}_G/\bar{m}_B & 1 \\ 0 & \bar{s}_G/\bar{s}_B & 1 \end{pmatrix} \begin{pmatrix} \bar{r}(\lambda) \\ \bar{g}(\lambda) \\ \bar{b}(\lambda) \end{pmatrix} = \begin{pmatrix} k_l \bar{l}(\lambda) \\ k_m \bar{m}(\lambda) \\ k_s \bar{s}(\lambda) \end{pmatrix}\quad (5)$$

Note that the constants  $k_l$ ,  $k_m$ , and  $k_s$  remain unknown. Their values are typically chosen to scale the three cone fundamentals to meet some side criterion: for example, so that  $k_l \bar{l}(\lambda)$ ,  $k_m \bar{m}(\lambda)$ , and  $k_s \bar{s}(\lambda)$  peak at unity. Smith and Pokorny<sup>53</sup> assume that  $k_l \bar{l}(\lambda) + k_m \bar{m}(\lambda)$  sum to  $V(\lambda)$ , the luminous efficiency function. Care should be taken when drawing conclusions that depend on the scaling chosen.

The five unknowns in the left of Eq. (5) can be estimated by fitting linear combinations of CMFs to cone spectral sensitivity measurements made in dichromatic observers and in normal observers under special conditions that isolate the responses of single cone types. They can also be estimated by comparing

color matches made by normal and dichromatic observers. Estimates from dichromats depend on the “loss,” “reduction,” or “König” assumption that dichromatic observers lack one of the three cone types, but retain two that are identical in spectral sensitivity to the normal counterparts.<sup>35,54</sup> The identity of the two remaining cone types means that dichromats accept all color matches set by normal trichromats. The loss hypothesis now has a firm empirical foundation, because it has become possible to sequence and identify the photopigment opsin genes of normal, dichromatic and monochromatic observers.<sup>55,56</sup> As a result, individuals who conform to the loss assumption can be selected by genetic analysis. Thanks to the longer wavelength part of the visible spectrum being effectively dichromatic, because of the insensitivity of the S cones to longer wavelength lights, the unknown value,  $\bar{s}_G / \bar{s}_B$ , can also be derived directly from normal color-matching data (see Refs. 57 and 58 for details).

Several authors have estimated LMS cone spectral sensitivities using the loss hypothesis.<sup>8,53,59–66</sup> Figure 6 shows estimates by Smith and Pokorny<sup>53</sup> and Stockman and Sharpe.<sup>46</sup> The Smith-Pokorny estimates are a transformation of the Judd-Vos corrected CIE 1931 functions (see earlier). The Stockman-Sharpe estimates are a transformation of the Stiles and Burch 10° (see earlier) adjusted to 2° (see Ref. 21 for further information).



**FIGURE 6** S-, M-, and L-cone spectral sensitivity estimates of Stockman and Sharpe<sup>46</sup> (colored lines) compared with the estimates of Smith and Pokorny<sup>53</sup> (dashed black lines). The lower inset shows the lens pigment optical density spectrum (black line) and the macular pigment optical density spectrum (magenta line) from Stockman and Sharpe.<sup>46</sup> Note the logarithmic vertical scale—commonly used in such plots to emphasize small sensitivities. (Based on Fig. 5 of Stockman.<sup>206</sup>)



## Limits of Color-Matching Data

Specifying a stimulus using tristimulus values depends on having an accurate set of color-matching functions. The CMFs and cone fundamentals discussed in preceding sections are designed to be representative of a standard observer under typical viewing conditions. A number of factors limit the precision to which a standard color space can predict the individual color matches. We describe some of these factors below. Wyszecki and Stiles<sup>8</sup> provide a more detailed treatment.

For most applications, standard calculations are sufficiently precise. However, when high precision is required, it is necessary to tailor a set of color-matching functions to the individual and observing conditions of interest. Once such a set of color-matching functions or cone fundamentals is available, the techniques described in other sections may be used to compute corresponding color coordinates.

Standard sets of color-matching functions are summaries or means of color-matching results for a number of color-normal observers. There is small but systematic variability between the matches set by individual observers, and this variability limits the precision to which standard color-matching functions may be taken as representative of any given color-normal observer. A number of factors underlie the variability in color matching. Stiles and Burch carefully measured color-matching functions for 49 observers using 10° fields.<sup>33,34</sup> Webster and MacLeod analyzed individual variation in these color-matching functions.<sup>67</sup> They identified five primary factors that drive the variation in individual color matches. These are macular pigment density, lens pigment density, photopigment optical density, amount of rod intrusion into the matches, and variability in the absorption spectra of the L, M, and S cone photopigments.

**Macular Pigment Density** Light must pass through the ocular media before reaching the photoreceptors. At the fovea this includes the macula lutea, which contains macular pigment. This pigment absorbs lights of shorter wavelengths covering a broad spectral region centered on 460 nm (see inset of Fig. 6). There are large individual differences in macular pigment density, with peak densities at 460 nm ranging from 0.0 to about 1.2.<sup>68–70</sup>

**Lens Pigment Density** Light is focused on the retina by the cornea and the yellow pigmented crystalline lens. The lens pigment absorbs light mainly of short wavelengths (see inset of Fig. 6). Individual differences in lens pigment density range by as much as  $\pm 25$  percent of the mean density in young observers (<30 years old).<sup>71</sup> Lens pigment also increases with age,<sup>72,73</sup> resulting in systematic differences in color-matching functions between populations of different ages.<sup>74</sup>

**Photopigment Optical Density** The axial optical density of the photopigment in the photoreceptor outer segment depends on several factors, including the underlying photopigment extinction or absorbance spectra, outer segment length, and the photopigment concentration within the outer segment. All these factors can vary between individuals,<sup>75–82</sup> and within individuals. Photoreceptor outer segment length, and thus axial photopigment optical density, decreases with retinal eccentricity.<sup>83,84</sup> Although changes in photopigment optical density are typically neglected, they can become important under circumstances where very intense adapting fields (which dilute the photopigment by bleaching) are employed or where fixation is eccentric. See section “Adjusting Cone Spectral Sensitivities for Individual Differences” for corrections that account for changes in photopigment optical density.

**Variability in Photopigment  $\lambda_{\max}$**  Genetic and behavioral evidence shows that there are multiple versions of the human L- and M-cone photopigments.<sup>56,85–89</sup> This multiplicity is known as cone polymorphism. The most common genetic polymorphism is the substitution of alanine for serine at position 180 of the L-cone photopigment gene. This substitution produces a shift in the L-cone photopigment spectral sensitivity of several nanometers, with the A180 variant shifted toward shorter wavelengths relative to the S180 variant (see Ref. 89 for a review of shift estimates). In applications where precise knowledge of an individual’s cone spectral sensitivities is important, genotyping can now help provide key information.<sup>46,90</sup> Some individuals possess more than one variant of the L- or M- cone photopigment gene.<sup>55,91–93</sup>



**Color-Deficient Observers** A class of color-deficient individuals, known as anomalous red-green trichromats, are trichromatic but set color matches substantially different from color-normal observers. Anomalous red-green trichromacy is caused by the spectral sensitivity of either the L- or the M-cone photopigment being shifted from its normal location to an anomalous position that lies closer to the location of the spectral sensitivity function of the remaining normal M- or L-cone photopigment (for a review, see Ref. 89). These shifts result from the inheritance of hybrid LM- or ML-cone photopigment opsin genes, which are fusion genes produced by intragenic crossing over, containing the coding sequences of both L- and M-cone pigment genes. Measurements of the absorbance spectrum peaks of the hybrid pigments made *in vitro*<sup>87,94</sup> and *in vivo*<sup>95,96</sup> reveal a wide range of possible anomalous spectra that lie between the normal L- and M-cone spectra. The peak absorbances of the LM hybrid pigments cluster within about 8 nm of the peak absorbance of the normal M-cone pigment, while those of the ML hybrid pigments cluster within about 12 nm of the peak absorbance of the normal L-cone pigment (see Table 1 of Ref. 97). In protanomalous trichromats, one of the two polymorphic variants of the normal L-cone pigment has been replaced with a hybrid LM pigment, whereas in deuteranomalous trichromats one of the two polymorphic variants of the normal M-cone pigment has been replaced with a hybrid ML pigment.

Our development of colorimetric calculations in Sec. 10.5 can be used to tailor color specification in a particular application for color anomalous individuals, if their color-matching functions are known. Estimates of the cone sensitivities of color anomalous observers are available.<sup>47,98</sup> Estimates of the A180 and S180 variants of the Stockman and Sharpe 2° functions are tabulated in Table A of Ref. 99. Details of how to adjust cone fundamentals for different  $\lambda_{\max}$  values are discussed in section “Adjusting Cone Spectral Sensitivities for Individual Differences” (see also section “Photopigment Optical Density Spectra” of Ref. 21).

Some individuals require only two primaries in the color-matching experiment (i.e., they are dichromats) or in rare cases only one primary (i.e., they are monochromats). Dichromats, like anomalous trichromats, are referred to as color deficient. Monochromats are, however, truly color blind (except for rod-cone interactions at mesopic levels in single cone monochromats<sup>100</sup>). Most forms of monochromacy and dichromacy can be understood by assuming that the individual lacks one or more of the normal three types of cone photopigment.<sup>35,101</sup> Individuals who lack the L-, M-, or S-cone photopigments are known, respectively, as protanopes, deuteranopes, or tritanopes. Protanopes and deuteranopes are much more common than tritanopes.<sup>89</sup> Some protanopes and deuteranopes have only one of the two normal longer wavelength cone photopigments, and so are true loss dichromats. Some, however, have a single hybrid ML- or LM-cone photopigment, which is intermediate in spectral position between M and L, while others have two cone photopigments with identical or nearly identical spectral sensitivities. For dichromatic and monochromatic individuals with normal cone photopigments (i.e., those without hybrid photopigments), the use of standard color coordinates will produce acceptable results, since a match for all three cone types will also be a match for any subset of these types. In very rare cases, an individual has no cones at all and his vision is mediated entirely by rods. His visual matches can be predicted by the CIE scotopic luminosity function [see Table I (4.3.2) of Ref. 8].

Simple standard tests exist for identifying color-deficient and color-anomalous individuals. These include the Ishihara pseudoisochromatic plates,<sup>102</sup> the Farnsworth 100 hue test,<sup>103</sup> and the Rayleigh match.<sup>104</sup> For coverage of the available clinical tests see Ref. 105. Genetic analysis may also be used to identify the variants of cone pigments likely to be expressed by a given individual.<sup>46</sup>

**Retinal Inhomogeneity** Most standard colorimetric systems are based on color-matching experiments where the bipartite field was either 2° or 10° in diameter and viewed foveally. The distribution of photoreceptors is not homogeneous across the retina, however, and both macular pigment and photopigment optical density decline with eccentricity. Thus, CMFs that are accurate for the fovea do not necessarily describe color matching in the extra fovea. The CIE 1964 10° XYZ color-matching functions are designed for situations where the colors being judged subtend a large visual angle. Stockman and Sharpe<sup>46</sup> provide both 2° and 10° cone fundamentals.

Another consideration is that the absence of S cones in approximately the central 25-min diameter of vision makes color matches confined to that small region tritanopic.<sup>106–109</sup>

**Rod Intrusion** Both outside the fovea and at low light levels, rods can play a role in color-matching. Under conditions where rods play a role, there is a shift in the color-matching functions due to the contribution of rod signals. Wyszecki and Stiles<sup>8</sup> discuss approximate methods for correcting standard sets of color-matching functions when rods intrude into color vision.

**Chromatic Aberrations** By some standards, even the small (roughly 2°) fields used as the basis of most color coordinate systems are rather coarse. The optics of the eye contain chromatic aberrations which cause different wavelengths of light to be focused with different accuracy. These aberrations can cause a shift in the color-matching functions if the stimuli being matched have fine spatial structure. Two stimuli which are metameric at low spatial frequencies may no longer be so at high spatial frequencies. Such effects can be quite large.<sup>110–112</sup> It is possible to correct color coordinates for chromatic aberration if enough side information is available. Such correction is rare in practice but can be important for stimuli with fine spatial structure. Some guidance is available from the literature.<sup>111,113</sup> Another strategy available in the laboratory is to correct the stimulus for the chromatic aberration of the eye.<sup>114</sup>

## Adjusting Cone Spectral Sensitivities for Individual Differences

**Adjustments from Corneal to Photoreceptor Sensitivities** Cone spectral sensitivities and CMFs are measured with respect to light entering the observers' cornea. However, between the cornea and photoreceptor, the light passes through the pigmented crystalline lens, and in the fovea through the macula lutea. Both of these filters markedly reduce the observers' sensitivity to short-wavelength lights (see Fig. 6).

In the first part of this section, we describe how to adjust the cone spectral sensitivities back to their values at the photoreceptor. A related adjustment is to correct cone spectral sensitivities and CMFs for individual differences in lens and macular pigment densities.

The calculation of photoreceptor sensitivities is straightforward given the lens [ $d_{\text{lens}}(\lambda)$ ] and macular [ $d_{\text{mac}}(\lambda)$ ] density spectra, as well as the respective scaling constants,  $k_{\text{lens}}$  and  $k_{\text{mac}}$ , by which each should be multiplied. Beginning with the *quantal* corneal spectral sensitivity of, for example, the L cones [ $\bar{I}(\lambda)$ ], the filtering by the lens pigment [ $k_{\text{lens}} d_{\text{lens}}(\lambda)$ ] and the macular pigment [ $k_{\text{mac}} d_{\text{mac}}(\lambda)$ ] is removed:

$$\log_{10}[\bar{I}_r(\lambda)] = \log_{10}[\bar{I}(\lambda)] + k_{\text{lens}} d_{\text{lens}}(\lambda) + k_{\text{mac}} d_{\text{mac}}(\lambda) \quad (6)$$

to give  $\bar{I}_r(\lambda)$ , the spectral sensitivity of the cones at the photoreceptor. The mean or standard  $d_{\text{lens}}(\lambda)$  and  $d_{\text{mac}}(\lambda)$  spectra that are assumed appropriate for the Stockman and Sharpe 2° cone fundamentals are tabulated in Table 2 of their paper.<sup>46</sup> These densities correspond to a macular density of 0.35 at 460 nm, and a lens density of 1.765 at 400 nm. For the standard 2° observer, the values of  $k_{\text{lens}}$  and  $k_{\text{mac}}$  are set to 1, but should be adjusted appropriately for individual observers or groups of observers with different lens and macular densities. For the mean 10° observer of Stockman and Sharpe, the values of  $k_{\text{lens}}$  and  $k_{\text{mac}}$  are assumed to be 1 and 0.27, respectively.

To calculate back from photoreceptor to corneal sensitivities, the filtering by the lens and macular pigments is added back:

$$\log_{10}[\bar{I}(\lambda)] = \log_{10}[\bar{I}_r(\lambda)] - k_{\text{lens}} d_{\text{lens}}(\lambda) - k_{\text{mac}} d_{\text{mac}}(\lambda) \quad (7)$$

Again,  $k_{\text{lens}}$  and  $k_{\text{mac}}$  should be adjusted as appropriate.

Macular pigment density can be estimated psychophysically from the differences between spectral sensitivities measured centrally and peripherally (in the macular-free area). Note, however, that such estimates can be affected by other changes between the two locations, such as photopigment optical density (see Fig. 2.5 of Ref. 21). Relative estimates of lens density can be obtained psychophysically by measuring spectral sensitivities in a macular-free area of the retina, and then comparing them with

mean spectral sensitivity data. Typically, rod spectral sensitivities are measured<sup>115</sup> and then compared with mean rod data, such as the data for 50 observers measured by Crawford<sup>72</sup> to obtain the mean standard rod spectral sensitivity function,  $V'(\lambda)$ . Absolute lens density estimates can be obtained by comparing spectral sensitivities with photopigment spectra. See Ref. 21 for discussion.

**Adjustments for Photopigment Optical Density** As noted above, decreases and increases in photopigment optical density result in a narrowing or broadening, respectively, of the cone spectral sensitivity curves. Corrections for these changes are most easily applied to the cone fundamentals.

The photopigment optical density or absorbance spectra  $[\bar{l}_{OD}(\lambda)]$  can be calculated from photoreceptor spectral sensitivity  $[\bar{l}_r(\lambda)]$  given the value of  $D_{\text{peak}}$ , the peak optical density of the photopigment, thus:

$$\bar{l}_{OD}(\lambda) = \frac{-\log_{10}[1 - \bar{l}_r(\lambda)]}{D_{\text{peak}}} \quad (8)$$

Note that  $\bar{l}_r(\lambda)$  should be scaled before applying Eq. (8) for  $\bar{l}_{OD}(\lambda)$  to peak at 1. Stockman and Sharpe<sup>46</sup> assume L-, M-, and S-cone  $D_{\text{peak}}$  of 0.5, 0.5, and 0.4, respectively, for their mean 2° observer, and values of 0.38, 0.38, and 0.3 for their mean 10° observer.

The spectral sensitivity at the photoreceptor,  $\bar{l}_r(\lambda)$ , can be calculated from the normalized photopigment optical density spectrum,  $\bar{l}_{OD}(\lambda)$ , by the inversion of Eq. (8) (see Ref. 116):

$$\bar{l}_r(\lambda) = 1 - 10^{-D_{\text{peak}}\bar{l}_{OD}(\lambda)} \quad (9)$$

Calculations from corneal spectral sensitivities to retinal photopigment optical densities ignore changes in spectral sensitivity that may result from the structure of the photoreceptor or other ocular structures and pigments (unless they are incorporated in estimates of the lens or macular pigment density spectra).

Photopigment optical density can be estimated from the differences between spectral sensitivities or color matches obtained when the concentration of the photopigment is dilute and those obtained when it is in its normal concentration. This can be achieved psychophysically by comparing data obtained under bleached versus unbleached conditions or for obliquely versus axially presented lights. Bleaching measurements yield mean peak optical density values in the range 0.3 to 0.6, those that depend on oblique presentation in the range 0.7 to 1.0 and objective measures in the range 0.35 to 0.57. See Ref. 21 for discussion.

**Adjustments for Changes in Photopigment  $\lambda_{\text{max}}$**  Adjustments in the spectral position of the photopigment spectra can be affected by shifting them along an appropriate spectral scale before applying Eq. (7) to restore the prereceptor filtering, the appropriate scale being one that preserves the shapes of photopigment spectra, in general, as  $\lambda_{\text{max}}$  changes (i.e., their shape should be invariant). An early proposal was by Dartnall<sup>117</sup> who proposed a “nomogram” or template shape for photopigment spectra that was invariant when shifted along frequency or wavenumber ( $1/\lambda$ , in units of  $\text{cm}^{-1}$ ) scale. Shape invariance, however, is better preserved when spectra are plotted as a function of log frequency or log wavenumber  $[\log(1/\lambda)]$ ,<sup>118–120</sup> which is equivalent to log wavelength  $[\log(\lambda)]$  or normalized frequency  $(\lambda_{\text{max}}/\lambda)$ . Barlow<sup>121</sup> has also proposed an abscissa of the fourth root of wavelength  $(\sqrt[4]{\lambda})$ . A recent nomogram proposed by Govardovskii et al.<sup>122</sup> is seeing considerable use. See also Eq. (8) of Ref. 46 for human photopigment nomograms. Linear wavelength scales ( $\lambda$ ) should not be used to shift pigment templates unless the spectral shift is quite small.

## Opponent and Contrast Spaces

Cone coordinates are useful because they make explicit the responses of the initial physiological mechanisms thought to mediate color vision. A number of investigators have begun to use representations that attempt to represent the responses of subsequent postreceptor mechanisms. Two basic

ideas underlie these representations. The first is the general opponent processing model described in companion chapter (Chap. 11) in this volume. We call representations based on this idea opponent color spaces. The second idea is that stimulus contrast is more relevant than stimulus magnitude.<sup>123</sup> We call spaces that are based on this second idea modulation or contrast color spaces. Some color spaces are both opponent and contrast color spaces.

*Cone contrast space* To derive coordinates in the cone contrast color space, the stimulus is first expressed in terms of its cone coordinates. The cone coordinates of a white point are then chosen. Usually these are the cone coordinates of a uniform adapting field or the spatio-temporal average of the cone coordinates of the entire image sequence. The cone coordinates of the white point are subtracted from the cone coordinates of the stimulus and the resulting differences are normalized by the corresponding cone coordinates of the white point.

*The DKL color space* Derrington, Krauskopf, and Lennie<sup>124</sup> introduced an opponent modulation space that is now widely used. This space is closely related to the chromaticity diagram suggested by MacLeod and Boynton<sup>125</sup> (see also Ref. 126). To derive coordinates in the DKL color space, the stimulus is first expressed in cone coordinates. As with cone contrast space, the cone coordinates of a white point are then subtracted from the cone coordinates of the stimulus of interest. The next step is to reexpress the resulting difference as tristimulus values with respect to a new choice of primaries that are thought to isolate the responses of post-receptoral mechanisms.<sup>127,128</sup> The three primaries are chosen so that modulating two of them does not change the response of the photopic luminance mechanism (see section “Brightness Matching and Photometry”). The color coordinates corresponding to these two primaries are often called the constant B and constant R and G coordinates. Modulating the constant R and G coordinates of a stimulus modulates only the S cones. Modulating the constant B coordinate modulates both the L and M cones but keeps the S-cone response constant. Because the constant R and G coordinates are not allowed to change the response of the photopic luminance mechanism, the DKL color space is well-defined only if the S cones do not contribute to luminance. The third primary of the space is chosen so that it has the same relative cone coordinates as the white point. The coordinate corresponding to this third primary is called the luminance coordinate. Flitcroft<sup>110</sup> and Brainard<sup>129</sup> provide detailed treatments of the DKL color space.

*Caveats* The basic ideas underlying the use of opponent and modulation/contrast color spaces seem to be valid. On the other hand, there is not a general agreement about how signals from cones are combined into opponent channels, about how this combination depends on adaptation, or about how adaptation affects signals originating in the cones. Since a specific model of these processes is implicit in any opponent or modulation/contrast color space, coordinates in these spaces must be treated carefully. This is particularly true of contrast spaces, where the relation between the physical stimulus and coordinates in the space depends on the choice of white point. As a consequence, radically different stimuli can have identical coordinates in a contrast space. For example, 100 percent contrast monochromatic intensity gratings are all represented by the same coordinates in contrast color spaces, independent of their wavelength. Nonetheless, such stimuli appear very different to human observers. Identity of coordinates in a contrast color space does not imply identity of appearance across different choices of white points. See Ref. 129 for more extended discussion.

## Visualizing Color Data

A challenge facing today’s color scientist is to produce and interpret graphical representations of color data. Because color coordinates are three-dimensional, it is difficult to plot them on a two-dimensional page. Even more difficult is to represent a dependent measure of visual performance as a function of color coordinates. We discuss several approaches.

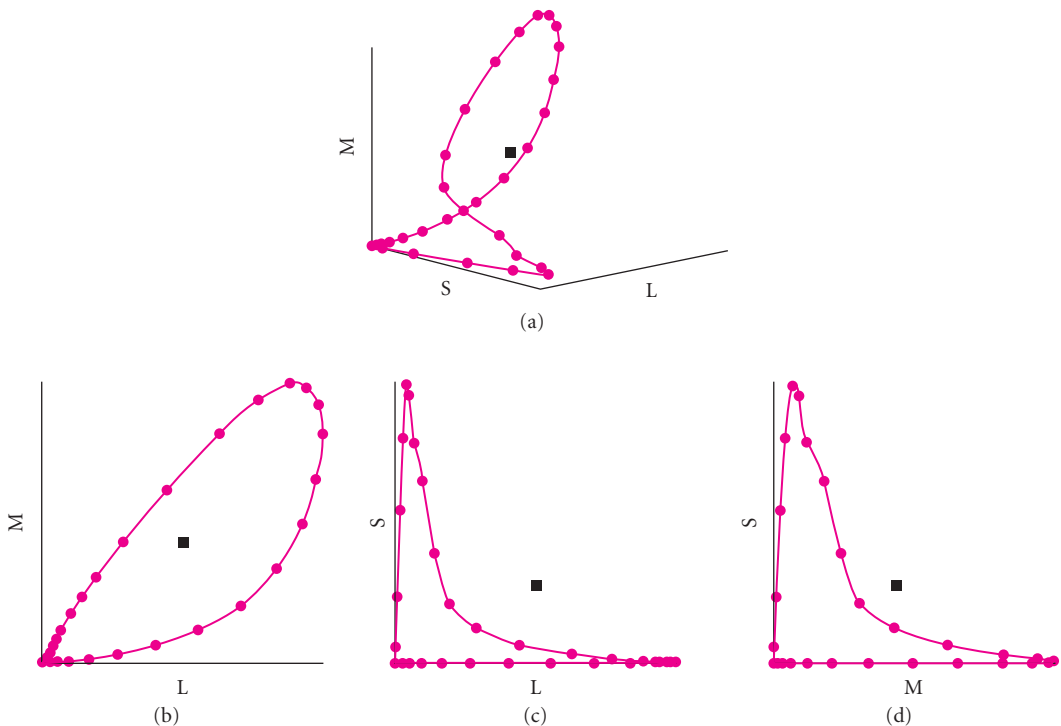
*Three-Dimensional Approaches* One strategy is to plot the three-dimensional data in perspective. In many cases the projection viewpoint may be chosen to provide a clear view of the regularities of

interest in the data. In Fig. 7a the spectrum locus is shown in the LMS tristimulus space. The three-dimensional structure of the data may be emphasized by the addition of various monocular depth cues to such figures, such as shading or drop lines. A number of computer graphics packages now provide facilities to aid in the preparation of three-dimensional perspective plots. Often these programs allow variation of the viewpoint and automatic inclusion of monocular depth cues.

Computer display technology also provides promise for improved methods of viewing three-dimensional data. For example, it is now possible to produce computer animations that show plots that vary over time. Such plots have the potential for representing multidimensional data in a manner that is more comprehensible to a human viewer than a static plot. Other interesting possibilities include the use of stereo depth cues and color displays. Online publication is making the use of such technologies more widely available for archival purposes.

Another approach to showing the three-dimensional structure of color data is to provide multiple two-dimensional views, as in a draftsman's sketch. This is illustrated in Fig. 7.

**Chromaticity Diagrams** A second strategy for plotting color data is to reduce the dimensionality of the data representation. One common approach is through the use of chromaticity coordinates. Chromaticity coordinates are defined so that any two lights with the same relative color coordinates have identical chromaticity coordinates. That is, the chromaticity coordinates of a light are invariant with respect to intensity scaling. Because chromaticity coordinates have one fewer degree of freedom than color coordinates, they can be described by just two numbers and plotted in a plane. We call a plot of chromaticity coordinates a chromaticity diagram. A chromaticity diagram eliminates all information about the intensity of a stimulus.



**FIGURE 7** Three-dimensional views of color data. The figure shows the color coordinates of an equal energy spectrum in color space defined by the human cone sensitivities (connected closed circles) and the color coordinates of CIE daylight D65 (closed squares). (a) The data in perspective. (b, c, and d) Three two-dimensional views of the same data.

There are many ways to normalize color coordinates to produce a set of chromaticity coordinates. In general, the chromaticity coordinates  $[r(\lambda), g(\lambda), \text{and } b(\lambda)]$  of the spectrum locus are related to the CMFs  $[\bar{r}(\lambda), \bar{g}(\lambda), \text{and } \bar{b}(\lambda)]$  as follows:

$$\begin{aligned} r(\lambda) &= \frac{\bar{r}(\lambda)}{\bar{r}(\lambda) + \bar{g}(\lambda) + \bar{b}(\lambda)} \\ g(\lambda) &= \frac{\bar{g}(\lambda)}{\bar{r}(\lambda) + \bar{g}(\lambda) + \bar{b}(\lambda)} \quad \text{and} \\ b(\lambda) &= \frac{\bar{b}(\lambda)}{\bar{r}(\lambda) + \bar{g}(\lambda) + \bar{b}(\lambda)} \end{aligned} \quad (10)$$

Given  $r(\lambda) + g(\lambda) + b(\lambda) = 1$ , only  $r(\lambda)$  and  $g(\lambda)$  are typically plotted, since  $b(\lambda)$  is  $1 - [r(\lambda) + g(\lambda)]$ .

For the special case of the 1931 CMFs, we have:

$$\begin{aligned} x(\lambda) &= \frac{\bar{x}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \quad \text{and} \\ y(\lambda) &= \frac{\bar{y}(\lambda)}{\bar{x}(\lambda) + \bar{y}(\lambda) + \bar{z}(\lambda)} \end{aligned} \quad (11)$$

Figure 8 shows the spectrum locus in the 1931 CIE  $x, y$  chromaticity space with an approximate representation of the colors associated with each coordinate.

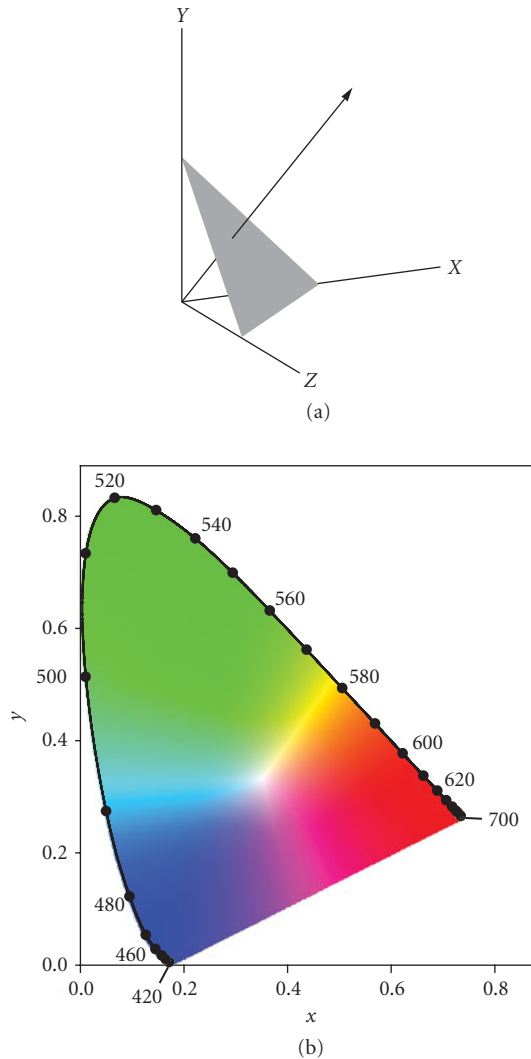
Neither  $r, g$  nor  $x, y$  chromaticity diagrams provide a strong visual connection between the data representation and the underlying cone mechanisms. For this reason, there is increasing use of chromaticity diagrams defined by the cone fundamentals. Figure 9 shows the spectrum locus plotted in  $l, m$  chromaticity coordinates.

A useful property of most chromaticity diagrams is that the chromaticity coordinates of the mixture of two lights is always a weighted combination of chromaticity coordinates of the individual lights. This is easily verified for the CIE 1931 chromaticity diagram by algebraic manipulation. Thus the chromaticity of a mixture of lights will plot somewhere on the chord connecting the chromaticities of the individual lights. Wyszecki and Stiles<sup>8</sup> review a number of standard chromaticity diagrams not discussed here.

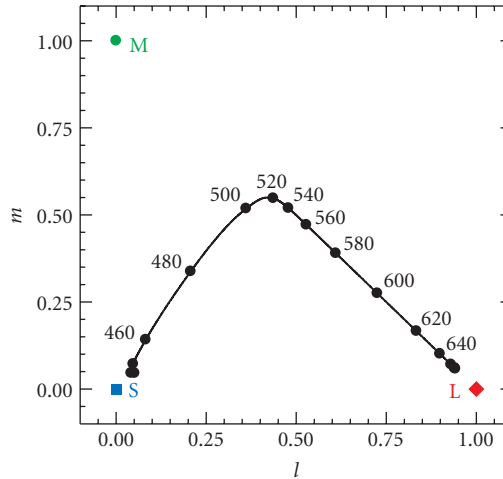
Implicit in the use of chromaticity coordinates is the assumption that scalar multiplication of the stimuli does not affect the visual performance being plotted. If the overall intensity of the stimuli matter, then the use of chromaticity coordinates can obscure important regularities. For example, the shape of color discrimination contours (see “Color Discrimination” in Sec. 10.6 and the Chap. 11) depends on how the overall intensity of the stimuli covaries with their chromaticities. Yet these contours are often plotted on a chromaticity diagram. This practice can lead to misinterpretation of the discrimination data. We recommend that plots of chromaticity coordinates be treated with some caution.

**Functions of Wavelength** Color data are often represented as functions of wavelength. The wavelength spectrum parameterizes a particular path through the three-dimensional color space. The exact path depends on how overall intensity covaries with wavelength. For an equal energy spectrum, the path is illustrated by Fig. 7.

Wavelength representations are particularly useful in situations where knowing the value of a function for the set of monochromatic stimuli provides a complete characterization of performance. Color-matching functions, for example, are usefully plotted as functions of wavelength because these functions may be used to predict the tristimulus values of any light. Plots of detection threshold



**FIGURE 8** CIE 1931  $xy$  chromaticity diagram. (a) A perspective view of the CIE 1931 XYZ tristimulus space. The ray shows a locus of points with constant chromaticity coordinates. The actual chromaticity coordinates for each ray are determined by where the ray intersects the plane described by the equation  $X + Y + Z = 1$ . This plane is indicated. The  $X$  and  $Y$  tristimulus values at the point of intersection are the  $x$  and  $y$  chromaticity coordinates for the ray. (b) The chromaticity coordinates of an equal energy spectrum with the interior colored to provide a rough indication of the color appearance of a stimulus of each chromaticity when viewed in a neutral context.



**FIGURE 9** Spectrum locus (continuous line) and selected wavelengths (filled circles) plotted in the Stockman and Sharpe<sup>46</sup> 2°  $l, m$  cone chromaticity space. The L- (red diamond), M- (green circle), and S- (blue square) cone fundamentals plot at (1,0), (0,1), and (0,0), respectively.

versus wavelength, on the other hand, cannot be used to predict the detection threshold for arbitrary lights.<sup>130</sup> Just as the chromaticity diagram tends to obscure the potential importance of manipulating the overall intensity of light, wavelength representations tend to obscure the potential importance of considering mixtures of monochromatic lights.

## Colorimetric Measurements

To apply the formulas described in this chapter, it is often necessary to measure the colorimetric properties of stimuli. The most general approach is to measure the full spectral power distribution of the stimuli. Often, however, it is not necessary to know the full spectral power distribution; knowledge of the tristimulus values (in some standard color space) is sufficient. For example, the color space transformations summarized in Table 3 depend on the full spectral power distributions of the primaries only through their tristimulus values.

Specialized instruments, called colorimeters, can measure tristimulus values directly. These instruments typically operate using the same principles as photometers with the exception that they have three calibrated filters rather than just one. Each filter mimics the spectral shape of one of the color-matching functions. Wyszecki and Stiles discuss the basics of colorimeter design.<sup>8</sup> Colorimeters are generally less expensive than radiometers and are thus an attractive option when full spectral data are not required.

Two caveats are worth noting. First, it is technically difficult to design filters that exactly match a desired set of color-matching functions. Generally, commercial colorimeters are calibrated so that they give accurate readings for stimuli with broadband spectral power distributions. For narrow band stimuli (e.g., the light emitted by the red phosphor of many color monitors) the reported readings may be quite inaccurate. Second, most colorimeters are designed according to the CIE 1931 standard. This may not be an optimal choice for the purpose of predicting the matches of an average human observer.



**TABLE 3** Color Space Transformations

		Spectral Functions Known	
Source	Destination	$\mathbf{M}$	Notes
Primaries $\mathbf{P}_1$	CMFs $\mathbf{T}_2$	$\mathbf{M} = \mathbf{T}_2 \mathbf{P}_1$	
CMFs $\mathbf{T}_1$	Primaries $\mathbf{P}_2$	$\mathbf{M} = (\mathbf{T}_1 \mathbf{P}_2)^{-1}$	
Primaries $\mathbf{P}_1$	Primaries $\mathbf{P}_2$	$\mathbf{M} = (\mathbf{TP}_2)^{-1}(\mathbf{TP}_1)$	$\mathbf{T}$ is any set of CMFs.
CMFs $\mathbf{T}_1$	CMFs $\mathbf{T}_2$	$\mathbf{T}_2 = \mathbf{MT}_1$	Use regression to find $\mathbf{M}$ .
One Space Specified in Terms of Other			
Known Tristimulus Coordinates		How to Construct $\mathbf{M}$	
Source primaries known in destination space.		Put them in columns of $\mathbf{M}$ .	
Source CMFs known in destination space.		Put them in rows of $\mathbf{M}^{-1}$ .	
Destination primaries known in source space.		Put them in columns of $\mathbf{M}^{-1}$ .	
Destination CMFs known in source space.		Put them in rows of $\mathbf{M}$ .	

CMFs stands for color matching functions.  
 The table summarizes how to form the matrix  $\mathbf{M}$  that transforms color coordinates between two spaces.

## 10.5 MATRIX REPRESENTATIONS AND CALCULATIONS

### Introduction

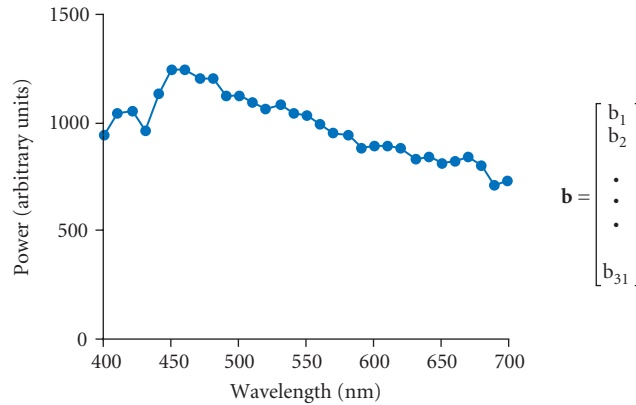
In the remainder of the chapter we move away from the conventional representation of colorimetric data and formulae as continuous functions of wavelength to their representation as vectors and matrices. Matrix algebra greatly simplifies the implementation of colorimetry on digital computers. Although a discrete representation provides only samples of the underlying function of wavelength, the information loss caused by this sampling can be made arbitrarily small by sampling at smaller intervals.

**Notation for Matrix Calculations** The conventional notation used in colorimetry does not lend itself easily to matrix and vector representations, and at the risk of some confusion between the notation used in Secs. 10.3 “Fundamentals of Colorimetry” and 10.4 “Color Coordinate Systems” and that used here and in Sec. 10.6 “Topics,” we now switch notational conventions. Table 2 provides a glossary of the major symbol usage for the matrix formulation. The following notational conventions are used: (a) scalars are denoted with italic symbols, (b) vectors are denoted with lowercase bold symbols, and (c) matrices are denoted with uppercase bold symbols. Symbols used in the appendix are generic.

### Stimulus Representation

**Vector Representation of Spectral Functions** Suppose that spectral power density has been measured at  $N_\lambda$  discrete sample wavelengths  $\lambda_1 \dots \lambda_{N_\lambda}$ , each separated by an equal wavelength step  $\Delta\lambda$ . As shown in Fig. 10, we can represent the measured spectral power distribution using an  $N_\lambda$  dimensional column vector  $\mathbf{b}$ . The  $n$ th entry of  $\mathbf{b}$  is simply the measured power density at the  $n$ th sample wavelength multiplied by  $\Delta\lambda$ . Note that the values of the sample wavelengths  $\lambda_1 \dots \lambda_{N_\lambda}$  and wavelength step  $\Delta\lambda$  are not explicit in the vector representation. These values must be provided as side information when they are required for a particular calculation. In colorimetric applications, sample wavelengths are typically spaced evenly throughout the visible spectrum at steps of 1, 5, or 10 nm. We follow the convention that the entries of  $\mathbf{b}$  incorporate  $\Delta\lambda$ , however, so that we need not represent  $\Delta\lambda$  explicitly when we approximate integrals over wavelength.

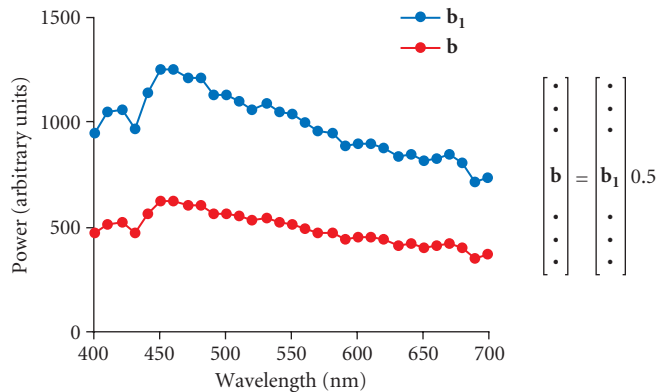
**Manipulation of Light** Intensity scaling is an operation that changes the overall power of a light at each wavelength without altering the relative power between any pair of wavelengths.



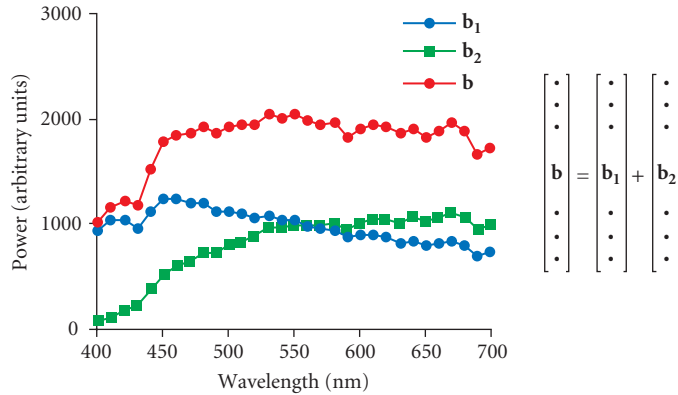
**FIGURE 10** The vector representation of functions of wavelength. The plot shows a spectral power distribution measured at 10-nm intervals between 400 and 700 nm. Each point on the plot represents the power at a single sample wavelength. The vector  $\mathbf{b}$  on the right depicts the vector representation of the same spectral power distribution. The  $n$ th entry of  $\mathbf{b}$  is simply the measured power density at the  $n$ th sample wavelength times  $\Delta\lambda$ . Thus  $b_1$  is derived from the power density at 400 nm,  $b_2$  is derived from the power density at 410 nm, and  $b_{31}$  is derived from the power density at 700 nm.

The superposition of two lights is an operation that produces a new light whose power at each wavelength is the sum of the power in the original lights at the corresponding wavelength. The effects of both manipulations may be expressed using matrix algebra.

We use scalar multiplication to represent intensity scaling. If a light  $\mathbf{b}_1$  is scaled by a factor  $a$ , then the result  $\mathbf{b}$  is given by the equation  $\mathbf{b} = \mathbf{b}_1 a$ . The expression  $\mathbf{b}_1 a$  represents a vector whose entries are obtained by multiplying the entries of the vector  $\mathbf{b}_1$  by the scalar  $a$ . Similarly, we use vector addition to represent superposition. If we superimpose two lights  $\mathbf{b}_1$  and  $\mathbf{b}_2$ , then the result  $\mathbf{b}$  is given by the equation  $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$ . The expression  $\mathbf{b}_1 + \mathbf{b}_2$  represents a vector whose entries are obtained by adding the entries of the vectors  $\mathbf{b}_1$  and  $\mathbf{b}_2$  component by component. Figures 11 and 12 depict both of these operations.



**FIGURE 11** Representation of intensity scaling. Suppose that light  $\mathbf{b}$  is created by reducing the power in light  $\mathbf{b}_1$  by a factor of 0.5 at each wavelength. The result is shown graphically in the plot. The vector representation of the same relation is given by the equation  $\mathbf{b} = \mathbf{b}_1 0.5$ .



**FIGURE 12** Representation of superposition. Suppose that light  $\mathbf{b}$  is created by superimposing two lights  $\mathbf{b}_1$  and  $\mathbf{b}_2$ . The result is shown graphically in the plot. The vector representation of the same relation is given by the equation  $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$ .

**Linear Models for Spectral Functions** Intensity scaling and superposition may be used in combination to produce a wide range of spectral functions. Suppose that we have a set of  $N_b$  lights that we can individually scale and superimpose. Let the vectors  $\mathbf{b}_1 \dots \mathbf{b}_{N_b}$  represent the spectral power distributions of these lights. In this case, we can produce any spectral power distribution  $\mathbf{b}$  that has the form

$$\mathbf{b} = \mathbf{b}_1 a_1 + \dots + \mathbf{b}_{N_b} a_{N_b} \quad (12)$$

Now suppose we know that a spectral function  $\mathbf{b}$  is constrained to have the form of Eq. (12) where the vectors  $\mathbf{b}_1 \dots \mathbf{b}_{N_b}$  are known. Then we can specify  $\mathbf{b}$  completely by providing the values of the scalars  $a_1 \dots a_{N_b}$ . If the number of primaries  $N_b$  is less than the number of sample wavelengths  $N_\lambda$ , then this specification is more efficient (i.e., requires fewer numbers) than specifying the entries of  $\mathbf{b}$  directly. We say that the spectral functions that satisfy Eq. (12) are described by (or lie within) a linear model. We call  $N_b$  the dimension of the linear model. We call the vectors  $\mathbf{b}_1 \dots \mathbf{b}_{N_b}$  the basis vectors for the model. We call the scalars  $a_1 \dots a_{N_b}$  required to construct any particular spectral function the model weights for that function.

**Matrix Representation of Linear Models** Equation (12) can be written using vector and matrix notation. Let  $\mathbf{B}$  be an  $N_\lambda$  by  $N_b$  dimensional matrix whose columns are the basis vectors  $\mathbf{b}_1 \dots \mathbf{b}_{N_b}$ . We call  $\mathbf{B}$  the basis matrix for the linear model. The composition of the basis matrix is shown pictorially on the left of Fig. 13. Let  $\mathbf{a}$  be an  $N_b$  dimensional vector whose entries are the weights  $a_1 \dots a_{N_b}$ . Figure 13 also depicts the vector  $\mathbf{a}$ . Using  $\mathbf{B}$  and  $\mathbf{a}$  we can re-express Eq. (12) as the matrix multiplication

$$\mathbf{b} = \mathbf{B}\mathbf{a} \quad (13)$$

The equivalence of Eqs. (12) and (13) may be established by direct expansion of the definition of matrix multiplication (see App. A). A useful working intuition for matrix multiplication is that the effect of multiplying a matrix times a vector (e.g.,  $\mathbf{B}\mathbf{a}$ ) is to produce a new vector (e.g.,  $\mathbf{b}$ ) that is a weighted superposition of the columns of the matrix (e.g.,  $\mathbf{B}$ ), where the weight for column  $\mathbf{b}_i$  is given by the  $i$ th weight,  $a_i$ , in  $\mathbf{a}$ .

**Use of Linear Models** When we know that a spectral function is described by a linear model, we can specify it by using the weight vector  $\mathbf{a}$ . The matrix  $\mathbf{B}$ , which is determined by the basis vectors, specifies the side information necessary to convert the vector  $\mathbf{a}$  back to the discrete wavelength

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_{N_b} \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{N_b} \end{bmatrix} \quad \begin{bmatrix} \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{a} \end{bmatrix}$$

**FIGURE 13** Vector representation of linear models. The matrix  $\mathbf{B}$  represents the basis vectors of the linear model. The vector  $\mathbf{a}$  represents the model weights required to form a particular spectral power distribution  $\mathbf{b}$ . The relation between  $\mathbf{b}$ ,  $\mathbf{a}$ , and  $\mathbf{B}$  is given by Eq. (13) and is depicted on the right of the figure.

representation  $\mathbf{b}$ . When we represent spectral functions in this way, we say that we are representing the functions within the specified linear model.

Representing spectral functions within a small-dimensional linear model places strong constraints on the form of functions. As the dimension of the model grows, linear models can represent progressively wider classes of spectral power distributions. In many cases of interest, there is prior information that allows us to assume that spectra are indeed described by a linear model. A common example of this situation is the light emitted from a computer controlled color monitor. Such a monitor produces different spectral power distributions by scaling the intensity of the light emitted by three different types of phosphor (see Chap. 22). Thus the emitted light lies within a three-dimensional linear model whose basis vectors are given by the emission spectra of the monitor's phosphors. Linear model constraints also turn out to be useful for describing naturally occurring surface and illuminant spectra.

Note that representing spectral functions within linear models is a generalization of, rather than an alternative to, the more traditional wavelength representation. To understand this, we need to only note that we can choose the basis vectors of the linear model to be discrete delta functions centered at each sample wavelength. We refer to this special choice of basis vectors as the identity basis or wavelength basis. We refer to the corresponding linear model as the identity model. For the identity model, the basis matrix  $\mathbf{B}$  is the (square)  $N_\lambda$  by  $N_\lambda$  identity matrix, where  $N_\lambda$  is the number of sample wavelengths. The identity matrix contains ones along its main diagonal and zeros elsewhere. Multiplying the identity matrix times any vector simply results in the same vector. From Eq. (13), we can see that when  $\mathbf{B}$  is the identity matrix, the representation of any light  $\mathbf{b}$  within the linear model is simply  $\mathbf{b} = \mathbf{a}$ .

**Sampling the Visible Spectrum** To use a discrete representation for functions of wavelength, it is necessary to choose a sampling range and sampling increment. Standard practice varies considerably. The Commission Internationale de l'Éclairage (International Commission on Illumination, commonly referred to as the CIE) provides recommendations on the sampling of the visible spectrum.<sup>10</sup> For many applications, using 5-nm increments between 380 and 780 nm is sufficient, and coarser sampling at 10 nm between 400 and 700 nm is not uncommon. In cases where a subset of the spectral data required for a calculation is not available, interpolation or extrapolation may be used to estimate the missing values.

## Vector Representation of Colorimetric Data

**The Basic Color-Matching Experiment** In the maximum saturation experiment color matching (see Fig. 3), observer's task is to adjust the spectral power distribution  $\mathbf{b}_m$  of the matching light on one side of the bipartite field to match the appearance of the spectral power distribution  $\mathbf{b}_t$  of the

test light on the other side of the field. As described above, the matching light's spectral power distribution is described completely by a three-dimensional linear model whose basis vectors are the primary lights' spectral power distributions. The tristimulus values of a test light are precisely the linear model weights required to form the matching light. We denote the primary spectral power distributions by the vectors  $\mathbf{p}_1 \dots \mathbf{p}_3$ . The associated linear model matrix  $\mathbf{P}$  contains these vectors in its three columns. We denote the tristimulus values of a light using the three-dimensional vector  $\mathbf{t}$ . Thus we can use the tristimulus values of any test light  $\mathbf{b}$  to reconstruct a matching light  $\mathbf{P}\mathbf{t}$  such that:

$$\mathbf{b} \sim \mathbf{P}\mathbf{t} \quad (14)$$

We emphasize that in general that  $\mathbf{P}\mathbf{t}$  will not be equal to  $\mathbf{b}$ .

*Grassmann's Laws Revisited* In vector notation, the proportionality law states

$$\begin{aligned} \text{if} & \quad \mathbf{b}_1 \sim \mathbf{b}_2 \\ \text{then} & \quad \mathbf{b}_1 a \sim \mathbf{b}_2 a \end{aligned} \quad (15)$$

where  $a$  is a scalar that represents any intensity scaling. The additivity law states

$$\begin{aligned} \text{if} & \quad \mathbf{b}_1 \sim \mathbf{b}_2 \quad \text{and} \quad \mathbf{b}_3 \sim \mathbf{b}_4 \\ \text{then} & \quad \mathbf{b}_1 + \mathbf{b}_3 \sim \mathbf{b}_2 + \mathbf{b}_4 \end{aligned} \quad (16)$$

The proportionality law allows us to determine the relation between the tristimulus values of a light and the tristimulus values of a scaled version of that light. Suppose that  $\mathbf{b} \sim \mathbf{P}\mathbf{t}$ . Applying the proportionality law, we conclude that for any scalar  $a$ , we have  $\mathbf{b}a \sim (\mathbf{P}\mathbf{t})a$ . Because matrix multiplication is associative, we can conclude that

$$\begin{aligned} \text{if} & \quad \mathbf{b} \sim \mathbf{P}\mathbf{t} \\ \text{then} & \quad \mathbf{b}a \sim \mathbf{P}(\mathbf{t}a) \end{aligned} \quad (17)$$

This means that the tristimulus values of a light  $\mathbf{b}a$  may be obtained by scaling the tristimulus values  $\mathbf{t}$  of the light  $\mathbf{b}$ . A similar argument shows that the additivity law determines the relation between the tristimulus values of two lights and the tristimulus values of their superposition

$$\begin{aligned} \text{if} & \quad \mathbf{b}_1 \sim \mathbf{P}\mathbf{t}_1 \quad \text{and} \quad \mathbf{b}_2 \sim \mathbf{P}\mathbf{t}_2 \\ \text{then} & \quad \mathbf{b}_1 + \mathbf{b}_2 \sim \mathbf{P}(\mathbf{t}_1 + \mathbf{t}_2) \end{aligned} \quad (18)$$

*Implication of Grassmann's laws* If the tristimulus values of the basis vectors for a linear model are known, then Grassmann's laws allow us to determine the tristimulus values of any light within the linear model. Let  $\mathbf{t}_1 \dots \mathbf{t}_{N_b}$  be the tristimulus values corresponding to the model basis vectors and let  $\mathbf{T}_B$  be the 3 by  $N_b$  matrix whose columns are  $\mathbf{t}_1 \dots \mathbf{t}_{N_b}$ . For any light  $\mathbf{b}$  within the linear model, we can write that  $\mathbf{b} = \mathbf{B}\mathbf{a}$  where  $\mathbf{a}$  now represents the vector of weights to be applied to each basis vector to produce  $\mathbf{b}$ . By expanding this matrix product and applying Eqs. (17) and (18), it is possible to show that the tristimulus values of  $\mathbf{b}$  are given by the matrix product

$$\mathbf{t} = \mathbf{T}_B \mathbf{a} \quad (19)$$

Equation (19) is very important. It tells how to compute the tristimulus values for any light within a linear model from the tristimulus values for each of the basis vectors. Thus a small number of color matches (one for each of the basis vectors) allow us to predict color matches for a large number of lights, that is, any light within the linear model.

*Color-Matching Functions and Cone Fundamentals* Let  $\mathbf{T}$  be the matrix of tristimulus values for the basis vectors of the identity model (i.e., equal energy monochromatic lights). In this case  $\mathbf{T}$  has

dimensions 3 by  $N_\lambda$ , where  $N_\lambda$  is the number of sample wavelengths. Each column of  $\mathbf{T}$  is the tristimulus values for a monochromatic light. Within the identity model, the representation of any light  $\mathbf{b}$  is simply  $\mathbf{b}$  itself. From Eq. (19) we conclude directly that the tristimulus values for any light are given by

$$\mathbf{t} = \mathbf{T}\mathbf{b} \quad (20)$$

Once we know the tristimulus values for a set of monochromatic lights centered at each of the sample wavelengths, we can use Eq. (20) to compute the tristimulus values of any light. Equation (20) is the matrix algebra version of Eq. (2).

We can regard each of the rows of  $\mathbf{T}$  as a function of wavelength, and in doing so we can identify these as the standard color-matching functions obtained with respect to the primaries used in the matching experiment.

We can similarly represent the spectral sensitivity functions of the three classes of cones by the rows of a 3 by  $N_\lambda$  matrix  $\mathbf{R}$ . Let  $\mathbf{r}$  be a three-dimensional vector whose entries are the cone quantal absorption rates of an arbitrary light represented by  $\mathbf{b}$ . We can compute the absorption rates through the matrix equation

$$\mathbf{r} = \mathbf{R}\mathbf{b} \quad (21)$$

This computation accomplishes the wavelength-by-wavelength multiplication and summation for each cone class.

We use the term cone coordinates to refer to the vector  $\mathbf{r}$ . We can relate cone coordinates to tristimulus values in a straightforward manner. Suppose that in a color-matching experiment performed with primaries  $\mathbf{P}$  we find that a light  $\mathbf{b}$  has tristimulus values  $\mathbf{t}$ . From our mechanistic explanation of color matching in terms of the cones, we have that

$$\mathbf{r} = \mathbf{R}\mathbf{b} = \mathbf{R}\mathbf{P}\mathbf{t} \quad (22)$$

Recall that the matrix  $\mathbf{P}$  holds the power spectrum of the three primaries used in a matching experiment. If we define the matrix  $\mathbf{M}_{T,R} = (\mathbf{R}\mathbf{P})$ , we see that the tristimulus values of a light are related to its cone coordinates by the linear transformation

$$\mathbf{r} = \mathbf{M}_{T,R}\mathbf{t} \quad (23)$$

By comparing Eq. (20) with Eq. (22) and noting that these equations hold for any light  $\mathbf{b}$ , we derive

$$\mathbf{R} = \mathbf{M}_{T,R}\mathbf{T} \quad (24)$$

Equation (24) has the key implication, which we took advantage of in Sec. 10.4 “Color Coordinate Systems” [Eqs. (3) and (4)], that the color-matching experiment determines the cone sensitivities up to a free linear transformation, the matrix  $\mathbf{M}_{T,R}$  of Eq. (24).

## Transformations between Color Spaces

Because of the large number of color spaces currently in use, the ability to transform data between various color spaces is of considerable practical importance. The derivation of such transformations depends on what is known about the source and destination color spaces. Below we discuss cases where both the source and destination color space are derived from the same underlying observer (i.e., when the source and destination color spaces both predict identical color matches). Table 3 summarizes these transformations. When the source and destination color spaces are characterized by a different underlying observer (e.g., if they are based on different CMFs) the transformation is more difficult and often cannot be done exactly. We discuss possible approaches in section “Color Cameras and Other Visual Systems.”

*Source Primaries and Destination Color-Matching Functions Known* Let  $\mathbf{P}_1$  be the matrix representing the spectral power distributions of a set of known primaries, with one primary in each column. Let  $\mathbf{T}_2$  be the matrix representing a known set of color-matching functions (e.g., the CIE 1931 XYZ color-matching functions), with one function in each of its three rows. We would like to determine a transformation between the color coordinate system specified by  $\mathbf{P}_1$  and that specified by  $\mathbf{T}_2$ . For example, linearized frame buffer values input to a computer-controlled color monitor may be thought of as tristimulus values in a color space defined by the monitor's phosphor emission spectra. The transformation we seek allows computation of the CIE 1931 tristimulus values from the linearized frame buffer values.

We start by using Eq. (20) to compute the tristimulus values, with respect to  $\mathbf{T}_2$ , for the three primary lights specified by  $\mathbf{P}_1$ . Each of these primaries is contained in a column of  $\mathbf{P}_1$ , so that we may perform this calculation directly through the matrix multiplication

$$\mathbf{M}_{P,T} = \mathbf{T}_2 \mathbf{P}_1 \quad (25)$$

Let the matrix  $\mathbf{P}_2$  represent the destination primaries. We do not need to know these explicitly, only that they exist. The meaning of Eq. (25) is that

$$\mathbf{P}_1 \sim \mathbf{P}_2 \mathbf{M}_{P,T} \quad (26)$$

where we have generalized the symbol “ $\sim$ ” to denote a column-by-column visual match for the matrices on both sides of the relation. This relation follows because the columns of  $\mathbf{M}_{P,T}$  specify how the destination primaries should be mixed to match the source primaries. Equation (26) tells us that we can substitute the three lights represented by the columns of  $\mathbf{P}_2 \mathbf{M}_{P,T}$  for the three lights represented by the columns of  $\mathbf{P}_1$  in any color matching experiment. In particular, we may make this substitution for any light  $\mathbf{b}$  with tristimulus values  $\mathbf{t}_1$  in the source color coordinate system. We have

$$\mathbf{b} \sim \mathbf{P}_1 \mathbf{t}_1 \sim \mathbf{P}_2 \mathbf{M}_{P,T} \mathbf{t}_1 \quad (27)$$

By inspection, this tells us that the three-dimensional vector

$$\mathbf{t}_2 = \mathbf{M}_{P,T} \mathbf{t}_1 \quad (28)$$

is the tristimulus values of  $\mathbf{b}$  in the destination color coordinate system.

Equation (28) provides us with the means to transform tristimulus values from a coordinate system where the primaries are known to one where the color-matching functions are known. The transformation matrix  $\mathbf{M}_{P,T}$  required to perform the transformation depends only on the known primaries  $\mathbf{P}_1$  and the known color-matching functions  $\mathbf{T}_2$ . Given these,  $\mathbf{M}_{P,T}$  may be computed directly from Eq. (25).

*Source Color-Matching Functions and Destination Primaries Known* A second transformation applies when the color-matching functions in the source color space and the primaries in the destination color space are known. This will be the case, for example, when we wish to render a stimulus specified in terms of CIE 1931 tristimulus values on a calibrated color monitor.

Let  $\mathbf{T}_1$  represent the known color-matching functions and  $\mathbf{P}_2$  represent the known primaries. By applying Eq. (28) we have that the relation between source tristimulus values and the destination tristimulus values is given by  $\mathbf{t}_1 = \mathbf{M}_{T,P} \mathbf{t}_2$ . This is a system of linear equations that we may solve to find an expression for  $\mathbf{t}_2$  in terms of  $\mathbf{t}_1$ . In particular, as long as the matrix  $\mathbf{M}_{T,P}$  is nonsingular, we can convert tristimulus values using the relation

$$\mathbf{t}_2 = \mathbf{M}_{T,P}^{-1} \mathbf{t}_1 \quad (29)$$

where we define

$$\mathbf{M}_{T,P} = (\mathbf{M}_{P,T})^{-1} = (\mathbf{T}_1 \mathbf{P}_2)^{-1} \quad (30)$$

**Source and Destination Primaries Known** A third transformation applies when the primaries of both the source and destination color spaces are known. One application of this transformation is to generate matching stimuli on two different calibrated monitors.

Let  $\mathbf{P}_1$  and  $\mathbf{P}_2$  represent the two sets of primaries. Let  $\mathbf{T}$  represent a set of color-matching functions for any human color coordinate system. (There is no requirement that the color-matching functions be related to either the source or the destination primaries. For example, the CIE 1931 XYZ color-matching functions might be used.) To do the conversion, we simply use Eq. (28) to transform from the color coordinate system described by  $\mathbf{P}_1$  to the coordinate system described by  $\mathbf{T}$ . Then we use Eq. (29) to transform from the coordinates system described by  $\mathbf{T}$  to the coordinate system described by  $\mathbf{P}_2$ . The overall transformation is given by

$$\mathbf{t}_2 = \mathbf{M}_{P_2P} \mathbf{t}_1 = (\mathbf{M}_{T,P_2})(\mathbf{M}_{P_1,T}) \mathbf{t}_1 = (\mathbf{TP}_2)^{-1} (\mathbf{TP}_1) \mathbf{t}_1 \quad (31)$$

It should not be surprising that this transformation requires the specification of a set of color-matching functions. These color-matching functions are the only source of information about the human observer in the transformation equation.

**Source and Destination Color-Matching Functions Known** Finally, it is sometimes of interest to transform between two color spaces that are specified in terms of their color-matching functions. An example is transforming between the space defined by the Stiles and Burch 10° color-matching functions<sup>34</sup> and the space defined by the Stockman and Sharpe 10° cone fundamentals.<sup>46</sup>

Let  $\mathbf{T}_1$  and  $\mathbf{T}_2$  represent the source and destination color-matching functions. Our development above assures us that there is some three-by-three transformation matrix, call it  $\mathbf{M}_{T,T}$  that transforms color coordinates between the two systems. Recall that the columns of  $\mathbf{T}_1$  and  $\mathbf{T}_2$  are themselves tristimulus values for corresponding monochromatic lights. Thus  $\mathbf{M}_{T,T}$  must satisfy

$$\mathbf{T}_2 = \mathbf{M}_{T,T} \mathbf{T}_1 \quad (32)$$

This is a system of linear equations where the entries of  $\mathbf{M}_{T,T}$  are the unknown variables. This system may be solved using standard regression methods. Once we have solved for  $\mathbf{M}_{T,T}$ , we can transform tristimulus values using the equation

$$\mathbf{t}_2 = \mathbf{M}_{T,T} \mathbf{t}_1 \quad (33)$$

The transformation specified by Eq. (33) will be exact as long as the two sets of color-matching functions  $\mathbf{T}_1$  and  $\mathbf{T}_2$  characterize the performance of the same observer. One sometimes wishes, however, to transform between two color spaces that are defined with respect to different observers. For example, one might want to convert CIE 1931 XYZ tristimulus values to Judd-Vos modified tristimulus values. Although the regression procedure described here will still produce a transformation matrix in this case, the result of the transformation is not guaranteed to be correct.<sup>4</sup> We will return to this topic in section “Color Cameras and Other Visual Systems.”

**Interpreting the Transformation Matrix** It is useful to interpret the rows and columns of the matrices derived in the preceding sections. Let  $\mathbf{M}$  be a matrix that maps the color coordinates from a source color space to a destination color space. Both source and destination color spaces are associated with a set of primaries and a set of color-matching functions. From our derivations above, we can conclude that the columns of  $\mathbf{M}$  are the coordinates of the source primaries in the destination color space [see Eq. (25)] and the rows of  $\mathbf{M}$  provide the destination color-matching functions with respect to the linear model whose basis functions are the primaries of source color space (see subsection “Use of Linear Modes” in Sec. 10.5). Similarly, the columns of  $\mathbf{M}^{-1}$  are the coordinates of the destination primaries in the source color-matching space and the rows of  $\mathbf{M}^{-1}$  are the source color-matching functions with respect to the linear model whose basis functions are the primaries of the destination color space. Thus in many cases it is possible to construct the matrix  $\mathbf{M}$  without full knowledge of the spectral functions. This can be of practical importance. For example, monitor manufacturers often specify the CIE 1931 tristimulus values of their monitors’ phosphors. In addition, colorimeters that measure tristimulus values directly are often more readily available than spectral radiometers.



**Transforming Primaries and Color-Matching Functions** We have shown that color coordinates in any two color spaces may be related by applying a linear transformation  $\mathbf{M}$ . The converse is also true. If we pick any nonsingular linear transformation  $\mathbf{M}$  and apply it to a set of color coordinates we have defined a new color space that will successfully predict color matches. The color-matching functions for this new space will be given by  $\mathbf{T}_2 = \mathbf{M}\mathbf{T}_1$ . A set of primaries for the new space will be given by  $\mathbf{P}_2 = \mathbf{P}_1\mathbf{M}^{-1}$ . These derived primaries are not unique. Any set of primaries that match the constructed primaries will also work.

The fact that new color spaces can be constructed by applying linear transformations has an important implication for the study of color. If we restrict attention to what we may conclude from the color-matching experiment, we can only determine the psychological representation of color up to a free linear transformation. There are two attitudes one can take toward this fact. The conservative attitude is to refrain from making any statements about the nature of color vision that depend on a particular choice of color space. The other is to appeal to experiments other than the color-matching experiment to choose a privileged representation.

## 10.6 TOPICS

### Surfaces and Illuminants

As shown in Fig. 1, the light reaching the eye is often formed when light from an illuminant reflects from a surface. Illuminants and surfaces are of interest in color reproduction applications involving inks, paints, and dyes and in lighting design applications.

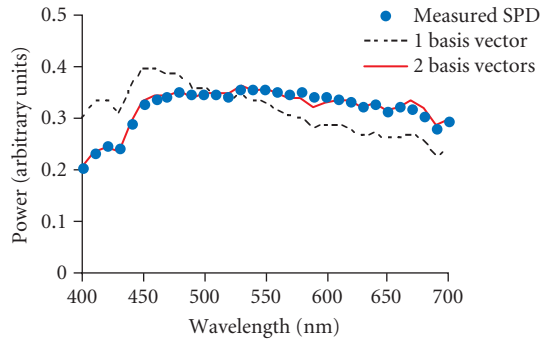
**Reflection Model** Illuminants are specified by their spectral power distributions. We will use the vector  $\mathbf{e}$  to represent illuminant spectral power distributions. In general, the interaction of light with matter is quite complex. For many applications, however, a rather simple model is acceptable. Using this model, we describe a surface by its surface reflectance function. The surface reflectance function specifies, for each sample wavelength, the fraction of illuminant power that is reflected to the observer. We will use the vector  $\mathbf{s}$  to represent surface reflectance spectra. Each entry of  $\mathbf{s}$  gives the reflectance measured at a single sample wavelength. Thus the spectral power distribution  $\mathbf{b}$  of the reflected light is given by the wavelength-by-wavelength product of the illuminant spectral power distribution and the surface reflectance function.

The most important consideration neglected in this formulation is viewing geometry. The relation between the radiant power emitted by a source of illumination, the material properties of a surface, and the radiant power reaching an observer can depend strongly on the viewing geometry. In our formulation, these geometrical factors must be incorporated implicitly into the specification of the illuminant and surface properties, so that any actual calculation is specific to a particular viewing geometry. Moreover, the surface reflectance must be understood as being associated with a particular image location, rather than with a particular object. A topic of current research in computer graphics is to find accurate and efficient ways to specify illuminants and surfaces for rendering,<sup>131,132</sup> and parallel work in human vision seeks to understand how the reflectance of spatially complex objects is perceived.<sup>133–136</sup> A second complexity that we neglect is fluorescence.

**Computing the Reflected Light** The relation between the surface reflectance function and the reflected light spectral power distribution is linear if the illuminant spectral power distribution is held fixed. We form the  $N_\lambda$  by  $N_\lambda$  diagonal illuminant matrix  $\mathbf{E}$  whose diagonal entries are the entries of  $\mathbf{e}$ . This leads to the relation  $\mathbf{b} = \mathbf{E}\mathbf{s}$ . By substituting into Eq. (20), we arrive at an expression for the tristimulus values of the light reflected from a surface:

$$\mathbf{t} = (\mathbf{TE})\mathbf{s} \quad (34)$$

where  $\mathbf{T}$  holds the colour matching functions in its three rows. The matrix  $(\mathbf{TE})$  in this equation plays exactly the same role as the color-matching functions do in Eq. (20). Any result that holds for spectral

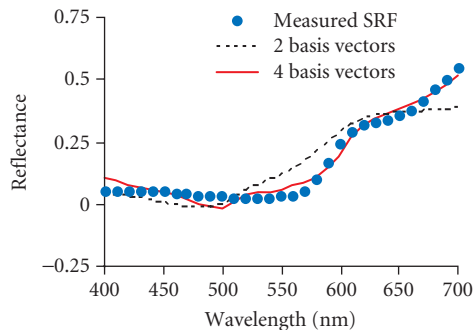


**FIGURE 14** The figure shows a daylight spectral power distribution and its approximation using the CIE linear model for daylight. For this particular illuminant, only two basis functions were required to provide a very good fit.

power distributions may thus be directly extended to a result for surface reflectance functions when the illuminant is known and held fixed (see subsection “Color Coordinates of Surfaces” in Sec. 10.6).

**Linear Model Representations for Surfaces and Illuminants** Judd, MacAdam, and Wyszecki<sup>137</sup> measured the spectral power distributions of a large number of naturally occurring daylights. They determined that a four-dimensional linear model provided a good description of their spectral measurements. Consideration of their results and other daylight measurements led the CIE to standardize a three-dimensional linear model for natural daylights.<sup>10</sup> Figure 14 depicts a daylight spectral power distribution measured by the first author and its approximation using the first two basis vectors of the CIE linear model for daylight.

Cohen<sup>138</sup> analyzed the reflectance spectra of a large set of Munsell papers<sup>139,140</sup> and concluded that a four-dimensional linear model provided a good approximation to the entire data set. Maloney<sup>141</sup> reanalyzed these data, plus a set of natural spectra measured by Krinov<sup>142</sup> and confirmed Cohen’s conclusion. More recently, reflectance measurements of the spectra of additional Munsell papers and natural objects<sup>143,144</sup> have been described by small-dimensional linear models. Figure 15 shows a measured surface reflectance spectrum (red cloth, measured by the first author) and its approximation using Cohen’s four-dimensional linear model.



**FIGURE 15** The figure shows a measured surface reflectance function and a fit to it using Cohen’s four-dimensional linear model.<sup>138</sup>

It is not yet clear why natural illuminant and surface spectra are well-approximated by small-dimensional linear models nor how general this conclusion is. Maloney<sup>141</sup> provides some speculations. None the less, the assumption that natural spectra do lie within small-dimensional linear models seems reasonable in light of the currently available evidence. This assumption makes possible a number of interesting practical calculations, as we illustrate in some of the following sections.

**Determining a Linear Model from Raw Spectral Data** Given a set of spectral measurements, it is possible, for any integer  $N_b$ , to find the  $N_b$  dimensional linear model that best approximates the spectral data set (in a least squares sense). Suppose that the data set consists of  $N_{\text{meas}}$  spectra, each of which is represented at  $N_\lambda$  sample wavelengths. Let  $\mathbf{X}$  be an  $N_\lambda$  by  $N_{\text{meas}}$  data matrix whose columns represent the individual spectral measurements. The goal of the calculation is to determine the  $N_\lambda$  by  $N_b$  matrix  $\mathbf{B}$  and an  $N_b$  by  $N_{\text{meas}}$  matrix of coefficients  $\mathbf{A}$  such that the linear model approximation  $\tilde{\mathbf{X}} = \mathbf{B}\mathbf{A}$  is the best least squares approximation to the data matrix  $\mathbf{X}$  over all possible choices of  $\mathbf{B}$  and  $\mathbf{A}$ .

The process of finding the matrix  $\mathbf{B}$  is called one mode components analysis.<sup>145,146</sup> It is very closely related to the principle components analysis technique discussed in most multivariate statistics texts.<sup>147</sup> One mode 1 components analysis may be accomplished numerically through the use of the singular value decomposition.<sup>27,148</sup> We define the singular value decomposition in Sec. 10.7 “Appendix—Matrix Algebra.” To see how the singular value decomposition is used to determine an  $N_b$  dimensional linear model for  $\mathbf{X}$ , consider Fig. 16. Figure 16a depicts the singular value decomposition of an  $N_\lambda$  by  $N_{\text{meas}}$  matrix  $\mathbf{X}$  for the case  $N_{\text{meas}} > N_\lambda$ , where the two matrices  $\mathbf{D}$  and  $\mathbf{V}^T$  have been collapsed. This form makes it clear that each column of  $\mathbf{X}$  is given by a linear combination of the columns of  $\mathbf{U}$ . Furthermore, for each column of  $\mathbf{X}$ , the weights needed to combine the columns of  $\mathbf{U}$  are given by the corresponding column of the matrix  $\mathbf{D}\mathbf{V}^T$ . Suppose we choose an  $N_b$  dimensional linear model  $\mathbf{B}$  for the data in  $\mathbf{X}$  by extracting the first  $N_b$  columns of  $\mathbf{U}$ . In this case, it should be clear that we can form an approximation  $\tilde{\mathbf{X}}$  to the data  $\mathbf{X}$  as shown in Fig. 16b. Because the columns of  $\mathbf{U}$  are orthogonal, the matrix  $\mathbf{A}$  consists of the first  $N_b$  rows of  $\mathbf{D}\mathbf{V}^T$ . The accuracy of the approximation depends on how important the columns of  $\mathbf{U}$  excluded from  $\mathbf{B}$  were to the original expression for  $\mathbf{X}$ . Under certain assumptions, it can be shown that choosing  $\mathbf{B}$  as above produces a linear model that minimizes the squared error of the approximation, for any choice of  $N_b$ .<sup>145</sup> Thus computing the singular value decomposition of  $\mathbf{X}$  allows us to find a good linear model of any desired dimension for  $N_b < N_\lambda$ . Computing linear models from data is quite feasible on modern desktop computers.

Although the above procedure produces the linear model that provides the best least squares fit to a data set, there are a number of additional considerations that should go into choosing a linear

$$\begin{aligned} \left[ \begin{array}{c} \mathbf{X} \end{array} \right] &= \left[ \begin{array}{c} \mathbf{U} \end{array} \right] \left[ \begin{array}{c} \mathbf{D}\mathbf{V}^T \end{array} \right] \\ &\text{(a)} \\ \left[ \begin{array}{c} \tilde{\mathbf{X}} \end{array} \right] &= \left[ \begin{array}{c} \mathbf{B} \end{array} \right] \left[ \begin{array}{c} \mathbf{A} \end{array} \right] \\ &\text{(b)} \end{aligned}$$

**FIGURE 16** (a) The figure depicts the singular value decomposition (SVD) of an  $N_\lambda$  by  $N_{\text{meas}}$  matrix  $\mathbf{X}$  for the case  $N_{\text{meas}} > N_\lambda$ . In this view we have collapsed the two matrices  $\mathbf{D}$  and  $\mathbf{V}^T$ . To determine an  $N_b$  dimensional linear model  $\mathbf{B}$  for the data in  $\mathbf{X}$  we let  $\mathbf{B}$  consist of the first  $N_b$  columns of  $\mathbf{U}$ . (b) The linear model approximation of the data is given by  $\tilde{\mathbf{X}} = \mathbf{B}\mathbf{A}$ , where  $\mathbf{A}$  consists of the first  $N_b$  rows of  $\mathbf{D}\mathbf{V}^T$ .

model. First, we note that the choice of linear model is not unique. Any nonsingular linear combination of the columns of  $\mathbf{B}$  will produce a linear model that provides an equally good account of the data. Second, the least squares error measure gives more weight to spectra with large amplitudes. In the case of surface spectra, this means that the more reflective surfaces will tend to drive the choice of basis vectors. In the case of illuminants, the more intense illuminants will tend to drive the choice. To avoid this weighting, the measured spectra are sometimes normalized to unit length before performing the singular value decomposition. The normalization equalizes the effect of the relative shape of each spectrum in the data set.<sup>141</sup> Third, it is sometimes desired to find a linear model that best describes the variation of a data set around its mean. To do this, the mean of the data set should be subtracted before performing the singular value decomposition. When the mean of the data is subtracted, one mode components analysis is identical to principle components analysis. Finally, there are circumstances where the linear model will be used not to approximate spectra but rather to approximate some other quantity (e.g., color coordinates) that depend on the spectra. In this case, more general techniques, closely related to those discussed here, may be used.<sup>149</sup>

**Approximating a Spectrum with Respect to a Linear Model** Given an  $N_b$  dimensional linear model  $\mathbf{B}$ , it is straightforward to find the representation of any spectrum with respect to the linear model. Let  $\mathbf{X}$  be a matrix representing the spectra of functions to be approximated. These spectra do not need to be members of the data set that was used to determine the linear model. To find the matrix of coefficients  $\mathbf{A}$  such that  $\tilde{\mathbf{X}} = \mathbf{BA}$  best approximates  $\mathbf{X}$  we use simple linear regression. Regression routines to solve this problem are provided as part of any standard matrix algebra software package.

**Digital Image Representations** If in a given application illuminants and surfaces may be represented with respect to small-dimensional linear models, then it becomes feasible to use point-by-point representations of these quantities in digital image processing. In typical color image processing, the image data at each point are represented by three numbers at each location. These numbers are generally tristimulus values in some color space. In calibrated systems, side information about the color-matching functions or primary spectral power distributions that define the color space is available to interpret the tristimulus values. It is straightforward to generalize this notion of color images by allowing the images to contain  $N_b$  numbers at each point and allowing these numbers to represent quantities other than tristimulus values.<sup>7</sup> For example, in representing the image produced by a printer, it might be advantageous to represent the surface reflectance at each location.<sup>150</sup> If the gamut of printed reflectances can be represented within a small-dimensional linear model, then representing the surface reflectance functions with respect to this model would not require much more storage than a traditional color image.<sup>7</sup> The basis functions for the linear model only need be represented once, not at each location. But by representing reflectances rather than tristimulus values, it becomes possible to compute what the tristimulus values reflected from the printed image would be under any illumination. We illustrate the calculation in the next section. Because of the problem of metamerism (see subsection “Computing the Reflected Light”), this calculation is not possible if only the tristimulus values are represented in the digital image. To avoid this limitation, hyperspectral images record full spectra at each image location.<sup>151–153</sup>

**Simulation of Illuminated Surfaces** Consider the problem of producing a signal on a monitor that has the same tristimulus values as a surface under a variety of different illuminants. The solution to this problem is straightforward and is useful in a number of applications. These include rendering digitally archived paintings,<sup>154,155</sup> generating stimuli for use in psychophysics,<sup>156</sup> and producing photorealistic computer-generated imagery.<sup>17</sup> We show the calculation for the data at a single image location. Let  $\mathbf{a}$  be a representation of the surface reflectance with respect to an  $N_b$  dimensional linear model  $\mathbf{B}$ . Let  $\mathbf{E}$  represent the illuminant spectral power distribution in diagonal matrix form. Let  $\mathbf{T}$  represent the color-matching functions for a human observer, and  $\mathbf{P}$  represent the primary phosphor spectral power distributions for the monitor on which the surface will be rendered. We wish to determine tristimulus values  $\mathbf{t}$  with respect to the monitor primaries so that the light emitted from the monitor will appear identical to the light reflected from the simulated surface under the

simulated illuminant. From Eqs. (13) (cast as  $\mathbf{s} = \mathbf{B}\mathbf{a}$ ), (34), (29), and (30) we can write directly the desired rendering equation

$$\mathbf{t} = [(\mathbf{TP})^{-1}(\mathbf{TE})\mathbf{B}]\mathbf{a} \quad (35)$$

The rendering matrix  $[(\mathbf{TP})^{-1}(\mathbf{TE})\mathbf{B}]$  has dimensions 3 by  $N_b$  and maps the surface weights directly to monitor tristimulus values. It is quite general, in that we may use it for any calibrated monitor and any choice of linear models. It does not depend on the particular surface being rendered and may be computed once for an entire image. Because the rendering matrix is of small dimension, rendering of this sort is feasible, even for very large images. As discussed in subsection “Transformations between Color Spaces” in Sec. 10.5, it may be possible to determine the matrix  $\mathbf{M}_{TP} = (\mathbf{TP})^{-1}$  directly. A similar shortcut is possible for the matrix  $(\mathbf{TE})\mathbf{B}$ . Each column of this matrix is the tristimulus values of one linear model basis vector under the illuminant specified by the matrix  $\mathbf{E}$ .

**Color Coordinates of Surfaces** Our discussion thus far has emphasized describing the color coordinates of lights. In many applications of colorimetry, it is desirable to describe the color properties of reflective objects. One efficient way to do this, as described above, is to use linear models to describe the full surface reflectance functions. Another possibility is to specify the color coordinates of the light reflected from the surface under standard illumination. This method allows the assignment of tristimulus values to surfaces in an orderly fashion.

The CIE has standardized several illuminant spectral power distributions that may be used for this purpose (see the next section). Using the procedures defined above, one can begin with the spectral power distribution of the illuminant and the surface reflectance function and from there calculate the desired color-matching coordinates.

The relative size of the tristimulus values assigned to a surface depend on its spectral reflectance function and on the illuminant chosen for specification. To factor the intensity of the illuminant out of the surface representation, the CIE specified a normalization of the color coordinates for use with 1931 XYZ tristimulus values. This normalization consists of multiplying the computed tristimulus values by the quantity  $100/Y_0$ , where  $Y_0$  is the Y tristimulus value for the illuminant.

The tristimulus values of a surface provide enough information to match the surface when it is viewed under the illuminant used to compute those coordinates. It is important to bear in mind that two surfaces that have the same tristimulus values under one illuminant do not necessarily share the same tristimulus values under another illuminant. A more complete description can be generated using the linear model approach described above.

**Standard Sources of Illumination** The CIE has standardized a number of illuminant spectral power distributions.<sup>157</sup> These were designed to be typical of various common viewing conditions and are useful as specific choices of illumination when the illuminant cannot be measured directly. CIE Illuminant A is designed to be representative of tungsten-filament illumination. CIE Illuminant D65 is designed to be representative of average daylight. Other CIE standard daylight illuminants may be computed using the CIE principle components of daylight as basis vectors and the formulas specified by the CIE.<sup>10</sup> Spectra representative of fluorescent lamps and other artificial sources are also available.<sup>8,10</sup>

## Metamerism

**Recovering Spectral Power Distributions from Tristimulus Values** It is not possible in general to recover a spectral power distribution from its tristimulus values. If some prior information about the spectral power distribution of the color signal is available, however, then recovery may be possible. Such recovery is of most interest in applications where direct spectral measurements are not possible and where knowing the full spectrum is important. For example, the effect of lens chromatic aberrations on cone quantal absorption rates depends on the full spectral power distribution.<sup>110</sup>

Suppose the spectral power distribution of interest is known to lie within a three-dimensional linear model. We may write  $\mathbf{b} = \mathbf{B}\mathbf{a}$ , where the basis matrix  $\mathbf{B}$  has dimensions  $N_\lambda$  by 3. Let  $\mathbf{t}$  be the

tristimulus values of the light with respect to a set of color-matching functions  $\mathbf{T}$ . We can conclude that  $\mathbf{a} = (\mathbf{TB})^{-1}\mathbf{t}$ , which implies

$$\mathbf{b} = \mathbf{B}(\mathbf{TB})^{-1}\mathbf{t} \quad (36)$$

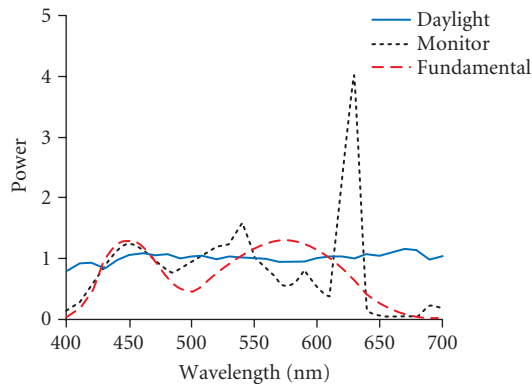
When we do not have a prior constraint that the signal belongs to a three-dimensional linear model, we may still be able to place some linear model constraint, of dimension higher than three, on the spectral power distribution. For example, when we know that the signal was produced by the reflection of daylight from a natural object, it is reasonable to assume that the color signal lies within a linear model of dimension that may be as low as nine.<sup>158</sup> In this case, we can still write  $\mathbf{b} = \mathbf{Ba}$ , but we cannot apply Eq. (36) directly because the matrix  $(\mathbf{TB})$  will be singular. To deal with this problem, we can choose a reduced linear model with only three dimensions. We then proceed as outlined above, but substitute the reduced model for the true model. This will lead to an estimate for the actual spectral power distribution  $\mathbf{b}$ . If the reduced linear model provides a reasonable approximation to  $\mathbf{b}$ , the estimation error may be quite small. The estimate will have the property that it is a metamer of  $\mathbf{b}$ . The techniques described above for finding linear model approximations may be used to choose an appropriate reduced model.

**Finding Metamers of a Light** It is often of interest to find metamers of a light. We discuss two approaches here. Wyszecki and Stiles<sup>8</sup> treat the problem in considerable detail.

*Using a linear model* If we choose any three-dimensional linear model we can combine Eq. (36) with the fact the fact that  $\mathbf{t} = \mathbf{Tb}$  [Eq. (20)] to compute a pair of metameric spectral power distributions  $\mathbf{b}$  and  $\hat{\mathbf{b}}$

$$\hat{\mathbf{b}} = \hat{\mathbf{B}}(\hat{\mathbf{T}}\mathbf{B})^{-1}\mathbf{Tb} \quad (37)$$

Each choice of  $\hat{\mathbf{B}}$  will lead to a different metamer. Figure 17 shows a number of metameric spectral power distributions generated in this fashion.



**FIGURE 17** The figure shows three metameric color signals with respect to the CIE 1931 standard observer. The three metamers were computed using Eq. (37). The initial spectral power distribution  $\mathbf{b}$  (not shown) was an equal energy spectrum. Three separate linear models were used: one that describes natural daylights, one typical of monitor phosphor spectral power distributions, and one that provides Cohen’s “fundamental metamer.”<sup>207</sup>

*Metameric blacks* Another approach to generating metamers is to note that there will be some spectral power distributions  $\mathbf{b}_0$  that have the property  $\mathbf{T}\mathbf{b}_0 = \mathbf{0}$ . Wyszecki referred to such distributions as metameric blacks, since they have the same tristimulus values as no light at all.<sup>159</sup> Grassmann's laws imply that adding a metameric black  $\mathbf{b}_0$  to any light  $\mathbf{b}$  yields a metamer of  $\mathbf{b}$ . Given a linear model  $\mathbf{B}$  with dimension greater than three it is possible to find a second linear model  $\mathbf{B}_0$  such that (a) all lights that lie in  $\mathbf{B}_0$  also lie in  $\mathbf{B}$  and (b) all lights in  $\mathbf{B}_0$  are metameric blacks. We determine  $\mathbf{B}_0$  by finding a linear model for the null space of the matrix  $\mathbf{T}\mathbf{B}$ . The null space of a matrix consists of all vectors that are mapped to  $\mathbf{0}$  by the matrix. Finding a basis for the null space of a matrix is a standard operation in numerical matrix algebra. If we have a set of basis vectors  $\mathbf{N}_0$  for the null space of  $\mathbf{T}\mathbf{B}$ , we can form  $\mathbf{B}_0 = \mathbf{B}\mathbf{N}_0$ . This technique provides a way to generate a large list of metamers for any given light  $\mathbf{b}$ . We choose a set of weights  $\mathbf{a}$  at random and construct  $\mathbf{b}_0 = \mathbf{B}_0\mathbf{a}$ . We then add  $\mathbf{b}_0$  to  $\mathbf{b}$  to form a metamer. To generate more metamers, we simply repeat with new choices of weight vector  $\mathbf{a}$ .

*Surface and Illuminant Metamerism* The formal similarity between Eq. (20) (which gives the relation between spectral power distributions and tristimulus values) and Eq. (34) (which gives the relation between surface reflectance functions and tristimulus values when the illuminant is known) makes it clear that our discussion of metamerism can be applied to surface reflectance spectra. Two physically different surfaces will appear identical if the tristimulus values of the light reflected from them is identical. This fact can be used to good purpose in some color reproduction applications. Suppose that we have a sample surface or textile whose color we wish to reproduce. It may be that we are not able to reproduce the sample's surface reflectance function exactly because of various limitations in the available color reproduction technology. If we know the illuminant under which the reproduction will be viewed, we may be able to determine a reproducible reflectance function that is metameric to that of the desired sample. This will give us a sample whose color appearance is as desired. Applications of this sort make heavy use of the methods described earlier to determine metamers.

But what if the illuminant is not known or if it is known to vary? In this case there is an additional richness to the topic of determining metamers. We can pose the problem of finding surface reflectance functions that will be metameric to a desired reflectance under multiple specified illuminants or under all of the illuminants within some linear model. The general methods developed here have been extended to analyze this case.<sup>158,160</sup> Similar issues arise in lighting design, where we desire to produce an artificial light whose color-rendering properties match those of a specified light (such as natural daylight). When wavelength by wavelength matching of the spectra is not feasible, it may still be possible to find a spectrum so that the light reflected from surfaces within a linear model is identical for the two light sources. Because of the symmetric role of illuminants and surfaces in reflection, this problem may be treated by the same methods as used for surface reproduction.

## Color Cameras and Other Visual Systems

We have treated colorimetry from the point of view of specifying the spectral information available to a human observer. We have developed our treatment, however, in such a way that it may be applied to handle other visual systems. Suppose that we wish to define color coordinates with respect to some arbitrary visual system with  $N_{\text{device}}$  photosensors. This visual system might be an artificial system based on a color camera or scanner, a nonhuman biological visual system, or the visual system of a color anomalous human observer. We assume that the sensitivities of the visual system's photosensors are known up to a linear transformation. Let  $\mathbf{T}_{\text{device}}$  be an  $N_{\text{device}}$  by  $N_{\lambda}$  matrix whose entries are the sensitivities of each of the device's sensors at each sample wavelength. We can compute the responses of these sensors to any light  $\mathbf{b}$ . Let  $\mathbf{t}_{\text{device}}$  be a vector containing the responses of each sensor type to the light. Then we have  $\mathbf{t}_{\text{device}} = \mathbf{T}_{\text{device}}\mathbf{b}$ . We may use  $\mathbf{t}_{\text{device}}$  as the device color coordinates of  $\mathbf{b}$ .

*Transformation between Color Coordinates of Different Visual Systems* Suppose that we have two different visual systems and we wish to transform between the color coordinates of each.



A typical example might be trying to compute the CIE 1931 XYZ tristimulus values of a light from the responses of a color camera. Let  $N_s$  be the number of source sensors, with sensitivities specified by  $\mathbf{T}_s$ . Similarly, let  $N_d$  be the number of destination sensors with sensitivities specified by  $\mathbf{T}_d$ . For any light  $\mathbf{b}$  we know that the source device color coordinates are given by  $\mathbf{t}_s = \mathbf{T}_s \mathbf{b}$  and the destination device color coordinates  $\mathbf{t}_d = \mathbf{T}_d \mathbf{b}$ . We would like to transform between  $\mathbf{t}_s$  and  $\mathbf{t}_d$  without direct knowledge of  $\mathbf{b}$ .

If we can find an  $N_d$  by  $N_s$  matrix  $\mathbf{M}$  such that  $\mathbf{T}_d = \mathbf{M} \mathbf{T}_s$  then it is easy to show that the matrix  $\mathbf{M}$  may be used to compute the destination device color coordinates from the source device color coordinates through  $\mathbf{t}_d = \mathbf{M} \mathbf{t}_s$ . We have already considered this case (in a less general form) in subsection “Transformations between Color Spaces.” The extension here is that we allow the possibility that the dimensions of the two color coordinate systems differ. When a linear transformation between  $\mathbf{T}_s$  and  $\mathbf{T}_d$  exists, it can be found by standard regression methods.

Horn demonstrated that when no exact linear transformation between  $\mathbf{T}_s$  and  $\mathbf{T}_d$  exists, it is not in general possible to transform between the two sets of color coordinates.<sup>4</sup> The reason for this is that there will always exist a pair of lights that have the same color coordinates for the source device but different color coordinates for the destination device. The transformation will therefore be incorrect for at least one member of this pair. When no exact linear transformation exists, it is still possible to make an approximate transformation. One approach is to use linear regression to find the best linear transformation  $\mathbf{M}$  between the two sets of color-matching functions in a least squares sense. This transformation is then applied to the source color coordinates as if it were exact.<sup>4</sup> Although this is an approximation, in many cases the results will be acceptable. In the absence of prior information about the spectral power distribution of the original light  $\mathbf{b}$ , it is a sensible approach.

A second possibility is to use prior constraints on the spectral power distribution of the light to guide the transformation.<sup>22</sup> Suppose that we know that the light is constrained to lie within an  $N_b$  dimensional linear model  $\mathbf{B}$ . Then we can find the best linear transformation  $\mathbf{M}$  between the two matrices  $\mathbf{T}_s \mathbf{B}$  and  $\mathbf{T}_d \mathbf{B}$ . This transformation may then be used to transform the source color coordinates to the destination color coordinates. It is easy to show that the transformation will be exact if  $\mathbf{T}_d \mathbf{B} = \mathbf{M} \mathbf{T}_s \mathbf{B}$ . Otherwise it is a reasonable approximation that takes the linear model constraint into account.

A number of recent papers present more elaborated methods for color correction.<sup>161–163</sup>

**Computational Color Constancy** An interesting application is the problem of estimating surface reflectance functions from color coordinates. This problem is of interest for two reasons. First, it appears that human color vision makes some attempt to perform this estimation, so that our percept of color is more closely associated with object surface properties than with the proximal properties of the light reaching the eye. Second, an artificial system that could estimate surface properties would have an important cue to aid object recognition. In the case where the illuminant is known, the problem of estimating surface reflectance properties is the same as the problem of estimating the color signal, because the illuminant spectral power distribution can simply be incorporated into the sensor sensitivities. In this case the methods outlined in the preceding section for estimating color signal spectral properties can be used.

The more interesting case is where both the illuminant and the surface reflectance are unknown. In this case, the problem is more difficult. Considerable insight has been gained by applying linear model constraints to both the surface and illuminant spectral power distributions. A number of approaches have been developed for recovering surface reflectance functions or otherwise achieving color constancy.<sup>158,164–172</sup> Each approach differs (1) in the additional assumptions that are made about the properties of the image and (2) in the sophistication of the model of illuminant surface interaction and scene geometry used. A thorough review of all of these methods is beyond the scope of this chapter. It is instructive, however, to review one of the simpler methods, that of Buchsbaum.<sup>165</sup> See Ebner<sup>173</sup> for a discussion of many current algorithms.

Buchsbaum assumed that in any given scene, the average reflectance function of the surfaces in the scene is known. This is commonly called the “gray world” assumption. He also assumed that the illuminant was diffuse and constant across the scene and that the illuminants and surfaces in the scene are described by linear models with the same dimensionality as the number of sensors.



Let  $\mathbf{S}_{\text{avg}}$  be the spectral power distribution of the known average surface, represented in diagonal matrix form. Then it is possible to write the relation between the space average of the sensor responses and the illuminant as

$$\mathbf{t}_{\text{avg}} = \mathbf{T}\mathbf{S}_{\text{avg}}\mathbf{B}_e\mathbf{a}_e \quad (38)$$

where  $\mathbf{a}_e$  is a vector containing the weights of the illuminant within the linear model representation  $\mathbf{B}_e$ . Because we assume that the dimension  $N_e = N_p$ , the matrix  $\mathbf{T}\mathbf{S}_{\text{avg}}\mathbf{B}_e$  will be square and typically may be inverted. From this we recover the illuminant as  $\mathbf{e} = \mathbf{B}_e(\mathbf{T}\mathbf{S}_{\text{avg}}\mathbf{B}_e)^{-1}\mathbf{t}_{\text{avg}}$ . If we let  $\mathbf{E}$  represent the recovered illuminant in matrix form, then at each image location we can write

$$\mathbf{t} = \mathbf{TEB}_s\mathbf{a}_s \quad (39)$$

where  $\mathbf{a}_s$  is a vector containing the weights of the surface within the linear model representation  $\mathbf{B}_s$ . Proceeding exactly as we did for the illuminant, we may recover the surface reflectance from this equation.

Although Buchsbaum's method depends on rather strong assumptions about the nature of the scene, subsequent algorithms have shown that these assumptions can be relaxed.<sup>22,166,172</sup> Several authors treat the relation between computational color constancy and the study of human vision.<sup>174–178</sup>

## Color Discrimination

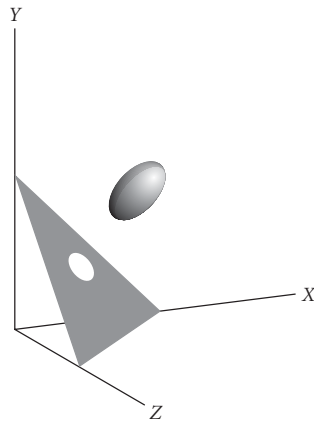
**Measurement of Small Color Differences** Our treatment so far has not included any discussion of the precision to which observers can judge identity of color appearance. To specify tolerances for color reproduction, it would be helpful to know how different the color coordinates of two lights must be for an observer to reliably distinguish between them. A number of techniques are available for measuring human ability to discriminate between colored lights. We review these briefly here as an introduction to the topic of uniform color spaces. A more extended discussion of color discrimination and its relation to postreceptoral mechanisms is presented in the companion chapter (Chap. 11).

One experimental method, employed in seminal work by MacAdam,<sup>179,180</sup> is to examine the variability in individual color matches. That is, if we have observers set matches to the same test stimulus, we will discover that they do not always set exactly the same values. Rather, there will be some trial-to-trial variability in the settings. MacAdam and others<sup>181,182</sup> used the sample covariance of the individual match tristimulus values as a measure of observers' color discrimination.

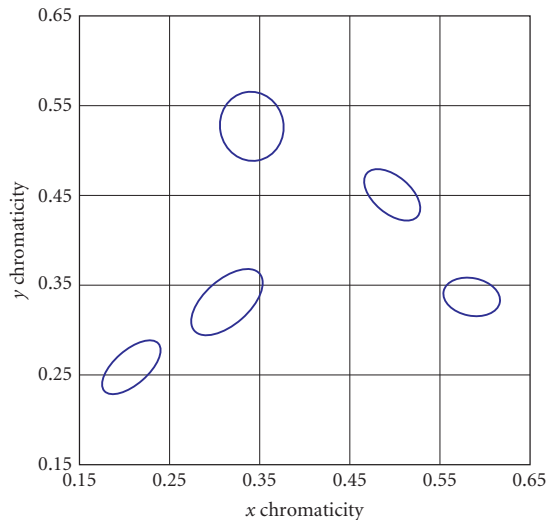
A second approach is to use more direct psychophysical methods (see Chap. 3) to measure observers' thresholds for discriminating between pairs of colored lights. Examples include increment threshold measurements for monochromatic lights<sup>183</sup> and thresholds measured systematically in a three-dimensional color space.<sup>184,185</sup>

Measurements of small color differences are often summarized with isodiscrimination contours. An isodiscrimination contour specifies the color coordinates of lights that are equally discriminable from a common standard light. Figure 18 shows an illustrative isodiscrimination contour. Isodiscrimination contours are often modeled as ellipsoids<sup>184,185</sup> and the figure is drawn to the typical ellipsoidal shape. The well-known MacAdam ellipses<sup>179</sup> are an example of representing discrimination data using the chromaticity coordinates of a cross-section of a full three-dimensional isodiscrimination contour (see the legend of Fig. 18). Chapter 11 provides a more extensive discussion of possible models of discrimination contours. Under some experimental conditions, the measured contour may not be ellipsoidal.

**CIE Uniform Color Spaces** Figure 19 shows chromaticity plots of theoretical isodiscrimination contours. A striking feature of the plots is that the size and shape of the contours depends on the standard stimulus. For this reason, it is not possible to predict whether two lights will be discriminable solely on the basis of the Euclidean distance between their color coordinates. The heterogeneity of the isodiscrimination contours must also be taken into account.



**FIGURE 18** Isodiscrimination contour. The plotted ellipsoid shows a hypothetical isodiscrimination contour in the CIE XYZ color space. This contour represents color discrimination performance for the standard light whose color coordinates are located at the ellipsoid's center. Isodiscrimination contours such as the one shown are often summarized by a two-dimensional contour plotted on a chromaticity diagram (see Fig. 19). The two-dimensional contour is obtained from a cross section of the full contour, and its shape can depend on which cross section is used. This information is not available directly from the two-dimensional plot. A common criterion for choice of cross section is isoluminance. The ellipsoid shown in the figure is schematic and does not represent actual human performance.



**FIGURE 19** Isodiscrimination contours plotted in the chromaticity diagram. These were computed using the CIE  $L^*a^*b^*$  uniform color space and  $\Delta E_{ab}^*$  difference metric. They provide an approximate representation of human performance. For each standard stimulus, the plotted contour represents the color coordinates of lights that differ from the standard by  $15 \Delta E_{ab}^*$  units but that have the same luminance as the standard. The choice of  $15 \Delta E_{ab}^*$  units magnifies the contours compared to those that would be obtained in a threshold experiment. The contours shown are a projection of isodiscrimination contours computed for isoluminant color differences. The luminance of the white point used in the CIELAB computations was set at  $1000 \text{ cd/m}^2$ , while the discriminations were around stimuli with a luminance of  $500 \text{ cd/m}^2$ .

The CIE<sup>10</sup> provides formulas that may be used to predict the discriminability of colored lights. The most recent recommendations are based on the CIE 1976  $L^*a^*b^*$  (CIELAB) color coordinates. These are obtained by a nonlinear transformation from CIE 1931 XYZ color coordinates. The transformation stretches the XYZ color space so that the resulting Euclidean distance between color coordinates provides an approximation to the how well lights may be discriminated. The  $L^*a^*b^*$  system is referred to as a uniform color space. There is also a CIE 1976  $L^*u^*v^*$  (CIELUV) system, but this is now less widely used than the  $L^*a^*b^*$  system and its derivatives.

*Transformation to CIELAB coordinates* The CIE 1976  $L^*a^*b^*$  color coordinates of a light may be obtained from its CIE XYZ coordinates according to the equations

$$L^* = \begin{cases} 116 \left( \frac{Y}{Y_n} \right)^{1/3} - 16 & \frac{Y}{Y_n} > 0.008856 \\ 903.3 \left( \frac{Y}{Y_n} \right) & \frac{Y}{Y_n} \leq 0.008856 \end{cases} \quad (40)$$

$$a^* = 500 \left[ f \left( \frac{X}{X_n} \right) - f \left( \frac{Y}{Y_n} \right) \right]$$

$$b^* = 200 \left[ f \left( \frac{Y}{Y_n} \right) - f \left( \frac{Z}{Z_n} \right) \right]$$

where the function  $f(s)$  is defined as

$$f(s) = \begin{cases} (s)^{1/3} & s > 0.008856 \\ 7.787(s) + \frac{16}{116} & s \leq 0.008856 \end{cases} \quad (41)$$

The quantities  $X_n$ ,  $Y_n$ , and  $Z_n$  in the equations above are the tristimulus values of a white point. Little guidance is available as to how to choose an appropriate white point. In the case where the lights being judged are formed when an illuminant reflects from surfaces, the tristimulus values of the illuminant may be used. In the case where the lights being judged on a computer-controlled color monitor, the sum of the tristimulus values of the three monitor phosphors stimulated at their maximum intensity may be used.

*Distance in CIELAB space* The Euclidean distance between the  $L^*a^*b^*$  coordinates of two lights provides a rough guide to their discriminability. The symbol  $\Delta E_{ab}^*$  is used to denote distance in the uniform color space and is defined as

$$\Delta E_{ab}^* = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2} \quad (42)$$

where the various quantities on the right represent the differences between the corresponding coordinates of the two lights. Roughly speaking, a  $\Delta E_{ab}^*$  value of 1 corresponds to a color difference that can just be reliably discerned by a human observer under optimal viewing conditions. A  $\Delta E_{ab}^*$  value of 3 is sometimes used as an acceptable tolerance in industrial color reproduction applications.

The CIE color difference measure  $\Delta E_{ab}^*$  provides only an approximate guide to the discriminability between two lights. There are a number of reasons why this is so. The first is that the relatively simple nonlinear transformation between CIE  $XYZ$  and CIE  $L^*a^*b^*$  coordinates does not completely capture the empirical data on color discrimination between two samples. In part this is because the formulae were designed to predict not only discrimination data but also certain suprathreshold judgments of color appearance.<sup>186</sup> Second, color discrimination thresholds depend heavily on factors other than the tristimulus values. These factors include the adapted state of the observer,<sup>183</sup> the spatial and temporal structure of the stimulus,<sup>187–189</sup> and the task demands placed on the observer.<sup>190–193</sup> Therefore, the complete specification of a uniform color space must incorporate these factors. The CIE has now recommended a more involved method of computing small color differences from the CIE  $L^*a^*b^*$  coordinates that attempts to provide better prediction of small color differences.<sup>10,194</sup> The resultant computed difference is referred to as  $\Delta E_{00}$ . The details of the computation of  $\Delta E_{00}$  are provided and discussed in a CIE technical report.<sup>194</sup> The reader considering using  $\Delta E_{00}$  is encouraged to study Wyszecki and Stiles's<sup>8</sup>(pp. 584–586) insightful discussion of color vision models.

### Effect of Errors in Color-Matching Functions

Given that there is some variation between different standard estimates of color-matching functions, between the color-matching functions of different individuals, and between the color-matching functions that mediate performance for different viewing conditions, it is of interest to determine whether the magnitude of this variation is of practical importance. There is probably no general method for making this determination, but here we outline one approach.

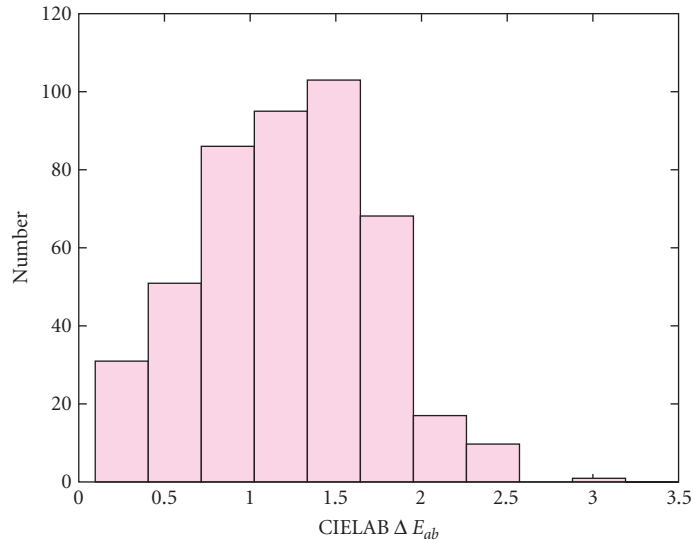
Consider the case of rendering a set of illuminated surfaces on a color monitor. If we know the spectral power distribution of the monitor's phosphors, it is possible to compute the appropriate weights on the monitor phosphors to produce a light metameric to each illuminated surface. The computed weights will depend on the choice of color-matching functions. Once we know the weights, however, we can find the  $L^*a^*b^*$  coordinates of the emitted light. This suggests the following method to estimate the effect of differences in color-matching functions. First we compute the  $L^*a^*b^*$  coordinates of surfaces rendered using the first set of color-matching functions. Then we compute the corresponding coordinates when the surfaces are rendered using the second set of color-matching functions. Finally, we compute the  $\Delta E_{ab}^*$  difference between corresponding sets of coordinates. If the  $\Delta E_{ab}^*$  are large, then the differences between the color-matching functions are important for the rendering application.

We have performed this calculation for a set of 462 measured surfaces<sup>139,140</sup> rendered under CIE Illuminant  $D_{65}$ . The two sets of color-matching functions used were the 1931 CIE  $XYZ$  color-matching functions and the Judd-Vos modified  $XYZ$  color-matching functions. The monitor phosphor spectral power distributions were measured by the first author. The results are shown in Fig. 20. The plot shows a histogram of the differences. The median difference is 1.2 units. This difference is quite close to discrimination threshold and for many applications, the differences between the two sets of color-matching functions will probably not be of great consequence.

### Brightness Matching and Photometry

The foundation of colorimetry is the human observer's ability to judge identity of color appearance. It is sometimes of interest to compare certain perceptual attributes of lights that do not, as a whole, appear identical. In particular, there has been a great deal of interest in developing formulas that predict when two lights with different relative spectral power distributions will appear equally bright. Colorimetry provides a partial answer to this question, since two lights that match in appearance must appear equally bright. Intuitively, however, it seems that it should be possible to set the relative intensities of any two lights so that they match in brightness.

In a heterochromatic brightness matching experiment, observers are asked to scale the intensity of a matching light until its brightness matches that of an experimentally controlled test light. Although



**FIGURE 20** Effect of changes in color-matching functions. The plot shows a histogram of the  $\Delta E_{ab}^*$  differences between two sets of lights, each of which is a monitor rendering of the same set of illuminated surfaces. The two renderings were computed using different sets of color-matching functions. The white point used in the CIELAB transformations was the XYZ coordinates of the illuminant used to compute the renderings, CIE D65.

observers can perform the heterochromatic brightness matching task, they often report that it is difficult and their matches tend to be highly variable.<sup>8</sup> Moreover, the results of brightness-matching experiments are not additive.<sup>43,195</sup> For photometry to be as practicable as radiometry, the measured luminous efficiency of any mixture of lights must equal the sum of the luminous efficiencies of the component lights. Such additivity is known as obedience to Abney's law.<sup>196,197</sup> For this reason, more indirect methods for equating the overall effectiveness of lights at stimulating the visual system have been developed.<sup>8,195,198–203</sup> The most commonly used method is that of flicker photometry. In a flicker photometric experiment, two lights of different spectral power distributions are presented alternately at the same location. At moderate flicker rates (about 20 Hz), subjects are able to adjust the overall intensity of one of the lights to minimize the apparent flicker. The intensity setting that minimizes apparent flicker is taken to indicate that the two lights match in their effectiveness as visual stimuli. Two lights equated in this way are said to be equiluminant or to have equal luminance.

Because experiments for determining when lights have the same luminance obey linearity properties similar to Grassmann's laws, it is possible to determine a luminous efficiency function that allows the assignment of a luminance value to any light. A luminous efficiency function specifies, for each sample wavelength, the relative contribution of that wavelength to the overall luminance. We can represent a luminous efficiency function as an  $N_\lambda$  dimensional row vector  $\mathbf{v}$ . Each entry of the matrix specifies the relative luminance of light at the corresponding sample wavelength. The luminance  $\nu$  of an arbitrary spectral power distribution  $\mathbf{b}$  may be computed by the equation

$$\nu = \mathbf{vb} \quad (43)$$

The CIE has standardized four luminous efficiency functions by definition. The most commonly used of these is the standard photopic luminous efficiency function  $V(\lambda)$ . This is identical by definition to the 1931 XYZ color-matching function  $\bar{y}\lambda$ . For lights that subtend more than  $4^\circ$  of visual angle, a luminous efficiency function  $V_{10}(\lambda)$  given by the 1964  $10^\circ$  XYZ color-matching functions is

preferred. More recently, the Judd-Vos modified color matching function has been made a supplemental standard.<sup>204</sup> A final standard luminous efficiency function is available for use at low light levels when the rods are the primary functioning photoreceptors. A new luminous efficiency function will be incorporated into the new CIE proposal for a set of physiologically relevant color-matching functions. The notation  $V_\lambda$  or  $V(\lambda)$  is often used in the literature to denote luminous efficiency functions. Note that Eq. (43) allows the computation of luminance in arbitrary units. Ref. 8 discusses standard measurement units for luminance.

It is important to note that luminance is a construct derived from flicker photometric and related experiments. As such, it does not directly predict when two lights will be judged to have the same brightness. The relation between luminance and brightness is quite complicated.<sup>8,205</sup>

It is also worth noting that there is considerable individual variation in flicker photometric judgments, even among color-normal observers. For this reason, it is a common practice in psychophysical experiments to use flicker photometry to determine isoluminant stimuli for individual subjects with the stimuli of interest.

## 10.7 APPENDIX—MATRIX ALGEBRA

This appendix provides a brief introduction to matrix algebra. The development emphasizes the aspects of matrix algebra that are used in this chapter and is somewhat idiosyncratic. In addition, we do not prove any of the results we state. Rather, our intention is to provide the reader unfamiliar with matrix algebra with enough information to make this chapter self-contained.

### Basic Notions

**Vectors and Matrices** A vector is a list of numbers. We use lowercase bold letters to represent vectors. We use single subscripts to identify the individual entries of a vector. The entry  $a_i$  refers to the  $i$ th number in  $\mathbf{a}$ . We call the total number of entries in a vector its dimension.

A matrix is an array of numbers. We use uppercase bold letters to represent matrices. We use dual subscripts to identify the individual entries of a matrix. The entry  $a_{ij}$  refers to the number in the  $i$ th row and  $j$ th column of  $\mathbf{A}$ . We sometimes refer to this as the  $ij$ th entry of  $\mathbf{A}$ . We call the number of rows in a matrix its row dimension. We call the number of columns in a matrix its column dimension. We generally use the symbol  $N$  to denote dimensions.

Vectors are a special case of matrices where either the row or the column dimension is 1. A matrix with a single column is often called a column vector. A matrix with a single row is often called a row vector. By convention, all vectors used in this chapter should be understood to be column vectors unless explicitly noted otherwise.

It is often convenient to think of a matrix as being composed of vectors. For example, if a matrix has dimensions  $N_r$  by  $N_c$ , then we may think of the matrix as consisting of  $N_c$  column vectors, each of which has dimension  $N_r$ .

**Addition and Multiplication** A vector may be multiplied by a number. We call this scalar multiplication. Scalar multiplication is accomplished by multiplying each entry of the vector by the number. If  $\mathbf{a}$  is a vector and  $b$  is a number, then  $\mathbf{b} = \mathbf{ab} = \mathbf{ba}$  is a vector whose entries are given by  $c_j = ba_j$ .

Two vectors may be added together if they have the same dimension. We call this vector addition. Vector addition is accomplished through entry-by-entry addition. If  $\mathbf{a}$  and  $\mathbf{b}$  are vectors with the same dimension, the entries of  $\mathbf{c} = \mathbf{a} + \mathbf{b}$  are given by  $c_j = a_j + b_j$ .

Two matrices may be added if they have the same row and column dimensions. We call this matrix addition. Matrix addition is also defined as entry by entry addition. Thus if  $\mathbf{A}$  and  $\mathbf{B}$  are matrices with the same dimension, the entries of  $\mathbf{C} = \mathbf{A} + \mathbf{B}$  are given by  $c_{ij} = a_{ij} + b_{ij}$ . Vector addition is a special case of matrix addition.

A column vector may be multiplied by a matrix if the column dimension of the matrix matches the dimension of the vector. If  $\mathbf{A}$  has dimensions  $N_r$  by  $N_c$  and  $\mathbf{b}$  has dimension  $N_c$ , then  $\mathbf{c} = \mathbf{A}\mathbf{b}$  is an  $N_r$  dimensional vector. The  $i$ th entry of  $\mathbf{c}$  is related to the entries of  $\mathbf{A}$  and  $\mathbf{b}$  by the equation:

$$c_i = \sum_{j=1}^{N_c} a_{ij} b_j \quad (44)$$

It is also possible to multiply a matrix  $\mathbf{B}$ , by another matrix,  $\mathbf{A}$  on the left, if the column dimension of  $\mathbf{A}$  matches the row dimension of  $\mathbf{B}$ . If  $\mathbf{A}$  has dimensions  $N_r$  by  $N$  and  $\mathbf{B}$  has dimensions  $N$  by  $N_c$ , then  $\mathbf{C} = \mathbf{A}\mathbf{B}$  is an  $N_r$  by  $N_c$  dimensional matrix. The  $ik$ th entry of  $\mathbf{C}$  is related to the entries of  $\mathbf{A}$  and  $\mathbf{B}$  by the equation:

$$c_{ik} = \sum_{j=1}^N a_{ij} b_{jk} \quad (45)$$

By comparing Eqs. (44) and (45) we see that multiplying a matrix by a matrix is a shorthand for multiplying several vectors by the same matrix. Denote the  $N_c$  columns of  $\mathbf{B}$  by  $\mathbf{b}_1, \dots, \mathbf{b}_{N_c}$  and the  $N_c$  columns of  $\mathbf{C}$  by  $\mathbf{c}_1, \dots, \mathbf{c}_{N_c}$ . If  $\mathbf{C} = \mathbf{A}\mathbf{B}$ , then  $\mathbf{c}_k = \mathbf{A}\mathbf{b}_k$  for  $k = 1, \dots, N_c$ .

It is possible to show that matrix multiplication is associative. Suppose we have three matrices  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  whose dimensions are such that the matrix products  $(\mathbf{A}\mathbf{B})$  and  $(\mathbf{B}\mathbf{C})$  are both well-defined. Then  $(\mathbf{A}\mathbf{B})\mathbf{C} = \mathbf{A}(\mathbf{B}\mathbf{C})$ . We often write  $\mathbf{ABC}$  to denote either product. Matrix multiplication is not commutative. Even when both products are well-defined, it is not in general true that  $\mathbf{BA}$  is equal to  $\mathbf{AB}$ .

**Matrix Transposition** The transpose of an  $N_r$  by  $N_c$  matrix  $\mathbf{A}$  is an  $N_c$  by  $N_r$  matrix  $\mathbf{B}$  whose  $ij$ th entry is given by  $b_{ij} = a_{ji}$ . We use the superscript “ $T$ ” to denote matrix transposition:  $\mathbf{B} = \mathbf{A}^T$ . The identity  $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$  always holds.

**Special Matrices and Vectors** A diagonal matrix  $\mathbf{D}$  is an  $N_r$  by  $N_c$  matrix whose entries  $d_{ij}$  are zero if  $i \neq j$ . That is, the only nonzero entries of a diagonal matrix lie along its main diagonal. We refer to the nonzero entries of a diagonal matrix as its diagonal entries.

A square matrix is a matrix whose row and column dimensions are equal. We refer to the row and column dimensions of a square matrix as its dimension.

An identity matrix is a square diagonal matrix whose diagonal entries are all one. We use the symbol  $\mathbf{I}_N$  to denote the  $N$  by  $N$  identity matrix. Using Eq. (45) it is possible to show that for any  $N_r$  by  $N_c$  matrix  $\mathbf{A}$ ,  $\mathbf{A}\mathbf{I}_{N_c} = \mathbf{I}_{N_r}\mathbf{A} = \mathbf{A}$ .

An orthogonal matrix  $\mathbf{U}$  is a square matrix that has the property that  $\mathbf{U}^T \mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I}_N$ , where  $N$  is the dimension of  $\mathbf{U}$ .

A zero vector is a vector whose entries are all zero. We use the symbol  $\mathbf{0}_N$  to denote the  $N$  dimensional zero vector.

## Linear Models

**Linear Combinations of Vectors** Equation (44) is not particularly intuitive. A useful way to think about the effect of multiplying a vector  $\mathbf{b}$  by matrix  $\mathbf{A}$  is as follows. Consider the matrix  $\mathbf{A}$  to consist of  $N_c$  column vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$ . Then from Eq. (44) we have that the vector  $\mathbf{c} = \mathbf{A}\mathbf{b}$  may be obtained by the operations of vector addition and scalar multiplication by

$$\mathbf{c} = \mathbf{a}_1 b_1 + \dots + \mathbf{a}_{N_c} b_{N_c} \quad (46)$$

where the numbers  $b_1, \dots, b_{N_c}$  are the entries of  $\mathbf{b}$ . Thus the effect of multiplying a vector by a matrix is to form a weighted sum of the columns of the matrix. The weights that go into forming the sum are the entries of the vector. We call any expression that has the form of the right hand side of Eq. (46) a linear combination of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$ .

**Independence and Rank** Consider a collection of vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$ . If no one of these vectors may be expressed as a linear combination of the others, then we say that the collection is independent. We define the rank of a collection of vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$  as the largest number of linearly independent vectors that may be chosen from that collection. We define the rank of a matrix  $\mathbf{A}$  to be the rank of the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$  that make up its columns. It may be proved that the rank of a matrix is always less than or equal to the minimum of its row and column dimensions. We say that a matrix has full rank if its rank is exactly equal to the minimum of its row and column dimensions.

**Linear Models** When a vector  $\mathbf{c}$  has the form given in Eq. (46), we say that  $\mathbf{c}$  lies within a linear model. We call  $N_c$  the dimension of the linear model. We call the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$  the basis vectors for the model. Thus an  $N_c$  dimensional linear model with basis vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$  contains all vectors  $\mathbf{c}$  that can be expressed exactly using Eq. (46) for some choice of numbers  $b_1, \dots, b_{N_c}$ . Equivalently, the linear model contains all vectors  $\mathbf{c}$  that may be expressed as  $\mathbf{c} = \mathbf{A}\mathbf{b}$  where the columns of the matrix  $\mathbf{A}$  are the vectors  $\mathbf{a}_1, \dots, \mathbf{a}_{N_c}$  and  $\mathbf{b}$  is an arbitrary vector. We refer to all vectors within a linear model as the subspace defined by that model.

## Simultaneous Linear Equations

**Matrix and Vector Form** Matrix multiplication may be used to express a system of simultaneous linear equations. Suppose we have a set of  $N_r$  simultaneous linear equations in  $N_c$  unknowns. Call the unknowns  $b_1, \dots, b_{N_c}$ . Conventionally we would write the equations in the form

$$\begin{aligned} a_{11}b_1 + \dots + a_{1N_c}b_{N_c} &= c_1 \\ a_{21}b_1 + \dots + a_{2N_c}b_{N_c} &= c_2 \\ &\dots \\ a_{N_r1}b_1 + \dots + a_{N_rN_c}b_{N_c} &= c_{N_r} \end{aligned} \quad (47)$$

where the  $a_{ij}$  and  $c_i$  represent known numbers. From Eq. (44) it is easy to see that we may rewrite Eq. (47) as a matrix multiplication

$$\mathbf{A}\mathbf{b} = \mathbf{c} \quad (48)$$

In this form, the entries of the vector  $\mathbf{b}$  represent the unknowns. Solving Eq. (48) for  $\mathbf{b}$  is equivalent to solving the system of simultaneous linear equations of Eq. (47).

**Solving Simultaneous Linear Equations** A fundamental topic in linear algebra is finding solutions for systems of simultaneous linear equations. We will rely on several basic results in this chapter, which we state here.

When the matrix  $\mathbf{A}$  is square and has full rank, it is always possible to find a unique matrix  $\mathbf{A}^{-1}$  such that  $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_{N_r}$ . We call the matrix  $\mathbf{A}^{-1}$  the inverse of the matrix  $\mathbf{A}$ . The matrix  $\mathbf{A}^{-1}$  is also square and has full rank. Algorithms exist for determining the inverse of a matrix and are provided by software packages that support matrix algebra.

When the matrix  $\mathbf{A}$  is square and has full rank, a unique solution  $\mathbf{b}$  to Eq. (48) exists. This solution is given simply by the expression  $\mathbf{b} = \mathbf{A}^{-1}\mathbf{c}$ . When the rank of  $\mathbf{A}$  is less than its row dimension, then there will not in general be an exact solution to Eq. (48). There will, however, be a unique vector  $\mathbf{b}$  that is the best solution in a least squares sense. We call this the least squares solution to Eq. (48). Finding the least squares solution to Eq. (48) is often referred to as linear regression. Algorithms exist for performing linear regression and are provided by software packages that support matrix algebra.

A generalization of Eq. (48) is the matrix equation

$$\mathbf{A}\mathbf{B} = \mathbf{C} \quad (49)$$



where the entries of the matrix  $\mathbf{B}$  are the unknowns. From our interpretation of matrix multiplication as a shorthand for multiple multiplications of a vector by a matrix, we can see immediately that this type of equation may be solved by applying the above analysis in a columnwise fashion. If  $\mathbf{A}$  is square and has full rank, then we may determine  $\mathbf{B}$  uniquely as  $\mathbf{A}^{-1}\mathbf{C}$ . When the rank of  $\mathbf{A}$  is less than its row dimension, we may use regression determine a matrix  $\mathbf{B}$  that satisfies Eq. (49) in a least squares sense.

It is also possible to solve matrix equations of the form  $\mathbf{BA} = \mathbf{C}$  where the entries of  $\mathbf{B}$  are again the unknowns. An equation of this form may be converted to the form of Eq. (49) by applying the transpose identity (see subsection “Matrix Transposition”). That is, we may find  $\mathbf{B}$  by solving the equation  $\mathbf{A}^T\mathbf{B}^T = \mathbf{C}^T$  if  $\mathbf{A}^T$  meets the appropriate conditions.

**Null Space** When the rank of a matrix  $\mathbf{A}$  is greater than its row dimension  $N_r$ , it is possible to find nontrivial solutions to the equation

$$\mathbf{A}\mathbf{b} = \mathbf{0}_{N_r} \tag{50}$$

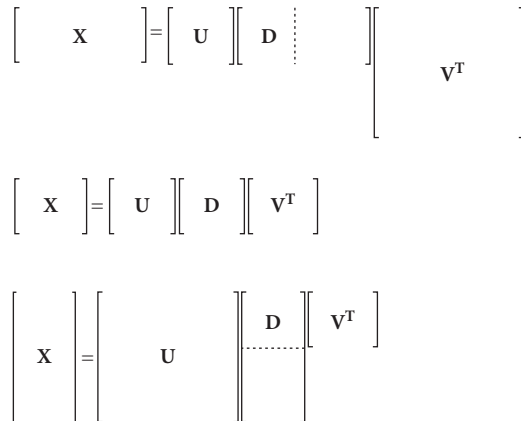
Indeed, it is possible to determine a linear model such that all vectors contained in the model satisfy Eq. (50). This linear model will have dimension equal to the difference between  $N_r$  and the rank of the matrix  $\mathbf{A}$ . The subspace defined by this linear model is called the null space of the matrix  $\mathbf{A}$ . Algorithms to find the basis vectors of a matrix’s null space exist and are provided by software packages that support matrix algebra.

### Singular Value Decomposition

The singular value decomposition allows us to write any  $N_r$  by  $N_c$  matrix  $\mathbf{X}$  in the form

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{51}$$

where  $\mathbf{U}$  is an  $N_r$  by  $N_r$  orthogonal matrix,  $\mathbf{D}$  is an  $N_r$  by  $N_c$  “diagonal” matrix, and  $\mathbf{V}$  is an  $N_c$  by  $N_c$  orthogonal matrix.<sup>148</sup> The diagonal entries of  $\mathbf{D}$  are guaranteed to be nonnegative. Some of the diagonal entries may be zero. By convention, the entries along this diagonal are arranged in decreasing order. We illustrate the singular value decomposition in Fig. 21. The singular value decomposition has a large number of uses in numerical matrix algebra. Routines to compute it are generally provided as part of software packages that support matrix algebra.



**FIGURE 21** The figure depicts the singular value decomposition (SVD) of an  $N$  by  $M$  matrix  $\mathbf{X}$  for three cases:  $N_c > N_r$ ,  $N_c = N_r$ , and  $N_c < N_r$ .

## 10.8 REFERENCES

1. J. von Kries, "Chromatic Adaptation," 1902, reprinted in *Sources of Color Vision*, D. L. MacAdam, (ed.), MIT Press, 1970, Cambridge, MA, pp. 109–119.
2. R. M. Evans, *An Introduction to Color*, Wiley, New York, 1948.
3. G. Wyszecki, "Color Appearance," in *Handbook of Perception and Human Performance: Sensory Processes and Perception*, K. R. Boff, L. Kaufman, and J. P. Thomas, (eds.), John Wiley & Sons, New York, 1986, pp. 9.1–9.56.
4. B. K. P. Horn, "Exact Reproduction of Colored Images," *Computer Vision, Graphics and Image Processing* **26**:135–167 (1984).
5. R. W. G. Hunt, *The Reproduction of Colour*, 4th ed., Fountain Press, Tolworth, England, 1987.
6. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley, Reading, MA, 1998.
7. D. H. Brainard and B. A. Wandell, "Calibrated Processing of Image Color," *Color Research and Application* **15**:266–271 (1990).
8. G. Wyszecki and W. S. Stiles, *Color Science—Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., John Wiley & Sons, New York, 1982.
9. V. C. Smith and J. Pokorny, "Color Matching and Color Discrimination," in *The Science of Color*, 2nd ed., S. K. Shevell, (ed.), Optical Society of America; Elsevier Ltd, Oxford, 2003, pp. 103–148.
10. CIE, *Colorimetry*, 3rd edition, Publication 15.2004, Bureau Central de la CIE, Vienna, 2004.
11. CIE, *Fundamental Chromaticity Diagram with Physiological Axes—Part 1, Publication 170–1*, Bureau Central de la CIE, Vienna, 2006.
12. D. H. Krantz, "Color Measurement and Color Theory: I. Representation Theorem for Grassmann Structures," *Journal of Mathematical Psychology* **12**:283–303 (1975).
13. P. Suppes, D. H. Krantz, R. D. Luce, and A. Tversky, *Foundations of Measurement*, Academic Press, San Diego, 1989, Vol. II.
14. D. L. MacAdam, *Sources of Color Science*, MIT Press, Cambridge, MA, 1970.
15. W. D. Wright, "The Origins of the 1931 CIE System," in *Human Color Vision*, 2nd ed., P. K. Kaiser and R. M. Boynton, (eds.), Optical Society of America, Washington, D.C., 1996, pp. 534–543.
16. J. D. Mollon, "The Origins of Modern Color Science," in *The Science of Color*, 2nd ed., S. K. Shevell, (ed.), Optical Society of America; Elsevier Ltd, Oxford, 2003, pp. 1–39.
17. R. Hall, *Illumination and Color in Computer Generated Imagery*, Springer-Verlag, New York, 1989.
18. D. B. Judd and G. Wyszecki, *Color in Business, Science, and Industry*, John-Wiley and Sons, New York, 1975.
19. G. S. Brindley, *Physiology of the Retina and the Visual Pathway*, 2nd ed., Williams and Wilkins, Baltimore, 1970.
20. R. M. Boynton, "History and Current Status of a Physiologically Based System of Photometry and Colorimetry," *Journal of the Optical Society of America A* **65**:1609–1621 (1996).
21. A. Stockman and L. T. Sharpe, "Cone Spectral Sensitivities and Color Matching," in *Color Vision: From Genes To Perception*, K. R. Gegenfurtner and L. T. Sharpe, (eds.), Cambridge University Press, Cambridge, MA, 1999, pp. 53–87.
22. B. A. Wandell, "The Synthesis and Analysis of Color Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**:2–13 (1987).
23. S. Westland and C. Ripamonti, *Computational Colour Science using MATLAB*, John Wiley & Sons, 2004.
24. B. Noble and J. W. Daniel, *Applied Linear Algebra*, 2nd ed., Prentice-Hall, Inc., Englewood Cliffs, NJ, 1977.
25. G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.
26. W. K. Pratt, *Digital Image Processing*, John Wiley & Sons, New York, 1978.
27. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, MA, 1988.
28. G. Wyszecki and W. Stiles, *Color Science—Concepts and Methods, Quantitative Data and Formulas*, John Wiley & Sons, New York, 1967.
29. W. S. Stiles, "The Physical Interpretation of the Spectral Sensitivity Curve of the Eye," in *Transactions of the Optical Convention of the Worshipful Company of Spectacle Makers*, Spectacle Maker's Company, London, 1948, pp. 97–107.

30. D. E. Mitchell and W. A. H. Rushton, "Visual Pigments in Dichromats," *Vision Research* **11**:1033–1043 (1971).
31. W. D. Wright, "A Re-determination of the Trichromatic Coefficients of the Spectral Colours," *Transactions of the Optical Society* **30**:141–164 (1928–1929).
32. J. Guild, "The Colorimetric Properties of the Spectrum," *Philosophical Transactions of the Royal Society of London A* **230**:149–187 (1931).
33. W. S. Stiles, "Interim Report to the Commission Internationale de l'Éclairage Zurich, 1955, on the National Physical Laboratory's Investigation of Colour-matching" (with an Appendix by W. S. Stiles & J. M. Burch) *Optica Acta* **2**:168–181 (1955).
34. W. S. Stiles and J. M. Burch, "NPL Colour-matching Investigation: Final Report (1958)," *Optica Acta* **6**:1–26 (1959).
35. J. C. Maxwell, "On the Theory of Compound Colours and the Relations of the Colours of the Spectrum," *Philosophical Transactions of the Royal Society of London* **150**:57–84 (1860).
36. J. G. Grassmann, "Theory of Compound Colors," in MacAdam, D.L., (ed.), *Sources of Color Vision*, MIT Press: Cambridge, MA, 1970. (Originally published in *Annalen der Physik und Chemie*, **89**:69–84 1853).
37. CIE, *Commission Internationale de l'Éclairage Proceedings, 1931*, Cambridge University Press, Cambridge, MA, 1932.
38. D. B. Judd, "Report of U.S. Secretariat Committee on Colorimetry and Artificial Daylight," *Proceedings of the Twelfth Session of the CIE* **1**:11 (1951).
39. J. J. Vos, "Colorimetric and Photometric Properties of a 2° Fundamental Observer," *Color Research and Application* **3**:125–128 (1978).
40. ISO/CIE, *CIE Standard Colorimetric Observers*, Reference Number 10527, International Organization for Standardization, Geneva, 1991.
41. I. E. Abdou and W. K. Pratt, "Quantitative Design and Evaluation of Enhancement/Thresholding Edge Detectors," *Proceedings of the IEEE* **67**:753–763 (1979).
42. CIE, *Commission Internationale de l'Éclairage Proceedings, 1924*, Cambridge University Press, Cambridge, MA, 1926.
43. H. G. Sperling, "An Experimental Investigation of the Relationship Between Colour Mixture and Luminous Efficiency," in *Visual Problems of Colour*, vol. 1, Her Majesty's Stationery Office, London, 1958, pp. 249–277.
44. A. Stockman, L. T. Sharpe, and C. C. Fach, "The Spectral Sensitivity of the Human Short-wavelength Cones," *Vision Research* **39**:2901–2927 (1999).
45. V. C. Smith, J. Pokorny, and Q. Zaidi, "How do Sets of Color-matching Functions Differ?," in *Colour Vision*, J. D. Mollon and L. T. Sharpe, (eds.), Academic Press, London, 1983, pp. 93–105.
46. A. Stockman and L. T. Sharpe, "Spectral Sensitivities of the Middle- and Long-Wavelength Sensitive Cones Derived from Measurements in Observers of Known Genotype," *Vision Research* **40**:1711–1737 (2000).
47. P. DeMarco, J. Pokorny, and V. C. Smith, "Full-spectrum Cone Sensitivity Functions for X-chromosome-linked Anomalous Trichromats," *Journal of the Optical Society of America* **9**:1465–1476 (1992).
48. N. I. Speranskaya, "Determination of Spectrum Color Coordinates for Twenty-seven Normal Observers," *Optics and Spectroscopy* **7**:424–428 (1959).
49. O. N. Rood, *Modern Chromatics, with Applications to Art and Industry*, D. Appleton & Co., New York, 1879.
50. T. Young, "On the Theory of Light and Colours," *Philosophical Transactions of the Royal Society of London* **92**:12–48 (1802).
51. H. L. F. von Helmholtz, "On the Theory of Compound Colours," *Philosophical Magazine Series, 4* **4**:519–534 (1852).
52. J. C. Maxwell, "Experiments on Colours, as Perceived by the Eye, with Remarks on Colour-blindness," *Transactions of the Royal Society of Edinburgh* **21**:275–298 (1855).
53. V. Smith and J. Pokorny, "Spectral Sensitivity of the Foveal Cone Photopigments between 400 and 500 nm," *Vision Research* **15**:161–171 (1975).
54. J. C. Maxwell, "On the Theory of Colours in Relation to Colour-blindness. A letter to Dr. G. Wilson," *Transactions of the Royal Scottish Society of Arts* **4**:394–400 (1856).
55. J. Nathans, T. P. Piantanida, R. L. Eddy, T. B. Shows, and D. S. Hogness, "Molecular Genetics of Inherited Variation in Human Color Vision," *Science* **232**:203–210 (1986).
56. J. Nathans, D. Thomas, and D. S. Hogness, "Molecular Genetics of Human Color Vision: the Genes Encoding Blue, Green and Red Pigments," *Science* **232**:193–202 (1986).

57. M. M. Bongard and M. S. Smirnov, "Determination of the Eye Spectral Sensitivity Curves from Spectral Mixture Curves," *Doklady Akademii nauk S.S.S.R.* **102**:1111–1114 (1954).
58. A. Stockman, D. I. A. MacLeod, and N. E. Johnson, "Spectral Sensitivities of the Human Cones," *Journal of the Optical Society of America A* **10**:2491–2521 (1993).
59. A. König and C. Dieterici, "Die Grundempfindungen in Normalen und anomalen Farben Systemen und ihre Intensitäts-Vertheilung im Spectrum," *Z Psychol Physiol Sinnesorg* **4**:241–347 (1893).
60. P. J. Bouma, "Mathematical Relationship between the Colour Vision System of Trichromats and Dichromats," *Physica* **9**:773–784 (1942).
61. D. B. Judd, "Standard Response Functions for Protanopic and Deutanopic Vision," *Journal of the Optical Society of America* **35**:199–221 (1945).
62. D. B. Judd, "Standard Response Functions for Protanopic and Deutanopic Vision," *Journal of the Optical Society of America* **39**:505 (1949).
63. J. J. Vos and P. L. Walraven, "On the Derivation of the Foveal Receptor Primaries," *Vision Research* **11**:799–818 (1971).
64. O. Estévez, "On the Fundamental Database of Normal and Dichromatic Color Vision," Ph.D., Amsterdam University, 1979.
65. J. J. Vos, O. Estevez, and P. L. Walraven, "Improved Color Fundamentals Offer a New View on Photometric Additivity," *Vision Research* **30**:937–943 (1990).
66. A. Stockman, D. I. A. MacLeod, and J. A. Vivien, "Isolation of the Middle- and Long-wavelength Sensitive Cones in Normal Trichromats," *Journal of the Optical Society of America A* **10**:2471–2490 (1993).
67. M. A. Webster and D. I. A. MacLeod, "Factors Underlying Individual Differences in the Color Matches of Normal Observers," *Journal of the Optical Society of America A* **5**:1722–1735 (1988).
68. G. Wald, "Human Vision and the Spectrum," *Science* **101**:653–658 (1945).
69. R. A. Bone and J. M. B. Sparrock, "Comparison of Macular Pigment Densities in the Human Eye," *Vision Research* **11**:1057–1064 (1971).
70. P. L. Pease, A. J. Adams, and E. Nuccio, "Optical Density of Human Macular Pigment," *Vision Research* **27**:705–710 (1987).
71. D. van Norren and J. J. Vos, "Spectral Transmission of the Human Ocular Media," *Vision Research* **14**:1237–1244 (1974).
72. B. H. Crawford, "The Scotopic Visibility Function," *Proceedings of the Physical Society of London B* **62**:321–334 (1949).
73. F. S. Said and R. A. Weale, "The Variation with Age of the Spectral Transmissivity of the Living Human Crystalline Lens," *Gerontologia* **3**:213–231 (1959).
74. J. Pokorny, V. C. Smith, and M. Lutze, "Aging of the Human Lens," *Applied Optics* **26**:1437–1440 (1987).
75. M. Alpern, "Lack of Uniformity in Colour Matching," *Journal of Physiology* **288**:85–105 (1979).
76. H. Terstiege, "Untersuchungen zum Persistenz- und Koeffizientensatz," *Die Farbe* **16**:1–120 (1967).
77. S. S. Miller, "Psychophysical Estimates of Visual Pigment Densities in Red-green Dichromats," *Journal of Physiology* **223**:89–107 (1972).
78. P. E. King-Smith, "The Optical Density of Erythrolabe Determined by a New Method," *Journal of Physiology* **230**:551–560 (1973).
79. P. E. King-Smith, "The Optical Density of Erythrolabe Determined by Retinal Densitometry Using the Self-screening Method," *Journal of Physiology* **230**:535–549 (1973).
80. V. Smith and J. Pokorny, "Psychophysical Estimates of Optical Density in Human Cones," *Vision Research* **13**:1099–1202 (1973).
81. S. A. Burns and A. E. Elsner, "Color Matching at High Luminances: Photopigment Optical Density and Pupil Entry," *Journal of the Optical Society of America A* **10**:221–230 (1993).
82. T. T. J. M. Berendschot, J. van der Kraats, and D. van Norren, "Foveal Cone Mosaic and Visual Pigment Density in Dichromats," *Journal of Physiology* **492**:307–314 (1996).
83. C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson, "Human Photoreceptor Topography," *Journal of Comparative Neurology* **292**:497–523 (1990).

84. A. E. Elsner, S. A. Burns, and R. H. Webb, "Mapping Cone Photopigment Optical Density," *Journal of the Optical Society of America A* **10**:52–58 (1993).
85. M. Neitz, J. Neitz, and G. H. Jacobs, "Spectral Tuning of Pigments Underlying Red-green Color Vision," *Science* **252**:971–973 (1991).
86. S. C. Merbs and J. Nathans, "Absorption Spectra of Human Cone Photopigments," *Nature* **356**:433–435 (1992).
87. S. L. Merbs and J. Nathans, "Absorption Spectra of the Hybrid Pigments Responsible for Anomalous Color Vision," *Science* **258**:464–466 (1992).
88. J. Winderickx, D. T. Lindsey, E. Sanocki, D. Y. Teller, A. G. Motulsky, and S. S. Deeb, "Polymorphism in Red Photopigment Underlies Variation in Colour Matching," *Nature* **356**:431–433 (1992).
89. L. T. Sharpe, A. Stockman, H. Jägle, and J. Nathans, "Opsin Genes, Cone Photopigments, Color Vision, and Color Blindness," in *Color Vision: From Genes To Perception*, K. R. Gegenfurtner and L. T. Sharpe, (eds.), Cambridge University Press, Cambridge, MA, 1999, pp. 3–51.
90. D. H. Brainard, A. Roorda, Y. Yamauchi, J. B. Calderone, A. Metha, M. Neitz, J. Neitz, D. R. Williams, and G. H. Jacobs, "Functional Consequences of the Relative Numbers of L and M Cones," *Journal of the Optical Society of America A* **17**:607–614 (2000).
91. M. Drummond-Borg, S. S. Deeb, and A. G. Motulsky, "Molecular Patterns of X Chromosome-linked Color Vision Genes among 134 Men of European Ancestry," *Proceedings of the National Academy of Sciences* **86**:983–987 (1989).
92. J. Neitz, M. Neitz, and G. H. Jacobs, "More than 3 Different Cone Pigments among People with Normal Color Vision," *Vision Research* **33**:117–122 (1993).
93. J. P. Macke and J. Nathans, "Individual Variation in the Size of the Human Red and Green Pigment Gene Array," *Investigative Ophthalmology and Visual Science* **38**:1040–1043 (1997).
94. A. B. Asenjo, J. Rim, and D. D. Oprian, "Molecular Determinants of Human Red/Green Color Discrimination," *Neuron* **12**:1131–1138 (1994).
95. L. T. Sharpe, A. Stockman, H. Jägle, H. Knau, and J. Nathans, "L, M, and L-M Hybrid Cone Photopigments in Man: Deriving  $\lambda_{\max}$  from Flicker Photometric Spectral Sensitivities," *Vision Research* **39**:3513–3525 (1999).
96. M. Neitz, J. Neitz, and G. H. Jacobs, "Genetic Basis of Photopigment Variations in Human Dichromats," *Vision Research* **35**:2095–2103 (1995).
97. A. Stockman, L. T. Sharpe, S. Merbs, and J. Nathans, "Spectral Sensitivities of Human Cone Visual Pigments Determined *in vivo* and *in vitro*," in *Vertebrate Phototransduction and the Visual Cycle, Part B. Methods in Enzymology, Vol. 316*, K. Palczewski, (ed.), Academic Press, New York, 2000, pp. 626–650.
98. J. J. Kremers, T. Usui, H. P. Scholl, and L. T. Sharpe, "Cone Signal Contributions to Electrograms in Dichromats and Trichromats," *Investigative Ophthalmology and Visual Science* **40**:920–930 (1999).
99. A. Stockman, H. Jägle, M. Pirzer, and L. T. Sharpe, "The Dependence of Luminous Efficiency on Chromatic Adaptation," *Journal of Vision* **8**:1–26 (2008).
100. A. Reitner, L. T. Sharpe, and E. Zrenner, "Is Colour Vision Possible with Only Rods and Blue-sensitive Cones?," *Nature* **352**:798–800 (1991).
101. A. König, and C. Dieterici, "Die Grundempfindungen und ihre Intensitäts-Vertheilung im Spectrum," *Sitzungsberichte Akademie der Wissenschaften*, Berlin **1886**:805–829 (1886).
102. S. Ishihara, *Tests for Colour-Blindness*, Kanehara Shuppen Company, Ltd., Tokyo, 1977.
103. D. Farnsworth, "The Farnsworth-Munsell 100 Hue and Dichotomous Tests for Color Vision," *Journal of the Optical Society of America* **33**:568–578 (1943).
104. L. Rayleigh, "Experiments on Colour," *Nature* **25**:64–66 (1881).
105. J. B. Birch, *Diagnosis of Defective Colour Vision*, Oxford University Press, Oxford, 1993.
106. A. König, "Über den Menschlichen Sehpurpur und seine Bedeutung für das Sehen," *Academie der Wissenschaften Sitzungsberichte* **30**:577–598 (1894).
107. E. N. Willmer, "Further Observations on the Properties of the Central Fovea in Colour-blind and Normal Subjects," *Journal of Physiology* **110**:422–446 (1950).
108. L. C. Thomson and W. D. Wright, "The Colour Sensitivity of the Retina within the Central Fovea of Man," *Journal of Physiology* **105**:316–331 (1947).

109. D. Williams, D. I. A. MacLeod, and M. Hayhoe, "Foveal Tritanopia," *Vision Research* **21**:1341–1356 (1981).
110. D. I. Flitcroft, "The Interactions between Chromatic Aberration, Defocus and Stimulus Chromaticity: Implications for Visual Physiology and Colorimetry," *Vision Research* **29**:349–360 (1989).
111. D. R. Williams, N. Sekiguchi, W. Haake, D. H. Brainard, and O. Packer, "The Cost of Trichromacy for Spatial Vision," in *From Pigments to Perception*, B. B. Lee and A. Valberg, (eds.), Plenum Press, New York, 1991, pp. 11–22.
112. D. H. Marimont and B. A. Wandell, "Matching Color Images—the Effects of Axial Chromatic Aberration," *Journal of the Optical Society of America A* **11**:3113–3122 (1994).
113. L. N. Thibos, M. Ye, X. X. Zhang, and A. Bradley, "The Chromatic Eye—a New Reduced-eye Model of Ocular Chromatic Aberration in Humans," *Applied Optics* **31**:3594–3667 (1992).
114. I. Powell, "Lenses for Correcting Chromatic Aberration of the Eye," *Applied Optics* **20**:4152–4155 (1981).
115. K. H. Ruddock, "The Effect of Age Upon Colour Vision. II. Changes with Age in Light Transmission of the Ocular Media," *Vision Research* **5**:47–58 (1965).
116. A. Knowles and H. J. A. Dartnall, "The Photobiology of Vision," in *The Eye*, H. Davson, (ed.), Academic Press, New York, 1977, pp. 321–533.
117. H. J. A. Dartnall, "The Interpretation of Spectral Sensitivity Curves," *British Medical Bulletin* **9**:24–30 (1953).
118. R. J. W. Mansfield, "Primate Photopigments and Cone Mechanisms," in *The Visual System*, A. Fein and J. S. Levine, (eds.), Alan R. Liss, New York, 1985.
119. E. F. MacNichol, Jr., "A Unifying Presentation of Photopigment Spectra," *Vision Research* **26**:1543–1556 (1986).
120. T. D. Lamb, "Photoreceptor Spectral Sensitivities: Common Shape in the Long-Wavelength Spectral Region," *Vision Research* **35**:3083–3091 (1995).
121. H. B. Barlow, "What Causes Trichromacy? A Theoretical Analysis using Comb-filtered Spectra," *Vision Research* **22**:635–643 (1982).
122. V. I. Govardovskii, N. Fyhrquist, T. Reuter, D. G. Kuzmin, and K. Donner, "In Search of the Visual Pigment Template," *Visual Neuroscience* **17**:509–528 (2000).
123. J. Walraven, C. Enroth-Cugell, D. C. Hood, D. I. A. MacLeod, and J. L. Schnapf, "The Control of Visual Sensitivity. Receptor and Postreceptor Processes," in *Visual Perception: The Neurophysiological Foundations*, L. Spillman and J. S. Werner, (eds.), Academic Press, San Diego, 1990, pp. 53–101.
124. A. M. Derrington, J. Krauskopf, and P. Lennie, "Colour-opponent Cells in the Dorsal Lateral Geniculate Nucleus of the Macaque," *Journal of Physiology* **329**:22–23 (1982).
125. D. I. A. MacLeod and R. M. Boynton, "Chromaticity Diagram Showing Cone Excitation by Stimuli of Equal Luminance," *Journal of the Optical Society of America* **69**:1183–1186 (1979).
126. R. Luther, "Aus dem Gebiet der Farbreizmetrik," *Zeitschrift für Technische Physik* **8**:540–558 (1927).
127. J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal Directions of Color Space," *Vision Research* **22**:1123–1131 (1982).
128. J. Krauskopf, D. R. Williams, M. B. Mandler, and A. M. Brown, "Higher Order Color Mechanisms," *Vision Research* **26**:23–32 (1986).
129. D. H. Brainard, "Cone Contrast and Opponent Modulation Color Spaces," in *Human Color Vision*, 2nd ed., P. K. Kaiser and R. M. Boynton, Optical Society of America, Washington, D.C., 1996, pp. 563–579.
130. A. B. Poirson and B. A. Wandell, "The Ellipsoidal Representation of Spectral Sensitivity," *Vision Research* **30**:647–652 (1990).
131. R. Ramamoorthi and P. Hanrahan, "A Signal-processing Framework for Reflection," *ACM Transactions on Graphics* **23**:1004–1042 (2004).
132. R. O. Dror, A. S. Willsky, and E. H. Adelson, "Statistical Characterization of Real-world Illumination," *Journal of Vision* **4**:821–837 (2004).
133. R. W. Fleming, R. O. Dror, and E. H. Adelson, "Real-world Illumination and the Perception of Surface Reflectance Properties," *Journal of Vision* **3**:347–368 (2003).
134. R. W. Fleming, A. Torralba, and E. H. Adelson, "Specular Reflections and the Perception of Shape," *Journal of Vision* **4**:798–820 (2004).



135. B. Xiao and D. H. Brainard, "Surface Gloss and Color Perception of 3D Objects," *Visual Neuroscience* **25**: 371–385 (2008).
136. I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image Statistics and the Perception of Surface Qualities," *Nature* **447**:206–209 (2007).
137. D. B. Judd, D. L. MacAdam, and G. W. Wyszecki, "Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature," *Journal of the Optical Society of America* **54**:1031–1040 (1964).
138. J. Cohen, "Dependency of the Spectral Reflectance Curves of the Munsell Color Chips," *Psychonomic Science* **1**:369–370 (1964).
139. K. L. Kelly, K. S. Gibson, and D. Nickerson, "Tristimulus Specification of the Munsell Book of Color from Spectrophotometric Measurements," *Journal of the Optical Society of America* **33**:355–376 (1943).
140. D. Nickerson, "Spectrophotometric Data for a Collection of Munsell Samples," U.S. Department of Agriculture, 1957.
141. L. T. Maloney, "Evaluation of Linear Models of Surface Spectral Reflectance with Small Numbers of Parameters," *Journal of the Optical Society of America A* **3**:1673–1683 (1986).
142. E. L. Krinov, "Surface Reflectance Properties of Natural Formations," National Research Council of Canada: Technical Translation **TT-439** (1947).
143. J. P. S. Parkkinen, J. Hallikainen, and T. Jaaskelainen, "Characteristic Spectra of Munsell Colors," *Journal of the Optical Society of America* **6**:318–322 (1989).
144. T. Jaaskelainen, J. Parkkinen, and S. Toyooka, "A Vector-subspace Model for Color Representation," *Journal of the Optical Society of America A* **7**:725–730 (1990).
145. J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, Chichester, 1988.
146. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed., John Wiley and Sons, New York, 1971.
147. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1988.
148. J. M. Chambers, *Computational Methods for Data Analysis*, John Wiley and Sons, New York, 1977.
149. D. H. Marimont and B. A. Wandell, "Linear Models of Surface and Illuminant Spectra," *Journal of the Optical Society of America A* **9**:1905–1913 (1992).
150. B. A. Wandell and D. H. Brainard, "Towards Cross-media Color Reproduction," Presented at the OSA Applied Vision Topical Meeting, San Francisco, CA, 1989.
151. C. A. Parraga, G. Brelstaff, T. Troscianko, and I. R. Moorehead, "Color and Luminance Information in Natural Scenes," *Journal of the Optical Society of America A* **15**:563–569 (1998).
152. P. Longère and D. H. Brainard, "Simulation of Digital Camera Images from Hyperspectral Input," in *Vision Models and Applications to Image and Video Processing*, C. J. van den Branden Lambrecht (ed.), Kluwer Academic, Boston, 2001, pp. 123–150.
153. D. H. Foster, S. M. C. Nascimento, and K. Amano, "Information Limits on Neural Identification of Colored Surfaces in Natural Scenes," *Visual Neuroscience* **21**:1–6 (2004).
154. D. Saunders and A. Hamber, "From Pigments to Pixels: Measurement and Display of the Colour Gamut of Paintings," Proceedings of the SPIE: *Perceiving, Measuring, and Using Color* **1250**:90–102 (1990).
155. R. Berns, "Rejuvenating Seurat's Palette Using Color and Imaging Science: a Simulation," in *Seurat and the Making of La Grande Jatte*, R. L. Herbert, (ed.), University of California Press, 2004, pp. 214–227.
156. D. H. Brainard, D. G. Pelli, and T. Robson, "Display Characterization," in *Encyclopedia of Imaging Science and Technology*, J. Hornak, (ed.) Wiley, 2002, pp. 72–188.
157. ISO/CIE, *CIE Standard Colorimetric Illuminants*, Reference Number 10526, International Organization for Standardization, Geneva, 1991.
158. D. H. Brainard, B. A. Wandell, and W. B. Cowan, "Black Light: How Sensors Filter Spectral Variation of the Illuminant," *IEEE Transactions on Biomedical Engineering* **36**:140–149 (1989).
159. G. Wyszecki, "Evaluation of Metameric Colors," *Journal of the Optical Society of America* **48**:451–454 (1958).
160. S. A. Burns, J. B. Cohen, and E. N. Kuznetsov, "Multiple Metatmers: Preserving Color Matches Under Diverse Illuminants," *Color Research and Application* **14**:16–22 (1989).

161. G. D. Finlayson and M. S. Drew, "The Maximum Ignorance Assumption with Positivity," Presented at the 4th IS&T/SID Color Imaging Conference, Scottsdale, AZ, 1996, pp. 202–204.
162. J. A. S. Viggiano, "Minimal-knowledge Assumptions in Digital Still Camerage Characterization. I. Uniform Distribution, Toeplitz Correlation," Presented at the 9th IS&T/SID Color Imaging Conference, Scottsdale, AZ, 2001, pp. 332–336.
163. X. Zhang and D. H. Brainard, "Bayesian Color-correction Method for Non-colorimetric Digital Image Sensors," Presented at the 12th IS&T/SID Color Imaging Conference, Scottsdale, AZ, 2004, pp. 308–314.
164. M. H. Brill, "A Device Performing Illuminant-invariant Assessment of Chromatic Relations," *Journal of Theoretical Biology* **71**:473–478 (1978).
165. G. Buchsbaum, "A Spatial Processor Model for Object Colour Perception," *Journal of the Franklin Institute* **310**:1–26 (1980).
166. L. T. Maloney and B. A. Wandell, "Color Constancy: A Method for Recovering Surface Spectral Reflectances," *Journal of the Optical Society of America A* **3**:29–33 (1986).
167. H. C. Lee, "Method for Computing the Scene-illuminant Chromaticity from Specular Highlights," *Journal of the Optical Society of America A* **3**:1694–1699 (1986).
168. B. Funt and J. Ho, "Color from Black and White," Presented at the International Conference on Computer Vision, Tampa, FL, 1988, pp. 2–8.
169. B. V. Funt and M. S. Drew, "Color Constancy Computation in Near-Mondrian Scenes Using a Finite Dimensional Linear Model," Presented at the IEEE Conference on Computer Vision and Pattern Recognition, Ann Arbor, MI, 1988, pp. 544–549.
170. D. A. Forsyth, "A Novel Approach to Colour Constancy," Presented at the International Conference on Computer Vision, Tampa, FL, 1988, pp. 9–18.
171. G. D. Finlayson, P. H. Hubel, and S. Hordley, "Color by correlation," Presented at the IS&T/SID Fifth Color Imaging Conference, Scottsdale, AZ, 1997, pp. 6–11.
172. D. H. Brainard and W. T. Freeman, "Bayesian Color Constancy," *Journal of the Optical Society of America A* **14**:1393–1411 (1997).
173. M. Ebner, *Color Constancy*, Wiley, Chichester, UK, 2007.
174. M. D'Zmura and P. Lennie, "Mechanisms of Color Constancy," *Journal of the Optical Society of America A* **3**:1662–1672 (1986).
175. A. C. Hurlbert, "Computational Models of Color Constancy," in *Perceptual Constancy: Why Things Look As They Do*, V. Walsh and J. Kulikowski, (eds.), Cambridge University Press, Cambridge, MA, 1998, pp. 283–322.
176. L. T. Maloney, "Physics-Based Approaches to Modeling Surface Color Perception," in *Color Vision: From Genes to Perception*, K. T. Gegenfurtner and L. T. Sharpe, (eds.), Cambridge University Press, Cambridge, MA, 1999, pp. 387–416.
177. D. H. Brainard, J. M. Kraft, and P. Longère, "Color Constancy: Developing Empirical Tests of Computational Models," in *Colour Perception: Mind and the Physical World*, R. Mausfeld and D. Heyer, (eds.), Oxford University Press, Oxford, 2003, pp. 307–334.
178. D. H. Brainard, P. Longere, P. B. Delahunt, W. T. Freeman, J. M. Kraft, and B. Xiao, "Bayesian Model of Human Color Constancy," *Journal of Vision* **6**:1267–1281 (2006).
179. D. L. MacAdam, "Visual Sensitivities to Color Differences in Daylight," *Journal of the Optical Society of America* **32**:247–274 (1942).
180. D. L. MacAdam, "Colour Discrimination and the Influence of Colour Contrast on Acuity" *Documenta Ophthalmologica* **3**:214–233 (1949).
181. G. Wyszecki, "Matching Color Differences," *Journal of the Optical Society of America* **55**:1319–1324 (1965).
182. G. Wyszecki and G. H. Fielder, "New Color-matching Ellipses," *Journal of the Optical Society of America* **61**:1135–1152 (1971).
183. W. S. Stiles, "Color vision: The Approach Through Increment Threshold Sensitivity," *Proceedings National Academy of Sciences (USA)* **45**:100–114 (1959).
184. C. Noorlander and J. J. Koenderink, "Spatial and Temporal Discrimination Ellipsoids in Color Space," *Journal of the Optical Society of America* **73**:1533–1543 (1983).
185. A. B. Poirson, B. A. Wandell, D. Varner, and D. H. Brainard, "Surface Characterizations of Color Thresholds," *Journal of the Optical Society of America A* **7**:783–789 (1990).



186. A. R. Robertson, "The CIE 1976 Color-difference Formula," *Color Research and Application* **2**:7–11 (1977).
187. H. de Lange, "Research into the Dynamic Nature of the Human Fovea-cortex Systems with Intermittent and Modulated Light. I. Attenuation Characteristics with White and Coloured Light," *Journal of the Optical Society of America* **48**:777–784 (1958).
188. H. de Lange, "Research into the Dynamic Nature of the Human Fovea-cortex Systems with Intermittent and Modulated Light. II. Phase Shift in Brightness and Delay in Color Perception," *Journal of the Optical Society of America* **48**:784–789 (1958).
189. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Efficiency for Detecting Isoluminant and Isochromatic Interference Fringes," *Journal of the Optical Society of America A* **10**:2118–2133 (1993).
190. E. C. Carter and R. C. Carter, "Color and Conspicuousness," *Journal of the Optical Society of America* **71**:723–729 (1981).
191. A. B. Poirson and B. A. Wandell, "Task-Dependent Color Discrimination," *Journal of the Optical Society of America A* **7**:776–782 (1990).
192. A. L. Nagy and R. R. Sanchez, "Critical Color Differences Determined with a Visual Search Task," *Journal of the Optical Society of America A* **7**:1209–1217 (1990).
193. H. E. Smithson, S. S. Khan, L. T. Sharpe, and A. Sotckman, "Transitions between Color Categories Mapped with a Reverse Stroop Task," *Visual Neuroscience* **23**:453–460 (2006).
194. CIE, *Improvement to Industrial Colour-Difference Evaluation*, Publication 142, Bureau Central de la CIE, Vienna, 2001.
195. G. Wagner and R. M. Boynton, "Comparison of Four Methods of Heterochromatic Photometry," *Journal of the Optical Society of America* **62**:1508–1515 (1972).
196. W. Abney and E. R. Festing, "Colour Photometry," *Philosophical Transactions of the Royal Society of London* **177**:423–456 (1886).
197. W. Abney, *Researches in Colour Vision*, Longmans, Green, London, 1913, p. 418.
198. A. Dresler, "The Non-additivity of Heterochromatic Brightness," *Transactions of the Illuminating Engineering Society (London)* **18**:141–165 (1953).
199. R. M. Boynton and P. Kaiser, "Vision: The Additivity Law Made to Work for Heterochromatic Photometry with Bipartite Fields," *Science* **161**:366–368 (1968).
200. S. L. Guth, N. J. Donley, and R. T. Marrocco, "On Luminance Additivity and Related Topics," *Vision Research* **9**:537–575 (1969).
201. Y. Le Grand, "Spectral Luminosity," in *Visual Psychophysics, Handbook of Sensory Physiology*, D. Jameson and L. H. Hurvich, (eds.), Springer-Verlag, Berlin, 1972, pp. 413–433.
202. J. Pokorny, V. C. Smith, and M. Lutze, "Heterochromatic Modulation Photometry," *Journal of the Optical Society of America A* **6**:1618–1623 (1989).
203. P. Lennie, J. Pokorny, and V. C. Smith, "Luminance," *Journal of the Optical Society of America A* **10**:1283–1293 (1993).
204. CIE, *CIE 1988 2° Spectral Luminous Efficiency Function for Photopic Vision*, Publication 86, Bureau Central de la CIE, Vienna, 1990.
205. J. Pokorny and V. C. Smith, "Colorimetry and Color Discrimination," in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, (eds.), John Wiley & Sons, 1986.
206. A. Stockman, "Colorimetry," in *The Optics Encyclopedia: Basic Foundations and Practical Applications*, T. G. Brown, K. Creath, H. Kogelnik, M. A. Kriss, J. Schmit, and M. J. Weber, (eds.), Wiley-VCH, Berlin, 2003, pp. 207–226.
207. J. B. Cohen and W. E. Kappauf, "Metameric Color Stimuli, Fundamental Metamers, and Wyszecki's Metameric Blacks: Theory, Algebra, Geometry, Application," *American Journal of Psychology* **95**:537–564 (1982).

---

# COLOR VISION MECHANISMS

---

## Andrew Stockman

*Department of Visual Neuroscience  
UCL Institute of Ophthalmology  
London, United Kingdom*

## David H. Brainard

*Department of Psychology  
University of Pennsylvania  
Philadelphia, Pennsylvania*

---

### 11.1 GLOSSARY

---

**Achromatic mechanism.** Hypothetical psychophysical mechanisms, sometimes equated with the luminance mechanism, which respond primarily to changes in intensity. Note that achromatic mechanisms may have spectrally opponent inputs, in addition to their primary nonopponent inputs.

**Bezold-Brücke hue shift.** The shift in the hue of a stimulus toward either the yellow or blue invariant hues with increasing intensity.

**Bipolar mechanism.** A mechanism, the response of which has two mutually exclusive types of output that depend on the balance between its two opposing inputs. Its response is nulled when its two inputs are balanced.

**Brightness.** A perceptual measure of the apparent intensity of lights. Distinct from luminance in the sense that lights that appear equally bright are not necessarily of equal luminance.

**Cardinal directions.** Stimulus directions in a three-dimensional color space that silence two of the three “cardinal mechanisms.” These are the isolating directions for the L+M, L–M, and S–(L+M) mechanisms. Note that the isolating directions do not necessarily correspond to mechanism directions.

**Cardinal mechanisms.** The second-site bipolar L–M and S–(L+M) chromatic mechanisms and the L+M luminance mechanism.

**Chromatic discrimination.** Discrimination of a chromatic target from another target or background, typically measured at equiluminance.

**Chromatic mechanism.** Hypothetical psychophysical mechanisms that respond to chromatic stimuli, that is, to stimuli modulated at equiluminance.

**Color appearance.** Subjective appearance of the hue, brightness, and saturation of objects or lights.

**Color-appearance mechanisms.** Hypothetical psychophysical mechanisms that mediate color appearance, especially as determined in hue scaling or color valence experiments.

**Color assimilation.** The phenomenon in which the hue of an area is perceived to be closer to that of the surround than to its hue when viewed in isolation. Also known as the von Bezold spreading effect.

**Color constancy.** The tendency of objects to retain their color appearance despite changes in the spectral characteristics of the illuminant, or, more generally, despite changes in viewing context.

**Color contrast.** The change in the color appearance of an area caused by the presence of a colored surround. The color change, unlike assimilation, is usually complementary to the surround color.

**Color-discrimination mechanisms.** Hypothetical psychophysical mechanisms that determine performance in chromatic detection or discrimination tasks. Assumed in some models to correspond to cone-opponent mechanisms.

**Color spaces.** Representations of lights either in terms of the responses of some known or hypothetical mechanisms thought to underlie the perception of color (such as cone or postreceptoral mechanisms), or in terms of the projection of the lights onto stimulus-based vectors (such as monochromatic primaries or mechanism-isolating vectors).

**Color valence.** A measure of the color of a light in terms of the amount of a cancelling light required to null one of the hue sensations produced by that light. Thus, if a light appears red it is cancelled by light that appears green, and the amount of this green light is its red valence. In opponent-colors theory, color appearance depends on the relative red-green and blue-yellow valences.

**Cone contrast.** The contrast (or relative change in quantal or energy catch) presented to each cone photoreceptor:  $\Delta L/L$ ,  $\Delta M/M$ , and  $\Delta S/S$ .

**Cone contrast space.** A color space where the position along each axis represents the contrast of one cone class.

**Cone mechanisms.** Hypothetical psychophysical mechanisms, the performances of which are limited at the cone photoreceptors.

**Cone-opponent mechanism.** Hypothetical psychophysical mechanisms with opposed cone inputs.

**Derrington Krauskopf Lennie (DKL) space.** Color space, the axes of which are the stimulus strengths in each of the three cardinal mechanism directions. Closely related to the spaces proposed by Lüther<sup>85</sup> and MacLeod and Boynton.<sup>86</sup> In some accounts of this space the axes are defined in a different way, in terms of the three vectors that isolate each of the three cardinal mechanisms.

**Detection surface or contour.** Detection thresholds measured in many directions in color space form a detection surface. Confined to a plane, they form a contour. The terms threshold surface and threshold contour are synonymous with detection surface and detection contour, respectively.

**Field method.** A method in which the observer's sensitivity for detecting or discriminating a target is measured as a function of some change in context or in the adapted state of the mechanism of interest.

**First-site adaptation.** Adaptation, usually assumed to be cone-class specific, occurring at or related to the photoreceptor level.

**Habituation.** Loss of sensitivity caused by prolonged adaptation to chromatic and/or achromatic stimulus modulations, also known as contrast adaptation.

**Incremental cone-excitation space.** A color space in which the axes represent the deviations of each of the three classes of cones from a background. Deviations can be negative (decrements) as well as increments.

**Intensity.** Generic term to denote variation in stimulus or modulation strength when chromatic properties are held constant. In the particular context of modulations around a background, the vector length of a modulation may be used as a measure of intensity.

**Invariant hue.** A stimulus produces an invariant hue if that hue is independent of changes to stimulus intensity. Generally studied in the context of monochromatic stimuli.

**Isolating direction.** Direction in a color space that isolates the response of a single mechanism.

**Linear visual mechanisms.** Hypothetical mechanisms that behave linearly, usually with respect to the cone isomerization rates, but in some models with respect to the cone outputs after von Kries adaptation or contrast coding.

**Luminance.** A measure of the efficiency (or effectiveness) of lights often linked to the assumed output of the achromatic mechanism.

**Mechanism direction.** Stimulus color direction along which a specified mechanism is most sensitive. Note that the mechanism direction is not, in general, the same as the isolating direction for the same mechanism.

**Noise masking.** Threshold elevations caused by superimposing targets in noise.

**Nonlinear visual mechanisms.** Hypothetical mechanisms that behave nonlinearly either with respect to the cone inputs or with respect to their own (assumed) inputs.

**Opponent-colors theory.** A color theory that accounts for color appearance in the terms of the perceptual opposition of red and green (R/G), blue and yellow (B/Y), and dark and light (W/B).

**Pedestal effects.** Changes in sensitivity that occur when a target is superimposed on another stimulus, called the pedestal, which may have either identical or different spatio-chromatic-temporal characteristics to the target.

**Second-site desensitization.** Adaptation or sensitivity losses that act on the outputs of second-site cone-opponent and achromatic mechanisms, and thus on the combined cone signals processed by each mechanism.

**Test method.** A method in which the sensitivity for detecting or discriminating a target is measured as a function of some target parameter, such as wavelength, size, or temporal frequency.

**Threshold surface or contour.** Synonyms for detection surface or contour.

**Unique hues.** Hues that appear perceptually unmixed, such as unique blue and unique yellow (which appear neither red nor green).

**Unipolar mechanism.** A mechanism that responds to only one pole of bipolar cone-opponent excursions, thought to be produced by half-wave rectification of bipolar signals.

**Univariant mechanism.** A mechanism, in which the output varies unidimensionally, irrespective of the characteristics of its inputs.

**von Bezold spreading.** See Color assimilation.

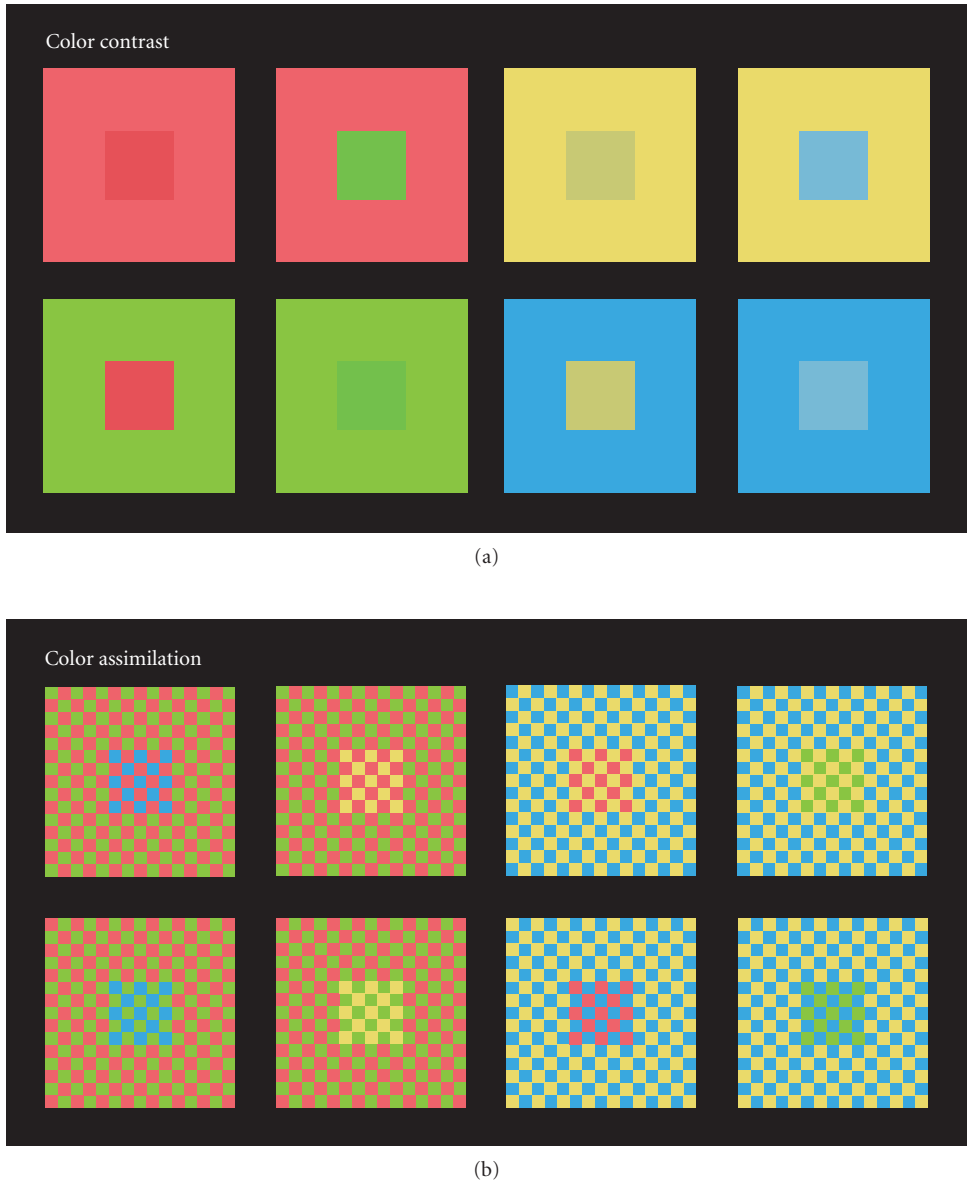
**von Kries adaptation.** Reciprocal sensitivity adjustment in response to changing light levels assumed to occur independently within each of the three cone mechanisms.

**Weber's law.**  $\Delta I/I = \text{constant}$ . The sensitivity to increments ( $\Delta I$ ) is inversely proportional to the adaptation level ( $I$ ).

## 11.2 INTRODUCTION

The first stage of color vision is now well understood (see Chap. 10). When presented in the same context under photopic conditions, pairs of lights that produce the same excitations in the long-, middle-, and short-wavelength-sensitive (L-, M-, and S-) cones match each other exactly in appearance. Moreover, this match survives changes in context and changes in adaptation, provided that the changes are applied equally to both lights. Crucially, however, while the match survives such manipulations, the shared appearance of the lights does not. Substantial shifts in color appearance can be caused both by changes in context and by changes in chromatic adaptation. The identity of lights matched in this way reflects univariance at the cone photoreceptor level, whereas their changed appearance reflects the complex activity of postreceptoral mechanisms acting on the outputs of the cone photoreceptors. Figure 1 shows examples of how color contrast and color assimilation can affect the color appearance of pairs of lights that are physically identical.

In addition to affecting color appearance, postreceptoral mechanisms play a major role in determining the discriminability of color stimuli. Indeed, measurements of color thresholds (detection and discrimination) are critical in guiding models of postreceptoral mechanisms. Models of color discrimination are also important in industrial applications, for instance, in the specification tolerances for color reproduction (see subsection “CIE Uniform Color Spaces” in Sec. 10.6, Chap. 10).



**FIGURE 1** (a) Color contrast: The pairs of smaller squares in each of the four vertical columns are physically the same, but their color appearances are very different. The differences arise because of the surrounding areas, which induce complementary color changes in the appearance of the central squares.<sup>473</sup> Comparable examples of color contrast have been produced by Akiyoshi Kitaoka,<sup>474</sup> which he attributed to Kasumi Sakai.<sup>475</sup> (b) Color assimilation or the von Bezold spreading effect:<sup>476</sup> The tiny squares that make up the checkerboard patterns in each of the four columns are identical, except in the square central areas. In those central areas, one of the checkerboard colors has been replaced by a third color. The replacement color is the same in the upper and lower patterns, but the colors of the checkers that it replaces are different. The result is that the replacement color is surrounded by a different color in the upper and lower patterns. Although the replacement color is physically the same in each column, it appears different because of the color of the immediately surrounding squares. Unlike color contrast, the apparent color change is toward that of the surrounding squares.

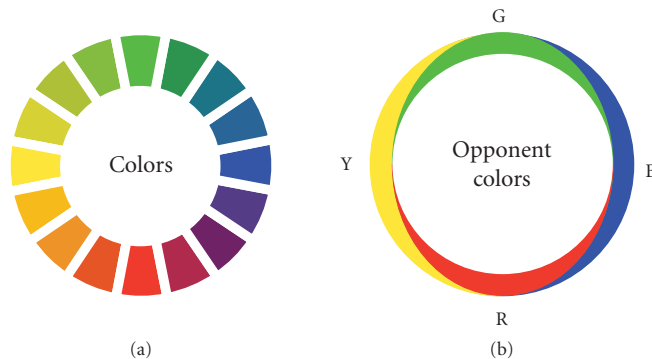
## The Mechanistic Approach

This chapter is about color vision after the photoreceptors. In the development, we adopt a mechanistic approach. The idea is to model color vision as a series of stages that act on the responses of the cones. Within the mechanistic approach, the central questions are: how many stages are needed, what are the properties of the mechanisms at each stage, and how are the mechanisms' outputs linked to measured performance? We focus on psychophysical (perceptual) data. Nonetheless, we are guided in many instances by physiological and anatomical considerations. For reviews of color physiology and anatomy, see, for example, Gegenfurtner and Kiper,<sup>1</sup> Lennie and Movshon<sup>2</sup>, and Solomon and Lennie.<sup>3</sup> A useful online resource is Webvision at <http://webvision.med.utah.edu/>.

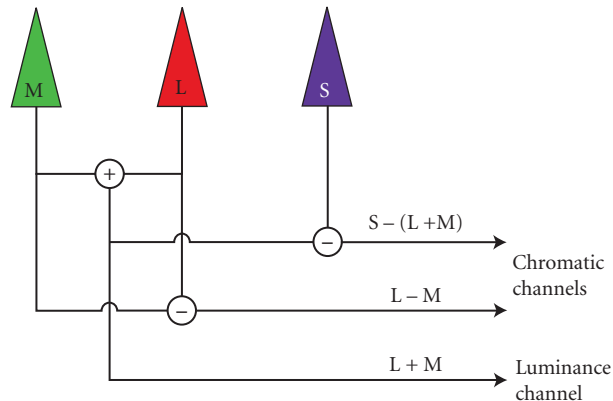
The distinction between color encoded at the photoreceptors and color encoded by postreceptoral mechanisms was anticipated by two theories that have dominated color vision research since the late nineteenth century. First, in the Young-Helmholtz trichromatic theory,<sup>4,5</sup> color vision is assumed to depend on the univariant responses of the three fundamental color mechanisms (see Chap. 10). Color vision is therefore trichromatic. Trichromacy allows us to predict which mixtures of lights *match*, but it does not address how those matches *appear*, nor the discriminability or similarity of stimuli that do not match.

Second, in Hering's<sup>6,7</sup> opponent colors theory, an early attempt was made to explain some of the phenomenological aspects of color appearance, and, in particular, the observation that under normal viewing conditions some combinations of colors, such as reddish-blue, reddish-yellow, and greenish-yellow, are perceived together, but others, such as reddish-green or yellowish-blue, are not. This idea is illustrated in Fig. 2. Hering proposed that color appearance arises from the action of three signed mechanisms that represent opposing sensations of red versus green, blue versus yellow, and light versus dark.<sup>6,7</sup> A consequence of this idea is that opposing or opponent pairs of sensations are exclusive, since they cannot both be simultaneously encoded. In this chapter, we will use the term "color-appearance mechanisms" to refer to model-constructs designed to account for the appearance of stimuli, and in particular the opponent nature of color appearance.

Early attempts to reconcile trichromacy with the opponent phenomenology of color appearance suggested that the color-appearance mechanisms reflect a postreceptoral stage (or "zone") of color processing that acts upon the outputs of the three Young-Helmholtz cone mechanisms. Modern versions of the two-stage theory explicitly incorporate the cone's characteristics as a first stage as well as



**FIGURE 2** Hering's opponent-colors diagram. A diagrammatic representation of opponent-colors theory. The ring on the left (a) shows a range of colors changing in small steps from green at the top clockwise to blue, red, yellow, and back to green. The ring on the right (b) shows the hypothetical contributions of each of the color-opponent pairs [red (R) vs. green (G), and blue (B) vs. yellow (Y)] to the appearance of the corresponding colors in (a). In accordance with opponent-colors theory, the opposed pairs of colors are mutually exclusive. (Redrawn from Plate 1 of Ref. 7).



**FIGURE 3** Model of the early postreceptoral stages of the visual system. The signals from the three cone types, S, M, and L, are combined to produce an achromatic or luminance channel,  $L+M$ , and two cone-opponent channels,  $L-M$  and  $S-(L+M)$ . Note that there is assumed to be no S-cone input to the luminance channel. (Based on Fig. 7.3 of Ref. 15).

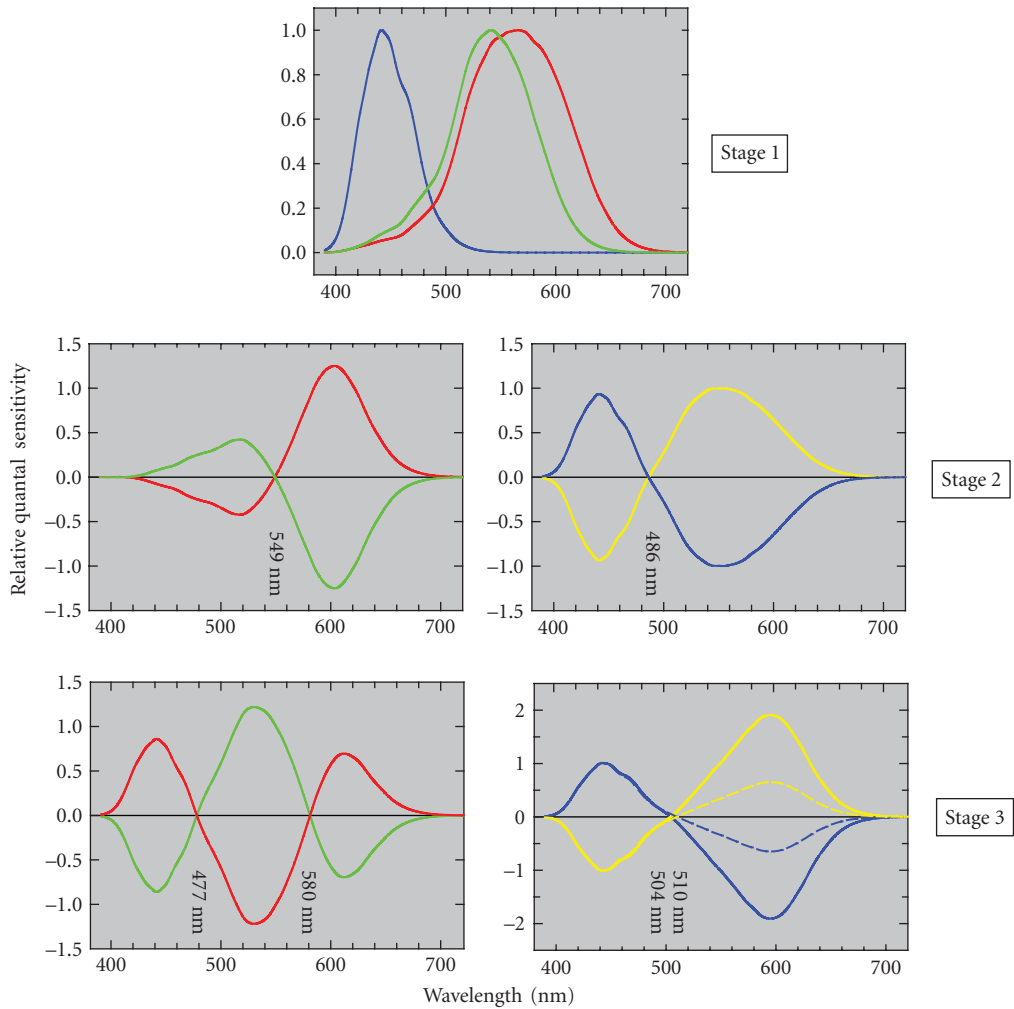
a second stage at which signals from the separate cone classes interact (e.g., Refs. 8–14). A familiar version of the two-zone model from Boynton<sup>15</sup> with chromatic,  $L-M$  and  $S-(L+M)$ , and achromatic,  $L+M$ , postreceptoral mechanisms is shown in Fig. 3.

Interestingly, the particulars of many modern two-stage models were formulated not to account for color appearance, but rather to explain threshold measurements of the detection and discrimination of visual stimuli. As Fig. 3 indicates, the opponent mechanisms in these models take on a simple form, represented as elementary combinations of the outputs of the three cone classes. We refer to opponent mechanisms that are postulated to explain threshold data as “color-discrimination mechanisms,” to distinguish them from color-appearance mechanisms postulated to explain appearance phenomena. Here we use the omnibus term color-discrimination to refer both to detection (where a stimulus is discriminated from a uniform background) and discrimination (where two stimuli, each different from a background, are discriminated from each other.)

The distinction between color-appearance and color-discrimination mechanisms is important, both conceptually and in practice. It is important conceptually because there is no a priori reason why data from the two types of experiments (appearance and threshold) need be mediated by the same stages of visual processing. Indeed, as we will see below, the theory that links measured performance to mechanism properties is quite different in the two cases. The distinction is important in practice because the mechanism properties derived from appearance and discrimination data are not currently well reconciled.

The discrepancy between color-discrimination and color-appearance mechanisms has been commented on by recent authors,<sup>11,14,16–22</sup> but the discrepancy is implicit in early versions of the three-stage Müller zone theories,<sup>23,24</sup> Judd’s version of which<sup>24</sup> was discussed again some years later in a more modern context (see Fig. 6. of Ref. 25). It is remarkable that models with separate opponent stages for the two types of data were proposed well before the first physiological observation of cone opponency in fish<sup>26</sup> and primate.<sup>27</sup>

Figure 4 illustrates a modern version of Judd’s three-stage Müller zone theory, which is described in more detail in the subsection “Three Stage Zone Models” in Sec. 11.6. The figure shows the spectral sensitivities of each of the three stages. The spectral sensitivities of Stage 1 correspond to the cone spectral sensitivities of Stockman and Sharpe,<sup>28</sup> those of Stage 2 to the spectral sensitivities of



**FIGURE 4** Version of the three-stage Müller zone model with updated spectral sensitivities. The panels show the assumed spectral sensitivities of the color mechanisms at Stages 1 (upper panel), 2 (middle panels), and 3 (lower panels). *Stage 1:* L- (red line), M- (green line), and S- (blue line) cone fundamental spectral sensitivities.<sup>28</sup> *Stage 2:* L-M (red line), M-L (green line), S-(L+M) (blue line), and (L+M)-S (yellow line) cone-opponent mechanism spectral sensitivities. *Stage 3:* R/G (red line), G/R (green line), B/Y (blue line), Y/B (yellow line) color-opponent spectral sensitivities. Our derivation of the cone-opponent and color-opponent spectral sensitivities is described in the subsection “Three-Stage Zone Models” in Sec. 11.6. The dashed lines in the lower right panel are versions of the B/Y and Y/B color-opponent spectral sensitivities adjusted so that the Y and B spectral sensitivity poles are equal in area. The wavelengths of the zero crossings of the Stage 2 and Stage 3 mechanisms are given in the figure. The spectral sensitivities of the achromatic mechanisms have been omitted.



color-discrimination mechanisms as suggested by threshold data, and those of Stage 3 to the spectral sensitivities of color-appearance mechanisms as suggested by appearance data.

Figure 4 sets the scene for this chapter, in which we will review the theory and data that allow derivation of the properties of color-discrimination and color-appearance mechanisms, and discuss the relation between the two. According to some commentators, one of the unsolved mysteries of color vision is how best to understand the relation between mechanisms referred to as Stages 2 and 3 of Fig. 4.

## Nomenclature

One unnecessary complication in the literature is that discrimination and appearance mechanisms are frequently described using the same names. Thus, the terms red-green (R/G), blue-yellow (B/Y), and luminance are often used to describe both types of mechanisms. We will attempt in this chapter to maintain a distinct nomenclature for distinct mechanisms.

It is now accepted that cones should be referred to as long-, middle-, and short-wavelength-sensitive (L-, M-, and S-), rather than red, green, and blue, because the color descriptions correspond neither to the wavelengths of peak cone sensitivity nor to the color sensations elicited by the excitation of single cones.<sup>30</sup> However, it is equally misleading to use color names to refer to color-discrimination mechanisms. Stimulation of just one or other side of such a mechanism does not necessarily give rise to a simple color sensation. Indeed, current models of opponent color-discrimination mechanisms have the property that modulating each in isolation around an achromatic background produces in one case a red/magenta to cyan color variation and in the other a purple to yellow/green variation.<sup>31,32</sup> Consequently, the perception of blue, green, and yellow, and to a lesser extent red, requires the modulation of *both* cone-opponent discrimination mechanisms (see subsection “Color Appearance and Color Opponency” in Sec. 11.5). We therefore refer to chromatic color-discrimination mechanisms according to their predominant cone inputs: L–M and S–(L+M). Although this approach has the unfortunate consequence that it neglects to indicate smaller inputs, usually from the S-cones (see subsection “Sensitivity to Different Directions of Color Space” in Sec. 11.5.), it has the advantage of simplicity and matches standard usage in much of the literature. We refer to the nonchromatic color discrimination mechanism as L+M. Note also that this nomenclature is intended to convey the identity and sign of the predominant cone inputs to each mechanism, but not the relative weights of these inputs.

In contrast, the perception of pure or “unique” red, green, yellow, and blue is, by construction of the theory, assumed to result from the responses of a single opponent color-appearance mechanism, the response of the other mechanism or mechanisms being nulled or in equilibrium (see subsection “Opponent-Colors Theory” in Sec. 11.5). We refer to opponent color-appearance mechanisms as R/G and B/Y, according to the color percepts they are assumed to generate. We refer to the nonopponent appearance mechanism as brightness.

## Guiding Principles

Behavioral measurements of color vision reflect the activity of an inherently complex neural system with multiple sites of processing that operate both in series and in parallel. Moreover, these sites are essentially nonlinear. The promise of the mechanistic approach lies in two main areas. First, in terms of developing an overall characterization of postreceptoral color vision, the hope is that it will be possible to identify broad regularities in the behavior of the system that can be understood in terms of models that postulate a small number of relatively simple mechanism constructs. Second, in terms of using psychophysics to characterize the behavior of particular neural sites, and to link behavior to physiology, the hope is that specific stimulus conditions can be identified for which the properties of the site of interest dominate the measured performance. In the conceptual limit of complexity, where a mechanistic model explicitly describes the action of every neuron in a given visual pathway, these models can, in principle, predict performance. But as a practical matter, it

remains unclear the degree to which parsimonious mechanistic models derived from behavioral measurements will succeed. Given the complexity of the underlying neural system, it is apparent that the mechanistic approach is ambitious.

In this context, several points are worth bearing in mind. First, the concept of a psychophysical mechanism will have the most utility when it can be shown to have an existence that extends beyond the particular stimulus and task conditions from which it was derived. Thus, we regard a mechanism as a theoretical construct whose usefulness depends on the range of data it can explain parsimoniously. This is a broader consideration than those sometimes used to determine the value of a mechanism.<sup>33,34</sup> In Secs. 11.3 and 11.4, we review two broad mechanism concepts that satisfy this criterion: opponency and adaptation.

Second, while some psychophysical techniques may emphasize the contribution of particular stages of the system, it is, with the exception of color matching, a simplification to ignore the contributions of earlier and/or later stages. For example, an often made but usually implicit assumption is that the cone quantal absorption rates are transmitted directly to postreceptoral mechanisms, as if the photoreceptors had no role other than to pass on a linear copy of their inputs. An interesting future direction for mechanistic models is to account for interactions between different stages of processing more fully.

Finally, in spite of our concerns, we have written this chapter primarily from the point of view of psychophysics. We readily acknowledge that over the past 50 years, results from molecular genetics, anatomy, and physiology have helped propel many psychophysical theories from intelligent speculation to received wisdom, particularly in the cases of cone spectral sensitivities and early retinal visual processing. Physiology and anatomy have also provided important evidence about color coding at different levels of the early visual system (e.g., Refs. 35–37), crucial information that is hard to obtain psychophysically. Obversely, as the increasing complexity of the visual system at higher levels of processing begins to limit the utility of a mechanistic psychophysical approach, it also constrains how much can be understood from a knowledge of the properties of the component neurons in the processing chain, the majority of which have yet to be characterized.

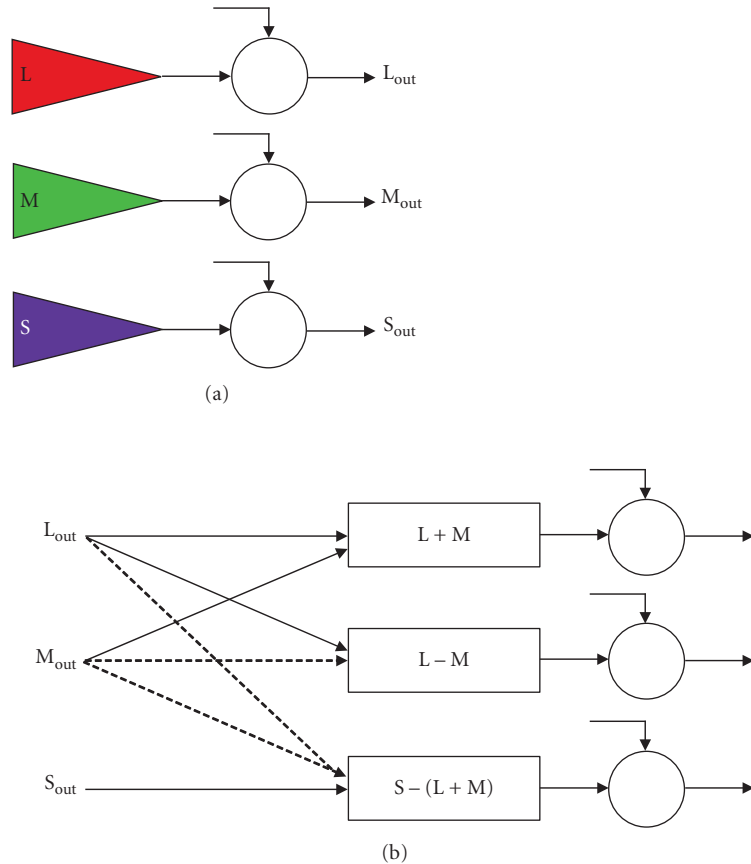
## Chapter Organization

The rest of this chapter is organized as follows: In the next two sections, we describe how mechanism concepts allow us to understand important features of both color discrimination and color appearance. Our treatment here is illustrative. We review a small selection of experimental data, and emphasize the logic that leads from the data to constraints on the corresponding models. In Sec. 11.3 we discuss discrimination data; in Sec. 11.4 we turn to appearance data. Two broad mechanistic concepts emerge from this initial review: those of color opponency and of adaptation. The initial review provides us with a basic model for postreceptoral color vision. Although this model is clearly oversimplified, it serves as the point of departure for much current work, and understanding it is crucial for making sense of the literature. The remainder of the chapter (Secs. 11.5 and 11.6) is devoted to a discussion of advanced issues that push the boundaries of the basic model.

## 11.3 BASICS OF COLOR-DISCRIMINATION MECHANISMS

### What Is a Mechanism?

Given the central role of the mechanism concept in models of color vision, one might expect that this concept is clearly and precisely defined, with a good consensus about its meaning. Our experience, however, is that although there are precise definitions of what constitutes a mechanism,<sup>33,34</sup> these are generally specific to a particular model and that the term is often used fairly loosely. In keeping with this tradition, we will proceed to discuss mechanisms through examples without attempting a rigorous definition.



**FIGURE 5** Basic mechanisms. (a) First-stage cone mechanisms: L, M, and S. The cone outputs are subject to some form of gain control (open circles), which can, in principle, be modified by signals from the same or from different cone mechanisms. (b) Second-stage color-discrimination mechanisms:  $L+M$ ,  $L-M$ , and  $S-(L+M)$ . The inputs to these mechanisms are the adapted cone outputs from the cone mechanisms. As with the cone mechanisms, the outputs of the second-stage mechanism are subject to some gain control, which can be modified by signals from the same or from different second-stage mechanisms. Dashed arrows indicate inhibitory outputs.

Figure 5 illustrates the broad mechanism concept. Figure 5a shows the L, M, and S cones. Each of these cones is a mechanism and satisfies several key properties. First, we conceive of each cone mechanism as providing a spatiotemporal representation of the retinal image, so that the mechanism output may be regarded as a spatial array that changes over time. Second, at a single location and time, the output of each cone mechanism is univariate and conveys only a single scalar quantity. This means that the output of a single cone mechanism confounds changes in the relative spectrum of the light input with the overall intensity of the spectral distribution. Third, the relation between mechanism input and output is subject to adaptation. This is indicated in the figure by the open circle along the output pathway for each cone mechanism. Often the adaptation is characterized as a gain control, or a gain control coupled with a term that subtracts steady-state input (e.g., Refs. 38 and 39).

Key to the mechanism approach is that the behavior of the mechanism can be probed experimentally with adaptation held approximately fixed. This may be accomplished, for example, by presenting brief, relatively weak, flashes on spatially uniform backgrounds. In this case, the state of adaptation is taken to be determined by the background alone; that is, it is assumed that the weak flashes do not measurably perturb the state of adaptation and thus that the state of adaptation is independent of the flash.

The arrows pointing into each circle in Fig. 5 indicate that the signals that control the state of adaptation can be quite general, and need to be specified as part of the model. In the classical (von Kries) mechanism concept, however, these signals are assumed to arise entirely within the array of signals from the same cone mechanism.<sup>33</sup> For example, the gain (or adaptation) applied to the L-cone signals at a location would be taken to be determined entirely by the array of L-cone responses, and be independent of the responses of M and S cones. Adaptation that acts on the output of the cones is referred to as “first-site” adaptation.

Figure 5*b* shows a simple model of three second-stage color-discrimination mechanisms with the same basic wiring diagram as Fig. 3. These mechanisms have a similar structure to the cone mechanisms, in the sense that they are assumed to provide a spatial array of univariate responses. In addition, as with the cone mechanisms, the second-stage mechanisms can adapt. An important difference between these second-stage mechanisms and the cone mechanisms is the nature of their input. Where the cone mechanisms transduce light into a neural signal, the second-stage mechanisms take the adapted cone responses as their input. As shown in the Fig. 5, each of the postreceptoral mechanisms can be thought of as computing a weighted combination of cone signals. The three mechanisms are labeled by the manner in which they combine cone inputs. The first mechanism, (L+M), which is often referred to as the luminance mechanism, adds inputs from the L and M cones. The second mechanism, (L–M), takes a difference between L and M cone signals. And the third mechanism, S–(L+M), takes a difference between S-cone signals and summed L- and M-cone signals.

Although not necessary as part of the mechanism definition, making the input to the mechanism be a linear function of the output of the preceding stage simplifies the model, and enables it to make stronger predictions; whether the cone combinations are actually linear is an empirical question. Adaptation that acts on the outputs of the second-stage postreceptoral mechanisms is usually referred to as “second-site” adaptation or desensitization.

The cone mechanisms used in color vision models represent the action of a well-defined neural processing stage (the cone photoreceptors). The connection between postreceptoral mechanisms and neurons is not as tight. Although the cone inputs to different classes of retinal ganglion cells are similar to those used in models, these models often do not incorporate important features of real ganglion cells, such as their spatial receptive field structure (but see subsection “Multiplexing Chromatic and Achromatic Signals” in Sec. 11.5), cell-to-cell variability in cone inputs, and the fact that ganglion cells come in distinct ON and OFF varieties that each partially rectify their output (but see “Unipolar vs. Bipolar Chromatic Mechanisms” in Sec. 11.6).

## Psychophysical Test and Field Methods

Broadly speaking, two principal psychophysical techniques have been used to investigate the properties of color-discrimination mechanisms.<sup>40–42</sup> In the “test” (or “target”) sensitivity method, the observer’s sensitivity for detecting or discriminating a target is measured as a function of some target parameter, such as its wavelength, size, or temporal frequency. As noted above, implicit in the use of this method is the assumption that the presentation of the target does not substantially alter the properties or sensitivity of the detection mechanism, so that targets are usually presented against a background and kept near visual threshold.

In the “field” sensitivity method, the observer’s sensitivity for detecting or discriminating a target is measured as a function of some change in the adaptive state of the mechanism. Field methods complement test methods by explicitly probing how contextual signals control adaptation. On the assumption that the control of adaptation occurs solely through signals within the mechanism mediating detection, the spectral properties of that mechanism can be investigated by, for example,

superimposing the target on a steady adapting field and changing the adapting field chromaticity and/or radiance, by habituating the observer to backgrounds temporally modulated in chromaticity and/or luminance just prior to the target presentation, or by superimposing chromatic and/or luminance noise on the target.

Both test and field methods have obvious limitations. In the test method, it is difficult to ensure that a target is detected by a single mechanism when the target parameters are varied, with the result that multiple mechanisms typically mediate most sets of test sensitivity measurements. As we discuss below, assigning different portions of chromatic detection contours to different mechanisms is problematic (see subsection “Sensitivity to Different Directions of Color Space” in Sec. 11.5). In the field method, it is often easier to ensure that a target is detected by a single mechanism. However, the assumption that adaptation is controlled entirely by signals from within the mechanism mediating detection is a strong one, and interpretation of results becomes much more complicated under conditions where this assumption is not secure. More generally, without an explicit and accurate model about the properties of the mechanism, both test and field sensitivity data may be uninterpretable.

## How Test Measurements Imply Opponency

In the next three sections, we introduce color-discrimination mechanisms. These are characterized by opponent recombination of cone signals at a second site (as shown in Fig. 5b), by adaptation in cone specific pathways (first-site adaptation, Fig. 5a), and by adaptation after the opponent recombination (second-site adaptation, Fig. 5b).

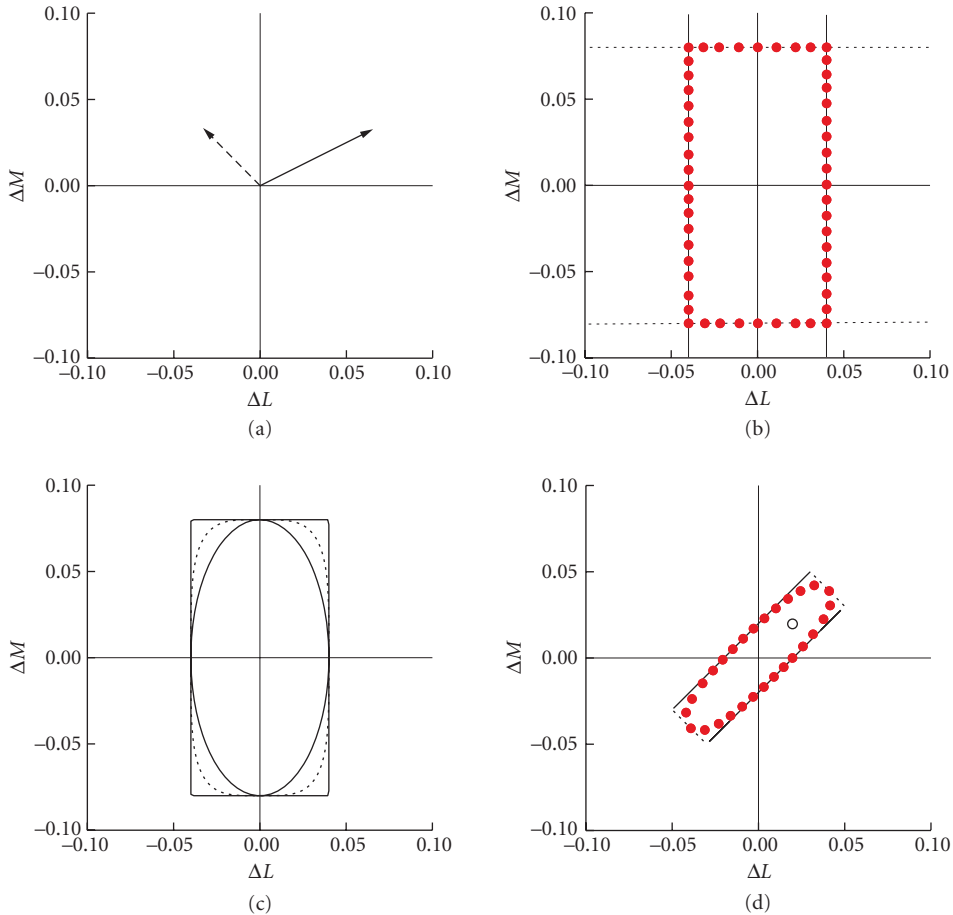
We begin with evidence from test sensitivity measurements that some sort of opponent recombination occurs. In the canonical test sensitivity experiment, a small test stimulus is presented against a uniform background. We denote the L-, M-, and S-cone excitations of the background by  $L_b$ ,  $M_b$ , and  $S_b$ . If we fix the background, the test stimulus may be characterized by how much the cone excitations it produces deviate from the background. Denote these deviations by  $\Delta L$ ,  $\Delta M$ , and  $\Delta S$ , so that the overall cone excitations of the test are given by  $L = L_b + \Delta L$ ,  $M = M_b + \Delta M$ , and  $S = S_b + \Delta S$ . Note that the deviations  $\Delta L$ ,  $\Delta M$ , and  $\Delta S$  may be positive or negative.

For any background, we can consider a parametric family of test stimuli whose L-, M-, and S-cone deviations are in the same proportion. That is, we can define a *test color direction* by a triplet of deviations  $\Delta L_d$ ,  $\Delta M_d$ , and  $\Delta S_d$  normalized so that  $\sqrt{\Delta L_d^2 + \Delta M_d^2 + \Delta S_d^2} = 1$ . Test stimuli that share the same color direction have the form  $\Delta L = c\Delta L_d$ ,  $\Delta M = c\Delta M_d$ , and  $\Delta S = c\Delta S_d$ . We refer to the constant  $c$  as the *intensity* of the test stimulus along the given color direction. Figure 6a illustrates the test color direction and intensity concept for two vectors in the  $\Delta L$ ,  $\Delta M$  plane.

If the background and color direction of a test stimulus are held fixed, an experimenter can vary the test intensity and determine the psychophysical threshold for detection. This is the lowest intensity at which the observer can just see the test, and its value can be found experimentally using a variety of procedures.<sup>43</sup> The experiment can then be repeated for different choices of color direction, and the threshold intensity can be determined for each one.

Figure 6b plots one way in which the data from a threshold experiment might come out. For simplicity, we assume that  $\Delta S_d$  was set to zero, and show data only in the  $\Delta L$ ,  $\Delta M$  plane. Each point in the plot represents the results of a threshold measurement for one color direction. The set of threshold points together trace out a “detection contour” (or “threshold contour”), since each point on the contour leads to an equal level of detection performance (the “threshold”). The two points that lie directly on the  $\Delta L$  axis represent threshold for increments and decrements that drive only the L cones and leave the M cones silent; that is, these targets produce zero change in the M cones. Similarly two points that lie directly on the  $\Delta M$  axis represent increments and decrements that produce only changes in the M cones. The other points on the contour show thresholds obtained when L- and M-cone signals are covaried in various ratios.

Understanding how threshold contours inform us about color-discrimination mechanisms is a key idea in the theory we present in this chapter. It is useful to begin by asking how the contour would come out if the visual system had a single color-discrimination mechanism consisting of, say,



**FIGURE 6** Basic thresholds. (a) Two vectors plotted in the  $\Delta L$ ,  $\Delta M$  plane that represent two lights of different color directions and different intensities. (b) Incremental and decremental thresholds determined by an L-cone mechanism (vertical black lines) or by an M-cone mechanism (black, horizontal dotted lines). The points define the contour expected if threshold is determined independently and with no interactions between L- and M-cone detection mechanisms. (c) The joint detection contours when the L- and M-cone detection mechanisms show different degrees of summation. The inner, middle, and outer contours are for summation exponents of  $k = 2$ , 4, and 1000, respectively. (d) Idealized detection contours (red solid points) for thresholds determined by chromatic L–M (solid lines at  $45^\circ$ ) and achromatic L+M (dotted lines at  $-45^\circ$ ) mechanisms. The single open circle shows the threshold expected in the  $\Delta L_d = \Delta M_d$  direction if the threshold was determined by the cone mechanisms (i.e., by the thresholds along the axes  $\Delta L_d = 0$  and  $\Delta M_d = 0$ ).

only the L cones. In this case, only the  $\Delta L$  component of the test modulation would affect performance, and the threshold contour would consist of two lines, both parallel to the  $\Delta M$  axis. One line would represent threshold for incremental tests with a positive  $\Delta L$  value, while the other would represent threshold for decremental tests with a negative  $\Delta L$  value. This hypothetical detection contour is shown as two solid vertical lines in Fig. 6b. If there were no other mechanisms, the threshold contour would simply continue along the extensions of the solid vertical lines. The contour would not be closed, because modulations along the M-cone isolating direction produce  $\Delta L$  values of zero and thus are invisible to the L-cone mechanism.

We can also consider a visual system with only an M-cone discrimination mechanism. By reasoning analogous to that used for the L-cone case above, the threshold contour for this system would consist of the two horizontal dotted lines also shown in the figure.

Finally, we can consider a system with independent L- and M-cone mechanisms and a threshold determined when either its  $\Delta L$  or  $\Delta M$  component reached the threshold for the corresponding mechanism. This would lead to a closed rectangular detection contour formed by the intersecting segments of the solid vertical and dotted horizontal lines shown in Fig. 6b. The threshold data plotted as solid red circles in the panel correspond to this simple model.

When signals from both L- and M-cone mechanisms mediate detection, the measured detection contour would be expected not to reach the corners of the rectangular contour shown in Fig. 6b. This is because even if the L- and M-cone mechanisms are completely independent, detection is necessarily probabilistic, and in probabilistic terms the likelihood that L, M, or both L and M together will signal the test is greater than the likelihood of either cone mechanism doing so alone even when the mechanisms are independent, especially when L and M are both near threshold. This “probability summation” will reduce thresholds in the corners where L and M signals are of similar detectability, rounding them. A number of simple models of how mechanisms outputs might combine in the presence of neural noise predict such rounding, with the exact shape of the predicted contour varying across models.

A convenient parametric form for a broad class of summation models is that test threshold is reached when the quantity  $\Delta = \sqrt[k]{\sum_{i=1}^N |\Delta_i|^k}$  reaches some criterion value. In this expression,  $N$  is the number of mechanism whose outputs are being combined,  $\Delta_i$  is the output of the  $i$ th mechanism, and the exponent  $k$  determines the form of summation. When  $k = 2$ , the quantity  $\Delta$  represents the Euclidean vector length of the mechanism outputs. When  $k \rightarrow \infty$ ,  $\Delta$  represents the output of the mechanism whose output is greatest.

Figure 6c shows expected threshold contours for the output of the L and M mechanisms for three values of  $k$ . The outer rectangular contour (solid lines) shows the contour for  $k = 1000$ . Here there is effectively no summation, and the contour has the rectangular shape shown in Fig. 6b. The inner contour (solid line) shows the case for  $k = 2$ . Here the contour is an ellipse. The middle contour (dotted line) shows the case for  $k = 4$ . Determining the value of  $k$  that best fits experimental data is a topic of current interest, as the answer turns out to have important implications for how to reason from sensitivity contours to mechanism properties (see subsection “Sensitivity to Different Directions of Color Space” in Sec. 11.5). Here, however, the key point is that we expect actual experimental data to show more rounded threshold contours than the one depicted in Fig. 6b. Under conditions where sensitivity is determined by the output of the L- and M-cone mechanisms, the expected shape of the threshold contour is a closed form whose major axes are aligned with the  $\Delta L$  and  $\Delta M$  axes.

Figure 6d shows an idealized representation of how the data actually come out when thresholds are measured against a neutral background (see also Fig. 20). Actual data deviate clearly from the predictions shown in Fig. 6c, which were based on the assumption that the color-discrimination mechanisms are the responses of the L and M cones. Instead, the data lie near an elongated ellipsoid whose axes are rotated intermediate to the  $\Delta L$  and  $\Delta M$  axes. The deviations from the predictions of the Fig. 6c are large and robust. In particular, note the location of the open circle in the figure. This circle shows an upper bound on threshold in the  $\Delta L_d = \Delta M_d$  color direction; if the color-discrimination mechanisms were the L and M cones and they operated without summation, then threshold would be given by the open circle. If there was also summation between L- and M-cone mechanisms, then threshold would lie between the open circle and the origin, with the exact location depending on the degree of summation. Thresholds in this direction vastly exceed the bound shown by the open circle. This observation demonstrates unequivocally that the outputs of different cone mechanisms must be postreceptorally recombined.

A natural interpretation of the threshold contour shown in Fig. 6d is indicated by the rotated rectangle shown on the plot. The parallel solid lines represent the threshold contour that would be observed if detection were mediated by a single mechanism that computed as its output the difference between the L- and M-cone excitations to the test, and if threshold depended on  $|\Delta L - \Delta M|$  reaching some criterion threshold level. Similarly, the dotted lines represent the threshold contours of a mechanism that sums the L- and M-cone excitations of the test. The idealized threshold data



shown is thus consistent with the first two color-discrimination mechanisms illustrated in Fig. 5*b*, with some amount of summation between mechanism outputs accounting for the quasi-ellipsoidal shape of the overall contour.

Logic similar to that shown, applied to data where  $\Delta S$  is varied, leads us to postulate a third color-discrimination mechanism whose output is given by a weighted opposition of  $\Delta S$  on the one hand and  $\Delta L$  and  $\Delta M$  on the other. Figure 21, later, shows detection contours in the equiluminant plane partially determined by the S-(L+M) mechanism.

Although it is straightforward to predict detection contours if the spectral properties of the underlying detection mechanisms and the interactions between them are known, it is harder, as we shall see below, to infer unequivocally the exact mechanism properties (e.g., the exact weights on signals from each cone class) from the contours.

## First-Site Adaptation

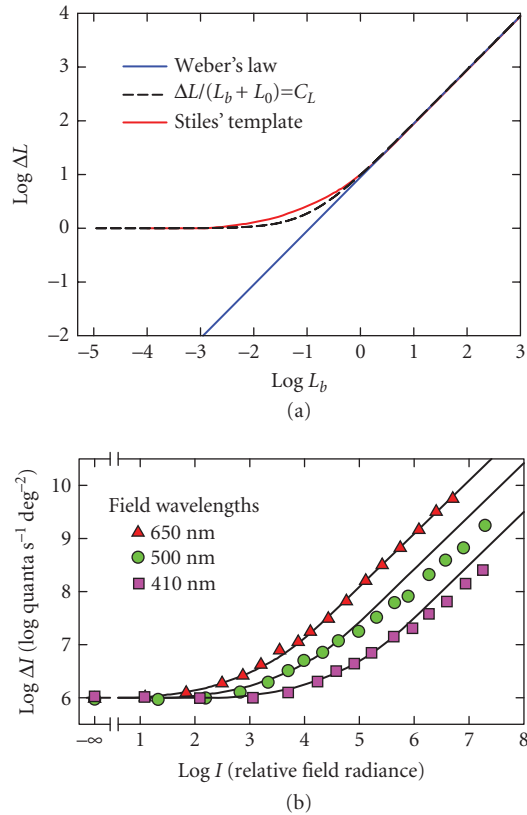
*Weber's law and contrast coding* The measurements described in the previous section assess sensitivity under conditions where the state of adaptation was held fixed. We now turn to how input signals from the cones depend on the conditioning or adapting background. The key idea here is that the cone excitations are converted to a contrast representation. Here we use L cones as an example and take the L-cone contrast of the test,  $C_L$ , to be given by  $C_L = \Delta L/L_b$ , where  $\Delta L$  is (as above) the difference between the L-cone excitations produced by the test and background, and  $L_b$  is the L-cone excitation produced by the background. Similar expressions apply for the M and S cones.

The conversion of raw cone excitations (photoisomerizations) to a contrast code implies an important adaptive function. The range of cone excitation rates encountered in the environment can be greater than  $10^6$ . This range greatly exceeds the limitations imposed by the individual neurons in the visual pathway, many of which have dynamic ranges of no more than about  $10^2$  from the level of noise—their spontaneous firing rate in the absence of light stimulation—to their response ceiling.<sup>44,45</sup> If there were no adaptation, the cone visual response would often convey no useful information, either because it was too small to rise above the noise at the low end of the stimulus range or because it was saturated at the maximum response level and thus unable to signal changes. For adaptation to protect the cone-mediated visual system from overload as the light level increases, the primary mechanisms of sensitivity regulation are likely to be early in the visual pathways, most likely in the cone photoreceptors themselves.<sup>46,48</sup> Indeed, the molecular mechanisms of adaptation acting within the photoreceptor are now fairly well understood.<sup>49–52</sup>

The idea that signals leaving the cones are effectively converted to a contrast code makes a specific prediction about how thresholds should depend on the background. Suppose we consider test stimuli that only stimulate the L cones, and that we manipulate the L-cone component of the background,  $L_b$ . Because of the cone-specific adaptation, increasing  $L_b$  will produce a proportional decrease in the contrast signal fed to the postreceptoral detection mechanisms. On the assumption that the noise limiting discrimination behavior remains constant across the change of background (see “Sites of Limiting Noise” in Sec. 11.3), this decrease means that the differential signal,  $\Delta L$ , required to bring a test stimulus to threshold will increase in proportion to  $L_b$ . This sort of behavior is often described as obedience to Weber's law; predictions of Weber's law models are shown by the blue line in Fig. 7*a*. Figure 7*a* shows a log-log plot of increment threshold  $\Delta L$  against background component  $L_b$ . The predicted thresholds increase with a slope of unity on the log-log plot.

A feature of contrast-coding predictions that is clearly not realistic is that the threshold  $\Delta L$  should decrease toward zero in complete darkness. A generalized and more realistic form of contrast coding is given by  $C_L = \Delta L/(L_b + L_o)$ , where  $L_o$  is a constant. The dashed line in Fig. 7*a* shows a prediction of this form, which has the feature that thresholds approach a constant nonzero value as  $L_b$  decreases below  $L_o$ . The value of  $L_o$  may be thought of as the hypothetical excitation level—sometimes called a “dark light”—produced internally within the visual system in the absence of light that limits performance at low background levels.





**FIGURE 7** Weber's law and increment thresholds. (a) Predictions of three first-site adaptation models for test lights detected by an L-cone mechanism as a function of background intensity. (b) Increment threshold data attributed to the L-cone mechanism ( $\pi_3$ , see subsection "Stiles'  $\pi$  Mechanisms" in Sec. 11.5) replotted from Fig. 2 (Observer: EP) of Sigel and Pugh<sup>54</sup> obtained using a 200-ms duration, 667-nm target presented on background fields of 650 nm (red triangles), 500 nm (green circles), and 410 nm (purple squares). The solid lines aligned with each data set at low background radiances are Stiles' standard template [Table 1(7.4.3) of Ref. 53]. The horizontal position of the data is arbitrary.

Stiles' template shape [see Table 1(7.4.3) of Ref. 53], shown as the red line in Fig. 7a, has a form similar to that predicted by generalized contrast coding (dashed line). This template was derived to account for increment threshold data obtained under a variety of conditions (see subsections "Sensitivity to Spectral Lights" and "Stiles'  $\pi$  Mechanisms" in Sec. 11.5). Figure 7b shows increment threshold data measured by Sigel and Pugh,<sup>54</sup> for three different background wavelengths under conditions where the L-cone mechanism is thought to dominate detection. These data fall approximately along Stiles' template, although deviations are clear at high background radiances for two of the three spectral backgrounds shown. The deviations are consistent with contributions of postreceptoral mechanisms to detection performance.<sup>54,55</sup> Data like these speak to the difficulties of distinguishing the contributions of different mechanisms from even the simplest data. Stiles, for example, accounted for the same sort of deviations by proposing an additional cone mechanism.<sup>33</sup>

*The signals reaching the second site* Although increment threshold data suggest a roughly Weber-type gain control, they do not by themselves determine whether the signal transmitted to the second site is of the contrast form  $\Delta L/(L_b + L_o)$ , as we have written above, or of the simpler form  $L/(L_b + L_o)$ , where to recap  $L = \Delta L + L_b$ . We cannot, in other words, determine from detection thresholds alone whether the adapted signal generated by the background is partially or wholly subtracted from the signal transmitted to the postreceptoral mechanisms.

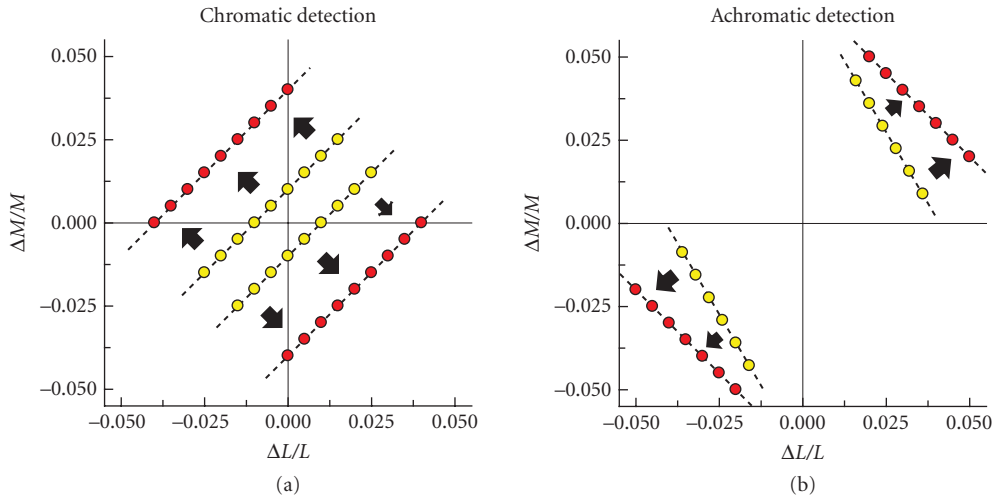
To understand the ambiguity, define the gain  $g$  as  $g = 1/(L_b + L_o)$ . In a gain control only model (with no background subtraction), the signal transmitted from the L cones in response to the background plus test would be  $L' = gL$ , where  $L$  represents the combined cone excitations to the background plus test. Similarly, under this model, the gain adjusted response to the background alone would be  $L'_b = gL_b$ . If we assume that threshold requires the difference in response to background plus test on the one hand and background alone on the other to reach some criterion level, then the threshold will depend on  $(L' - L'_b) = g(L - L_b) = (L - L_b)/(L_b + L_o) = \Delta L/(L_b + L_o) = C$ , which is the contrast form. Thus, predictions about detection threshold are independent of whether a constant is subtracted from the signals leaving the cones.

To deduce the need for a subtractive term (i.e., to differentiate contrast-like coding from gain control alone) from threshold data requires discrimination experiments, in which threshold is measured not just for detecting a test against the background, but also for detecting the test (often referred to as a probe in this context) presented against an incremental flash, both of which are presented against a steady background. In these experiments, the threshold intensity for the probe depends on the flash intensity, and changes in this dependence with the background may be used to infer the nature of the adaptive processes. We will not review these experiments or their analysis here; other papers provide detailed descriptions (e.g., Refs. 56–59). The conclusion drawn by these authors is that to account for steady-state adaptation a subtractive term is required in addition to multiplicative gain control.

## Second-Site Adaptation

The final piece of the basic model of color-discrimination mechanisms is a second stage of adaptation that modifies the output of the second-stage mechanisms (Fig. 5*b*). First-site adaptation is postulated to remove the effects of steady, uniform backgrounds by converting signals to a contrast representation; second-site adaptation is postulated to be driven by contrast signals in the background (see subsection “Field Sensitivities” in Sec. 11.5). Before proceeding, however, we should add two caveats. First, even with uniform, steady fields, first-site adaptation is usually incomplete. Were it complete, detection thresholds expressed as contrasts would be independent of background chromaticity, which they clearly are not (see Fig. 8 and subsection “Field Sensitivities” in Sec. 11.5). Second, first-site adaptation is not instantaneous. If it were, then in the extreme case of complete first-site adaptation being local to the photoreceptor as well as being instantaneous, neither local nor global changes in contrast would be transmitted to the second-site. In the less extreme case of complete, instantaneous adaptation being more spatially extended, then global changes in contrast would not be transmitted—with the result that large uniform field or Ganzfeld flicker should be invisible, which is not the case.<sup>60</sup> To understand the basic logic that connects threshold experiments to second-site contrast adaptation or desensitization, we first consider the effects of steady backgrounds and then the effects of habituation.

*Second-site desensitization by steady fields* Because first-site adaptation is incomplete, steady chromatic backgrounds can desensitize second-site mechanisms. The effect of this type of desensitization on detection contours is illustrated in Fig. 8. The colored circles in Fig. 8*a* show segments of detection contours in the  $\Delta L/L$ ,  $\Delta M/M$  plane of cone contrast space for two background chromaticities. The contours have a slope of 1, which is consistent with detection by a chromatic L–M mechanism with equal and opposite L- and M-cone contrast weights. If the background chromaticity is changed and adaptation at the first site follows Weber’s law, then the contours plotted in cone contrast units should not change. What happens in practice, when for example, a field is changed



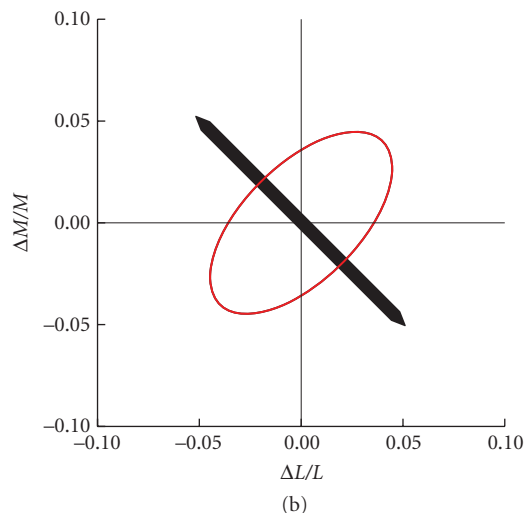
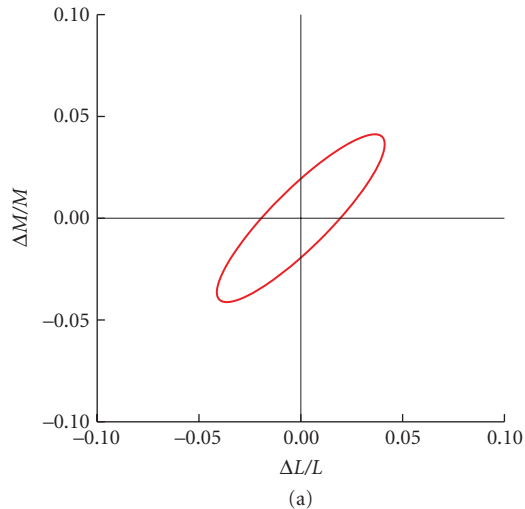
**FIGURE 8** Changes in contrast threshold caused by second-site adaptation to steady fields. (a) Hypothetical contours for detection mediated by the L–M mechanism plotted in cone contrast space before (yellow circles) and after (red circles) second-site desensitization (caused by, e.g., a change in background chromaticity from yellow to red). The L–M mechanism is assumed to have equal and opposite L-cone contrast weights at its input, and this equality is assumed to be unaffected by the desensitization. (b) Hypothetical detection contours for detection by the L+M mechanism before (yellow circles) and after (red circles) a change in background chromaticity from yellow to red. Initially, the L+M mechanism is assumed to have an L-cone contrast weight 1.7 times greater than the M-cone contrast weight. Changing the field to red suppresses the L-cone contrast weight, causing a rotation of the contour, as indicated by the arrows and the red circles.<sup>63</sup>

in chromaticity from spectral yellow to red, is that the contours move outward,<sup>61</sup> as indicated by the arrows and the red circles in Fig. 8a. (Actual data of this type are shown in Fig. 28.)

The constant slope of the L–M detection contours with changing background chromaticity is consistent with Weber’s law operating at the first site. The loss of sensitivity in excess of Weber’s law is most naturally interpreted as a second-site desensitization that depends on background chromaticity and acts on the joint L–M signals. Experimental evidence also supports the idea that first-site adaptation is in the Weber regime under conditions where supra-Weber desensitization is observed.<sup>62</sup>

The effect of second-site chromatic adaptation on the L+M luminance mechanism is different. The yellow circles show partial detection contours with a slope of  $-1.7$ , which is consistent with detection by an achromatic L+M mechanism with an L-cone contrast weight 1.7 times greater than the M-cone contrast weight (this choice is arbitrary, but, in general, the L-cone weight is found to be greater than the M-cone weight, see subsection “Luminance” in Sec. 11.5). If the background chromaticity is again changed from spectral yellow to red, the detection contours rotate, as indicated by the arrows and the yellow circles.<sup>63</sup> In this case, the rotation is due to a suppression by the long-wavelength field of the L-cone contribution to luminance detection relative to the M-cone contribution.<sup>63,64</sup> Given first-site adaptation that follows Weber’s law independently in the L- and the M-cones, first-site adaptation should affect L- and M-cone inputs equally in all postreceptoral mechanisms. The fact that the L–M measurements show an equal effect on the L- and M-cone contrast inputs, whereas the L+M measurements show a differential effect, is most easily explained by positing adaptation effects at a second-site. The idea that the cone inputs to luminance and chromatic mechanisms undergo adaptation that is specific to each type of mechanism was framed by Ahn and MacLeod.<sup>65</sup> The same idea is also explicit in the earlier work of Stromeyer, Cole, and Kronauer.<sup>61,63</sup> The effects of steady fields on the L+M mechanism are discussed further in subsection “Achromatic Direction and Chromatic Adaptation” in Sec. 11.5.

*Second-site habituation* The effects of steady backgrounds at the second site are comparatively small because the background signals are attenuated by adaptation at the first site. Second-site effects can be enhanced by temporally modulating the chromaticity or luminance of the background as in an habituation experiment. Figure 9a shows an idealized threshold contour for detection thresholds



**FIGURE 9** Habituation predictions. (a) Hypothetical threshold contour under conditions where detection is mediated by second-site mechanisms. The contour is plotted in a cone contrast representation, and shows the contour in the  $\Delta L/L$ ,  $\Delta M/M$  contrast plane. The S-cone contrast is taken to be zero, so that thresholds are mediated by the L+M and L-M mechanisms. The contour was computed on the assumption that the gain of the L-M mechanism was four times that of the L+M mechanism, using a summation model with an exponent of 2. (b) Corresponding contour after habituation has reduced the gain of the L-M mechanism by a factor of 2 and left the L+M mechanism gain unchanged. The habituating direction is indicated by the black arrow. The result is the threshold contour becomes more elongated along the negative diagonal in the cone contrast plot.

in the  $\Delta L/L$ ,  $\Delta M/M$  plane obtained against a uniform background. This is the same type of data shown in Fig. 6*d*, but plotted on cone contrast axes. Detection is assumed to be mediated by L–M and L+M mechanisms with L–M being four times more sensitive to cone contrast than L+M. The elliptical contour reflects a summation exponent of 2, which we assume here for didactic purposes. If the same experiment is repeated after the subject has been habituated to contrast modulated in the L–M direction (as indicated by the black arrow Fig. 9*b*), the basic second-site model predicts that thresholds will be elevated for test stimuli that are modulated in that direction. On the other hand, for test stimuli modulated in the L+M direction, thresholds should be unaffected. Finally, for stimuli in intermediate directions, thresholds will be elevated to the extent that the L–M opponent mechanism contributes to detection. Figure 9*b* shows the predicted threshold contour, computed on the assumption that the L–M habituation cuts the gain of the L–M mechanism in half while leaving that of the L+M mechanism unaffected.

The data shown in Fig. 9 may be replotted by taking the difference of the post- and prehabituation threshold stimuli, for each color direction. Figure 10*a* shows an habituation effect plot of this sort. In this plot, the distance from the origin to the contour in each direction shows the threshold increase produced by the L–M habituation for test stimuli modulated in that color direction. The radius is large in the L–M test direction, and drops to zero for the L+M test direction.

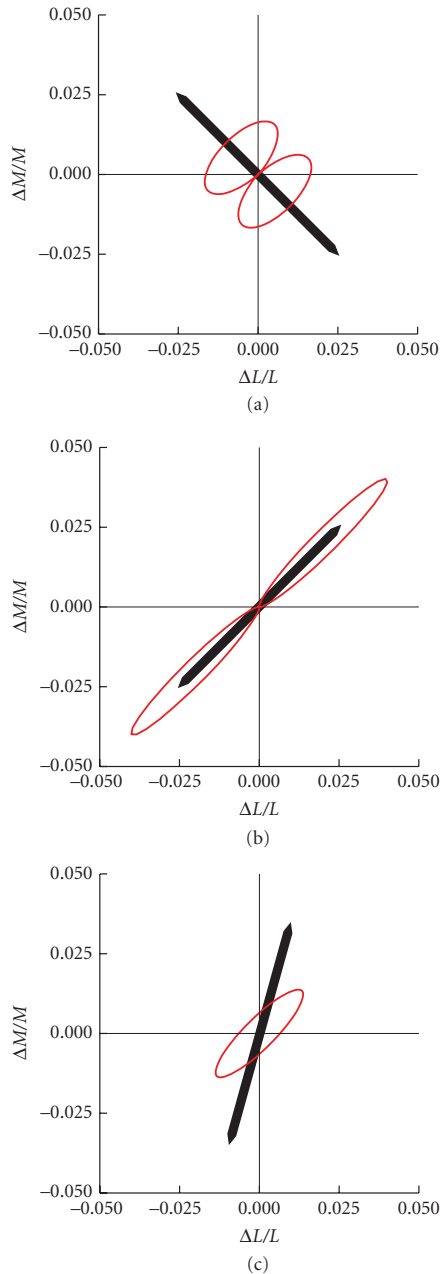
Figure 10*b* and *c* show idealized habituation effect plots for habituation in the L+M direction and to an intermediate color direction, as shown by the black arrows. What the figure makes clear is that the effect of habituation on detection threshold should have a characteristic signature for each habituation direction.

Figure 11 replots the parts of Fig. 10 on axes that represent not cone contrast but rather stimulus directions that isolate the L+M and L–M discrimination mechanisms. These axes have been normalized so that a unit step along each axis direction corresponds to a threshold stimulus (without any habituation) for the corresponding mechanism. We discuss in more detail in subsection “Color Data Representations” in Sec. 11.5 the advantages and disadvantages of using an opponent representation of this type, but introduce it here to allow direct comparison of the idealized predictions of the basic model to the data reported by Krauskopf, Williams, and Heeley.<sup>14</sup> These authors made measurements of how habituation in different color directions affected threshold contours. Their data for observer DRW, as shown in Fig. 29 in subsection “Habituation or Contrast Adaptation Experiments” in Sec. 11.5, agree qualitatively with the predictions shown in Fig. 11, and provide fairly direct evidence to support habituation at a second site. As with other aspects of the basic model, we return below to discuss more subtle features of the data that deviate from its predictions.

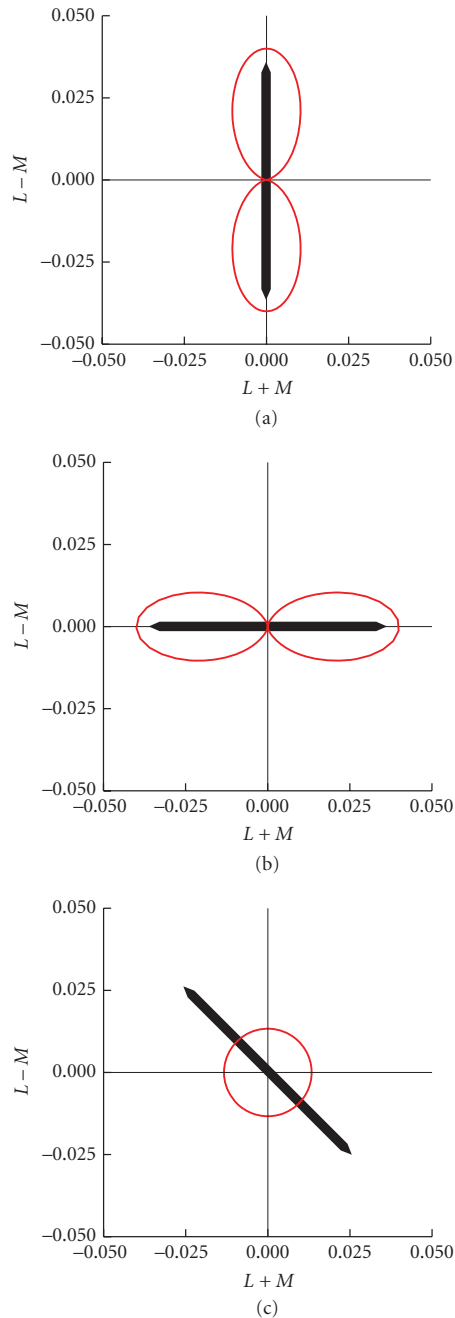
## Sites of Limiting Noise

Before proceeding to color-appearance mechanisms, we pause to discuss a fundamental theoretical issue. This is the question of how the properties of a mechanism at an early stage of the visual processing chain could directly mediate observed detection thresholds, given that the signals from such a site must pass through many additional stages of processing before a perceptual decision is made. This question parallels one often asked in the context of colorimetry, namely, how it is that the very first stage of light absorption can mediate the overall behavior observed in color matching (see Chap. 10). Both questions have the same underlying answer, namely, that information lost at an early stage of visual processing cannot be restored at subsequent stages. The difference is the nature of the information loss. In the case of color matching, the information loss occurs because of the univariant nature of phototransduction. In the case of color thresholds, the information loss occurs because of noise in the responses of visual mechanisms. These sources of noise, relative to the signal strength, set the limits on visual detection and discrimination.

The link between thresholds and noise is well developed under the rubric of the theory of signal detection.<sup>43,66–68</sup> Here we introduce the basic ideas. Consider a test stimulus of intensity  $c$  in the  $\Delta L_d = 1$ ,  $\Delta M_d = 0$  color direction. Only the L-cone mechanism responds to this stimulus, so for this initial example we can restrict attention to the L-cone response. If we present this same stimulus over many trials, the response of the L-cone mechanism will vary from trial-to-trial. This is indicated



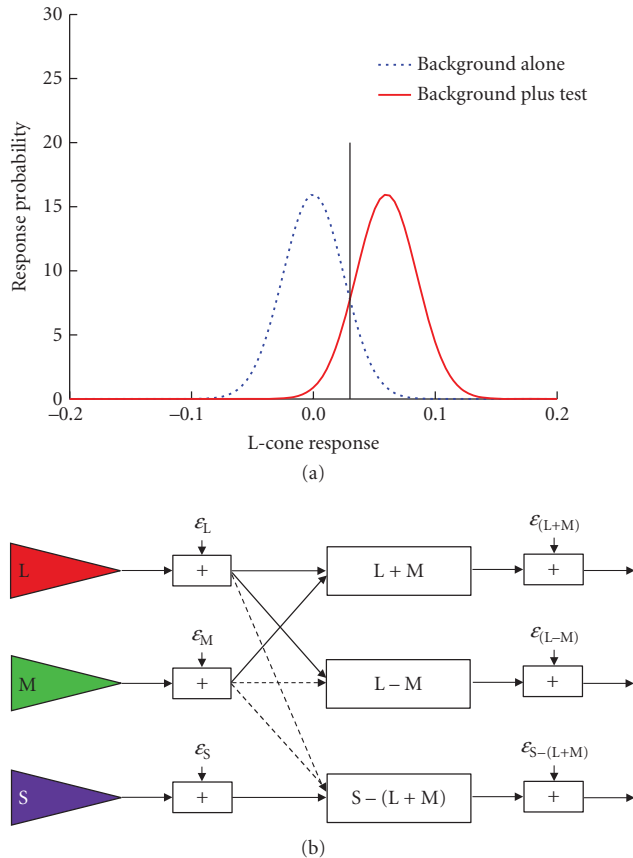
**FIGURE 10** Changes in contrast threshold caused by habituation. It is conventional to report the results of a habituation experiment using a plot that shows the change in threshold caused by habituation. (a) Depiction of the hypothetical data shown in Fig. 9 in this fashion. For each direction in the cone contrast plot, the distance between the origin and the contour shows the increase in threshold caused by habituation for tests in that color direction. For the example shown, there is no change in threshold in the L+M contrast direction, and the contour returns to the origin for this direction. (b) Changes in threshold that would be produced by a habituating stimulus that decreases the gain of the L+M mechanism by a factor of 2 and leaves the gain of the L–M mechanism unchanged. (c) Changes in threshold that would be produced by a habituating stimulus that decreases the gain of both mechanisms equally by a factor of 1.33. The habituating directions are indicated by the black arrows.



**FIGURE 11** Changes in L+M and L-M sensitivity caused by habituation. Panels (a) to (c) replot the threshold changes shown in the corresponding panels of Fig. 10. Rather than showing the effects in cone-contrast space, here the results are shown in a space where the axes correspond to the L+M and L-M color directions. In addition, the units of each of these rotated axes have been chosen so that detection threshold for tests along the axis is 1. This representation highlights the canonical pattern of results expected when habituation along each axis reduces sensitivity for only the corresponding mechanism, and where habituation along the 45° direction reduces sensitivity equally for the two mechanisms. The habituating directions are indicated by the black arrows.

by the probability distribution (solid red line) in Fig. 12a. Although the mean response here is 0.06, sometimes the response is larger and sometimes smaller. Similarly, the mechanism response to the background alone will also fluctuate (probability distribution shown as dotted blue line).

It can be shown<sup>43,69</sup> that optimal performance in a detection experiment occurs if the observer sets an appropriate criterion level and reports “background alone” if the response falls below this criterion



**FIGURE 12** Thresholds and noise. (a) The theory of signal detection’s account of how noise determines thresholds, for a single mechanism. When the background alone is presented, the mechanism’s responses vary somewhat from trial-to-trial because of additive response noise. The response variation is illustrated by the probability distribution centered on zero and shown by the dotted blue line. When the background and test are presented together, the average response increases but there is still trial-to-trial variability. This is illustrated by the probability distribution shown by the solid red line. When the two types of trial are presented equally often and the magnitudes of the costs associated with correct and incorrect responses are equal, it can be shown that the observer maximizes percent correct by reporting that the test was present whenever the mechanism response exceeds the criterion value shown in the figure by the vertical line. Different costs and prior probabilities simply move the location of this criterion. Even when this optimal criterion is used, there will still be some incorrect trials, those on which the response to background alone exceeds the criterion and those on which the response to background and test falls below the criterion. Threshold is reached when the test is intense enough that the observer is correct on a sufficient percentage of trials. How much the average response to background plus test must exceed that to background alone is determined by the magnitude of the noise. (b) The two-stage model of color-discrimination mechanisms, drawn in a manner that emphasizes that noise ( $\epsilon$ ) may be added both to the output of the first-site mechanisms and to the output of the second-site mechanisms. A full signal detection theoretic model of detection takes into account the response gain at each stage (which determines the separation of the average response to background and to background plus test) and the magnitude of noise at each stage. It also replaces the simple criterion shown in (a) with an optimal multivariate classifier stage.<sup>70</sup>



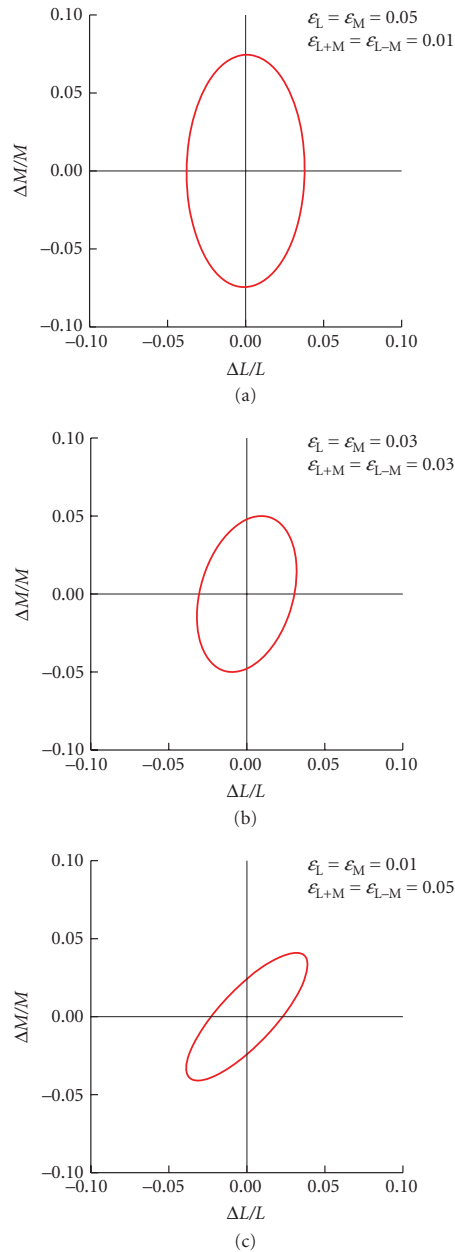
and “background plus test” otherwise. The plot shows the location of the optimal criterion (vertical line) for the special case when background-plus-test and background-alone events are presented equally often and when the benefits associated with both sorts of correct response (hits and correct rejections) have the same magnitude as the costs associated with both sorts of incorrect response (false alarms and misses). Note that even the optimal strategy will not lead to perfect performance unless the intensity of the test is large enough that the two distributions become completely separated. In this sense, the magnitude of response noise (which determines the widths of the distributions), relative to the effect of the test on the mean mechanism response, is what limits performance. More generally, the theory of signal detection allows computation of the probability of the observer performing correctly in the detection experiment as a function of the noise magnitude and test intensity.

With this idea in mind, we can turn to understanding how threshold measurements can reveal properties of early mechanisms, despite later processing by subsequent mechanisms. Figure 12*b* shows the basic model of color-discrimination mechanisms from Fig. 5, but without the adaptation stages shown explicitly. What is shown here, however, is that noise ( $\epsilon$ ) is added to the response of each mechanism. At the first stage, the cone responses are noisy. The noisy cone responses are then transformed at the second-site, and more noise is added. Thus the responses of the second-site mechanisms are subject both to noise that propagates from the first site, in addition to noise injected directly at the second site. For simplicity, we consider the noise added to each mechanism at each site to be independent; note that even with this assumption the noise at the opponent site will be correlated across the three color-discrimination mechanisms as a result of the way the first-site noise is transformed.

This model allows us to compute the expected shape of the threshold contours for a visual system that makes optimal use of the output of the second-site mechanisms. The computation is more involved than the simple unidimensional signal detection example illustrated in Fig. 12*a*, because optimal performance makes use of the output of all three second-site mechanisms and must take into account the correlated nature of the noise at this stage. Nonetheless, the ideas underlying the computation are the same as in the simple example, and the theory that allows the computation is well worked out.<sup>70</sup> We can thus ask how the expected discrimination contours, computed on the basis of the output at the second site, vary with the properties of the noise added at each stage.

Figure 13 shows threshold contours computed for three choices of noise. Figure 13*a* shows a case where the standard deviation of the noise injected at the first site is 5 times larger than that at the second site. The derived contour has the properties expected of detection by cone mechanisms, with the major axes of the contour aligned with the L- and M-cone contrast axes (see Fig. 6*b* and *c*). Figure 13*c* on the other hand, shows the opposite situation where the noise injected at the second site is 5 times larger than that at the first site. Simply changing the noise magnitude has major effects on the observed contour; the shape of the contour shown in Fig. 13*c* is the expected result for detection mediated by second-site mechanisms (see Fig. 6*d*). Finally, Fig. 13*b* shows an intermediate case, with equal noise injected at the two sites. Here the result is intermediate between the two other cases.

This example lets us draw a number of important conclusions. First, the result in Fig. 13*a* shows explicitly how, in a two-stage visual system, threshold contours can reveal properties of the first-stage mechanism. The necessary condition is that the noise added before the transformation to the second stage be large enough to dominate the overall performance. When a situation like this holds, we say that the first stage represents the *site of limiting noise*. As long as subsequent processing does not add significant additional noise, it is the site of limiting noise whose properties will be reflected by threshold measurements. Although Fig. 13*a* shows an example where the cones were the site of limiting noise, the fact that detection contours often have the shape shown in Fig. 13*c* suggests that often the second site limits performance, and that subsequent processing at sites past the second site do not add significant additional noise. As an aside, we note that much of threshold psychophysics is concerned with arranging auxiliary stimulus manipulations that manipulate mechanism properties so as to move the site of limiting noise from one stage of processing to another, or to shift it across mechanisms within a stage (e.g., to change the relative contribution of L+M and L–M mechanisms to performance), with the goal of enabling psychophysical probing of individual mechanisms at different stages of processing. Generally, the stimulus manipulations are thought to



**FIGURE 13** Threshold predictions from signal detection theoretical model. The figure shows threshold contours predicted using the model shown in Fig. 12b, for various choices of noise amplitude. In all cases, only the L+M and L–M mechanisms were considered. The gain on L-cone contrast was set at 2, while the gain on M-cone contrast was set at 1. The gain of the L+M mechanism was set at 1, while that of the L–M mechanism was set at 4. The noise added at each site was assumed to be Gaussian, and for each test color direction and choice of noise magnitude, the test intensity that led to optimal classification performance of 75 percent was determined. (a) Resulting contour when the noise at the first site is five times that of the noise at the second site. In this case, the detection contour has the shape expected for the first-site mechanisms, because little additional information is lost at the second-site mechanisms. (c) Contour when the magnitude of the noise at the second site is five times that of the first. In this case, the properties of the second-site mechanisms dominate measured performance. (b) Contour when the noise at the two sites is of comparable magnitude, and here the result is itself intermediate between that expected from the properties of first- and second-site mechanisms.

act by changing the gains applied to mechanism outputs. This changes the effect of the stimulus relative to the noise, rather than the noise magnitude itself.

A second important point can be drawn from Fig. 13*b*. It is not unreasonable to think that noise at different stages of visual processing will be of a comparable magnitude, and Fig. 13*b* shows that in such cases observed performance will represent a mixture of effects that arise from mechanisms properties at different stages. This is a more formal development of a point we stressed in the introduction, namely, that in spite of our attempt to attribute performance under any given experimental conditions primarily to the action of a small number of mechanisms, this is likely to be a highly simplified account. Eventually, it seems likely that models that explicitly account for the interactions between mechanisms at different stages will be needed. The theory underlying the production of Fig. 13, which explicitly accounts for the role of noise at each stage, would be one interesting approach toward such models.

Finally, note that the development of Fig. 13 is in the tradition of ideal-observer models, where performance is derived from an explicit model of a set of stages of visual processing, followed by the assumption that further processing is optimal.<sup>71-74</sup> Although the assumption of optimal processing is unlikely to be exactly correct, ideal-observer models are useful because they can provide theoretical clarity and because cases where the data deviate from the predictions of such a model highlight phenomena that require further exploration and understanding. For example, the rounded shape of threshold contours, which we initially described using a descriptive vector length account, emerges naturally from the ideal-observer analysis. Indeed, for the ideal observer the predicted shape of the threshold contours is essentially ellipsoidal (summation exponent of 2). This prediction is another reason why careful assessment of the observed shape of threshold contours is of theoretical interest.

In subsection “Noise-Masking Experiments” in Sec. 11.5, we consider the effects of adding external noise to the system.

## 11.4 BASICS OF COLOR-APPEARANCE MECHANISMS

We now turn to consider how a different class of observations can inform us about mechanisms. These are measurements that assess not the observer’s ability to detect or discriminate stimuli, but rather measurements of how stimuli look to observers. More specifically, we will consider measurements of color appearance.

As with threshold measurements, color-appearance measurements may be divided into two classes, those that assess the appearance of test stimuli as some function of those stimuli (test measurements) and those that assess how the color appearance of a test stimulus depends on the context in which it is viewed (field measurements). We review how both sorts of measurements can inform a model of color-appearance mechanisms. This treatment parallels our development for color-discrimination mechanisms in order to highlight similarities and differences between models that account for the two types of data.

### Appearance Test Measurements and Opponency

As described in the introduction, one of the earliest suggestions that the output of the cone mechanisms were recombined at a second site came from Hering’s observation that certain pairs of color sensations (i.e., red and green, blue and yellow) are mutually exclusive. Hurvich and Jameson elaborated this observation into a psychophysical procedure known as hue cancellation (see subsection “Spectral Properties of Color-Opponent Mechanisms” in Sec. 11.5). Hue cancellation can be performed separately for the red/green opponent pair and the blue/yellow opponent pair.

In the red/green hue-cancellation experiment, the observer is presented with a test stimulus, which is usually monochromatic, and asked to judge whether its appearance is reddish or greenish. The results may be used to derive the properties of an opponent mechanism, under the “linking

hypothesis” that when the mechanism’s response is of one sign the stimulus appears reddish, while when it is of the opposite sign the stimulus appears greenish. That is, color-appearance measurements are used to derive mechanism properties on the assumption that the mechanism output provides an explicit representation of the sensation experienced by the observer.

To understand the hue-cancellation procedure in more detail, imagine that the experimenter picks a fixed monochromatic reference light that in isolation appears greenish to the observer. The experimenter then chooses a series of monochromatic test lights of unit intensity, each of which appears reddish in isolation. The observer judges mixtures of each test light and the reference light. If the mixture appears reddish, the intensity of the greenish reference light is increased. If the mixture appears greenish, the intensity of the greenish reference light is decreased. Repeating this procedure allows the experimental determination of a balance or equilibrium point for the red-green mechanism, where the result of the mixture appears neither reddish nor greenish. (Depending on the test stimulus, the balanced mixture will appear either as yellow or blue or achromatic.) Colored stimuli that appear neither reddish nor greenish are referred to as “unique yellow” or “unique blue.” The intensity of the green reference light in the balanced mixture may be taken as a measure of the amount of redness in the test stimulus, in the same way that the amount of weight added to one side of a balance scale indexes the weight on the other side when the balance is even (see Krantz’s “force table” analogy.<sup>75</sup>)

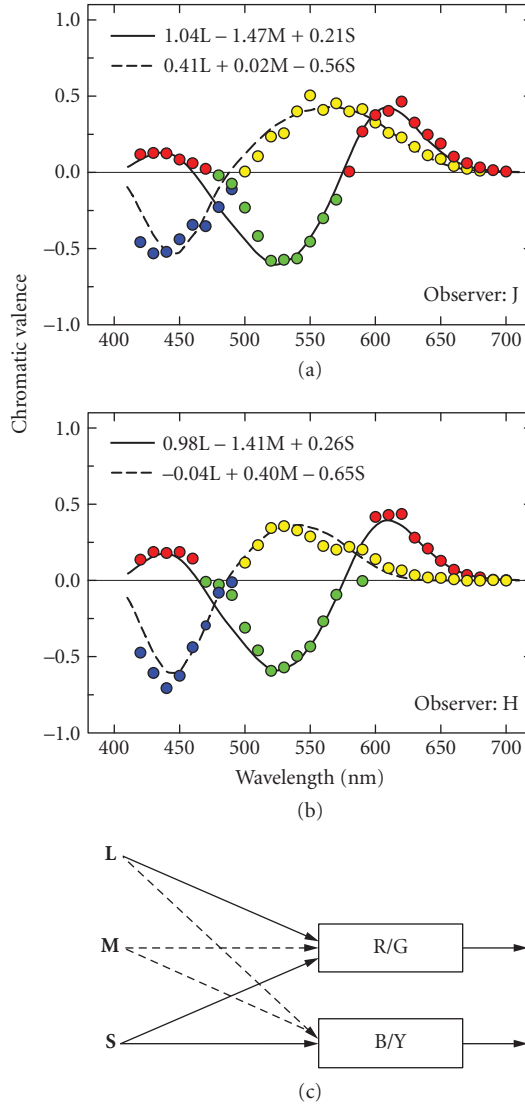
A similar procedure may be used to determine the amount of greenness in test stimuli that have a greenish component, by choosing an adjustable reference stimulus that appears reddish. Moreover, the units of redness and greenness may be equated by balancing the two reference stimuli against each other. Figure 14*a* and *b* show prototypical results of such measurements for spectral test lights, as well as results for an analogous procedure used to determine the amount of blueness and yellowness in the same set of spectral test lights. Such data are referred to as “chromatic valence” data.

Krantz<sup>75,76</sup> showed that if the amount of redness, greenness, blueness, and yellowness in any test light is represented by the output of two opponent color-appearance mechanisms, and if these mechanisms combine cone signals linearly, then the spectral hue-valence functions measured in the cancellation experiment must be a linear transformation of the cone spectral sensitivities. Although the linearity assumption is at best an approximation (see subsection “Linearity of Color-Opponent Mechanisms” in Sec. 11.5), Krantz’s theorem allows derivation of the cone inputs from the data shown in Fig. 14. Such fits are shown in Fig. 14*a* and *b*, and lead to the basic wiring diagram shown in Fig. 14*c*. This shows two opponent color-appearance mechanisms, labeled R/G and B/Y. The cone inputs to these mechanisms are similar to those for the opponent color-discrimination mechanisms, with one striking difference: the S cones make a strong R contribution to the R/G appearance mechanism. By contrast, the S-cone input to the L–M discrimination mechanism is small at most. Another difference, which we address later, is that the M-cone contribution to B/Y in some models is of the same sign as the S-cone contribution (see “Three-Stage Zone Models” in Sec. 11.6).

Although the hue-cancellation procedure may be used to derive properties of opponent-appearance mechanisms, it is less suited to deriving the properties of a third mechanism that signals the brightness of colored stimuli. This aspect of color appearance may be assessed using heterochromatic brightness matching, where subjects are asked to equate the brightness of two stimuli, or scaling methods, where observers are asked to directly rate how bright test stimuli appear (see subsection “Luminance and Brightness” in Sec. 11.5).

## Appearance Field Measurements and First-Site Adaptation

It is also possible to make field measurements using appearance judgments. A common method is known as asymmetric matching, which is a simple extension of the basic color-matching experiment (see Chap. 10). In the basic color-matching experiment, the observer adjusts a match stimulus to have the same appearance as a test stimulus when both are seen side-by-side in the same context. The asymmetric matching experiment adds one additional variable, namely that the test and match are each seen in separate contexts. As illustrated in Fig. 1, when an identical test patch is presented



**FIGURE 14** Chromatic valence data and wiring. (a) and (b) Valence data (colored symbols) replotted from Figs. 4 and 5 of Jameson and Hurvich<sup>279</sup> for observers “J” (a) and “H” (b) fitted with linear combinations of the Stockman and Sharpe<sup>28</sup> cone fundamentals (solid and dashed curves). The best-fitting cone weights are noted in the key. (c) Wiring suggested by this (a) and subsequent color valence data. In the diagram, the sign of the M-cone contribution to B/Y is shown as negative (see subsection “Spectral Properties of Color-Opponent Mechanisms” in Sec. 11.5).

against two different surrounds, it can appear quite different. In the asymmetric matching experiment, the observer compares test patches seen in two different contexts (e.g., against different adapting backgrounds) and adjusts one of them until it matches the other in appearance. The cone coordinates of the match test patches typically differ, with the shift being a measure of the effect of changing from one context to the other.

Figure 15a plots asymmetric matching data from an experiment of this sort reported by MacAdam.<sup>77</sup> The filled circles show the L- and S-cone coordinates of a set of test stimuli, seen in a region of the retina adapted to a background with the chromaticity of a typical daylight. The open circles show a subset of the corresponding asymmetric matches seen in a region of the retina adapted to a typical tungsten illuminant. The change in adapting background produces a change in test appearance, which observers have compensated for in their adjustments.

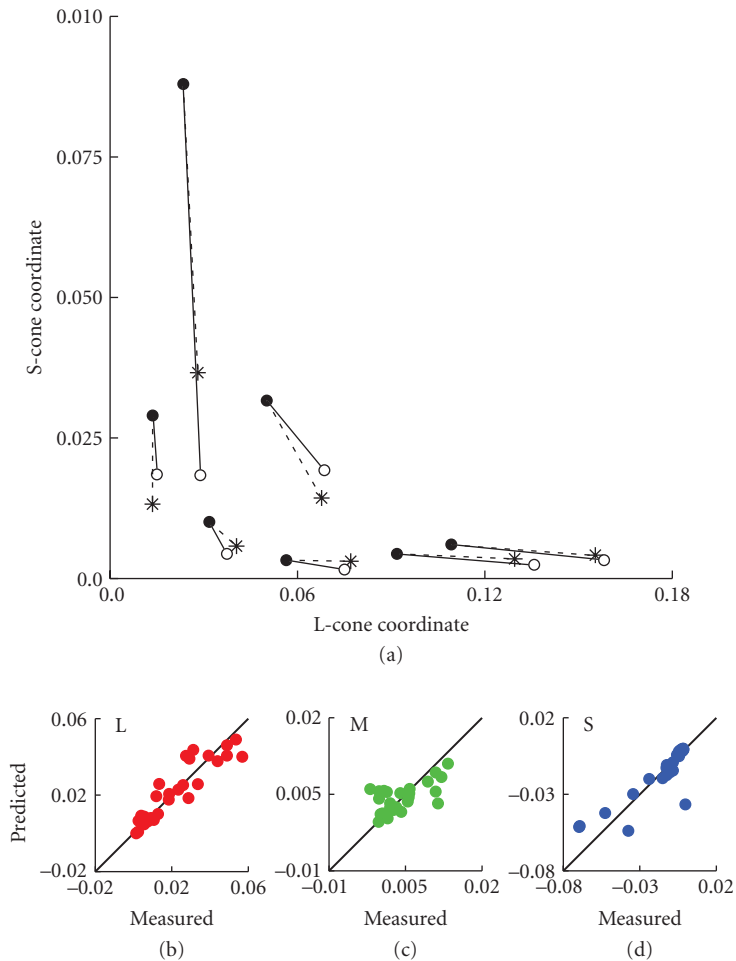
Asymmetric matching data can be used to test models of adaptation. For example, one can ask whether simple gain control together with subtractive adaptation applied to the cone signals can account for the matches. Let  $(L_1, M_1, S_1), (L_2, M_2, S_2), \dots, (L_N, M_N, S_N)$  represent the L-, M-, and S-cone coordinates of  $N$  test stimuli seen in one context. Similarly, let  $(L'_1, M'_1, S'_1), (L'_2, M'_2, S'_2), \dots, (L'_N, M'_N, S'_N)$  represent the cone coordinates of the matches set in the other context. If the data can be accounted for by cone-specific gain control together with subtractive adaptation, then we should be able to find gains  $g_L, g_M, g_S$ , and subtractive terms  $L_0, M_0, S_0$  such that  $L'_i \approx g_L L_i - L_0$ ,  $M'_i \approx g_M M_i - M_0$ , and  $S'_i \approx g_S S_i - S_0$  for all of the  $N$  matches. The asterisks plotted in Fig. 15a show predictions of this gain control model in the L-S cone plane. Figure 15b, c, and d show predicted shifts against the measured shifts for all three cone classes. The model captures the broad trends in the data, although it misses some of the detail. The first-site adaptation captured by this model is often taken to be the same adaptation revealed by sensitivity experiments.<sup>78</sup>

## Appearance Field Measurements and Second-Site Adaptation

Asymmetric matching may also be used to investigate the effect of contrast adaptation on color appearance. In a series of experiments, Webster and Mollon<sup>32,79</sup> investigated the effect of contrast adaptation (or habituation) on the color appearance of suprathreshold lights (see also Refs. 80 and 81). Observers habituated to a background field that was sinusoidally modulated at 1 Hz along various color directions. The habituating field was interleaved with a stimulus presented at the same location. The observer's task was to adjust the chromaticity and luminance of a match stimulus so that its appearance matched that of the test field. Webster and Mollon found that habituation affected the appearance of the test stimulus.

The motivation behind these experiments was to test whether or not the changes in color appearance caused by contrast adaptation were consistent with the type of second-site adaptation that characterizes sensitivity experiments, namely independent adaptation of three cardinal mechanisms with cone inputs: L-M, L+M, and S-(L+M). The prediction for this case is that the appearance effects should always be greatest along the color axis corresponding to the mechanism most adapted by the habituating stimulus, and least along the axis corresponding to the mechanism least adapted by the stimulus—and this pattern should be found whatever the chromatic axis of the habituating stimulus. Thus the *relative* changes in chromaticity and/or luminance caused by habituation should be roughly elliptical with axes aligned with the mechanism axes.

The results showed that the sensitivity changes did not consistently align with the axes of the cardinal mechanisms. Instead, the largest selective sensitivity losses in the equiluminant plane, for two out of three subjects, aligned with the adapting axis, while the smallest losses aligned with an axis approximately 90° away. Comparable results were found when the habituating and test stimuli varied in color and luminance in either the plane including the L-M and L+M+S axes or the plane including the S and L+M+S axes. Thus, contrast adaptation produces changes in color appearance that are, in general, selective for the habituating axis rather than selective for the purported underlying mechanism axes. Nevertheless, greater selectivity was found for habituation along the three cardinal axes, which suggests their importance. See Ref. 32 for further details.



**FIGURE 15** Analysis of asymmetric matching data. (a) Subset of the data from Observer DLM reported by MacAdam.<sup>77</sup> The closed circles show L- and S-cone coordinates of tests viewed on part of the retina adapted to a background with the chromaticity of daylight. The connected open circles show the asymmetric matches to these tests, when the matching stimulus was viewed on part of the retina adapted to the chromaticity of a tungsten illuminant. The star symbols show the predictions of a model that supposes first-site gain control and subtractive adaptation, operating separately within each cone mechanism. (b), (c), and (d) Model predictions for the full set of 29 matches set by Observer DLM. Each panel shows data for one cone class. The x axis shows the measured difference between the test and match, and represents the size of the asymmetric matching effect. The y axis shows the corresponding predicted difference. If the model were perfect, all of the data would lie along the positive diagonals. The model captures the broad trend of the data, but misses in detail.

Webster and Mollon<sup>32</sup> conclude that the changes in color appearance following habituation are inconsistent with models of color vision that assume adaptation in just three independent postreceptoral channels. Given, however, that the properties of the color-appearance mechanisms as revealed by test methods differ from those of color-discrimination mechanisms, this result is hardly surprising. Also note that comparisons between sensitivity measurements made for test stimuli near detection threshold and appearance measurements made for suprathreshold tests are complicated by the presence of response nonlinearities at various stages of processing. See Refs. 78 and 82 for discussion.

## 11.5 DETAILS AND LIMITS OF THE BASIC MODEL

Sections 11.3 and 11.4 introduce what we will refer to as a “basic model” of color-discrimination and color-appearance mechanisms. There is general agreement that the components incorporated in this model are correct in broad outline. These are (1) first-site, cone specific, adaptation roughly in accord with Weber’s law for signals of low temporal frequency, (2) a recombination of cone signals that accounts for opponent and nonopponent postreceptoral second-site signals, (3) second-site desensitization and habituation, and (4) the necessity for distinct mechanistic accounts of discrimination and appearance.

In the earlier sections, we highlighted example data that strongly suggest each feature of the basic model. Not surprisingly, those data represent a small subset of the available experimental evidence brought to bear on the nature of postreceptoral color vision. In this section we provide a more detailed review, with the goal of highlighting both data accounted for by the basic model, as well as data that expose the model’s limitations. Current research in the mechanistic tradition is aimed at understanding how we should elaborate the basic model to account parsimoniously for a wider range of phenomena, and at the end of this chapter we outline approaches being taken toward this end.

Sections “Test Sensitivities” through “Color Appearance and Color Opponency” follow the same basic organization as the introductory sections. We consider discrimination and appearance data separately, and distinguish between test and field methods. We begin, however, with a discussion of color data representations.

### Color Data Representations

Over the past 50 years, color vision research has been driven in part by the available technology. Conventional optical systems with spectral lights and mechanical shutters largely restricted the volume of color space that could be easily investigated to the 1/8 volume containing positive, incremental cone modulations. When such data are obtained as a function of the wavelength of monochromatic stimuli, a natural stimulus representation is to plot the results as a function of wavelength (e.g., Fig. 14a and b). This representation, however, does not take advantage of our understanding of the first stage of color processing, transduction of light by the cone photoreceptors. Moreover, with the availability of color monitors and other three-primary devices, stimuli are now routinely generated as a combination of primaries, which makes increments, decrements, and mixtures readily available for experimentation. Spectral plots are not appropriate for such stimuli, which instead are represented within some tristimulus space (see Chap. 10).

Our estimates of the human cone spectral sensitivities have become increasingly secure over the past 30 years (see Chap. 10). This allows stimulus representations that explicitly represent the stimulus in terms of the L-, M-, and S-cone excitations and which therefore connect data quite directly to the first-stage color mechanisms. Such representations are called “cone-excitation spaces,” and an example of this sort of representation is provided in Fig. 7 of Chap. 10.

An extension of cone-excitation space is one in which the cone excitations are converted to increments and decrements of each cone type relative to a background. Thus, the three axes are the changes in cone excitations:  $\pm\Delta L$ ,  $\pm\Delta M$ , and  $\pm\Delta S$ . We introduced such spaces in Fig. 6; they are useful when

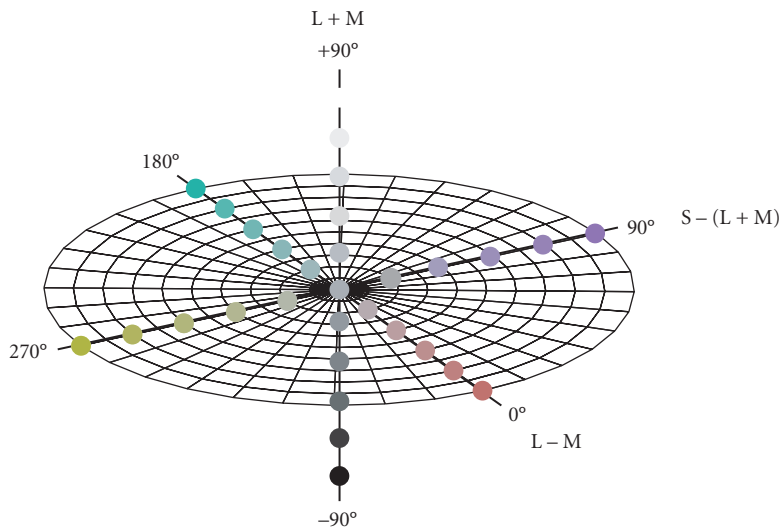


the stimulus is most naturally conceived as a modulation relative to a background, and where the properties of the background per se are of less interest. We will use the shorthand “incremental cone spaces” to refer to this type of representation, although clearly they allow expression of both increments and decrements.

Closely related to incremental cone spaces are “cone contrast spaces,” in which the incremental/decremental changes in cone excitations produced by the target are divided by the cone excitations produced by the background.<sup>61,83</sup> Here, the three axes are dimensionless cone contrasts:  $\Delta L/L_b$ ,  $\Delta M/M_b$ , and  $\Delta S/S_b$ . If Weber’s law holds independently for each cone type then  $\Delta L/L_b$ ,  $\Delta M/M_b$ , and  $\Delta S/S_b$  all remain constant. Thus a particular advantage of this space is that it factors out the gross effects of first-site cone-specific adaptation, which tends to follow Weber’s law at higher intensities (see subsection “First-Site Adaptation” in Sec. 11.3). Plots in cone contrast space help to emphasize desensitization that occurs after the receptors, and, to the extent that Weber’s law holds, provide an explicit representation of the inputs to postreceptoral mechanisms. (We introduced a cone contrast representation in Fig. 8 exactly for this reason.)

At lower adaptation levels, when adaptation falls short of Weber’s law (see Fig. 7) or under conditions where Weber’s law does not hold (such as at high spatial and temporal frequencies; see, e.g., Ref. 84), cone contrast space is less useful. However, for targets of low temporal and spatial frequencies, Weber’s law has been found to hold for chromatic detection down to low photopic levels.<sup>62</sup>

A final widely used color space is known as the “Derrington-Krauskopf-Lennie” (DKL) space<sup>36,85,86</sup> in which the coordinates represent the purported responses of the three second-site color-discrimination mechanism, L+M, L–M, and S–(L+M). In this context, modulation directions that change the response of one of these mechanisms while leaving the response of the other two fixed are referred to as “cardinal directions.”<sup>14</sup> The DKL representation was introduced in Fig. 11 and is further illustrated in Fig. 16. See Ref. 87 for further discussion of color spaces and considerations of when each is most appropriately used.



**FIGURE 16** Derrington-Krauskopf-Lennie (DKL) color space. The grid corresponds to the equiluminant plane, which includes the L–M ( $0^\circ$ – $180^\circ$ ) and S–(L+M) ( $90^\circ$ – $270^\circ$ ) cardinal mechanisms axes. The vertical axis is the achromatic L+M axis ( $-90^\circ$ – $+90^\circ$ ). The colors along each axis are approximate representations of the appearances of lights modulated along each cardinal axis from an achromatic gray at the center. Notice that the unique hues do not appear. The axes are labeled according to the mechanisms that are assumed to be uniquely excited by modulations along each cardinal axis. (Figure provided by Caterina Ripamonti.)

Although the conceptual ideas underlying the various color spaces discussed here are clear enough, confusion often arises when one tries to use one color space to represent constructs from another. The confusion arises because there are two distinct ways to interpret the axes of a three-dimensional color space. The first is in terms of the responses of three specified mechanisms, while the second is in terms of how the stimulus is decomposed in terms of the modulation directions that isolate the three mechanisms. To understand why these are different, note that to compute the response of a particular mechanism, one only needs to know its properties (e.g., that it is an L–M mechanism). But, in contrast, the stimulus direction that isolates that same mechanism is the one that silences the responses of the other two mechanisms. The isolating stimulus direction depends not on the properties of the mechanism being isolated but on those of the other two. For this reason, it is important to distinguish between the *mechanism direction*, which is the stimulus direction that elicits the maximum positive response of a given mechanism per unit intensity, and the *isolating direction* for the same mechanism, which produces no response in the other two mechanisms. For the triplet of mechanisms L+M, L–M, and S–(L+M), the mechanism and isolating directions align (by definition) in DKL space, but this property does not hold when these directions are plotted in the antecedent incremental cone space or cone contrast space. Note that in some descriptions of DKL space, the axes are defined in terms of the cone modulations that isolate mechanisms rather than in terms of the mechanism responses. These conflicting definitions of DKL space have led to a good deal of confusion. See Chap. 10 and Refs. 87–89 for further discussion of transformations between color spaces and the relation between mechanism sensitivities, mechanism directions, and isolating directions.

**Caveats** Crucial for the use of any color space based on the cone mechanisms either for generating visual stimuli or for interpreting the resulting data is that the underlying cone spectral sensitivities are correct. Our estimates of these are now secure enough for many applications, but there are uncertainties that may matter for some applications, particularly at short wavelengths. These are compounded by individual differences (see Chap. 10), as well as issues with some of the sets of color-matching functions used to derive particular estimates of cone spectral sensitivities (again see Chap. 10, and also Ref. 90 for further discussion).

Such concerns are magnified in the use of postreceptoral spaces like the DKL space, because as well as the cone spectral sensitivities needing to be correct, the rules by which they are combined to produce the spectral sensitivities of the postreceptoral mechanisms must also be correct. For example, to silence the L+M luminance mechanism it is necessary to specify exactly the relative weights of the M and L cones to this mechanism. However, these weights are uncertain. As discussed in subsection “Luminance” in Sec. 11.5, the M- and L-cone weights show large individual differences. Furthermore, the assumption typically used in practice, when constructing the DKL space, is that luminance is a weighted sum of M and L cones and that the S cones therefore do not contribute to luminance. This may not always be the case.<sup>91–94</sup> Luminous efficiency also varies with chromatic adaptation, the spatial and temporal properties of the stimulus, and the experimental task used to define luminous efficiency. Lastly, the standard 1924 CIE  $V(\lambda)$  luminous efficiency function, which determines candelas/m<sup>2</sup>, and has been used by several groups to define the spectral sensitivity of the luminance mechanism, seriously underestimates luminous efficiency at shorter wavelengths (see Ref. 90).

One final note of caution is that it is useful to keep in mind that independently representing and manipulating the responses of mechanisms at a particular stage of visual processing only makes sense if there are three or fewer mechanisms. If there are more than three photopic mechanisms at the second stage of color processing (see also subsection “Low-Level and Higher-Order Color-Discrimination Mechanisms” in Sec. 11.6), independently manipulating all of them is not possible because of the inherent three-dimensionality of the first stage.

These caveats are not meant to discourage entirely the use of physiologically motivated color spaces for data representation; on the contrary we believe such representations offer important advantages if used judiciously. It is crucial, however, to bear in mind that such spaces usually assume particular theories of processing, and that interpreting data represented within them should be done in the context of a clear understanding of the limitations of the corresponding theories.

## Test Sensitivities

The most direct method of studying the chromatic properties of color mechanisms is the test method. Variations of this method have been tailored to investigate the properties of different visual mechanisms, including cone, chromatic, and achromatic (or luminance) mechanisms.

**Sensitivity to Spectral Lights** Perhaps the most influential chromatic detection data have been the two-color threshold data of Stiles,<sup>41,42,95–98</sup> so called because the threshold for detecting a target or test field of one wavelength is measured on a larger adapting or background field usually of a second wavelength (or mixture of wavelengths). These data were collected within the context of an explicit model of color mechanisms, referred to as “ $\pi$ -mechanisms,” and how they adapt. Enoch<sup>99</sup> provides a concise review of Stiles’ work.

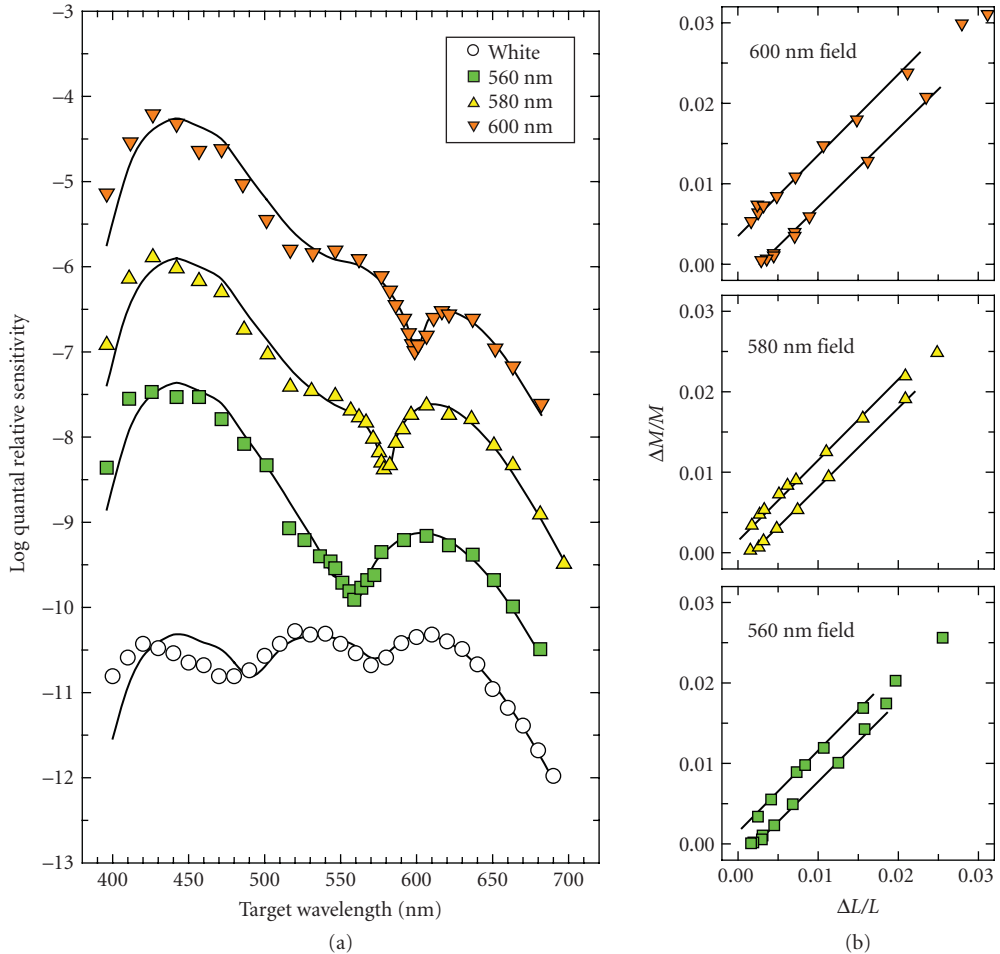
An important limitation of Stiles’ model is that it did not allow for opponency: the  $\pi$ -mechanisms had all-positive spectral sensitivities and interactions between them were required to be summative. In some instances, the deviations caused by opponency were accommodated in Stiles’ model by the postulation of new  $\pi$ -mechanisms, such as  $\pi_5'$  (see Refs. 54 and 55). As data in the two-color tradition accumulated, this limitation led to the demise of the  $\pi$ -mechanism framework and the development of the basic model we introduced in Sec. 11.3. Nonetheless, any correct model must be consistent with the experimental data provided by the two-color threshold technique, and we review some key aspects in the context of current thinking.

Two-color threshold test spectral sensitivity measurements obtained on steady backgrounds as a function of target wavelength have characteristic peaks and troughs that depend on the background wavelength (see Fig. 1 of Ref. 98). These undulations were interpreted as the envelope of the spectral sensitivities of the underlying cone mechanisms (or “ $\pi$ -mechanisms,” see Fig. 23), but crucially they are now known also to reflect opponent interactions between mechanisms.<sup>100</sup>

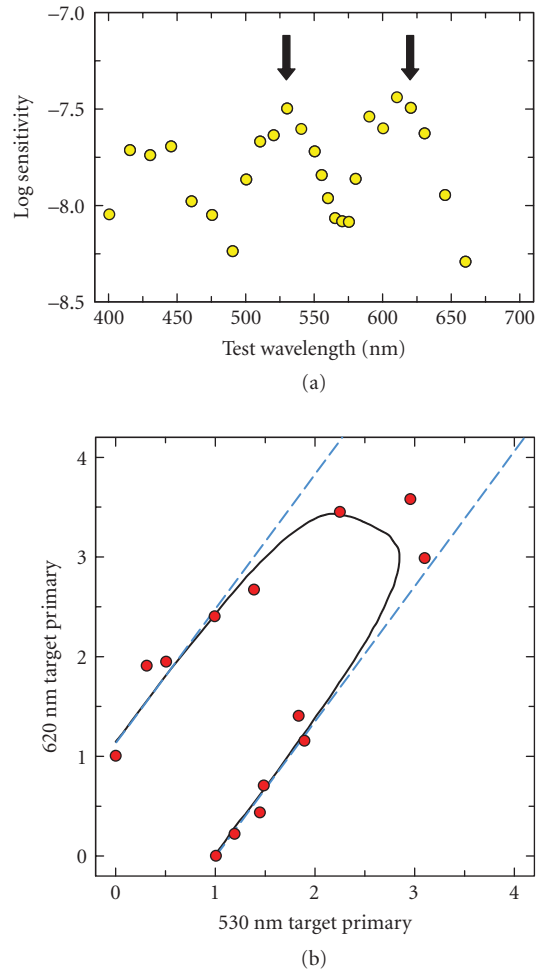
The shape of test spectral sensitivity functions depend not only on background wavelength but also on the type of target used in the detection task. Detection by second-site opponent color-discrimination mechanisms is generally favored by targets comprised mainly of low temporal and spatial frequency components, whereas detection by achromatic mechanisms, which sum cone signals, is favored by targets with higher temporal and spatial frequency components.<sup>83,101,102</sup> Stiles typically used a flashed target of  $1^\circ$  in visual diameter and 200 ms in duration, which favors detection by chromatic mechanisms. When Stiles’ two-color threshold experiments are carried out using brief, 10-ms duration targets, the results are more consistent with detection being mediated by achromatic mechanisms.<sup>103</sup>

Detection by chromatically opponent mechanisms is most evident in test spectral sensitivity functions measured using targets of low temporal and/or spatial frequency. Measured on neutral fields, chromatic detection is characterized by peaks in the spectral sensitivity curves at approximately 440, 530, and 610 nm that are broader than the underlying cone spectral sensitivity functions and separated by pronounced notches.<sup>102,104–108</sup> The so-called “Sloan notch”<sup>109</sup> corresponds to the loss of sensitivity when the target wavelength is such that the target produces no chromatic signal (e.g., when the L- and M-cone inputs to the L–M chromatic mechanism are equal), so that detection is mediated instead by the less-sensitive achromatic mechanism. By contrast, the broad peaks correspond to target wavelengths at which the target produces a large chromatic signal.<sup>101,102,105–108,110,111</sup> Examples of detection spectral sensitivities with strong contributions from opponent mechanisms are shown in Fig. 17a and Fig. 18a. Those shown as circles in Fig. 17 and Fig. 18 were measured on a white background. The other functions in Fig. 17 were measured on backgrounds of 560 nm (green squares), 580 nm (yellow triangles), and 600 nm (orange triangles).

The nature of the mechanisms responsible for the spectral sensitivity data in Fig. 17a can be seen clearly by replotting the data as cone contrasts, as illustrated in Fig. 17b for the three chromatic backgrounds (from Ref. 89). As indicated by the straight lines fitted to each set of data, the detection contours have slopes close to one in cone contrast space, which is consistent with detection by L–M chromatic mechanisms with equal cone contrast weights (see subsection “Sensitivity to Different Directions of Color Space” in Sec. 11.5). Note also that the contours move outward with increasing field wavelength. This is consistent with second-site desensitization (see Fig. 8). The dependence



**FIGURE 17** Sloan notch and chromatic adaptation. The three data sets in (a) are spectral sensitivity data replotted from Fig. 1 of Thornton and Pugh<sup>107</sup> measured on  $10^{10}$  quanta  $\text{sec}^{-1}$   $\text{deg}^{-2}$  fields of 560 nm (green squares), 580 nm (yellow triangles), and 600 nm (orange inverted triangles) using a long duration (one period of a 2-Hz cosine), large ( $3^\circ$  diameter, Gaussian-windowed) target. The lowest data set in (a) are data replotted from Fig. 4 (Observer JS) of Sperling and Harwerth<sup>105</sup> measured on 10,000 td, 5500 K, white field (open circles) using a 50-ms duration, 45-min diameter target. The data have been modeled (solid lines) by assuming that the spectral sensitivities can be described by  $b|aL - M| + c|M - aL| + d|S - e(L + 0.5M)|$ , where L, M, and S are the quantal cone fundamentals<sup>28</sup> normalized to unity peak, and  $a-e$  are the best-fitting scaling factors. These fits are comparable to the ones originally carried out by Thornton and Pugh<sup>107</sup> using different cone fundamentals. Notice that the Sloan notch (on colored backgrounds) coincides with the field wavelength. The detection contours in (b) have been replotted from Fig. 18.4B of Eskew, McLellan, and Giulianini.<sup>89</sup> The data are the spectral sensitivities measured on 560 nm (green squares), 580 nm (yellow triangles), and 600 nm (orange inverted triangles) fields already shown in (a), but transformed and plotted in cone-contrast space. Only data for target wavelengths  $\geq 520$  nm are shown. As indicated by the straight lines fitted to each set of data, the detection contours have slopes close to one in cone contrast space, which is consistent with detection by L-M chromatic mechanisms with equal L- and M-cone contrast weights.



**FIGURE 18** Test additivity. (a) Spectral sensitivity for the detection of a low frequency target measured on a 6000 td xenon (white) background (yellow circles). The shape is characteristic of detection by chromatic channels. The arrows indicate the wavelengths of the two primaries used in the test additivity experiment shown in (b). (b) Detection thresholds measured on the same field for different ratios of the two primaries. Data and model fits replotted from Fig. 2 of Thornton and Pugh.<sup>164</sup> The solid curve is the fit of a multi-mechanism model comprising an achromatic (L+M) and chromatic mechanisms (L–M and M–L with the same but opposite weights). The sensitivities of the chromatic mechanisms are shown by the dashed blue line. (See Ref. 164 for details.)

of the spectral position of the Sloan notch on background wavelength is discussed in subsection “Chromatic Adaptation and the Sloan Notch” in Sec. 11.5.

### Luminance

*Luminous efficiency* The target parameters (and the experimental task) can also be chosen to favor detection by achromatic mechanisms. For example, changing the target parameters from a 200-ms duration, 1° diameter flash to a 10 ms, 0.05° diameter flash presented on a white background changes the spectral sensitivity from a chromatic one to an achromatic one, with chromatic peaks at approximately 530- and 610-nm merging to form a broad peak near 555 nm.<sup>102</sup> Measurements that favor detection by achromatic mechanisms have been carried out frequently in the applied field of photometry as a means of estimating scotopic (rod), mesopic (rod-cone), or photopic (cone) “luminous efficiency” functions.

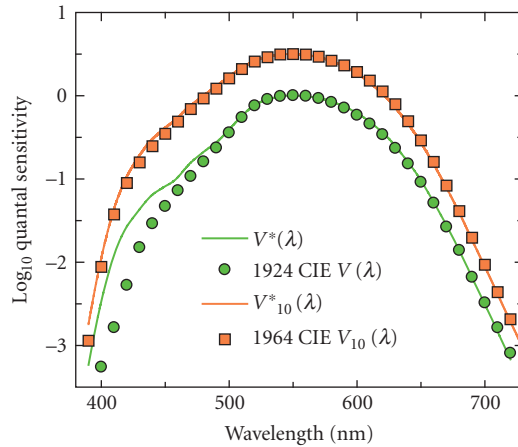
Luminous efficiency was introduced by the CIE (Commission Internationale de l’Éclairage) to provide a perceptual analog of radiance that could be used to estimate the visual effectiveness of lights. The measurement of luminous efficiency using tasks that favor achromatic detection is a practical solution to the requirement that luminous efficiency should be additive. Additivity, also known as obedience to Abney’s law,<sup>112,113</sup> is necessary in order for the luminous efficiency of a light of arbitrary spectral complexity to be predictable from the luminous efficiency function for spectral lights,  $V(\lambda)$ .

Some of the earlier luminous efficiency measurements<sup>114–117</sup> incorporated into the original 1924 CIE photopic luminous efficiency function,  $V(\lambda)$ , used techniques that are now known to fail the additivity requirement. Techniques that satisfy the additivity requirement, and depend mainly on achromatic or luminance mechanisms, include heterochromatic flicker photometry (HFP), heterochromatic modulation photometry (HMP), minimally distinct border (MDB), and minimum motion (MM). Details of these techniques can be found in several papers.<sup>28,53,90,118–124</sup>

In visual science,  $V(\lambda)$  or its variants, has often been assumed to correspond to the spectral sensitivity of the human postreceptoral achromatic or “luminance” mechanism, which is assumed to add positively weighted inputs from the L and M cones.<sup>125</sup> As a result, estimates of luminous efficiency have taken on an important theoretical role in studies of the opponent mechanisms, because they are used to produce stimulus modulations that produce no change in the purported luminance mechanism. In this context, the  $V(\lambda)$  standard is often overinterpreted, since it obscures a number of factors that affect actual measurements of luminous efficiency. These include (1) the strong dependence of luminous efficiency on the mean state of chromatic adaptation (see subsection “Achromatic Detection and Chromatic Adaptation” in Sec. 11.5), (2) the sizeable individual differences in luminous efficiency that can occur between observers (see below), and (3) the fact that luminous efficiency is affected both by the spatial properties of the target, as well by where on the retina it is measured.

A more fundamental problem is that the CIE photopic 1924  $V(\lambda)$  function seriously underestimates luminous efficiency at short wavelengths, because of errors made in its derivation. Consequently,  $V(\lambda)$  is seldom used in visual science, except for the derivation of troland values (with the result that luminance is often underestimated at short wavelengths). Attempts to improve  $V(\lambda)$ <sup>126,127</sup> have been less than satisfactory,<sup>28,90</sup> but are continuing in the form of a new estimate referred to as  $V^*(\lambda)$ .<sup>128,129</sup> The  $V^*(\lambda)$  (green line) and the CIE 1924  $V(\lambda)$  (green circles) luminous efficiency functions for centrally viewed fields of 2° in visual diameter, and the  $V_{10}^*(\lambda)$  (orange line) and the CIE 1964  $V_{10}(\lambda)$  (orange squares) luminous efficiency functions for centrally viewed fields of 10° in visual diameter can be compared in Fig. 19. The functions differ mainly at short wavelengths. The  $V^*(\lambda)$  and  $V_{10}^*(\lambda)$  functions have been recommended by the CIE for use in “physiologically relevant” colorimetric and photometric systems.<sup>130</sup>

In general, luminous efficiency functions can be approximated by a linear combination of the L-cone  $[\bar{l}(\lambda)]$  and M-cone  $[\bar{m}(\lambda)]$  spectral sensitivities, thus:  $V(\lambda) = a\bar{l}(\lambda) + \bar{m}(\lambda)$ , where  $a$  is the L-cone weight relative to the M-cone weight. Luminous efficiency shows sizeable individual differences even after individual differences in macular and lens pigmentation have been taken into account. For example, under neutral adaptation, such as daylight D65 adaptation, the L-cone weight in a group of 40 subjects, in whom 25-Hz HFP spectral sensitivity data were measured on a 3 log td white D65 background, varied from 0.56 to 17.75, while the mean HFP data were consistent with an L-cone weight of 1.89.<sup>128,129</sup>



**FIGURE 19** Luminous efficiency functions for  $2^\circ$  and  $10^\circ$  central viewing conditions. The lower pair of curves compare the CIE 1924  $2^\circ V(\lambda)$  function (green circles) with the  $2^\circ V^*(\lambda)$  function<sup>128,129</sup> (green line). The upper pair of curves compare the CIE 1964  $10^\circ V_{10}(\lambda)$  function (orange squares) with the  $10^\circ V^*(\lambda)$  function (orange line). The main discrepancies are found at short wavelengths.

*Cone numerosity* On the assumption that the variation in L:M cone contribution to luminous efficiency is directly related to the variation in the ratio of L:M cone numbers in the retina, several investigators have used luminous efficiency measurements to estimate relative L:M cone numerosity.<sup>125,131–141</sup> It is important to note, however, that luminous efficiency depends on factors other than relative L:M cone number, and is affected, in particular, by both first-site and second-site adaptation.<sup>63,64,142</sup> Indeed, the L:M contribution to luminous efficiency could in principle have little or nothing to do with the relative numbers of L and M cones, but instead reflect the relative L- and M-cone contrast gains set in a manner independent of cone number. Such an extreme view seems unlikely to hold, however. Population estimates of relative L:M cone numerosity obtained using luminous efficiency, correlate reasonably well with estimates obtained using direct imaging,<sup>30</sup> flicker ERG,<sup>143,144</sup> and other methods.<sup>134,137–139,145–148</sup> More precise delineation of the relation between luminous efficiency and relative L:M cone numerosity awaits further investigation. Note that an important piece of this enterprise is to incorporate measurements of individual variation in L- and M-cone spectral sensitivity.<sup>128,144,146</sup>

*Multiple luminance signals* As developed later in subsection “Multiplexing Chromatic and Achromatic Signals” in Sec. 11.5, cone-opponent, center-surround mechanisms that are also spatially opponent, such as parvocellular or P-cells, encode not only chromatic information, but also luminance information (see Fig. 38). The fact that P-cells transmit a luminance signal suggests that there may be two distinct “luminance” signals at the physiological level: one generated by P-cells and another by magnocellular or M-cells, a possibility which has been discussed in several papers by Ingling and his coworkers.<sup>149–152</sup> The need for a P-cell-based luminance system is also suggested by the fact that the retinal distribution of M-cells is too coarse to support the observed psychophysical spatial acuity for luminance patterns<sup>153</sup> (but see Ref. 154). Others have also suggested that multiple mechanisms might underlie luminance spectral sensitivity functions.<sup>124</sup>

Luminance mechanisms based on M-cell and P-cells should have different spatial and temporal properties. Specifically, the mechanism more dependent on M-cell activity, which we refer to as  $(L+M)_{M'}$ , should be more sensitive to high temporal frequencies, whereas the mechanism more dependent on



P-cell activity, which we refer to as  $(L+M)_p$ , should be more sensitive to high spatial frequencies. If, as has been suggested,<sup>149–152</sup>  $(L+M)_p$  is multiplexed with L–M, we might also expect the interactions between  $(L+M)_p$  and L–M to be different from those between  $(L+M)_M$  and L–M. Similarly, chromatic noise may have a greater effect on  $(L+M)_p$  than on  $(L+M)_M$ . Note, however, that if the stimulus temporal frequency is high enough for the spatially opponent surround to become synergistic with the center (i.e., when the surround delay is half of the flicker period), the P-cell will, in principle, become sensitive to any low spatial frequency luminance components of the stimulus.<sup>151,155</sup>

Psychophysical evidence for multiple luminance mechanisms is relatively sparse. Given the likely difference in the M- and L-cone weights into  $(L+M)_M$ , the conventional luminance mechanism, and into L–M, the chromatic mechanism [from which  $(L+M)_p$  is assumed to derive], the spectral sensitivities for spatial acuity tasks, which favor  $(L+M)_p$ , and for flicker tasks, which favor  $(L+M)_M$ , should be different. However, little difference in spectral sensitivity is actually found.<sup>122,156–158</sup> Ingling and Tsou<sup>152</sup> speculate that this similarity is found despite there being different luminance mechanisms, because the spatial acuity task may depend on P-cell centers rather than on centers and surrounds. Averaged over many cells, the P-cell *center* spectral sensitivity may be similar to the M-cell sensitivity, particularly if the latter depends on relative cone numerosity (see previous subsection).

Webster and Mollon<sup>17</sup> looked at the effect of correlated chromatic and luminance contrast adaptation on four different measures of “luminous efficiency.” They found that lightness settings and minimum-motion settings for 1-Hz counter-phase gratings were strongly biased by contrast adaptation, but that flicker settings and minimum-motion settings for 15-Hz counter-phase gratings were unaffected. On the face of it, these results are consistent with the idea that different luminance mechanisms mediate low- and high-temporal frequency tasks.

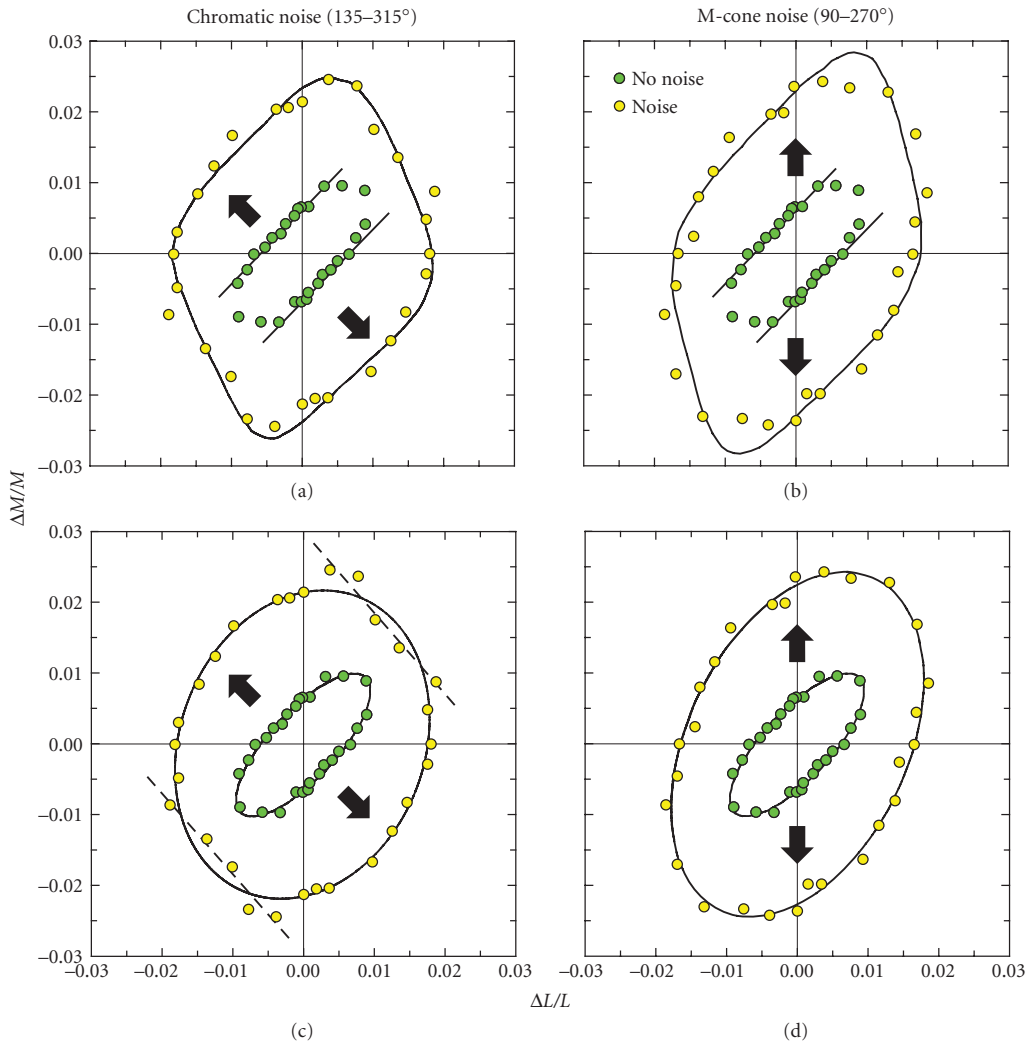
We speculate that some of the discrepancies found between the detection contours measured in the L,M plane might be related to detection being mediated by different luminance mechanisms under different conditions. Conditions chosen to favor the conventional luminance channel (e.g., the presence of chromatic noise and/or the use of targets of high temporal frequency) are likely to favor detection by  $(L+M)_M$ . Perhaps, under these conditions the  $(L+M)_M$  luminance contours appear as distinct segments, because there is relatively little threshold summation with L–M (e.g.,  $k = 4$ ). Examples of such segments are shown by the dashed lines in Fig. 20c. Conditions *not* specially chosen to favor the conventional luminance channel (e.g., the use of targets of high spatial and/or low temporal frequencies) may favor instead detection by  $(L+M)_p$ . Perhaps, under these conditions the luminance  $(L+M)_p$  contours form elliptical contours with L–M ( $k = 2$ ). Examples of such contours are shown in Fig. 20d. Noorlander, Heuts, and Koenderink,<sup>83</sup> however, reported that elliptical contours were found across all spatial and temporal frequencies.

**Sensitivity to Different Directions of Color Space** Before considering results obtained using trichromatic test mixtures, we first consider earlier results obtained using bichromatic mixtures.

*Sensitivity to bichromatic test mixtures* The use of spectral lights restricts the range of measurements to the spectrum locus in color space, thus ignoring the much larger and arguably environmentally more relevant volume of space that includes desaturated colors, achromatic lights, and extra-spectral colors (such as purples). Measuring detection sensitivity for mixtures of spectral targets probes the inner volume of color space. Studying these “bichromatic mixtures,” moreover, can provide a direct test of the additivity of a mechanism. For example, under conditions that isolate a univariant linear mechanism, pairs of lights should combine additively, so that the observer should be more sensitive to the combination than to either light alone. For a cone-opponent mechanism, on the other hand, pairs of lights that oppositely polarize the mechanism should combine subadditively, so that the observer is less sensitive to the mixture than to either component (as illustrated in Fig. 6d). If different mechanisms detect the two targets, then other potential interactions include probability summation or gating inhibition (winner takes all). Boynton, Ikeda, and Stiles<sup>100</sup> found a complex pattern of interactions between various test fields that included most of the possible types of interactions.

Several workers have demonstrated subadditivity for mixtures of spectral lights that is consistent with either L–M or S–(L+M) cone opponency.<sup>100,106,120,159–165</sup> Clear examples of subadditivity were





**FIGURE 20** Detection contours in the  $\Delta L/L$ ,  $\Delta M/M$  plane. Detection thresholds for a 1-cpd Gabor replotted from Fig. 5 of Giulianini and Eskew<sup>175</sup> obtained with (yellow circles) or without (green circles) noise made up of superposed flickering rings modulated along the  $\Delta L/L = -\Delta M/M$  (chromatic) axis (a and c) or the  $\Delta M/M$  (M-cone) axis (b and d). The arrows indicate the noise directions. (a) and (b). Model fits (solid lines) by Giulianini and Eskew<sup>175</sup> in which L–M chromatic detection is assumed to be mediated by a mechanism with relative cone contrast weights of 0.70  $\Delta L/L$  and 0.72  $\Delta M/M$  of opposite sign, while L+M achromatic detection is mediated by a mechanism with relative weights of 0.90  $\Delta L/L$  and 0.43  $\Delta M/M$  of the same sign. Without chromatic noise, the detection data are consistent with detection solely by the chromatic mechanism (as indicated by the parallel lines) except along the luminance axis. With added chromatic noise, the sensitivity of the chromatic mechanism is reduced, revealing the achromatic mechanism in the first and third quadrants. In the presence of noise, the data are consistent with detection by both chromatic and achromatic mechanisms (as indicated by the rounded parallelogram). The combination exponent for the two mechanisms,  $k$ , was assumed to be 4 (see subsection “How Test Measurements Imply Opponency” in Sec. 11.3). (c) and (d) Alternative fits to the data using elliptical contours, which provide a surprisingly good fit to the detection data, except in the achromatic direction with chromatic noise (c). The ellipse fitted to the yellow circles in (c) has the formula  $3332x^2 + 2021y^2 + 873xy = 1$  (major axis  $73.16^\circ$ ), while that fitted to the green circles has the formula  $27,191x^2 + 24,842y^2 + 43,005xy = 1$  (major axis  $46.56^\circ$ ), where  $x = \Delta L/L$  and  $y = \Delta M/M$ . Those data with noise that are poorly accounted for by the ellipse have been fitted with a straight line of best-fitting slope  $-1.14$  (dashed line). The ellipse fitted to the yellow circles in (d) has the formula  $3648x^2 + 1959y^2 + 1992xy = 1$  (major axis  $65.15^\circ$ ), while that fitted to the green circles has the formula  $26,521x^2 + 23,587y^2 + 38,131xy = 1$  (major axis  $47.20^\circ$ ).

obtained by Thornton and Pugh,<sup>164</sup> who presented targets chosen to favor chromatic detection on a white (xenon-arc) background. Figure 18a shows the test spectral sensitivity obtained on the white field characteristic of chromatic detection. The spectral positions of the test mixture primaries used in the additivity experiment are indicated by the arrows. Figure 18b shows the results of the text mixture experiment. Detection contours aligned with the parallel diagonal contours (dashed blue lines) is consistent with chromatic detection by an L–M mechanism, whereas detection along the 45° vector (parallel to the contours) is consistent with achromatic detection by an L+M mechanism. Thornton and Pugh<sup>164</sup> also used test primaries of 430 and 570 nm, which were chosen to favor detection by the S cones and by luminance (L+M), respectively. Although less clear than the L–M case shown here, they found evidence for inhibition between the 430- and 570-nm target consistent with detection by an S–(L+M) cone-opponent mechanism.

Guth and Lodge<sup>11</sup> and Ingling and Tsou<sup>12</sup> both used models that incorporated opponency to analyze bichromatic and spectral thresholds. Both found that the threshold data could be accounted for by two opponent-color mechanisms [which can be approximated to the basic cone-opponent discrimination mechanisms, L+M and S–(L+M)], and a single nonopponent mechanism (L+M). Both also developed models that could better account for “suprathreshold” data, such as color valence data, by modifying their basic models. These modifications took the form of an increase in the contribution of the S–(L+M) mechanism and either inhibition between the two cone-opponent mechanisms<sup>11</sup> or addition of an S-cone contribution to the L-cone side of the L–M mechanism.<sup>12</sup> These suprathreshold models are related to the color-appearance models discussed in subsection “Color Appearance and Color Opponency” in Sec. 11.5. It is worth noting that Guth and Lodge acknowledge that modulations of their cone-opponent discrimination mechanisms do not produce strictly red–green or yellow–blue appearance changes, but rather yellowish–red versus blue–green and greenish–yellow versus violet. That is, their threshold discrimination mechanisms do not account for unique hues (see subsection “Spectral Properties of Color-Opponent Mechanisms” in Sec. 11.5).

Kranda and King-Smith<sup>106</sup> also modeled bichromatic and spectral threshold data, but they required four mechanisms: an achromatic mechanism (L+M), two L–M cone-opponent mechanisms of opposite polarities (L–M and M–L), and an S-cone mechanism (S). The main discrepancy from the previous models is that they found no clear evidence for an opponent inhibition of S by L+M. Two opposite polarity, L–M mechanisms were required because Kranda and King-Smith assumed that only the positive lobe contributed to detection, whereas previous workers had assumed that both negative and positive lobes could contribute. This presaged the unipolar versus bipolar mechanism debate (see “Unipolar versus Bipolar Chromatic Mechanisms” in Sec. 11.6).

With the exception of Boynton, Ikeda, and Stiles,<sup>100</sup> most of the preceding studies used incremental spectral test lights, which confined the measurements to the eighth volume of color space in which all three cone signals increase. To explore other directions in color space, both incremental and decremental stimuli must be used.

*Detection contours in the L,M plane* The first explicit use of incremental and decremental stimuli in cone contrast space was by Noorlander, Heuts, and Koenderink,<sup>83</sup> who made measurements in the L,M plane. They fitted their detection contours with ellipses, and found that the alignment of the minor and major axes depended on temporal and spatial frequency. At lower frequencies, the minor and major axes aligned with the chromatic ( $\Delta M/M = -\Delta L/L$ ) and achromatic ( $\Delta M/M = \Delta L/L$ ) directions, respectively. By contrast, at higher frequencies, the reverse was the case. Such a change is consistent with relative insensitivity of chromatic mechanisms to higher temporal and spatial frequencies (see subsection “Spatial and Temporal Contrast Sensitivity Functions” in Sec. 11.5).

If the individual detection contours are truly elliptical in two-dimensional color space or ellipsoidal in three, then the vectors of the underlying linear detection mechanisms cannot be unambiguously defined. This ambiguity arises because ellipses or ellipsoids can be linearly transformed into circles or spheres, which do not have privileged directions.<sup>165</sup> Consequently, whether or not detection contours are elliptical or ellipsoidal has become an important experimental question. Knoblauch and Maloney<sup>166</sup> specifically addressed this question by testing the shapes of detection contours measured in two color planes. One plane was defined by modulations of the red and green phosphors of their display (referred to by them as the  $\Delta R, \Delta G$  plane), and the other by the modulations of the red and green phosphors together and the blue phosphor of their display (referred to as the  $\Delta Y, \Delta B$  plane). The

detection contours they obtained did not differ significantly in shape from an elliptical contour.<sup>166</sup> Note, however, that a set of multiple threshold contours, each ellipsoidal but measured under related conditions, can be used to identify mechanisms uniquely, given a model about how the change in conditions affects the contours (see, e.g., Ref. 167; discussed in subsection “Spatial and Temporal Contrast Sensitivity Functions” in Sec. 11.5).

Deviations from an ellipse, if they do occur, are most likely to be found under experimental conditions that maximize the probability that different portions of the contour depend on detection by different mechanisms. Such a differentiation is unlikely to be found under the conditions used by Knoblauch and Maloney.<sup>166</sup> At lower temporal frequencies (1.5-Hz sine wave or a Gaussian-windowed,  $\sigma = 335$  ms, 1.5-Hz cosine wave), the detection contour in the  $\Delta R, \Delta G$  phosphor space is likely to be dominated by the much more sensitive L–M mechanisms, whereas the contour in  $\Delta Y, \Delta B$  space is likely to depend upon detection by all three mechanisms [L–M, L+M, and S–(L+M)]. A comparable conclusion about the shape of threshold contours was reached by Poirson, Wandell, Varner, and Brainard,<sup>168</sup> who found that color thresholds could be described well by ellipses, squares, or parallelograms. Again, however, they used long duration (a Gaussian envelope lasting 1 s, corresponding to +5 to –5 standard deviations) and large (2° diameter) target that would favor detection by L–M.

The relatively high sensitivity of the L–M mechanism to foveally viewed, long-duration flashed targets<sup>169–171</sup> means that the L–M mechanism dominates the detection contours unless steps are taken to expose the L+M luminance mechanisms, such as by the use of flickering targets,<sup>63,172</sup> very small targets,<sup>173</sup> or chromatic noise (see subsection “Noise-Masking Experiments” in Sec. 11.5). Detection threshold contours measured in the L,M plane that expose both L–M and L+M contours are arguably better described by a parallelogram with rounded corners rather than an ellipse, with the parallel sides of positive slope corresponding to detection by the L–M chromatic mechanism and those of negative slope corresponding to detection by the L+M luminance mechanism.<sup>61,63,173,174</sup> Examples of detection contours that appear nonelliptical include parts of Figs. 3 and 4 of Ref. 63; Fig. 3, 5, and 7 of Ref. 175; and Fig. 2 of Ref. 176.

Figure 20 shows the detection data replotted from Fig. 5 of Ref. 175. The data are shown twice: once in the upper panels (a, b) and again in the lower panels (c, d). The data shown as green circles were measured without noise. The data shown as yellow circles in Fig. 20a and c were measured in the presence of chromatic noise, while those in Fig. 20b and d were measured in the presence of M-cone noise. The noise directions are indicated by the black arrows.

If, for the sake of argument, we accept that the mechanism vectors of the detection mechanisms can be unambiguously defined from detection contours (as illustrated in Fig. 20a and b), it is possible to estimate their cone weights and therefore their spectral sensitivities. The chromatic L–M detection contours in Fig. 20a and b obtained without noise (green circles) have been fitted with lines of parallel slopes of approximately 1.0, which indicates that the L–M chromatic mechanism responds to the linear difference of the L- and M-cone *contrast* signals;  $|c\Delta L/L - d\Delta M/M| = \text{constant}$ , where  $c = d$ .<sup>61,63,173,174</sup> The equality of the L- and M-cone weights, unlike the underlying relative L- and M-cone numerosities (e.g., Ref. 30), shows a remarkable lack of individual variability across observers (see Ref. 89). Similarly, color appearance settings (see subsection “Color Appearance and Color Opponency” in Sec. 11.5) in the form of unique yellow settings show much less variability than would be expected if unique yellow depended on the relative number of L and M cones in the retina.<sup>146</sup>

As suggested by the yellow circles in Fig. 20a and c, the use of a high frequency flickering target in chromatic noise exposes more of the L+M contour in the first and third quadrants. Unlike the L–M contour, the L+M contour has a negative slope, which indicates that the L+M achromatic chromatic mechanism responds to the linear sum of the L- and M-cone *contrast* signals;  $|a\Delta L/L + b\Delta M/M| = \text{constant}$ , where  $a$  is typically greater than  $b$ .<sup>61,63,89,173,174</sup>

*Mechanism interactions* The shapes of the detection contours when two or more postreceptoral mechanisms mediate threshold depends upon the way in which those mechanisms interact (see Fig. 6c). In principle, such interactions can range from complete summation, through probability summation to exclusivity (winner takes all) and inhibition (see Ref. 100). The usual assumption is that L–M, S–(L+M), and L+M are stochastically independent mechanisms that combine by probability summation. That is, if two or three mechanisms respond to the target, then the threshold will be lower than if only one responds to it. How much the threshold is lowered depends upon the

underlying frequency-of-seeing curve for each mechanism. Steep frequency-of-seeing curves result in a relatively small drop in threshold, whereas shallow curves result in a larger drop. Several groups have mimicked probability summation by using the summation rule introduced in “How Test Measurements Imply Opponency” in Sec. 11.3. The summation exponent can be varied to mimic the effects of probability summation for psychometric functions with different slopes (see Refs. 174 and 89). A closely related geometrical description of the threshold contours was used by Sankeralli and Mullen.<sup>172</sup> For the fits in Fig. 20*a* and *b*, the summation exponent was assumed to be 4, which produces a contour shaped like a rounded parallelogram.<sup>89,175</sup>

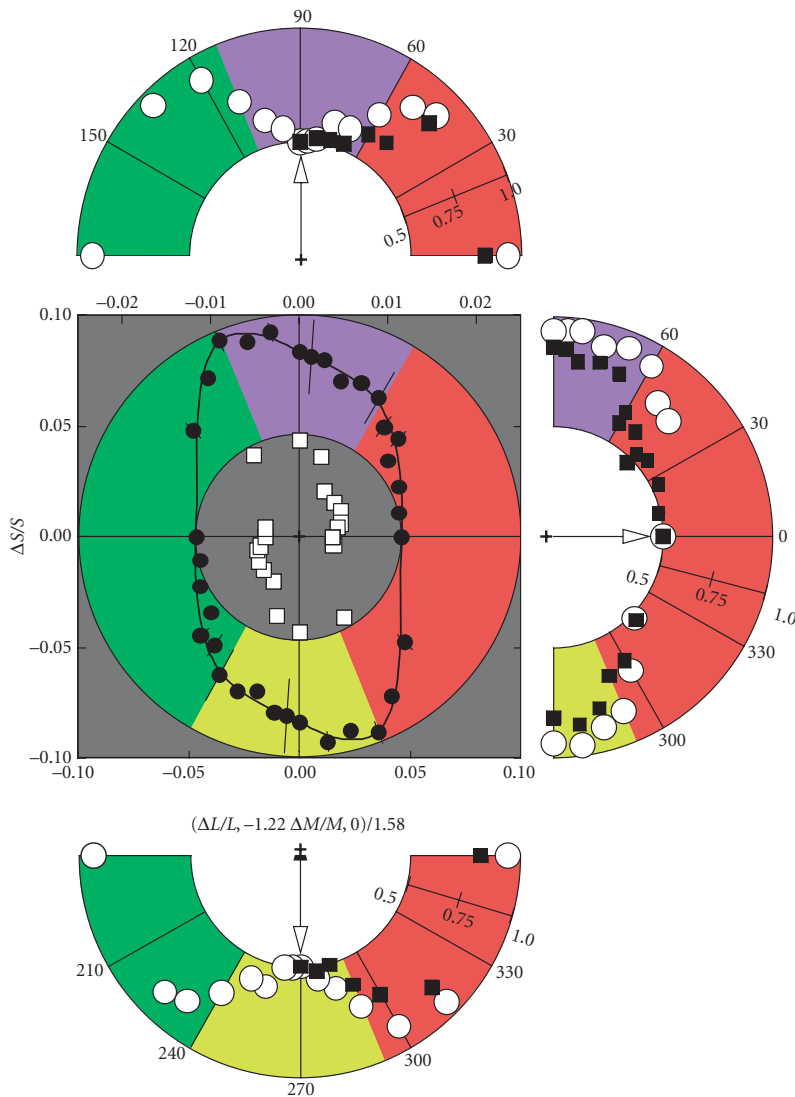
The superiority of a rounded parallelogram over an ellipse (summation exponent of 2) is not always clear. For example, Fig. 20*c* and *d* show the same detection data as the *a* and *b*. In *c* and *d*, the data have been fitted with ellipses. As can be seen, each set of data is well fitted by an ellipse, except perhaps for a small part of the contour measured in the presence of chromatic noise shown in Fig. 20*c*. In this region, we have fitted a straight line plotted as the dashed line. Changing the direction of the noise from L–M chromatic to M cone between Fig. 20*c* and *d* rotates the ellipse slightly anticlockwise. The L–M data along the opponent diagonals clearly follow an elliptical rather than a straight contour in both cases. As previously noted, when an elliptical representation underlies the data, it does not uniquely determine the inputs to the underlying mechanisms.

Support for the summation exponent of 4 used to fit the data in Fig. 20*a* and *b* is also lacking from estimates of the slopes of the underlying psychometric functions, which are consistent with much smaller exponents.<sup>171,177,178</sup>

*Detection in the other planes* An investigation of mechanisms other than L–M and L+M (as well as information about whether or not there are S-cone inputs into those mechanisms) requires the measurement of responses outside the L,M plane of color space. Two systematic investigations have been carried out that include S-cone modulations. Cole, Hine, and McIlhagga<sup>174</sup> made measurements on a roughly 1000 log td white background using a Gaussian-blurred 2°, 200-ms spot that favors chromatic detection. Their data were consistent with three independent mechanisms: an L–M mechanism with equal and opposite L- and M-cone inputs but no S-cone input; an L+M mechanism with unequal inputs and a small positive S-cone input, and an S–(L+M) mechanism. Sankeralli and Mullen<sup>172</sup> used three types of gratings with spatiotemporal properties chosen to favor either the chromatic L–M (1 cpd, 0 Hz), chromatic S–(L+M) (0.125 cpd, 0 Hz), or the achromatic L+M mechanism (1 cpd, 24 Hz). They confirmed that L–M had equal and opposite L- and M-cone inputs, but found a small S-cone input of about 2 percent added to either the L or the M cones for different observers. In addition, they found a small 5 percent S-cone input into the luminance mechanism that opposed the L+M input (see also Ref. 92). The S–(L+M) mechanism was assumed to have balanced opposed inputs.<sup>172</sup> Eskew, McLellan, and Giulianini<sup>89</sup> provide a separate analysis of both these sets of data (see Table 18.1 of Ref. 89).

Figure 21 shows chromatic detection data measured in the L–M,S plane by Eskew, Newton, and Giulianini.<sup>179</sup> The central panel shows detection thresholds with (filled circles) and without (open squares) L–M chromatic masking noise. Detection by the relatively insensitive S-cone chromatic mechanisms becomes apparent when detection by L–M chromatic mechanism is prevented by chromatic noise. The nearly horizontal portions of the black contour reflect detection by S–(L+M), while the vertical portions reflect detection by L–M (the model fit is actually of the unipolar mechanisms [L–M], [M–L], [S–(L+M)], and [(L+M)–S], see Ref. 34 for details; also “Unipolar versus Bipolar Chromatic Mechanisms” in Sec. 11.6).

The question of whether or not the S cones make a small contribution to the L–M detection mechanisms remains controversial. Boynton, Nagy, and Olsen<sup>180</sup> found that S cones show summation with L–M (+S with L–M, and –S with M–L) in chromatic difference judgments. Stromeyer et al.<sup>181</sup> showed that S-cone flicker facilitated detection and discrimination of L–M flicker, with +S facilitating L–M and –S facilitating M–L. Yet, the estimated S-cone contrast weight was only approximately 1/60th of the L- and M-cone contrasts weight into L–M (see p. 820 of Ref. 181). In general, in detection measurement the S-cone signal adds to redness (L–M) rather than greenness (M–L)—as expected from color-opponent theory (see Sec. 11.4), but the size of the contribution is much less than expected. Eskew and Kortick,<sup>20</sup> however, made both hue equilibria and detection measurements and estimated that the S-cone contrast weight was about 3 percent that of the L- and



**FIGURE 21** Detection and discrimination contours in the equiluminant plane. This figure is a reproduction of Fig. 5 from Eskew.<sup>34</sup> The central panel shows thresholds for the detection of Gabor patches with (filled circles) and without (open squares) L–M chromatic masking noise ( $0^\circ/180^\circ$  color direction). For clarity, the horizontal scale for the no-noise thresholds (open squares) has been expanded as indicated by the upper axis. Data are symmetric about the origin due to the symmetric stimuli. Detection by the relatively insensitive S-cone chromatic mechanisms becomes apparent when detection by the L–M chromatic mechanism(s) is suppressed by L–M chromatic noise. The black contour fitted to the filled circles is the prediction of a model in which detection is mediated by probability summation between four symmetric unipolar linear mechanisms [L–M], [M–L], [S–(L+M)], and [(L+M)–S], (see subsection “Unipolar versus Bipolar Chromatic Mechanisms” Sec. 11.6). The three outlying semicircular polar plots are discriminability data (filled squares), in which the angular coordinate corresponds to the same stimuli, at the same angles, as in the detection data. The radial coordinate gives the discriminability between 0.5 (chance) and 1.0 (perfect discriminability) of the given test color angle relative to the standard color angle indicated by the arrow. The open circles are predictions of a Bayesian classifier model that takes the outputs of the four mechanisms fitted to the detection data as its inputs. Colored regions indicate bands of poorly discriminated stimuli, and are redrawn on the detection plot for comparison. These results suggest that just four univariate mechanisms are responsible for detection and discrimination under these conditions. See Ref. 34 for further details. [Figure adapted by Eskew<sup>34</sup> from Fig. 4 (Observer JRN) of Ref. 179 Reprinted from *The Senses: A Comprehensive Reference, Volume 2: Vision II*, Fig. 5 from R. T. Eskew, Jr., “Chromatic Detection and Discrimination,” pp. 109–117. Copyright (2008), with permission from Elsevier.]

M-cone contrast weights for both tasks. Yet, although test methods support an S-cone contribution to L–M, field methods do not (see Ref. 34). Changes in the S-cone adaptation level have little effect on L–M detection<sup>182</sup> or on L–M mediated wavelength discrimination.<sup>183</sup>

Given the uncertainties about cone spectral sensitivities and luminous efficiency at short wavelengths, as well as the substantial individual differences in prereceptor filtering between observers (see, e.g., Ref. 90), evidence for small S-cone inputs should be treated with caution.

**Spatial and Temporal Contrast Sensitivity Functions** The test method can also be used to determine the temporal and spatial properties of visual mechanisms. These properties are traditionally characterized by temporal and spatial “contrast sensitivity functions” (CSFs). These are sometimes incorrectly called “modulation transfer functions” (MTFs)—incorrectly because MTFs require phase as well as amplitude specifications.

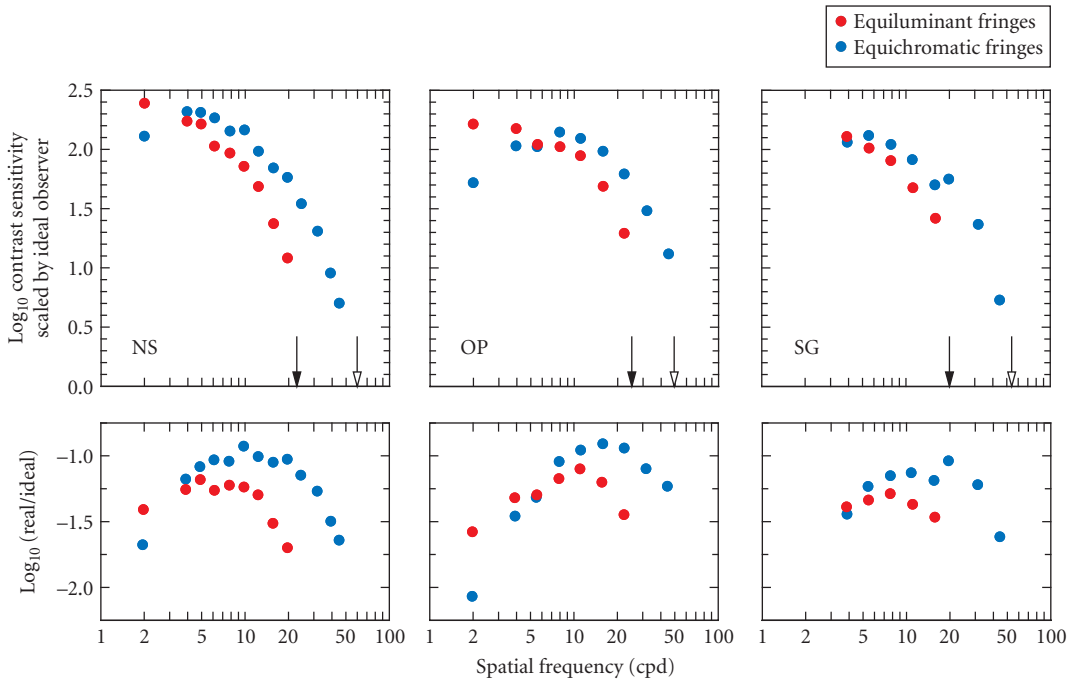
In a temporal CSF determination, the observer’s sensitivity for detecting sinusoidal flicker is measured as a function of temporal frequency. Generally, the chromatic mechanisms have a lowpass temporal frequency response with typically greater temporal integration and a poorer response to higher temporal frequencies than the more band-pass achromatic mechanism.<sup>94,102,184–190</sup> A lowpass CSF means that the temporal response is approximately independent of frequency at low frequencies but falls off at higher frequencies, whereas a band-pass CSF means that the response peaks at an intermediate frequency and falls off at both lower and higher frequencies.

On the grounds that different chromatic flicker frequencies can be discriminated near threshold, Metha and Mullen<sup>190</sup> suggest that at least two chromatic mechanisms with different temporal CSFs must underlie the measured chromatic CSF, one with a band-pass CSF and the other with a lowpass CSF. They argue that two mechanisms must operate, because a single *univariant* flicker mechanism will confound changes in contrast and flicker frequency, making frequency identification impossible. However, the assumption that flicker is encoded univariantly may be flawed. At low frequencies, flicker may be encoded as a moment-by-moment variation that matches the flicker frequency.<sup>191</sup>

In a spatial CSF determination, the observer’s sensitivity for detecting sinusoidal gratings is measured as a function of their spatial frequency. Generally, the chromatic mechanisms have a lowpass spatial frequency response with greater spatial integration and a poorer response to higher spatial frequencies than the usually more band-pass achromatic mechanisms.<sup>102,192–197</sup>

Spatial CSFs are degraded by sensitivity losses introduced by the optics of the eye. In one study, laser interference fringes were used to project gratings directly on the retina, thereby effectively bypassing the MTF of the eye’s optics.<sup>198,199</sup> However, because the red and green equiluminant interference fringes could not be reliably aligned to produce a steady artifact-free chromatic grating, they were drifted in opposite directions at 0.25 Hz. Thus, the gratings continuously changed from a red-green chromatic (equiluminant) grating to a yellow-black luminance (equichromatic) grating. The subject’s task was to set the contrast threshold separately for the different spatial phases, which was apparently possible.<sup>199</sup> One concern about this method, however, is that the close juxtaposition of chromatic and luminance gratings in time and space might produce facilitation or interference (see subsection “Pedestal Experiments” in Sec. 11.5). For S-cone detection, a single grating was used.

The results for equiluminant, chromatic (red circles) and equichromatic, achromatic (blue circles) fringes are shown for three observers in the upper panels of Fig. 22. As expected, there is a steeper fall-off in high spatial-frequency sensitivity for chromatic gratings than for luminance gratings. By estimating the sensitivity losses caused by known factors such as photon noise, ocular transmission, cone-aperture size, and cone sampling, Sekiguchi, Williams, and Brainard<sup>198</sup> were able to estimate an ideal observer’s sensitivity at the retina. Any additional losses found in the real data are then assumed to be due to neural factors alone. Estimates of the losses due to neural factors are shown in the lower panels of Fig. 22. Based on these estimates, the neural mechanism that detects luminance gratings has about 1.8 times the bandwidth of the mechanisms that detect chromatic gratings. Interestingly, the neural mechanisms responsible for the detection of equiluminant L–M gratings and S-cone gratings have strikingly similar (bandpass) contrast sensitivities, despite the very different spatial properties of their submosaics.<sup>200</sup> This commonality may have important implications for color processing (see “Three-Stage Zone Models” in Sec. 11.6).



**FIGURE 22** Spatial equiluminant and equichromatic contrast sensitivity functions. Top panels: Foveal CSFs for equiluminant (red circles) and equichromatic (blue circles) interference fringes for three observers. Filled and open arrows represent the foveal resolution limit for the equiluminant and the equichromatic stimuli, respectively. Bottom panels: The ratio of the ideal to the real observer's contrast sensitivity for detecting equiluminant (red circles) and equichromatic (blue circles) stimuli. (Based on Figure 9 of Ref. 199.)

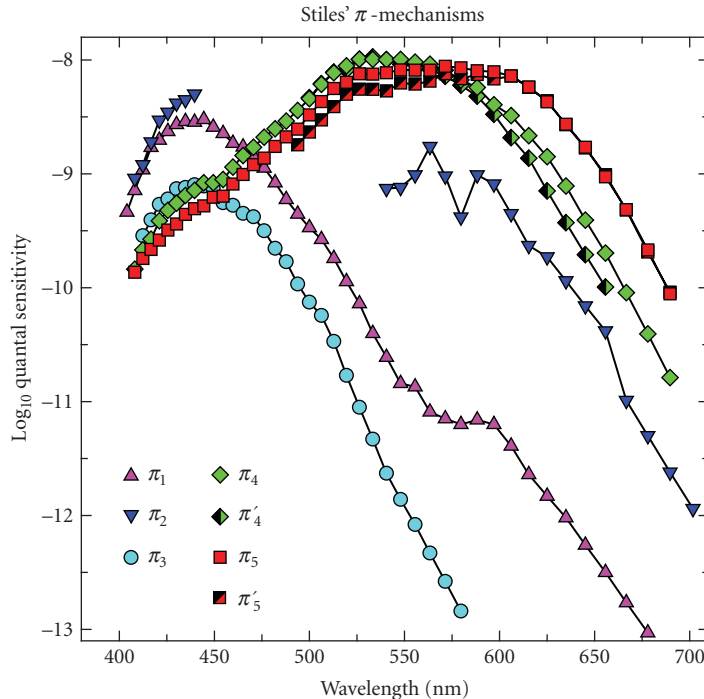
Poirson and Wandell<sup>167</sup> measured spatial CSFs along the two cardinal directions and along intermediate directions in color space; that is, they effectively measured threshold contours for a series of grating spatial frequencies. They fitted the contours with an ellipsoidal model, but added a constraint of pattern-color separability; that is, they assumed that the effect of pattern operated independently in each of three second-site mechanisms. This constraint resolves the mechanism ambiguity inherent in fitting ellipsoidal contours to the individual data, and allowed Poirson and Wandell to estimate the sensitivities of three underlying mechanisms by jointly analyzing the contours across spatial frequency. They found two opponent and one nonopponent mechanisms, with sensitivities consistent with the cardinal mechanisms, L+M, L-M, and S-(L+M) (see Fig. 6 of Ref. 167).

We next consider results obtained using the field method.

## Field Sensitivities

**Stiles'  $\pi$ -Mechanisms** Stiles used the field sensitivity method to define the spectral sensitivities of the  $\pi$ -mechanisms, which are illustrated in Fig. 23. The tabulated field sensitivities can be found in Table 2 (7.4.3) of Ref. 53 and Table B of Ref. 33. The field sensitivities are the reciprocals of the field radiances (in log quanta  $s^{-1} \text{ deg}^{-2}$ ) required to raise the threshold of each isolated  $\pi$ -mechanism by one log unit above its absolute threshold. The tabulated data are average data for four subjects: three females aged 20 to 30 and one male aged 51.



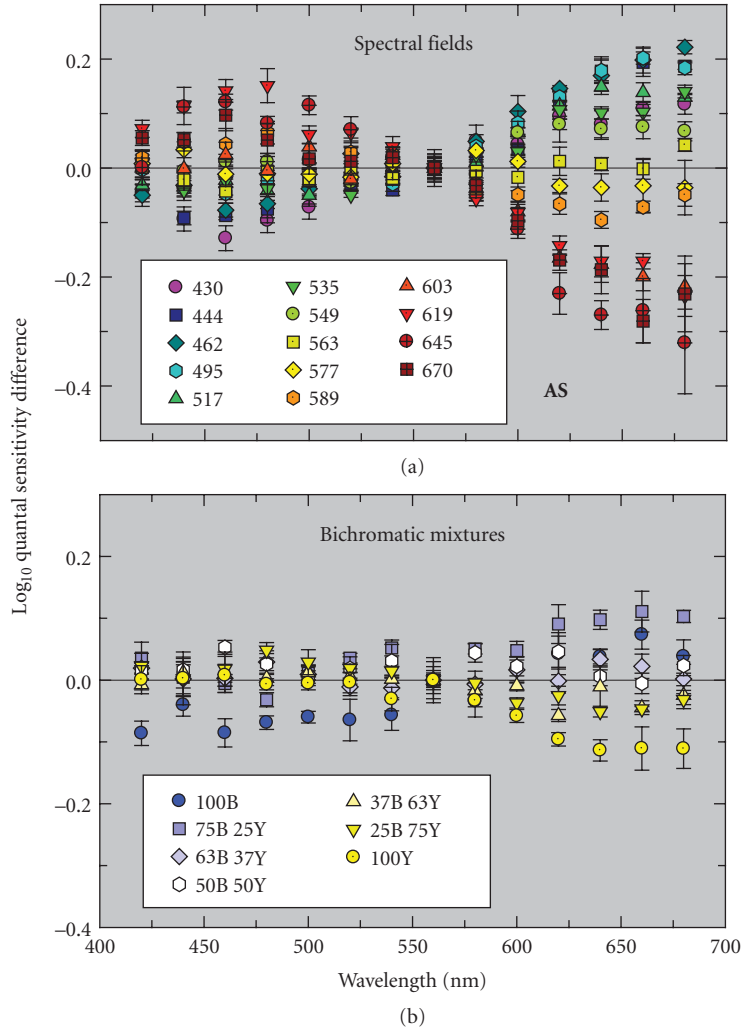


**FIGURE 23** Stiles'  $\pi$ -mechanisms. Field spectral sensitivities of the seven photopic  $\pi$ -mechanisms:  $\pi_1$  (purple triangles),  $\pi_2$  (dark blue inverted triangles),  $\pi_3$  (light blue circles),  $\pi_4$  (green diamonds),  $\pi'_4$  (green half-filled diamonds),  $\pi_5$  (red squares),  $\pi'_5$  (red half-filled squares). The field sensitivities are the reciprocals of the field radiances (in log quanta  $s^{-1} \text{ deg}^{-2}$ ) required to raise the threshold of each "isolated"  $\pi$ -mechanism by one log unit above its absolute threshold. The results were averaged across four subjects: three females aged 20 to 30 and one male aged 51. Details of the experimental conditions can be found on p. 12 of Stiles' book.<sup>33</sup> (Data from Table 2(7.4.3) of Wyszecki and Stiles.<sup>53</sup>)

Through field sensitivity measurements, and measurements of threshold versus background radiance for many combinations of target and background wavelength, Stiles identified seven photopic  $\pi$ -mechanisms: three predominately S-cone mechanisms ( $\pi_1$ ,  $\pi_2$ , and  $\pi_3$ ), two M-cone ( $\pi_4$  and  $\pi'_4$ ), and two L-cone ( $\pi_5$  and  $\pi'_5$ ). Although it has been variously suggested that some of the  $\pi$ -mechanisms might correspond to cone mechanisms,<sup>97,201–203</sup> it is now clear that all reflect some form of postreceptoral cone interaction. The surfeit of  $\pi$ -mechanisms and their lack of obvious correspondence to cone mechanisms or simple postreceptoral mechanisms led to the demise of Stiles' model. Its usefulness, however, was prolonged in the work of several investigators, who used failures of Stiles' model as a way of investigating and understanding the properties of chromatic mechanisms (see "Field Additivity" in this section). But, as noted in the parallel discussion in the context of test sensitivity, the potential ongoing value of the  $\pi$ -mechanism field sensitivities is that they provide an empirical database that may be used to constrain current models.

**Achromatic Detection and Chromatic Adaptation** Achromatic spectral sensitivity (or luminous efficiency) is strongly dependent on chromatic adaptation.<sup>63,64,142,204–210</sup> Figure 24a shows the changes in luminous efficiency caused by changing background field wavelength from 430 to 670 nm and Fig. 24b shows the changes caused by mixing 478- and 577-nm adapting fields in different luminance





**FIGURE 24** Dependence of luminous efficiency on chromatic adaptation. Changes in quantal luminous efficiency caused by changes in adapting field chromaticity plotted relative to the mean luminous efficiency. Data from Observer S1 of Stockman, Jägle, Pirzer and Sharpe.<sup>129</sup> Measurements were made using 25-Hz heterochromatic flicker photometry (HFP) on 1000-td backgrounds. (a) HFP changes on fourteen 1000-td spectral adapting fields from 430 to 670 nm and (b) changes on seven 1000 td bichromatic, 478(B) + 577(Y)-nm adapting field mixtures that varied from 100 percent B to 100 percent Y.

ratios. These changes are consistent with selective adaptation reducing L-cone sensitivity relative to M on longer wavelength fields, and reducing M-cone sensitivity relative to L on shorter wavelength fields. Between about 535 and 603 nm the selective changes are roughly consistent with Weber's law ( $\Delta I/I = \text{constant}$ ). Consequently, plotted in cone contrast space, the sensitivity contour for the achromatic mechanism should remain of constant slope between field wavelengths of 535 and 603 nm, *unlike* the example shown in Fig. 8b. For background wavelengths shorter than 535 nm,

on the other hand, the changes in spectral sensitivity are consistent with a relative loss of M-cone sensitivity in excess of that implied by Weber's law (for details, see Ref. 129). This is consistent with other evidence that also shows that the luminance contribution of the L- or M-cone type more sensitive to a given chromatic field can be suppressed by more than Weber's law implies.<sup>63,64,142</sup> Such supra-Weber suppression, illustrated in Fig. 8*b* for the L-cone case, causes the luminance sensitivity contours in cone contrast space to rotate away from the axis of the more suppressed cone.

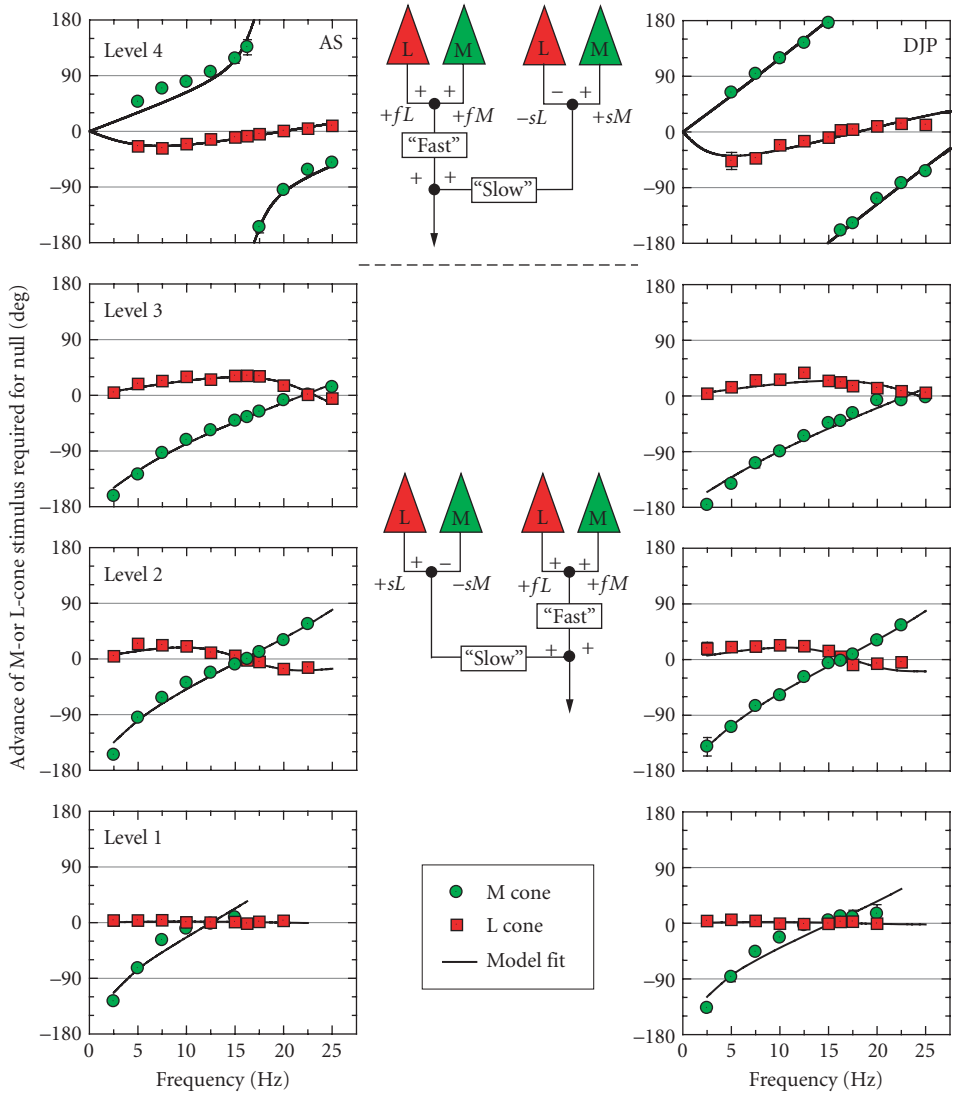
*Multiple cone inputs* There is now good psychophysical evidence that the standard model of the luminance channel as an additive channel with just fast L- and M-cone inputs (referred to here as  $+fL+fM$ ) is incomplete. According to this model, pairs of sinusoidally alternating lights that are "luminance-equated" and detected solely by the luminance channel should appear perfectly steady or nulled whatever their chromaticities (equating alternating lights to produce a null is known as heterochromatic flicker photometry or HFP). And, indeed, such nulls are generally possible, but sometimes only if moderate to large phase delays are introduced between the pairs of flickering lights.<sup>184,211–215</sup> Moreover, these large phase delays are often accompanied by substantial frequency-dependent changes in the spectral sensitivity of flicker detection and modulation sensitivity.<sup>216–218</sup> From the phase delays and sensitivity changes, it is possible to infer that, in addition to the faster  $+fM+fL$  signals, sluggish  $+sM-sL$ ,  $+sL-sM$ , and  $-sS$  signals also contribute to the luminance-signal nulls in a way that depends on adapting chromaticity and luminance.<sup>94,215–223</sup>

Figure 25 shows examples of large phase delays that are found for M- and L-cone-detected flickering lights on four increasingly intense 658 nm fields. In these plots,  $0^\circ$  means that the two lights cancelled when they were physically in opposite phase (i.e., when they were alternated), while  $\pm 180^\circ$  means that they cancelled when they were actually in the same phase. The differences between the plotted M- and L-cone phase delays therefore show the delays between the M- and L-cone signals introduced *within* the visual system. As can be seen, some of the phase delays are substantial even at moderately high temporal frequencies, particularly for the M-cone signals. Such delays are inconsistent with the standard model of luminance, which, except for phase differences caused by the selective adaptation of the L cones by the long-wavelength field, predicts that no phase adjustments should be required. Notice also the abrupt changes in phase *between* intensity levels 3 and 4.

The continuous lines in Fig. 25 are fits of a vector model, in which it was assumed that sluggish and fast signals both contribute to the resultant M- and L-cone signals, but in different ratios depending on the cone type and intensity level. At levels 1 to 3, the dominant slow and fast signals are  $+sL-sM$  and  $+fL+fM$ , respectively, as illustrated in the lower neural circuit diagram of Fig. 25, whereas at level 4 they are  $-sL+sM$  and  $+fL+fM$ , as illustrated in the upper neural circuit diagram. According to the model, the polarities of the both the slow M- and the slow L-cone signals reverse between levels 3 and 4.

Note that although the sluggish signals  $+sM-sL$  and  $+sL-sM$  are spectrally opponent, they produce an achromatic percept that can be flicker photometrically cancelled with luminance flicker, rather than producing an R/G chromatic percept.

*Chromatic Adaptation and the Sloan Notch* Thornton and Pugh<sup>107</sup> measured test spectral sensitivity functions on fields of 560, 580, and 600 nm using a target that strongly favors chromatic detection. As illustrated in Fig. 17*a*, they found that the local minimum in spectral sensitivity, or Sloan notch, coincides with the field wavelength (i.e., it occurs when the target and background wavelengths are the same; i.e., homochromatic). As discussed in subsection "Sensitivity to Spectral Lights" in Sec. 11.5, the notch is thought to correspond to the target wavelength that produces a minimum or null in the L-M chromatic channel, so that the less-sensitive, achromatic channel takes over detection. For the notch to occur at the homochromatic target wavelength means that adaptation to the field has shifted the zero crossing or null in the cone-opponent mechanism to the field wavelength. Such a shift is consistent with reciprocal (von Kries) adaptation occurring independently in both the M and the L cones. This type of adaptation is plausible given that Weber's law behavior for detection will have been reached on the fairly intense fields used by Thornton and Pugh.<sup>33</sup> If first-site adaptation had fallen short of the proportionality implied by Weber's Law, then the Sloan notch would not have shifted as far as the homochromatic target wavelength.



**FIGURE 25** Large phase delays in the “luminance” channel. Phase advances of M-cone (green circles) or L-cone (red squares) stimuli required to flicker-photometrically null a 656-nm target for observers AS (left panels) and DJP (right panels) measured on 658-nm backgrounds of 8.93 (*Level 1*), 10.16 (*Level 2*), 11.18 (*Level 3*), or 12.50 (*Level 4*)  $\log_{10}$  quanta  $s^{-1} \text{ deg}^{-2}$ . The M-cone stimuli were alternating pairs of L-cone-equated 540 and 650 nm targets; and the L-cone stimuli were pairs of M-cone-equated 650- and 550-nm targets. The continuous lines are fits of a model in which the L- and M-cone signals are assumed to be the resultant of a fast signal ( $f$ ) and a delayed slow signal ( $s$ ) of the same or opposite sign. The predominant signals at each level are indicated by the wiring diagrams in the central insets (see Ref. 223).

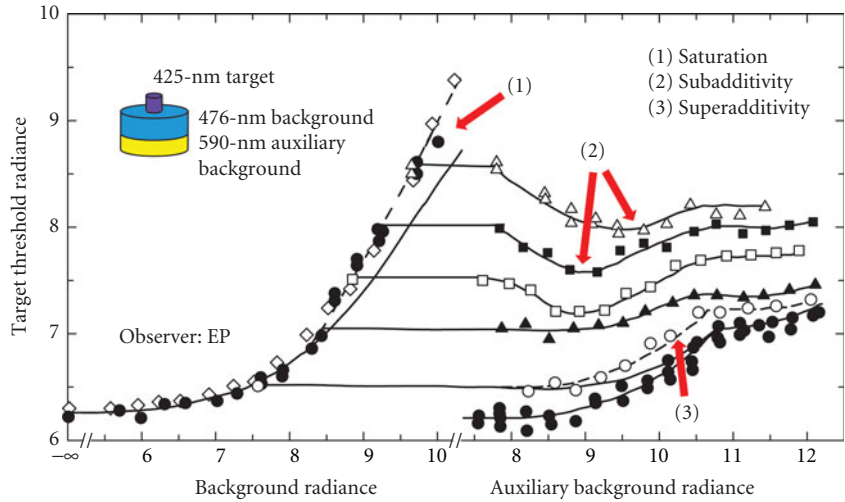
As noted in Sec. 11.3, the idea that the zero or null point of an opponent second site always corresponds to the background field wavelength is due to von Kries adaptation is clearly an oversimplification. Were that the case, chromatic detection contours measured in cone contrast units would be independent of field wavelength, which they clearly are not (see subsection “Detection Contours and Field Adaptation” in Sec. 11.5). Similarly, chromatic sensitivity would be independent of field chromaticity at intensities at which Weber’s law had been reached, which is also not the case in field additivity experiments (see next). In general, if first-site adaptation was precisely inversely proportional to the amount of background excitation, second-site opponent desensitization would play little or no role in adaptation to steady fields. It would, of course, still play an important role in contrast adaptation; that is, in adaptation to excursions around the mean adapting level.

**Field Additivity** By using mainly spectral backgrounds (but sometimes with fixed-wavelength auxiliary backgrounds), Stiles restricted his measurements to the spectrum locus of color space. Some of the most revealing failures of his model, however, occur in the interior volume of color space when two spectral backgrounds are added together in different intensity ratios. Given the expectation in Stiles’ model that  $\pi$ -mechanisms should behave univariantly, bichromatic field mixtures should raise the threshold of a  $\pi$ -mechanism simply in accordance with the total quantum catch generated by those fields in that mechanism. For a linear mechanism, the effects of mixed fields, in other words, should be additive whenever only a single  $\pi$ -mechanism is involved. However, failures of field additivity can be subadditive, the mixed fields have less than the expected effect, or superadditive, when they have more than the expected effect. Subadditivity is an indication of opponency, and cases of subadditivity are thus another way that field experiments provide evidence of opponency.

Arguably the most interesting failures of field additivity occur under conditions that isolate Stiles’ S-cone mechanisms,  $\pi_1$  and  $\pi_3$ , as illustrated in Fig. 26. On a blue, 476-nm background, the threshold for a 425 nm target rises faster than Weber’s law predicts (leftmost data set, dashed vs continuous lines), this is superadditivity and is consistent with S-cone saturation.<sup>224,225</sup> However, when a yellow, 590-nm background is added to the blue background, additivity typically fails. For lower 476-nm radiances (open circles, dashed vs continuous lines), superadditivity is found again,<sup>226</sup> whereas for higher 476-nm radiances (open and filled squares, open triangles) subadditivity is found.<sup>227,228</sup> Comparable failures of field additivity are found for the L-cone mechanism or  $\pi_5$ ,<sup>55,188</sup> but less clearly for the M-cone mechanisms or  $\pi_4$ .<sup>188,196,229,230</sup>

**First- and Second-Site Adaptation** Pugh and Mollon<sup>231</sup> developed a model to account for the failures of field additivity and other “anomalies” in the S-cone pathway, which has been influential. They assumed that the S-cone signal can be attenuated by gain controls at two sites. Gain at the first site is cone-specific, being controlled in the case of  $\pi_1$  and  $\pi_3$  solely by excitation in the S cones. By contrast, gain at the cone-opponent second-site is controlled by the net difference between signals generated by the S cones and those generated by the two other cone types. Excursions at the second site either in the S-cone direction or in the L+M-cone direction reduce the gain of that site, and thus attenuate the transmitted S-cone signal. Sensitivity at the second site is assumed to be greatest when the S-cone and the L+M-cone signals are balanced. Superadditivity occurs when attenuation at the first and second sites work together, for example on a blue spectral background. Subadditivity occurs when the addition of a background restores the balance at the second site, for example when a yellow background is added to a blue one. The original version of the Pugh and Mollon<sup>231</sup> model is formalized in their Eqs. (1) to (3). Another version of the model in which the second-site “gain” is implemented as a bipolar static nonlinearity (see Fig. 6 of Ref. 227) is illustrated in Fig. 27. The nonlinearity is a sigmoidal function that compresses the response and thus reduces sensitivity if the input is polarized either in the +S or -(L+M) direction or in the L+M or -S direction. The presence of a compressive nonlinearity in the output of mechanisms is now generally accepted as an important extension of the basic model. Such nonlinearities form a critical piece of models that aim to account for discrimination as well as detection data (see subsection “Pedestal Experiments” in Sec. 11.5).

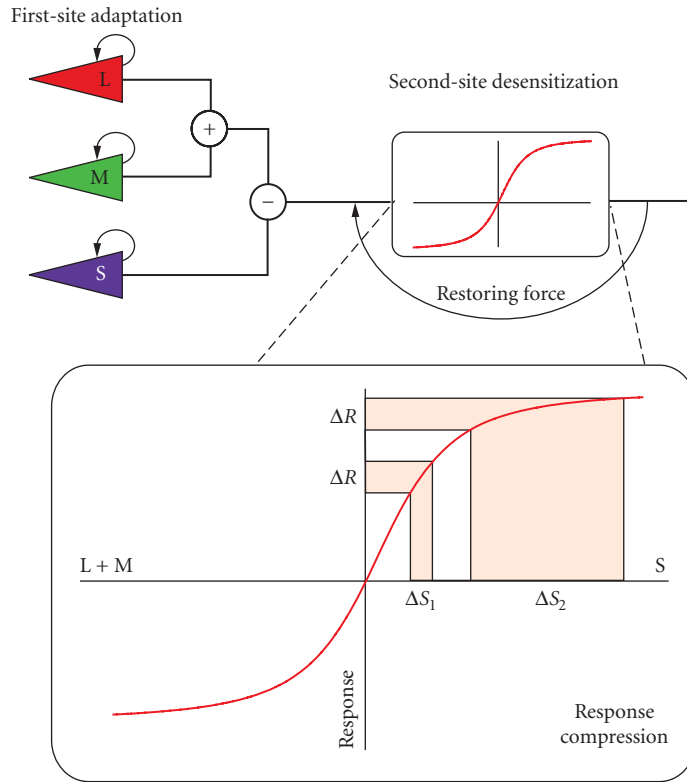
There is an inconsistency in the Pugh and Mollon scheme for steady-state adaptation. In order for first-site adaptation to be consistent with Weber’s law ( $\Delta I/I = \text{constant}$ ), sensitivity adjustment is assumed to be reciprocal—any increase in background radiance is compensated for by a reciprocal



**FIGURE 26** Field additivity failures. S-cone threshold data replotted from Fig. 1A of Pugh and Larimer.<sup>228</sup> The data are increment thresholds for a 425 nm, 200-ms duration, 1° diameter foveal target presented in the centre of 476-nm fields (filled circles and open diamonds, left curves), 590-nm fields (filled circles, right curves), and various bichromatic mixtures of the two. (A 590-nm field of at least  $8.0 \log_{10} \text{ quanta s}^{-1} \text{ deg}^{-2}$  was always added to the 476-nm field.) In the bichromatic mixture experiments a series of increasingly intense 590-nm fields, the intensities of which are indicated along the abscissa on the right, were added to a fixed intensity 476-nm field, the intensity of which corresponds to the abscissa associated with the dashed curve at target threshold radiance values corresponding to the data points at the lowest auxiliary background radiances. The shape of solid curves associated with the filled circles and open diamonds on the left and with the filled circles on the right is the standard Stiles threshold-versus-increment (tvi) template shape. The data provide examples of saturation (1), when the threshold elevation grows faster than Weber's law,<sup>224,225</sup> subadditivity (2), when the threshold elevation of the bichromatic mixture falls short of the additivity prediction,<sup>227</sup> and superadditivity (3), when the threshold elevation exceeds the additivity prediction.<sup>226</sup> Figure design based on Figs. 1A and 2 of Pugh and Larimer.<sup>228</sup> Units of radiance are  $\log \text{ quanta sec}^{-1} \text{ deg}^{-2}$ .

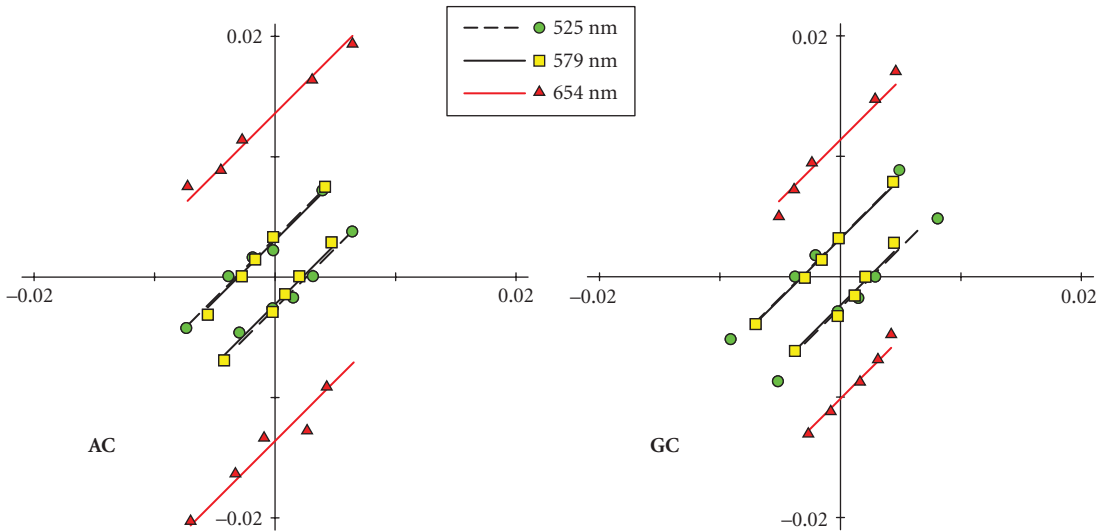
decrease in gain. Consequently, the signal that leaves the first site (and potentially reaches the second site) should be independent of background radiance (see also "First-Site Adaptation" in Sec. 11.3). However, for the Pugh and Mollon model (and comparable models of first- and second-site adaptation) to work requires that the cone signals reaching the second site continue to grow with background radiance. To achieve this in their original model, they decrease the S-cone gain at the second site in proportion to the background radiance raised to the power  $n$  (where  $n$  is 0.71, 0.75, or 0.82 depending on the subject) rather than in proportion to the background radiance as at the first site ( $n = 1$ ). Thus, the S-cone signal at the input to the second site continues to grow with background radiance, even though the signal at the output of the first site is constant. This scheme requires that the cone signal at second site somehow bypasses first-site adaptation, perhaps depending on signals at the border<sup>89</sup> or on higher temporal frequency components, which are not subject to Weber's law.<sup>52</sup>

A now mainly historical issue is how much adaptation occurs at the cone photoreceptors. Although some evidence has suggested that relatively little adaptation occurs within photoreceptors until close to bleaching levels,<sup>232,233</sup> other compelling evidence suggests that significant adaptation occurs at much lower levels.<sup>46–48</sup> Measures of adaptation in monkey horizontal cells show that light adaptation is well advanced at or before the first synapse in the visual pathway, and begins at levels as low as 15 td.<sup>234,235</sup> Moreover, psychophysically, local adaptation has been demonstrated to occur with the resolution of single cones.<sup>236–238</sup> The available evidence supports a receptor site of cone adaptation.



**FIGURE 27** Pugh and Mollon cone-opponent model. Version of the Pugh and Mollon model<sup>231</sup> in which the cone-opponent, second-site desensitization is implemented at a bipolar static nonlinearity.<sup>227</sup> The nonlinearity is a sigmoidal function that compresses the response as the input is polarized either in the +S [or  $-(L+M)$ ] direction or in the L+M or  $(-S)$  direction. The effect of the nonlinearity can be understood by assuming that the system needs a criterion change in response ( $\Delta R$ ) in order for a change in the input to be detected. At smaller excursions along the +S direction, a smaller change in the S-cone signal,  $\Delta S_1$ , is required to produce the criterion response than the change in signal,  $\Delta S_2$ , required at greater excursions. Sensitivity should be greatest when the polarization is balanced, and the system is in the middle of its response range.

**Detection Contours and Field Adaptation** Stromeyer, Cole, and Kronauer<sup>61</sup> measured cone contrasts in the L,M-cone plane (the plane defined by the  $\Delta L/L$  and  $\Delta M/M$  cone contrast axes, which includes chromatic L–M and achromatic L+M excursions) on spectral green, yellow, and red adapting backgrounds. Changing background wavelength moved the parallel chromatic contours for the detection of 200-ms flashes either outward (reducing sensitivity) or inward (increasing sensitivity). In general, background field wavelength strongly affects chromatic detection with sensitivity being maximal on yellow fields, declining slightly on green fields, and declining strongly on red fields. Since the slopes of the contours in contrast space are unaffected by these changes, the results again show that chromatic adaptation does not alter the relative weights of the M- and L-cone contrast inputs to the chromatic mechanism, but does change the sensitivity of the chromatically opponent second site. These data should be compared with the predictions shown in Fig. 8a. S-cone stimulation did not affect the L–M chromatic detection contours.<sup>61</sup>



**FIGURE 28** Second-site desensitization of L–M by steady fields. L–M detection contours in cone contrast space for two subjects measured by Chaparro, Stromeyer, Chen, and Kronauer<sup>62</sup> on 3000-td fields of 525 nm (green circles), 579 nm (yellow squares), and 654 nm (red triangles) replotted from their Fig. 4.  $\Delta M/M$  is plotted along the ordinate and  $\Delta L/L$  along the abscissa. The straight contours fitted to thresholds have slopes of 1, consistent with detection by an L–M mechanism with equal and opposite L- and M-cone contrast weights. Constant relative L- and M-cone weights are a characteristic of von Kries first-site adaptation. The displacement of the contour from the origin on the 654 nm field is characteristic of second-site adaptation.

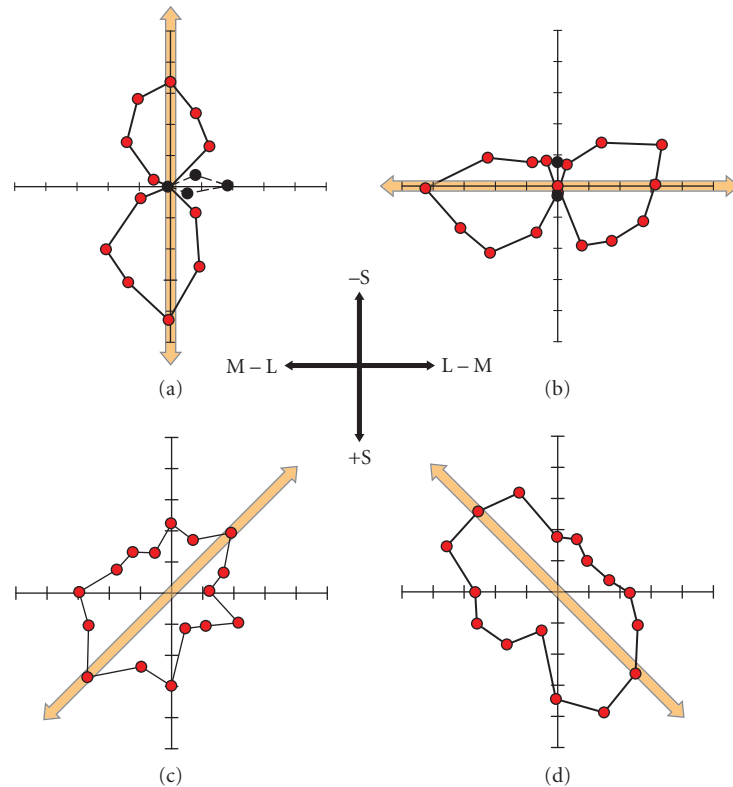
Figure 28 shows a clear example of the effect of changing the background wavelength of 3000 td fields from 525 to 579 to 654 nm from Fig. 4 of Ref. 62. Consistent with Stromeyer, Cole, and Kronauer,<sup>61</sup> the L–M chromatic detection contours have positive unity slopes, and move outward from the origin as the wavelength lengthens to 654 nm.

The relation between these detection contours and spectral sensitivity data is nicely illustrated by replotting the spectral sensitivity data as cone contrasts. Figure 17*b*, shows the spectral sensitivity data of Thornton and Pugh<sup>107</sup> replotted in cone contrast space by Eskew, McLellan, and Giulianini.<sup>89</sup> The replotted spectral sensitivity data have the same characteristics as the detection contours shown in Fig. 28.

**Habituation or Contrast Adaptation Experiments** The extent of second-site desensitization produced by steady backgrounds is fundamentally limited by first-site adaptation. A natural extension of the field sensitivity method is to temporally modulate the chromaticity or luminance of the background around a mean level, so partially “bypassing” all but the most rapid effects of first-site adaptation and enhancing the effects of second-site desensitization. Moreover, if the spectral properties of the postreceptoral detection mechanisms are known, one mechanism can be selectively desensitized by adapting (or habituating) to modulations along a cardinal axis, which are invisible or “silent” to the other mechanisms (see Ref. 239 for a review of silent substitution methods and their origins).

In an influential paper, Krauskopf, Williams, and Heeley<sup>14</sup> used habituation (the loss of sensitivity following prolonged exposure, in this case, to a temporally varying adapting stimulus) to investigate the properties of what they referred to as the “cardinal mechanisms” (see also Ref. 240). They used a 50 td, 2° diameter field made up of 632.5-, 514.5-, and 441.6-nm laser primaries. Following habituation to 1-Hz sinusoidal modulations (a 30-s initial habituation, then 5-s top-ups between test presentations) that were either chromatic (in the L–M, or S directions at equiluminance), or achromatic (L+M+S), or intermediate between these three directions, the detection sensitivity for Gaussian target pulses ( $\sigma = 250$  ms) was measured in various directions of color space.

Figure 29 shows the results for observer DRW. The red circles show the losses of sensitivity and the black circles the increase in sensitivity following habituation to stimuli modulated in the directions of the orange arrows shown in all panels. These results should be compared with the predictions shown in Fig. 11. In their original paper, Krauskopf, Williams, and Heeley<sup>14</sup> concluded that their results showed that habituation occurred independently along each of the three cardinal mechanism axes: L-M, S, and L+M. Thus, the sensitivity losses to habituation along one axis were mainly confined to targets modulated along the same axis, whereas the losses to habituation along axes intermediate to the cardinal ones involved losses in both mechanisms.



**FIGURE 29** Habituation. Data from Fig. 6 of Krauskopf, Williams, and Heeley<sup>14</sup> plotted in the canonical format shown in Fig. 11. These data are for stimuli that silence the L+M mechanism and reveal the properties of the L-M and S-(L+M) mechanisms plotted in a color space in which the horizontal axis corresponds to stimuli that only stimulate the L-M mechanism and the vertical axis corresponds to stimuli that only stimulate the S-(L+M) mechanism (by modulating S). (a) and (b) Results for habituation stimuli oriented along the horizontal and vertical axes, respectively. (c) and (d) Threshold changes for habituation stimuli along intermediate directions. The pattern of results is in qualitative agreement with the predictions developed in Figs. 9 to 11, and provides evidence for second-site adaptation in two opponent-discrimination mechanisms. The red circles plot increases in threshold following habituation, whereas the black circles plot cases where threshold was slightly decreased, rather than increased, by habituation.



Three years later, after a more rigorous analysis using Fourier methods, Krauskopf, Williams, Mandler, and Brown<sup>241</sup> concluded instead that the same data showed that habituation could desensitize multiple “higher-order” mechanisms tuned to intermediate directions in the equiluminant plane between the L–M and S cardinal directions. Yet, in terms of the magnitude of the selective losses, those along the intermediate axes are very much second-order effects compared with those found along the cardinal L–M or S axes. These second-order effects are not particularly compelling evidence for the existence of higher-order mechanisms, because they might also be due to more mundane causes, such as the visual distortion of the habituating stimuli or low-level mechanism interactions. The strong conclusion that can be drawn from the habituating experiments is that selective losses are largest when the habituating stimuli are modulated along the S, L–M, or L+M+S cardinal axes. The evidence for higher-order mechanisms along intermediate axes is secondary. Other workers have suggested that instead of there being multiple mechanisms, the cardinal mechanisms adapt to decorrelate the input signals so that their mechanism vectors rotate.<sup>242–244</sup>

**Noise-Masking Experiments** Another field method used to investigate visual mechanisms is the introduction of masking noise to raise detection threshold. The spectral properties of the underlying mechanisms can then be investigated by varying either the chromaticity and/or luminance of the target (which, if the noise is held constant, is strictly speaking a test method) or by varying the chromaticity and/or luminance of the noise. As with habituating stimuli, the noise stimuli can be fashioned to excite one or other receptor or postreceptor mechanism. However, whereas habituation is assumed to have little effect at the first site, noise masking can potentially raise thresholds at both the first and the second sites. Although it is possible to selectively excite different first-stage cone mechanisms by modulating the noise along cone-isolating vectors or to selectively excite different second-stage color-discrimination mechanisms by modulating the noise along cardinal vectors, each noise stimulus inevitably excites both first- and second-stage mechanisms. As described in section “Sites of Limiting Noise” in Sec. 11.3 with respect to internal noise (see Figs. 12 and 13), changing the balance of noise at the first and second sites has the potential of rotating the detection contours in color space. Consequently, the discovery of detection contours for some noise vectors that do not align with the assumed low-level mechanism axes does not necessarily imply the existence of high-order mechanisms.

Gegenfurtner and Kiper<sup>245</sup> measured thresholds for a 1.2-cpd Gabor patch ( $\sigma = 0.8^\circ$  in space and 170 ms in time) in the presence of noise modulated in different directions around white in the L,M plane of DKL space. The target was either modulated in the L–M or L+M+S directions or in an intermediate direction along which the target color appeared to change between bright red and dark green. For all three directions masking was maximal when the noise was in the target direction, and minimal when it was in the orthogonal direction, even if the target direction did not lie along a cardinal axis. These findings suggested to the authors that there must be multiple mechanisms, some tuned to axes other than the cardinal ones.

The mechanisms derived by Gegenfurtner and Kiper, however, combined cone inputs nonlinearly. In a linear mechanism, the effectiveness of the noise should depend on the cosine between the cardinal direction and the noise direction, but Gegenfurtner and Kiper<sup>245</sup> found mechanisms that were more narrowly tuned. Curiously, when square rather than Gabor patches were used, evidence for multiple mechanisms was not found.

The central findings of Gegenfurtner and Kiper have been contradicted in other papers, some of which used comparable techniques. Sankeralli and Mullen<sup>246</sup> found evidence for only three mechanisms in  $\Delta L/L$ ,  $\Delta M/M$ ,  $\Delta S/S$  cone contrast space using a 1-cpd semi-Gabor patch ( $\sigma = 1.4^\circ$  vertically, and  $4^\circ$  horizontally with sharp edges in space, and  $\sigma = 170$  ms in time). Mechanism tuning was found to be linear in cone contrast units with a cosine relationship between the noise effectiveness and the angle between the mechanism and noise vectors. Giulianini and Eskew<sup>175</sup> working in the  $\Delta L/L$ ,  $\Delta M/M$  cone contrast plane using 200-ms circular Gaussian blobs ( $\sigma = 1^\circ$ ) or horizontal Gabor patches (1 cpd,  $\sigma = 1^\circ$ ) similarly found evidence for only two mechanisms in that plane. Their results were consistent with a linear L–M mechanism with equal cone contrast weights, and a linear L+M mechanism with greater L-than M-cone weights.

Some of Giulianini and Eskew's<sup>175</sup> results from their Fig. 5 are shown in Fig. 20, for noise masks in the chromatic direction (Fig. 20*a* and *c*) and in the M-cone direction (Fig. 20*b* and *d*). Their interpretation

of the data is shown by the solid contours in Fig. 20*a* and *b*. Their model is consistent with detection by two mechanisms with roughly constant L–M contours across the two noise directions, but L+M contours that rotate slightly away from the M-cone axis in the presence of M-cone noise. However, as our elliptical fits in Fig. 20*c* and *d* show, other interpretations of their data are also plausible; most of the data in the presence of chromatic and M-cone noise can be described by ellipses, but the major axes of the ellipses do not align with the noise directions.

Subsequently, Eskew, Newton, and Giulianini<sup>179</sup> extended these measurements to include S-cone modulations in the equiluminant plane, and again found no evidence for higher-order mechanisms in detection or near-threshold discrimination tasks. Some of their results are shown in Fig. 21. These data seem more consistent with a rounded parallelogram than an ellipse. All three studies just described, in contradiction to Gegenfurtner and Kiper,<sup>245</sup> found that the directions at which noise masking was maximal more closely aligned with the expected mechanism directions than with the noise directions. However, see also Hansen and Gegenfurtner.<sup>247</sup>

In a related study, D’Zmura and Knoblauch<sup>248</sup> did find evidence for multiple mechanisms within the equiluminant plane of DKL space. They used “sectored” noise made up of noise centered on the signal direction combined with noise of various strengths in the orthogonal direction. Contrast thresholds for “yellow,” “orange,” “red,” and “violet” signals ( $\sigma = 2.4^\circ$  in space and 160 ms in time) were independent of the strength of the orthogonal noise, which suggests the existence of linear broad-band mechanisms tuned to each of the signal directions (for which the orthogonal direction is a null direction). Like Gegenfurtner and Kiper<sup>245</sup> multiple mechanisms were found, but they were broadly rather than narrowly tuned. D’Zmura and Knoblauch<sup>248</sup> argued that narrow tuning can result from off-axis looking in which subjects switch among multiple linear broad-band detection channels to reduce the noise and maximize the detectability of the signal.

An important caveat about D’Zmura and Knoblauch’s experiments is that they used the 1924  $V(\lambda)$  function to equate luminance (see p. 3118 of Ref. 248). Because  $V(\lambda)$  substantially underestimates luminous efficiency at short wavelengths, modulations of their blue phosphor will produce a luminance signal in their nominally equiluminant plane, which could complicate the interpretation of their results. For example, nominally S-cone modulations might have superimposed and cancelling L+M modulations, and some signal directions might be luminance-detected (and therefore artifactually independent of chromatic noise in the orthogonal direction). Nonetheless, Monaci et al.<sup>249</sup> using a related technique, but a better definition of luminous efficiency, came to the same conclusion.

Most recently, Giulianini and Eskew<sup>250</sup> used noise masking to investigate the properties of the S–(L+M) mechanism, and, in particular, whether or not the mechanism responds linearly to noises in different directions of cone contrast space. Though the effects of noise on the L–M mechanism are linear, for both polarities of the S–(L+M) mechanism (i.e., for +S and –S detection) its effects are nonlinear.<sup>250</sup>

## Chromatic Discrimination

**Basic Experiments** Chromatic discrimination is the ability to distinguish the difference between two test lights that differ in chromaticity. The prototypical set of chromatic discrimination data are the discrimination ellipses of MacAdam<sup>251</sup> originally plotted as CIE  $x, y$  chromaticity coordinates, and thus in a color space that is not especially conducive to understanding the underlying mechanisms. These ellipses were determined from the variability of repeated color matches made at 25 different reference chromaticities in the CIE color space. Le Grand<sup>252</sup> reanalyzed MacAdam’s data and found that much of the variability between the ellipses could be accounted for by considering variability along the tritan (S) dimension and along the cone-opponent (L–M) dimension. Interpretation of MacAdam’s data is complicated by the fact that each ellipse was measured for a different state of adaptation, namely that determined by the mean match whose variability was being assessed.

Boynton and Kambe<sup>253</sup> made a series of discrimination measurements in the equiluminant plane along the important S and L–M color directions. They found that discrimination along the S axis depended on only S-cone excitation level; specifically that  $\Delta S/S + S_0 = \text{constant}$ , where S is the background excitation and  $S_0$  is an “eigengrau” or “darklight” term. At high S-cone excitation levels, as  $S_0$  becomes less significant, Weber’s law holds (i.e.,  $\Delta S/S$  is approximately constant).<sup>254–256</sup> A similar

conclusion was reached by Rodieck<sup>257</sup> in his reanalysis of MacAdam's<sup>251</sup> data (see Figs. XXIII-10 and XXIII-11 of Ref. 257). When plotted in a normalized cone excitation diagram, chromaticity ellipses for individual subjects have similar shapes regardless of the reference chromaticity.<sup>258</sup>

**Chromatic Discrimination Near Detection Threshold** Under conditions that isolate the chromatic mechanisms, near-threshold lights can be indistinguishable over fairly extended spectral or chromaticity ranges. On a white background using 0.75° diameter, 500-ms duration test discs, Mullen and Kulikowski<sup>197</sup> found that there were four indistinguishable wavelength ranges that corresponded to the threshold lights appearing orange (above 585 nm), pale yellow (566–585 nm), green (c. 490–566 nm), and blue (c. 456–490 nm), and that there was weaker evidence for a fifth range that appeared violet (< c. 456 nm). Importantly, too, these boundaries were independent of the wavelength *differences* between pairs of lights, which suggests that each mechanism is univariant. These results suggest the existence of four or five distinct unipolar mechanisms at threshold.<sup>197</sup> Unlike the bipolar mechanisms (L–M) and S–(L+M), a unipolar mechanism responds with only positive sign. The bipolar (L–M) mechanism, for example, may be thought of as consisting of two unipolar mechanisms |L–M| and |M–L|, each of which is half-wave rectified (see “Unipolar versus Bipolar Chromatic Mechanisms” in Sec. 11.6).

Eskew, Newton, and Giulianini<sup>179</sup> looked at discrimination between near-threshold Gabor patches modulated in various directions in the equiluminant plane. Noise masking was used to desensitize the more sensitive L–M mechanism and thus reveal more of the S-cone mechanism, and perhaps higher-order mechanisms. Despite the presence of noise they found evidence for only four color regions within which lights were indiscriminable. These regions corresponded to detection by the four poles of the classical cone-opponent mechanisms |L–M|, |M–L|, |S–(L+M)|, and |(L+M)–S|. Some of their results are shown in Fig. 21. The four colors shown in the three outer arcs of the figure correspond to the four regions within which lights were indiscriminable. The data in each semicircular polar plot (filled squares) show the discriminability of a given chromatic vector from the standard vector shown by the arrow. The open circles are model predictions (see legend). There is no evidence here for higher-order color mechanisms. In a later abstract, Eskew, Wang, and Richters<sup>259</sup> reported evidence also from threshold level discrimination measurements for a fifth unipolar mechanism, which from hue-scaling measurements they identified as generating a “purple” hue percept (with the other four mechanisms generating blue, green, yellow, and red percepts).

Thus, the work of both Mullen and Kulikowski<sup>197</sup> and Eskew et al.<sup>179,259</sup> may be consistent with five threshold-level color mechanisms. By contrast, Krauskopf et al.<sup>241</sup> found evidence for multiple mechanisms. They measured the detection and discrimination of lights that were 90° apart in the equiluminant DKL plane and modulated either along cardinal or intermediate axes. No difference was found between data from the cardinal and noncardinal axes, which should not be the case if only cardinal mechanisms mediate discrimination. The authors argued that stimuli 90° apart modulated along the two cardinal axes should always be discriminable at threshold, because they are always detected by different cardinal mechanisms. In contrast, stimuli 90° apart modulated along intermediate axes will only be discriminable if they are detected by different cardinal mechanisms, which will not always be the case. Care is required with arguments of this sort, because angles between stimuli are not preserved across color space transformations (i.e., they depend on the choice of axes for the space and also on the scaling of the axes relative to each). More secure are conclusions based on an explicit model of the detection and discrimination process. Note that this caveat applies not just here but also to many studies that rely on intuitions about the angular spacing of stimuli relative to each other.

Clearly, colors that are substantially suprathreshold and stimulate more than one color mechanism should be distinguishable. However, the range of suprathreshold univariance can be surprisingly large. Calkins, Thornton, and Pugh<sup>111</sup> made chromatic discrimination and detection measurements in the red-green range from 530 to 670 nm both at threshold and at levels up to 8 times threshold. Measurements were made with single trough-to-trough periods of a 2-Hz raised-cosine presented on a yellow field of 6700 td and 578 nm. The observers' responses were functionally monochromatic either for wavelengths below the Sloan notch (between 530 and 560 nm) or for wavelengths above the Sloan notch (between 600 and 670 nm); that is, any two wavelengths within each range were indistinguishable at some relative intensity. The action spectra for indiscriminability above threshold

and for detection at threshold were consistent with an L–M opponent mechanism. Discrimination for pairs of wavelengths either side of the Sloan notch was very good, and between stimuli near the notch and stimuli further away from the notch was also good. The good discriminability near the notch is consistent with a diminution of  $|L-M|$  as the signals become balanced and with the intrusion of another mechanism, which between 575 and 610 nm has a spectral sensitivity comparable with the photopic luminosity function [although whether or not this corresponds to luminance or the L+M lobe of  $S-(L+M)$  is unclear]. For full details, see Ref. 111.

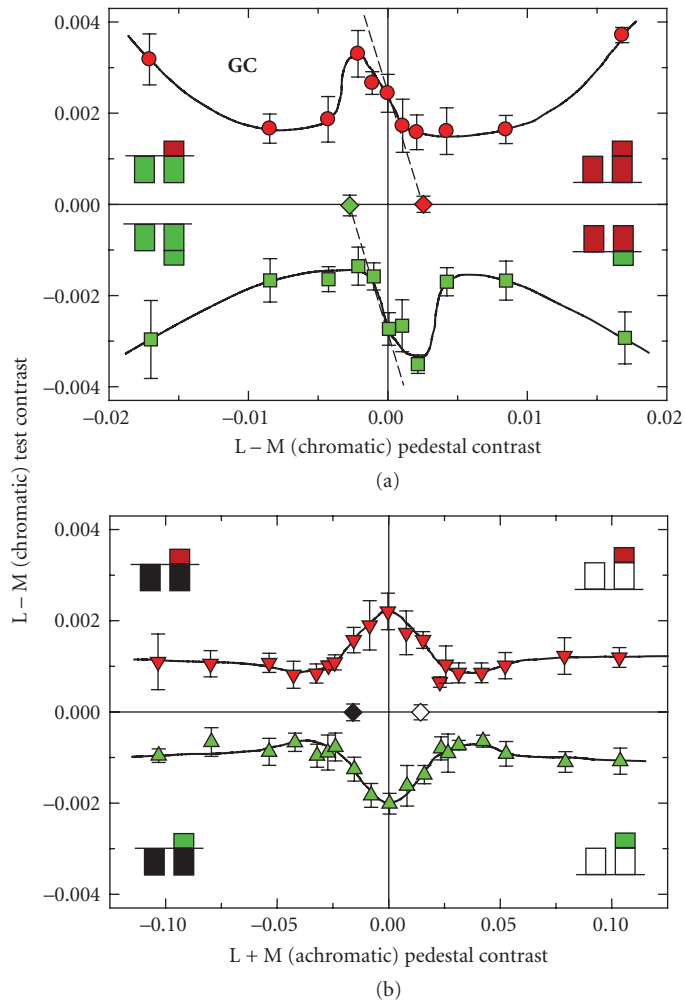
The results reviewed in this section are especially important because they suggest that each color mechanism is not only univariant, but behaves like a “labeled line.” As Calkins et al. note on their p. 2365 “Phenomenologically speaking, the discrimination judgments [in the monochromatic regions] are based upon variation in perceptual brightness.”

**Pedestal Experiments** Discrimination experiments, sometimes called “pedestal experiments” can be used to investigate the properties of mechanisms over a range of contrasts both above and below the normal detection threshold. In a typical pedestal experiment, an observer is presented with two identical but brief “pedestal” stimuli, separated either in space or in time, upon one of which a test stimulus is superimposed. The observer’s task is to discriminate in which spatial or temporal interval the test stimulus occurred. By presenting the test and pedestal stimuli to the same mechanism (sometimes called the “uncrossed” condition), it is possible to investigate the properties of chromatic and luminance mechanisms separately. By presenting the test and pedestal stimuli to different mechanisms (sometimes called the “crossed” condition), it is possible to investigate interactions between them. The presentation of the pedestals causes a rapid shift away from the mean adapting state. The transient signals produced by these shifts will be affected by instantaneous or rapid adaptation mechanisms, such as changes in the response that depend on static transducer functions, but not by more sluggish adaptation mechanisms.

When pedestal and test stimuli are both uncrossed and of the *same* sign, the discrimination of the test stimulus typically follows a “dipper” function as the pedestal contrast is increased. As the pedestal contrast is raised from below its contrast threshold, the threshold for detecting the test stimulus falls by a factor of about two or three, reaching a minimum just above the pedestal threshold, after which further increases in pedestal contrast hurts the detection of the test.<sup>260–265</sup> Facilitation has been explained by supposing there is a transducer (input-output) function that accelerates at lower contrasts, thus causing the difference in output generated by the pedestal and test-plus-pedestal to increase with pedestal contrast (e.g., Refs. 263, 266). Alternatively, it has been explained by supposing that the pedestal, because it is a copy of the test, reduces detection uncertainty.<sup>267</sup> The dipper function is found both for uncrossed luminance and uncrossed chromatic stimuli, and has similar characteristics for the two mechanisms.<sup>169,268,269</sup> When pedestals and test stimuli are uncrossed but of the *opposite* sign, increases in pedestal contrast first raise threshold—consistent with subthreshold cancellation between the pedestal and test. As the pedestal approaches its own threshold, however, discrimination threshold abruptly decreases, after which further increases in pedestal contrast hurt detection of the test for both uncrossed luminance and uncrossed chromatic conditions.<sup>169</sup> The increase followed by decrease in threshold has been referred to as a “bumper” function.<sup>270</sup> The dipper and bumper functions found under uncrossed conditions show clear evidence of subthreshold summation or cancellation. Figure 30a shows examples of the dipper (upper right and lower left quadrants) and bumper functions (upper left and lower right quadrants) for uncrossed L–M chromatic pedestals and targets from Cole, Stromeyer, and Kronauer.<sup>169</sup>

Presenting the pedestals and test stimuli to different mechanisms in the crossed conditions is a potentially powerful way of investigating how the luminance and chromatic mechanisms interact. For example, if the two mechanisms interact before the hypothesized accelerating transducer, then the crossed and uncrossed dipper and bumper functions should be comparable and should also show evidence for subthreshold interactions. If the chromatic and luminance mechanisms are largely independent until later stages of visual processing, then subthreshold effects should be absent, but given the uncertainty model suprathreshold facilitation might still be expected.

The results obtained with crossed test and pedestal stimuli, however, remain somewhat equivocal. Luminance square-wave gratings or spots have been shown to facilitate chromatic discrimination.<sup>194,271</sup>



**FIGURE 30** Pedestal effects. (a) Detection of chromatic L–M tests on chromatic L–M pedestals. As indicated by the icons in each quadrant, the data points are the thresholds for detecting red chromatic targets (red circles) on either red (upper right quadrant) or green (upper left quadrant) chromatic pedestals, or for detecting green chromatic targets (green squares) on either red (lower right quadrant) or green (lower left quadrant) pedestals. The diamonds show the thresholds for detecting the red (red diamond) or green (green diamond) chromatic pedestals alone. The dashed lines show predictions that the discrimination of the presence of the test depends only on the intensity of the test plus the pedestal (which implies that pedestal alone is ineffectual). (b) Detection of chromatic L–M tests on achromatic L+M pedestals. As indicated by the icons, the data are the thresholds for detecting red chromatic targets (red inverted triangles) on incremental (upper right quadrant) or decremental (upper left quadrant) luminance pedestals, or for detecting green chromatic targets (green triangles) on incremental (lower right quadrant) or decremental (lower left quadrant) luminance pedestals. The diamonds show the thresholds for detecting the incremental (open diamond) or decremental (filled diamond) luminance pedestals alone. (Data replotted from Figs. 4 and 5 of Cole, Stromeyer, and Kronauer.<sup>169</sup> Design of the figure based on Fig. 18.10 of Eskew, McLellan, and Giulianini.<sup>89</sup>)

Using sine-wave gratings, De Valois and Switkes<sup>265</sup> and Switkes, Bradley, and De Valois<sup>268</sup> found that crossed facilitation and masking was asymmetric: for luminance pedestals and chromatic tests they found facilitation and no masking, but for chromatic pedestals and luminance tests they found no facilitation and strong masking. This curious asymmetry has not always been replicated in subsequent measurements. Using spots, Cole, Stromeyer, and Kronauer<sup>169</sup> obtained equivalent results for crossed luminance pedestals and chromatic tests and for crossed chromatic pedestals and luminance tests. They found in both cases that increasing the contrast of the crossed pedestal had little effect on test threshold until the pedestal approached or exceeded its own threshold by a factor of two or more, and that further increases in pedestal contrast had little or no masking effect.<sup>169</sup> Figure 30*b* shows examples of their crossed luminance and chromatic results. Similarly, Mullen, and Losada<sup>269</sup> found facilitation for both types of crossed pedestals and tests when the pedestal was suprathreshold, but unlike Cole, Stromeyer, and Kronauer<sup>169</sup> they found crossed masking for higher pedestal contrasts. Chen, Foley, and Brainard,<sup>272</sup> however, obtained results that were more consistent with De Valois and Switkes<sup>265</sup> and Switkes, Bradley, and De Valois.<sup>268</sup> They used spatial Gabors (1 cpd, 1° standard deviation) with a Gaussian temporal presentation (40-ms standard deviation, 160-ms total duration), and investigated the effects of targets and pedestals presented in various directions of color space. They found masking for all target-pedestal pairs and facilitation for most target-pedestal pairs. The exceptions were that (1) equiluminant pedestals did not facilitate luminance targets, and (2) tritan (blue/yellow) pedestals facilitated only tritan targets.

So, what does all this mean? The fact that the crossed facilitation, when it is found, occurs mostly with suprathreshold pedestals<sup>169,268,269</sup> suggests that the interaction occurs after the stages that limit the thresholds in the chromatic and luminance mechanisms. This is surprising, given that center-surround L–M neurons respond to both chromatic and luminance gratings (see subsection “Multiplexing Chromatic and Achromatic Signals” in Sec. 11.5). This early independence is supported by findings with gratings that the crossed facilitation is independent of the spatial phase of the gratings.<sup>268,269</sup> However, the facilitation is not found with crossed dichoptic pedestals and tests. The finding that crossed facilitation survives if the luminance pedestal is replaced with a thin ring<sup>169</sup> suggests that it might be due to uncertainty reduction. However, yes-no psychometric functions and receiver operating characteristics show clearly that the facilitation of chromatic detection by luminance contours is inconsistent with the uncertainty model.<sup>177</sup> To explain their results, Chen, Foley, and Brainard<sup>273</sup> presented a model made up of three postreceptoral channels with separate excitatory inputs (producing facilitation), but each with divisive inhibition aggregated from all channels (producing masking).

Krauskopf and Gegenfurtner<sup>255</sup> performed a series of experiments in which chromatic discrimination was measured after an abrupt change in chromaticity away from an equal-energy-white adapting light (see also Refs. 274, 275). The change in chromaticity was effected by simultaneously presenting four 36-min diameter discs for 1 s. Three of the discs were of the same chromaticity, but the fourth was slightly offset from the others. The observer’s task was to identify which disc was different. The task can be thought of as the presentation of three stimuli defined by a “pedestal” color vector from the white point with the fourth stimulus being defined by the vector sum of the pedestal vector plus a “test” vector.<sup>89</sup> The experiments were carried out in the DKL equiluminant plane at a luminance of 37 cd/m<sup>2</sup>. The discrimination of a test vector along one of the cardinal axes was found to deteriorate if the pedestal vector was along the same axis, but was relatively unaffected if the pedestal vector was along the other axis. Color discrimination, therefore, depended primarily on the L–M and S–(L+M) mechanisms nominally isolated by modulations along the cardinal axes.<sup>255</sup> A more detailed determination of the shapes of chromatic discrimination ellipses for pedestal and test vectors along 16 color directions was also consistent with discrimination being determined by the cardinal mechanisms, but there was also evidence for another mechanism orientated along a blue-green to orange axis, from 135 to 315° in the DKL space.<sup>255</sup> This might reflect the existence of a higher-level mechanism, but it might also be consistent with other explanations. Maintaining the phosphors at 37 cd/m<sup>2</sup> [a photometric measure defined by the discredited CIE  $V(\lambda)$  function that underestimates luminous efficiency in the blue and violet] results in modulation of the blue phosphor at nominal “equiluminance” producing a luminance (L+M) axis close to the blue-green axis. The effect of such a signal is hard to predict, but could affect the sensitivity of the S–(L+M) mechanism by



cancelling S, or it could be detected by L+M. Another possibility is that the evidence for an additional mechanism might instead reflect a small S-cone contribution to L–M, which would rotate the L–M mechanism axis.<sup>89</sup>

## Color Appearance and Color Opponency

When viewed in isolation (i.e., in the “aperture” mode), monochromatic lights have characteristic color appearances. Across the visible spectrum hues vary from violet through blue, blue-green or cyan, green, yellow-green, yellow, orange, and red.<sup>276</sup> Yet, however compelling these colors might seem, they are private perceptions. Consequently, to investigate them, we must move away from the safer and more objective psychophysical realms of color matching, color detection, and color discrimination, and rely instead on observers’ introspections about their perceptual experiences (see Ref. 277 for a related discussion of “Class A” and “Class B” observations). Experiments that depend upon introspection are an anathema to some psychophysicists, yet they represent one of the few methods of gaining insights into how color is processed beyond the early stages of the visual system. As long as the limitations of the techniques are recognized, such experiments can be revealing. In addition, it is of interest to understand color appearance per se, and to do so requires the use of subjective methods in which observers report their experience.

Physiological measurements of the spectral properties of neurons in the retina and LGN are generally inconsistent with the spectral properties of color-appearance mechanisms, which suggests that color appearance is likely to be determined by cortical processes (see, e.g., Ref. 18). Relative to the input at the photoreceptors, the relevant neural signals that eventually determine color appearance will have undergone several transformations, likely to be both linear and nonlinear, before they reach the cortex. Despite this many models of color appearance are based on the assumption that appearance can be accounted for by mechanisms that linearly combine cone inputs (see Sec. 11.4). As we will see later, this is at best an approximation, at least for some aspects of color appearance. One reason for the failures of linearity may be that the suprathreshold test stimuli used to probe appearance are often adapting stimuli that intrude significantly on the measurements. That is, unlike in detection experiments where near-threshold tests can reasonably be assumed not to perturb the adapted state of the visual system very much, nonlinear adaptive processes cannot be ignored even for nominally steady-state color-appearance measurements, except perhaps under some contrived experimental conditions (see subsection “Linearity of Color-Opponent Mechanisms” in Sec. 11.5).

**Opponent-Colors Theory** Color appearance is often conceptualized in terms of Hering’s opponent-color theory; that is, in terms of the perceptual opposition of red and green (R/G), blue and yellow (B/Y), and dark and light. In many versions of the model, the opponent mechanisms are assumed, either explicitly or implicitly, to combine their inputs from cones in a linear fashion.<sup>11,12,18,19,29,76,278–282</sup> Stage 3 of Fig. 4 represents a version of the linear model derived by the simple combinations of cone-opponent mechanisms (see subsection “Three-Stage Zone Models” in Sec. 11.6 for details).

The assumption of linearity constrains and simplifies the predicted properties of the color-opponent mechanisms. For example, under this assumption each opponent mechanism is completely characterized by its spectral response. As a function of wavelength, the R/G opponent mechanism responds R at both short and long wavelengths, corresponding to the red constituent in the appearance of both short- and long-wavelength lights, and responds G at middle wavelengths. The B/Y mechanism responds B at shorter wavelengths and Y at longer wavelengths. Each response (R, G, Y, or B) is univariant, and the responses of opposed poles (R vs. G, and B vs. Y) are mutually exclusive. The colors of lights detected solely by the same pole of a given mechanism cannot be distinguished at threshold, but lights detected by different poles can. In the model, color appearance depends on the relative outputs of the color-opponent appearance mechanisms. The individual mechanisms signal one opponent color or the other (e.g., redness or greenness) in different strengths, or a null.

The wavelength at which the opponent response of a mechanism changes polarity is a zero crossing for that mechanism, at which its response falls to zero. These zero crossings correspond to the unique hues. Unique blue and yellow are the two zero crossings for the R/G mechanism, and unique green and red are the two zero crossings for the Y/B mechanism. Unique red does not appear in plots

of valence against wavelength because it is extra-spectral (i.e., it has no monochromatic metamer and can only be produced by a mixture of spectral lights). Another extra-spectral color is white, which corresponds to an “equilibrium color” for *both* the R/G and B/Y mechanisms. A property of opponent-color mechanisms is that both the R/G and the Y/B mechanisms do not respond to a suitably chosen broadband white stimulus (such as an equal energy white). Thus, the sum of the outputs of the positive and negative lobes of each mechanism in response to such stimuli should be zero.

Tests of the assumptions of linearity of opponent-colors mechanisms are discussed below (see subsection “Linearity of Color-Opponent Mechanisms” in Sec. 11.5). See also Hurvich<sup>283</sup> for more details of opponent colors theory.

**Spectral Properties of Color-Opponent Mechanisms** Several techniques have been used to investigate the spectral properties of the color-opponent mechanisms. Most rely on the assumption that opponent-color mechanisms can be nulled or silenced by lights or combinations of lights that balance opposing sensations. The null may be achieved experimentally by equating two lights so that their combination appears to be in equilibrium (see section “Appearance Test Measurements and Opponency” in Sec. 11.4). A closely related alternative is to find single spectral lights that appear to be in equilibrium for a particular mechanism (e.g., that appear neither red nor green, or neither blue nor yellow). Another related technique is hue scaling, in which lights are scaled according to how blue, green, yellow, or red they appear. Thus, lights that are rated 100 percent blue, green, yellow, or red are equilibrium or unique hues.

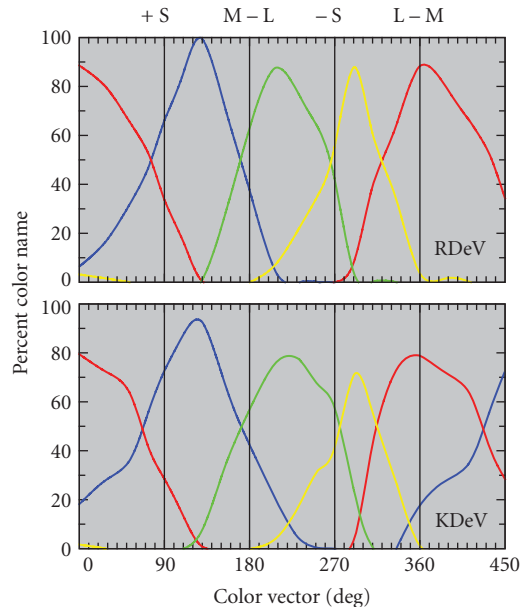
*Hue scaling* Despite the variety of color terms that could be used to describe the colors of lights, observers require only four—red, yellow, green, and blue—to describe most aspects of color appearance.<sup>6,7</sup> Other colors such as orange can be described as reddish-yellow, cyan as bluish-green, purple as reddish-blue, and so on. The need for just four color terms is consistent with opponent-colors theory. By asking observers to scale how blue, green, yellow, and blue spectral lights appear, hue scaling can be used to estimate the spectral response curves of the opponent mechanisms.<sup>21,284–286</sup> A cyan, for example, might be described as 50-percent green, 50-percent blue, and an orange as 60-percent red and 40-percent yellow. Figure 31 shows hue scaling data obtained by De Valois et al.,<sup>21</sup> who instead of using spectral lights used equiluminant modulations in DKL space, the implications of which are discussed below. Hue-scaling results have been reported to be consistent with hue-cancellation valence functions,<sup>286</sup> but inconsistencies have also been reported.<sup>287</sup>

*Color-opponent response or valence functions* As introduced in section “Appearance Test Measurements and Opponency,” the phenomenological color-opponency theory of Hering<sup>6</sup> was given a more quantitative basis by the hue-cancellation technique.<sup>279,280,288</sup> Jameson and Hurvich (see Fig. 14 for some of their data) determined the amount of particular standard lights that had to be added to a spectral test light in order to cancel the perception of either redness, greenness, blueness, or yellowness that was apparent in the test light, and so produced “valence” functions. Figure 32 shows additional R/G valence data from Ingling and Tsou<sup>12</sup> and B/Y valence data from Werner and Wooten.<sup>286</sup> Individual lobes of the curves show the amounts of red, green, blue, or yellow light required to cancel the perception of its opposing color in the test light. The zero crossings are the “unique” colors of blue, green, and yellow.

These types of judgments are, of course, highly subjective because the observer must abstract a particular color quality from color sensations that vary in more than one perceptual dimension. In valence settings, the color of the lights at equilibrium depends on the wavelength of the spectral test light. Lights that are in red/green equilibrium vary along a blue-achromatic-yellow color dimension, while those in yellow/blue equilibrium vary along a red-achromatic-green color dimension. To reduce the need for the observer to decide when a light is close enough to the neutral point to accept the adjustment, some authors have used techniques in which the observer judges simply whether a presented light appears, for example, reddish or greenish, to establish the neutral points of each mechanism.<sup>289</sup>

*Unique hues and equilibrium colors* The equilibrium points of the R/G color opponent mechanism correspond to the unique blue and unique yellow hues that appear neither red nor green,



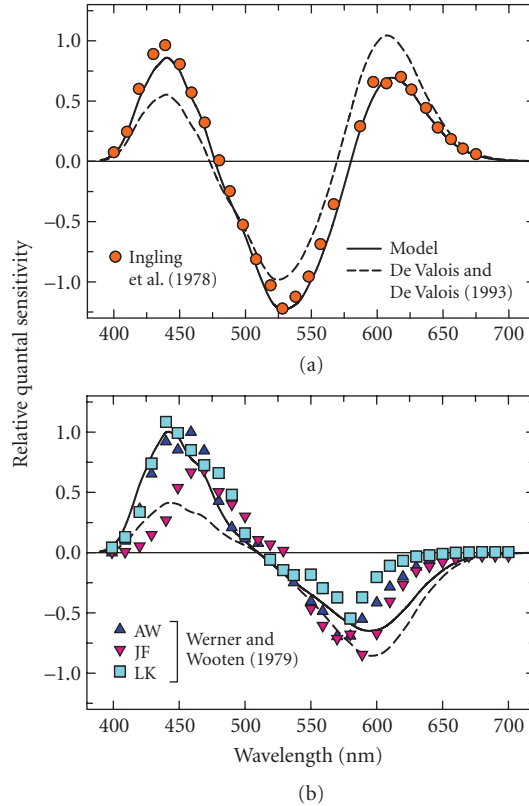


**FIGURE 31** Hue scaling. Data replotted from Fig. 3 (Observers: RDeV and KDeV) of De Valois, De Valois, Switkes, and Mahon.<sup>21</sup> The data show the percentage of times that a stimulus with a given color vector in the equiluminant plane of DKL space was called red (red lines), yellow (yellow lines), green (green lines), or blue (blue lines). The vertical lines indicate the cardinal axes at 90/450° (+S), 180° (M-L), 270° (-S), and 0/360° (L-M). With the exception perhaps of the red function for RDeV, the other hue naming functions do not align with the cardinal axes.

whereas those of the B/Y mechanism correspond to the unique green and unique red hues that appear neither blue nor yellow. These hues are “unique” in the sense that they appear phenomenologically unmixed.<sup>290</sup>

Measurements of the spectral positions of the unique hues have been made in several studies.<sup>281,290–296</sup> Kuehni<sup>297</sup> provides a useful summary of the unique hue settings from 10 relatively recent studies. For unique yellow, the mean and standard deviation after weighting by the number of observers in each study are 577.8 and 2.9 nm, respectively; for unique green, 527.2 and 14.9 nm; and for unique blue, 476.8 and 5.2 nm. Studies that generate colors either on monitors<sup>22,289,298</sup> or in print,<sup>299,300</sup> and therefore use desaturated, nonspectral colors, are not necessarily comparable with the results obtained using spectral lights, since perceived hue depends upon saturation.<sup>301</sup> The mean unique yellow, green, and blue wavelengths taken from Table of Kuehni<sup>297</sup> *excluding* those obtained by extrapolation from desaturated, nonspectral equilibrium lights are 576.3, 514.8, and 478.0 nm, respectively. Although unique hues show sizable individual differences, the settings within observers are generally quite precise.<sup>302,303</sup> Dimmick and Hubbard<sup>290</sup> review historical estimates.

Dimmick and Hubbard<sup>290</sup> reported, nearly 70 years ago, that unique blue and yellow and unique red and green are not complementaries (see also Ref. 304). By extending unique hue settings to include nonspectral colors it is possible to determine the equilibrium vectors of the opponent mechanisms in two dimensions or the equilibrium planes in three dimensions. Valberg<sup>305</sup> determined unique blues, greens, yellows, and reds as a function of saturation. Although not commented on in the original paper, two important features of his results were that (1) the unique red and green vectors plotted



**FIGURE 32** Color valence data. (a) R/G valence data (red circles) replotted from Fig. 2A of Ingling, Russell, Rea, and Tsou.<sup>314</sup> (b) B/Y valence data for observers AW (dark blue triangles), JF (purple inverted triangles) and LK (light blue squares) replotted from Fig. 9 of Werner and Wooten.<sup>286</sup> The continuous lines are the spectral sensitivity predictions of Stage 3 of the Müller-zone model outlined in the text (see section “Three-Stage Zone Models” in Sec. 11.6). Each set of valence data has been scaled to best fit the spectral sensitivity predictions. For comparison, the spectral sensitivity predictions of the model proposed by De Valois and De Valois<sup>18</sup> are also shown as dashed lines, in each case scaled to best fit the Müller-zone model predictions.

in a CIE  $x, y$  chromaticity diagram were not colinear, and (2) unique yellow and blue vectors, though continuous through the white point, were slightly curved. Burns et al.<sup>16</sup> in a comparable experiment confirmed that the unique green and red vectors were not colinear, and that the unique blue vector was curved, but found that the unique yellow and green vectors were roughly straight.

Chichilnisky and Wandell<sup>289</sup> located equilibrium boundary planes by presenting stimuli varying in both intensity and chromaticity on various backgrounds, and then asking subjects to classify them into red-green, blue-yellow, and white-black opponent-color categories. They found that the opponent-color classification boundaries were not coplanar. Wuerger, Atkinson, and Cropper<sup>22</sup> determined the null planes for the four unique hues using a hue-selection task in which subjects selected the equilibrium hue from an array of colors. They concluded that unique green and unique red planes in a space defined by the color monitor phosphors were not coplanar, consistent with the previous findings, but

that unique blue and unique yellow could form a single plane. However, they did not try to fit their data with curved surfaces, as might be suggested by previous work.

In summary, unique red and green are not colinear in two-dimensional color space or coplanar in three, and the unique blue “vector” or “plane” is curved. As we discuss below, failures of colinearity or coplanarity imply that either a bipolar color-opponent mechanism is not a single mechanism or it is a single but nonlinear mechanism, while curved vectors and planes imply failures of additivity within single mechanisms.

## Linearity of Color-Opponent Mechanisms

*Tests of linearity* For the spectral sensitivities of color-opponent mechanisms to be generalizable implicitly or explicitly, the color-opponent mechanisms must be linear and additive. This assumption is made in many accounts of zero crossings and/or in descriptions of valence functions as linear combinations of the cone fundamentals.<sup>11,18,19,24,29,76,278,279,306–308</sup>

Larimer, Krantz, and Cicerone<sup>307,309</sup> specifically tested whether the equilibrium colors behaved additively under dark adapted conditions. They tested two predictions of additivity: (1) that equilibrium colors should be intensity-invariant (the so-called “scalar invariance law”), and (2) that mixtures of equilibrium hues should also be in equilibrium (“the additivity law”). Another consequence of these predictions is that the chromatic response or valence function should be a linear combination of the cone fundamentals.<sup>75</sup> For R/G opponency, Larimer, Krantz, and Cicerone<sup>307</sup> found that the spectral position of the blue and yellow equilibrium hues were intensity-invariant over a range of 1 to 2 log units, and that mixtures of red/green equilibrium hues remained in red/green equilibrium. These results suggest that R/G color opponency is additive. Additivity for the R/G mechanism has also been reported in other studies<sup>310,311</sup> and is consistent with the R/G valence functions being linear combinations of the cone spectral sensitivities.<sup>12,286,312</sup>

Several studies, however, suggest that linearity fails for the R/G mechanism. The curvilinear unique blue vector<sup>16,305</sup> implies that mixtures of lights that individually appear unique blue and therefore in R/G equilibrium will not be in equilibrium when they are mixed—as is required for additivity. Similarly, the nonplanar red-green color categories<sup>289</sup> also imply failures of additivity. Ayama, Kaiser, and Nakatsue<sup>313</sup> carried out an additivity test for red valence and found that while additivity was found for some pairs of wavelength within separate R valence lobes (400 paired with 440 nm and 610 paired with 680 nm) failures were found between lobes (400 paired with 680 nm) for 3 out of 4 observers. Ingling et al.<sup>314</sup> found that the short-wavelength shape of the R valence lobe was dependent on the technique used to measure it. If the redness of the violet light was assessed by matching rather than by cancellation, the estimate of redness was as much as 30 times less. Ingling, Barley, and Ghani<sup>287</sup> analyzed previous R/G hue cancellation and hue-scaling data and found them to be inconsistent with the linear model.

By contrast, most evidence suggests that the B/Y color-opponent mechanism is nonlinear. Larimer, Krantz, and Cicerone<sup>309</sup> found that equilibrium green was approximately intensity-invariant, but that equilibrium red was not, becoming more bluish-red as the target intensity was increased. Moreover, although mixtures of B/Y equilibrium hues remained in equilibrium as their overall intensity changed, the equilibria showed an intensity-dependence. Other results also suggest that B/Y color opponency is nonlinear. The unique green and red vectors are not colinear in two dimensions of color space<sup>16,305</sup> and the unique green and unique red planes are not coplanar.<sup>22,289</sup> These failures of colinearity and coplanarity suggest that the B and the Y valence function have different spectral sensitivities, but the same neutral or balance point. The failure of additivity is consistent with the fact that the B/Y valence functions cannot be described by a linear combination of the cone fundamentals; instead some form of nonlinearity must be introduced.<sup>286,309</sup> Elzinga and de Weert<sup>315</sup> found failures of intensity-invariance for equilibrium mixtures and attributed the failures to a nonlinear (power) transform of the S-cone input. Ayama and Ikeda<sup>316</sup> found additivity failures for combinations of several wavelength pairs. Ingling, Barley, and Ghani<sup>287</sup> found that previous B/Y hue cancellation and hue-scaling data were inconsistent with the linear model. Knoblauch and Shevell<sup>317</sup> found that the B/Y nonlinearity might be related to the signs of the cone signals changing with luminance.

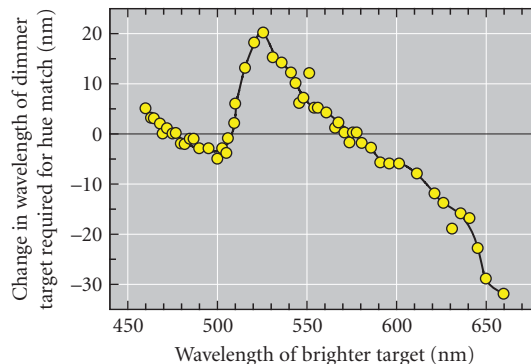
Interestingly, the B/Y mechanism behaves linearly in protanopes and deuteranopes, which suggests that the nonlinearity might depend in some way on L–M.<sup>318,319</sup>

*Why linearity?* This question of whether color-opponent mechanisms are linear or nonlinear was originally formulated when it was still considered plausible that cone signals somehow fed directly into color-opponent channels without modification. Given our current knowledge of receptor adaptation and postreceptor circuitry, that hope now seems somewhat optimistic. Receptor adaptation is an essentially nonlinear process, so that unless the adaptive states of all three cone types are held approximately constant (or they somehow change together as perhaps in the case of invariant hues), linearity will fail.

Tests that have found linearity have used elaborate experimental strategies to avoid the effects of adaptation. Larimer, Krantz, and Cicerone,<sup>307,309</sup> for example, used dark-adapted conditions and presented the stimuli for only 1 second once every 21 seconds. This protracted procedure succeeded in achieving additivity in the case of the red/green equilibrium experiments but not in the case of the blue/yellow ones.

It would be a mistake to conclude from results like these that color opponency and color appearance are, in general, additive with respect to the cone inputs. Experiments like those of Larimer, Krantz, and Cicerone<sup>307,309</sup> demonstrate additivity, but only under very specific conditions. While the results are important, because they provide information about the properties of the “isolated” postreceptor mechanisms, they are not necessarily relevant to natural viewing conditions under which the mechanisms may well behave nonlinearly with respect to the cone inputs.

**Bezold-Brücke Effect and Invariant Hues** Hue depends upon light intensity. Although a linear explanation of such hue changes has been suggested (see, e.g., Fig. 4 on p.74 of Hurvich<sup>283</sup>), it now seems clear that these changes reflect underlying nonlinearities in the visual response. These nonlinearities arise in part because the cone outputs are nonlinear functions of intensity, but they may also reflect inherent nonlinearities within the color-appearance mechanisms themselves and the way in which they combine cone signals. The dependence of hue on intensity is revealed clearly by the Bezold-Brücke effect,<sup>320,321</sup> which is illustrated in Fig. 33 using prototypical data from Purdy.<sup>322</sup> As the intensity of a spectral light is increased, those with wavelengths shorter than 500-nm shift in appearance toward that of an invariant blue hue of approximately 474 nm, while those with wavelengths



**FIGURE 33** Bezold-Brücke effect. Data replotted from Fig. 1 of Purdy<sup>322</sup> illustrating the changes in apparent hue as the light level increases, a phenomenon known as the Bezold-Brücke hue shift. The graph shows the change in the wavelength of a 100-td target from that of the 1000-td target required to match the hue of the 1000-td target plotted as a function of the wavelength of the more intense target.

longer than about 520 nm shift in appearance toward that of the invariant yellow of approximately 571 nm. Intermediate spectral colors shift to an invariant green of approximately 506 nm.

The invariant hues are usually assumed to coincide with the unique hues of color-opponent theory,<sup>283</sup> yet when the coincidence is tested, small discrepancies are found in some<sup>284,322,323</sup> but not all<sup>307,309</sup> studies (see also Ref. 296). Although invariant hues are often interpreted as being the zero crossings of cone-opponent mechanisms,<sup>136</sup> they are also consistent with models that incorporate just receptor adaptation.<sup>323,324</sup> As Vos<sup>323</sup> pointed out, if hue invariance and unique hues reflect different underlying processes, their approximate agreement may serve an important purpose.

**Color Appearance and Chromatic Adaptation** In the Bezold-Brücke effect, the level of adaptation is varied by increasing or decreasing the intensity of spectral lights of fixed wavelength. Under such conditions, as noted in the previous section, the appearances of some wavelengths are invariant with intensity. If, however, the state of chromatic adaptation is varied by superimposing the spectral lights on a chromatic background, their color appearance will change. In general, the appearances of targets superimposed on an adapting background move away from the appearance of the adapting wavelength. Thus, a target that appears yellow when viewed alone will appear reddish if superimposed on a middle-wavelength adapting background or greenish if superimposed on a long-wavelength one.<sup>325–327</sup> Similarly, the appearance of a superimposed target of the *same* wavelength as the adapting background will move toward the appearance of the equilibrium yellow hue. Indeed, on intense adapting fields of 560, 580, and 600 nm, Thornton and Pugh<sup>107</sup> showed that the spectral position of equilibrium yellow for a superimposed target coincided with the adapting field wavelength. However, this agreement breaks down on longer wavelength fields. Long-wavelength flashes presented on long-wavelength fields appear red, not equilibrium yellow.<sup>65,328,329</sup>

Historically, changes in color appearance with adaptation were thought to be consistent with von Kries adaptation,<sup>330</sup> in which the gain of each photoreceptor is attenuated in proportion to the background adaptation. However, von Kries adaptation cannot completely account for the changes in color appearance<sup>77,331–333</sup> (but see Ref. 334). In particular, asymmetric color matches, in which targets presented on different adapting backgrounds are adjusted to match in hue, do not survive proportional changes in overall intensity—as they should if the von Kries coefficient law holds.<sup>335</sup>

Jameson and Hurvich<sup>336</sup> proposed that the adapting field alters color appearance in two ways. First, by changing the cone sensitivities in accordance with von Kries adaptation, and second by an additive contribution to the appearance of the incremental mixture. For example, a red background tends to make an incremental target appear greener because it selectively adapts the L- cones, but it also tends to make it appear redder because it adds to the target. Color-appearance models of this sort remain in vogue (see Ref. 337). Figure 15 shows a case where a model with both gain control at the cones and an additive term accounts for much but not all of the variance in an asymmetric matching data set.

The details of this two-stage theory have been contentious, despite the fact that all authors are in essential agreement that much of the effect of the background is removed from the admixed background and test. Walraven,<sup>325,338</sup> using 1.5° diameter targets superimposed on 7° diameter backgrounds, determined the equilibrium mixtures of incremental 540 and 660 nm targets on various 660 nm backgrounds. He found that the equilibria were consistent with von Kries adaptation produced by the background, but that the color appearance of the background was completely discounted, so that it made no additive contribution to the appearance of the incremental target. By contrast, Shevell, using thin annuli or transiently presented 150-ms targets that were contiguous with the background, found that the background is not discounted completely and instead makes a small additive contribution to the target.<sup>326,327,339</sup> Walraven<sup>338</sup> suggested that Shevell's results reflected the fact that his conditions hindered the easy separation of the incremental target from the background. Yet, later workers came to similar conclusions as Shevell.<sup>340–342</sup> However, although the background is not entirely discounted, its additive contribution to color appearance is much less than would be expected from its physical admixture.<sup>327,343</sup> Thus, a discounting mechanism must be playing some role. Indeed, the failure of complete discounting is small compared to the magnitude of the additive term, so that the failure is very much a second-order effect.

**Color Appearance and Chromatic Detection and Discrimination** Several attempts have been made to link explicitly the cone-opponent effects evident in field sensitivity and field adaptation detection experiments with changes in color appearance measured under the same conditions.

On a white, xenon field, Thornton and Pugh<sup>164</sup> showed that the L–M cone-opponent detection implied by the Sloan notch is consistent with the suprathreshold color opponency implied by red–green hue equilibria. Thus, the wavelength of the notch, which is assumed to occur when the L–M cone-opponent signals are equal, coincided with an R/G equilibrium hue. A comparable convergence was found for 430- and 570-nm mixtures and the suprathreshold yellow–blue equilibrium locus.

S-cone thresholds on bichromatic mixtures of spectral blue and yellow fields are subadditive (see Fig. 26). According to the Pugh and Mollon<sup>231</sup> cone-opponent model, the greatest sensitization should occur when the opposing signals at the second site are in balance; that is, when the S- and L+M-cone signals are equal and opposite. In an attempt to link these cone-opponent effects to the color opponency of Hering, two groups have suggested that the background mixture that yields the lowest threshold, which is therefore the balance point of the cone-opponent at the second site, should also be the mixture that is in blue–yellow equilibrium (i.e., the mixture that appears neither yellow nor blue).

Pugh and Larimer<sup>228</sup> investigated the detection sensitivity of  $\pi_1/\pi_3$  on field mixtures that appeared to be in yellow–blue equilibrium. They reasoned that if such mixtures are also null stimuli for the cone-opponent S–(L+M) second site, then detection sensitivity should depend only on first-site adaptation. Their results were consistent with their hypothesis, since field mixtures in yellow–blue equilibrium never produced superadditivity, a characteristic assumed to be due to second-site adaptation (see “Field Additivity” in Sec. 11.5). However, Polden and Mollon<sup>227</sup> looked specifically at the relationship between the maximum sensitization and the contribution of the longer wavelength component of the field mixture and found that it approximately followed a  $V(\lambda)$  spectral sensitivity. By contrast, the comparable equilibrium hue settings for long-wavelength fields fell much more steeply with wavelength for fields longer than 590 nm, implying that different underlying spectral sensitivities operate in the two cases.

Rinner and Gegenfurtner<sup>344</sup> looked at the time course of adaptation for color appearance and discrimination and identified three processes with half lives of less than 10 ms, 40 to 70 ms, and 20 s. The slow and intermediate adaptation processes were common to color appearance and discrimination, but the fast process affected only color appearance. However, in an extensive study, Hillis and Brainard<sup>78</sup> compared the effects of chromatic adaptation on color discrimination and asymmetric color matching. Pedestal discrimination measurements made as a function of pedestal intensity on five different chromatic backgrounds were used to estimate the response-intensity curves on each background. These response-intensity curves were then used to predict pairs of light that should match on different pairs of backgrounds. The agreement they found between the predicted asymmetric matches and measured ones suggests that color appearance and discriminability on different uniform backgrounds are controlled by the same underlying mechanism.<sup>78</sup> A follow-up study reached the same conclusion about the effect of habituation on “unstructured” spatiotemporal contrast,<sup>82</sup> although this study did not probe performance with test stimuli likely to tap purported higher-order chromatic mechanisms.

Overall, the results of these experiments are somewhat equivocal, but the majority of experiments find some correspondence between the effects of adaptation on color detection and discrimination on the one hand, and its effects on color appearance on the other. Given, however, that the discrimination and appearance processes have very different spectral sensitivities, this correspondence must break down under some conditions of chromatic adaptation, at least if the signals that control the adaptation of mechanisms depend strongly on the output of those same mechanisms. Of note here is that Hillis and Brainard did find clear dissociations of the effect of adaptation on discrimination and on appearance when the stimuli contained sufficient spatial structure to be perceived as illuminated surfaces.<sup>345</sup> Elucidation of the nature of this dissociation awaits further investigation.

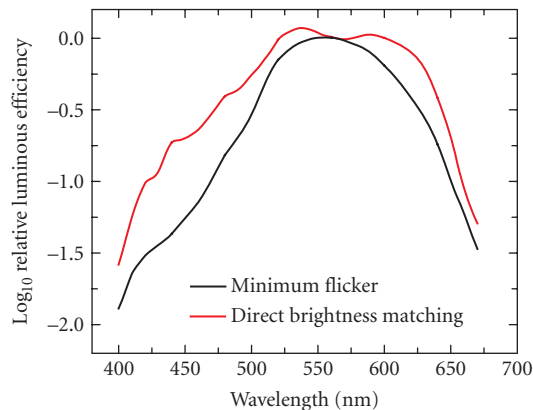
**Color Appearance and Habituation** The results of Webster and Mollon<sup>32,79</sup> were used as an example in subsection “Appearance Field Measurements and Second-Site Adaptation” in Sec. 11.4. Briefly, they

found that contrast adaptation produced changes in appearance that were selective for the habituating axis whether that axis was in the equiluminant plane or not. This finding is a potentially important difference between the effects of habituation on color appearance and its effects on detection, since only the appearance data provides clear evidence for higher-order mechanisms that are sensitive to both chromatic and luminance modulations. The detection data measured after habituation suggest that the chromatic and luminance mechanisms behave independently.<sup>241</sup> Determining conclusively whether this comparison across studies reflects a dissociation between habituation effects on thresholds and appearance, or instead represents a difference between the processing of near-threshold and suprathreshold test stimuli, could be resolved by applying the logic developed by Hillis and Brainard<sup>78,82</sup> to data collected for the full set of stimulus conditions studied by Webster and Mollon.<sup>32,79</sup>

Webster and Mollon<sup>32,79</sup> suggested two models that might account for their data. In the first, they replaced the three color mechanisms tuned to the three cardinal directions of color space with many mechanisms with tuning that varied according to a Gaussian distribution around each of the three cardinal directions. They found that they could account for the individual data by varying the standard deviations of the Gaussian distributions. In the second, they assumed that there were just three mechanisms, but that their tuning could be modified by adaptation. With adaptation, inhibition was assumed to build up between the mechanisms to decorrelate the input signals and reduce redundancy (see also Refs. 242–244). These models can also be applied to chromatic detection data.

**Luminance and Brightness** Brightness matching functions are broader than luminous efficiency functions measured using techniques such as HFP or MDB that produce additive results (see subsection “Sensitivity to Spectral Lights” in Sec. 11.5) and are thought to measure the sensitivity of the luminance channel.<sup>115,118,122,346</sup>

Figure 34 shows examples of a brightness matching function measured by direct brightness matching (red line) and luminous efficiency measured by flicker photometry (black line) replotted from Wagner and Boynton.<sup>122</sup> The brightness matching function is relatively more sensitive in the blue and orange spectral regions and slightly less sensitive in the yellow. These differences are usually attributed to a chromatic contribution to brightness but not to luminance.<sup>11,122,347</sup>



**FIGURE 34** Luminous efficiency and brightness. Mean matching data for three subjects replotted from Figs. 6 and 8 of Wagner and Boynton<sup>122</sup> obtained either using flicker photometry (black line) or using direct brightness matching (red line). The discrepancies are consistent with a chromatic contribution to brightness matching but not to flicker photometric matching.<sup>120,122</sup>



## Mechanisms of Color Constancy

The literature on mechanisms of color appearance is closely tied to a phenomenon known as “color constancy.” The need for color constancy stems from two observations. First, our normal use of color appearance is to associate colors with objects. Second, the stimulus reflected to the eye from a fixed object depends not only on the object’s intrinsic reflectance properties but also on the spectrum of the illumination. When the illuminant changes, so does the spectrum of the reflected light. If the visual system did not compensate for this change to stabilize the appearance of objects, it would not be possible to use color appearance as a reliable guide to object properties.

Empirically, color constancy is studied in much the same way as color appearance, using techniques of asymmetric matching, hue cancellation, or hue scaling. Papers published under the rubric of color constancy, however, have tended to employ stimulus conditions designed to model scenes consisting of illuminated objects and to emphasize context changes induced by manipulating the illuminant, rather than the simpler test-stimuli-on-background configurations favored for working out mechanistic accounts of color appearance.<sup>348–353</sup> The evidence supports the empirical generalization that when the illuminant is varied in a manner typical of the variations that occur in the natural environment, the visual system adapts to compensate for the physical change in reflected light, and thus to stabilize color appearance. That is, constancy in the face of naturally occurring illumination changes is good.<sup>352,353</sup> At the same time, it is important to note that constancy is not perfect: objects do not appear exactly the same across illumination changes. Moreover, other scene manipulations, such as those that change the reflectance of objects near the one of interest, also affect color appearance, even in rich scenes.<sup>354</sup> Figure 1 shows examples of this latter type of effect for simple stimulus configurations; these may be regarded as failures of constancy. Several recent reviews treat empirical studies of color constancy in more detail.<sup>355–357</sup>

Considerable theoretical attention has been devoted to modeling human color constancy. Much of this work is in the computational tradition, in which the theorist starts not with experimental data but rather by asking how constancy could in principle be achieved (or approximated), given the information about object reflectance properties that is available in the retinal image.<sup>358–363</sup>

With a few exceptions, work in the computational tradition does not connect explicitly with the mechanistic account of processing that we have developed in this chapter. Rather the work seeks to elucidate the sources of information in an image that could contribute to constancy. This is then used both to guide experiments<sup>354</sup> as well as in models whose purpose is to explicitly predict color appearance.<sup>364–367</sup> A particular promise of this approach is that it generalizes naturally as scene complexity increases, allowing (for example) predictions of the effects of manipulations in complex three-dimensional scenes using the principles developed for simpler scene configurations.<sup>368–370</sup>

A limitation of computational models is that they do not, in and of themselves, provide much insight about how neural mechanisms might act to provide constancy. Not surprisingly, then, there is also an important line of theoretical work that attempts to understand constancy in terms of the action of the first- and second-site mechanisms of adaptation that we have reviewed in this chapter.

Land’s retinex theory<sup>371–374</sup> may be understood as a computational model that incorporates first-site cone-specific gain control to stabilize the representation of object color. Several authors have shown that for natural illuminant changes, such first-site gain control provides excellent, although not perfect, compensation for the effect of illumination changes on the light reflected from natural objects as long as the gains are set correctly for the scene illumination.<sup>375–378</sup> Moreover, asymmetric matching data from experiments conducted using rich, naturalistic stimuli are well fit by postulating first-site gain control.<sup>352</sup> Webster and Mollon<sup>379</sup> (see also Ref. 380) extended this general line of thinking by showing that following first-site cone-specific adaptation with a second-site process that is sensitive to image contrast can improve the degree of compensation for the physical effect of illuminant changes. As we have reviewed in subsections “Appearance Field Measurements and Second Site Adaptation” in Sec. 11.4 and “Color Appearance and Habituation” in Sec. 11.5, it has also been established that contrast adaptation affects color appearance,<sup>32,79,379,381</sup> which provides a connection between human performance and the observation that contrast adaptation can help achieve color constancy.



A number of authors<sup>382–384</sup> have emphasized that it is clarifying, in the context of understanding constancy and more generally adaptation, to distinguish two questions. First, what parameters of visual processing (e.g., first-site gains, subtractive terms, second-site gains) are affected by contextual factors. That is, what can adapt? Second, what are the particular contextual features that act to set each of the adaptable parameters? Our understanding of these aspects is less developed. For example, it is not obvious how theories based on data obtained with spatially uniform adapting fields should be generalized to predict the adapted state for image contexts that have rich spatial and temporal structure. Simple hypotheses for generalization include the idea that a weighted average of the cone coordinates of the image might play the role of the uniform field, as might the most luminous image region. Ideas of this sort are implicit, for example, in Land's retinex theory. Explicit tests of hypotheses of this nature for rich scenes, however, have not revealed simple scene statistics sufficient to predict the visual system's state of adaptation.<sup>354</sup> Indeed, understanding what image statistics are used by the visual system to achieve constancy, and connecting the extraction of these statistics explicitly to mechanistic sites of adaptation is a primary focus of current research on constancy. In this regard, the computational work may provide critical guidance, as at the heart of each computational theory is an analysis of where in the retinal image the most valuable information about the illumination may be found. Zaidi<sup>385</sup> and Golz and MacLeod,<sup>386</sup> for example, translate analyses of image statistics that carry information about the illuminant into mechanistic terms.

## Color and Contours

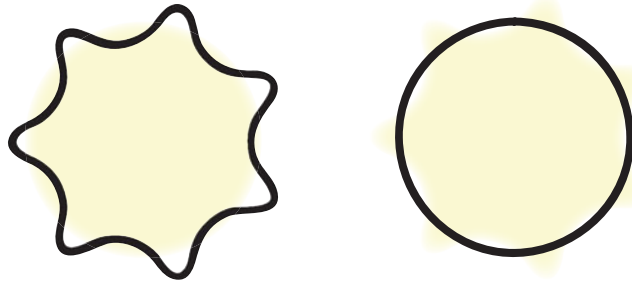
In the natural world, differently colored regions are usually delimited by contours across which there is also a luminance difference. If the luminance difference is removed, the contour becomes indistinct, and it becomes more difficult to discriminate the chromatic differences between the two regions. For example, Boynton, Hayhoe, and MacLeod<sup>387</sup> showed that color discrimination is severely impaired if the difference between two regions is dependent on only the S cones, and Eskew and Boynton<sup>388</sup> showed that the contour remains indistinct even for chromaticity differences up to twice threshold. For two regions separated by a border defined by S cones (i.e., if the regions are tritan pairs), "melting" of the border<sup>389</sup> and "chromatic diffusion" between the regions<sup>388</sup> have been reported.

Figure 35a shows two examples of chromatic regions filling in an area bordered by a luminance contour. These are versions of the "Boynton illusion" (see p. 287 of Ref. 15 and <http://www.yorku.ca/eye/boynton.htm>). The yellow areas, which are discriminated from the white background mainly by S cones, appear to fill-in the black borders as you move further away from the figure. The filling-in illustrates the tendency of luminance contours to constrain the spatial extent of signals mediated by the chromatic pathways.

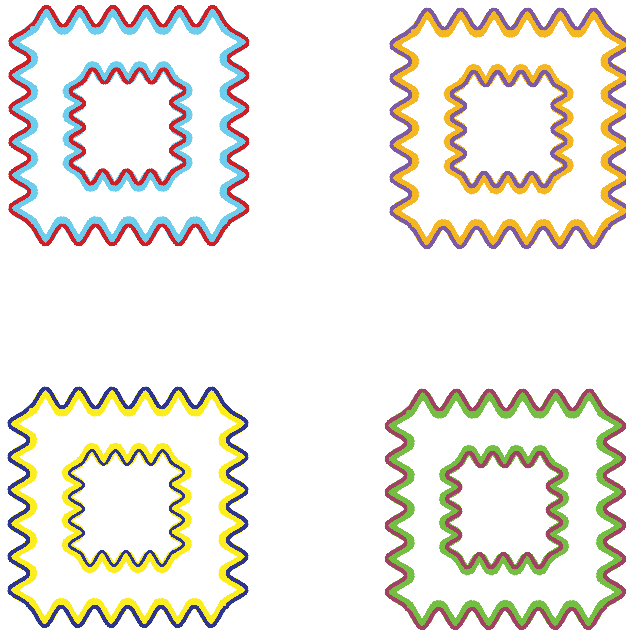
Figure 35b shows four examples of a related illusion, which is known as the watercolor illusion or effect.<sup>390–392</sup> If an area is delineated by a darker outer chromatic contour flanked by a brighter inner chromatic contour, then the brighter color spreads faintly into the inner area.<sup>393</sup> This faint coloration resembles a "watercolor" wash.

**The "Gap" Effect and Luminance Pedestals** The gap effect is a well-known phenomenon described by Boynton, Hayhoe, and MacLeod.<sup>387</sup> A small gap introduced between two juxtaposed test fields improves their discriminability if the fields differ only in S-cone excitation, impairs it if they differ only in luminance, and can result in a small improvement if they differ only in L- and M-cone excitation at constant luminance (i.e., in L–M excitation).<sup>387,394</sup> The small gap that separates the tritan pair and improves discriminability can be produced either by a luminance difference or by an L–M chromatic difference.<sup>395</sup>

The improvement in chromatic discriminability caused by gaps can be related to the improvement in chromatic sensitivity when a luminance pedestal or contour is coincident with the chromatic stimulus<sup>169,177,194,269,271,272</sup>—the so-called "crossed" pedestal facilitation described above (see subsection "Pedestal Experiments" in Sec. 11.5). Moreover, as also noted above, the crossed facilitation survives if the luminance pedestal is replaced by a thin ring<sup>169</sup> (see also Ref. 177). Montag<sup>394</sup> compared the gap



(a)



(b)

**FIGURE 35** Boynton and watercolor illusions. (a) Two examples of the Boynton illusion. In the left picture, a star-shaped black outline is superimposed on a circular yellow area, while in the right picture a circular black outline is superimposed on a star-shaped yellow area. As you move away from the picture, the yellow areas appear to fill the black outlines. These are variations of an illusion attributed to Robert M. Boynton by Peter Kaiser, see: <http://www.yorku.ca/eye/boynton.htm>. (b) Four examples of the watercolor illusion.<sup>390–392</sup> Each square is outlined by a pair of juxtaposed sinuous contours. The similar colors of the inner contour of each larger square and the outer contour of each smaller square fill in the intervening space with a faint color (which looks like a watercolor wash). Inspired by a version of the watercolor illusion designed by Akiyoshi Kitaoka, see: <http://www.psy.ritsumeai.ac.jp/~akitaoka/watercolorillusionsamples.jpg>.

effect with facilitation produced by adding thin lines of the same orientation to gratings at half-cycle intervals. He found that the detection of S-cone gratings was facilitated by placing dark lines at the midpoints between the peaks and troughs of the gratings (i.e., where the luminance of the grating is equal to the mean luminance of the stimulus—zero crossings of the spatially varying factor in the stimulus), but that the facilitation declined to zero as the lines were moved toward the peaks and troughs. A comparable but smaller effect was found for equiluminant L- and M-cone modulated gratings, while the detection of luminance gratings was always impaired. Montag suggested that the gap effect and pedestal enhancement may be complementary effects.<sup>394</sup> Indeed, Gowdy, Stromeyer, and Kronauer<sup>396</sup> reported that the facilitation of L–M detection by luminance pedestals is markedly enhanced if the luminance grating is square-wave rather than sinusoidal. They argued that the square-wave produces an abrupt border across which the L–M mechanism is somehow able to compare chromatic differences.

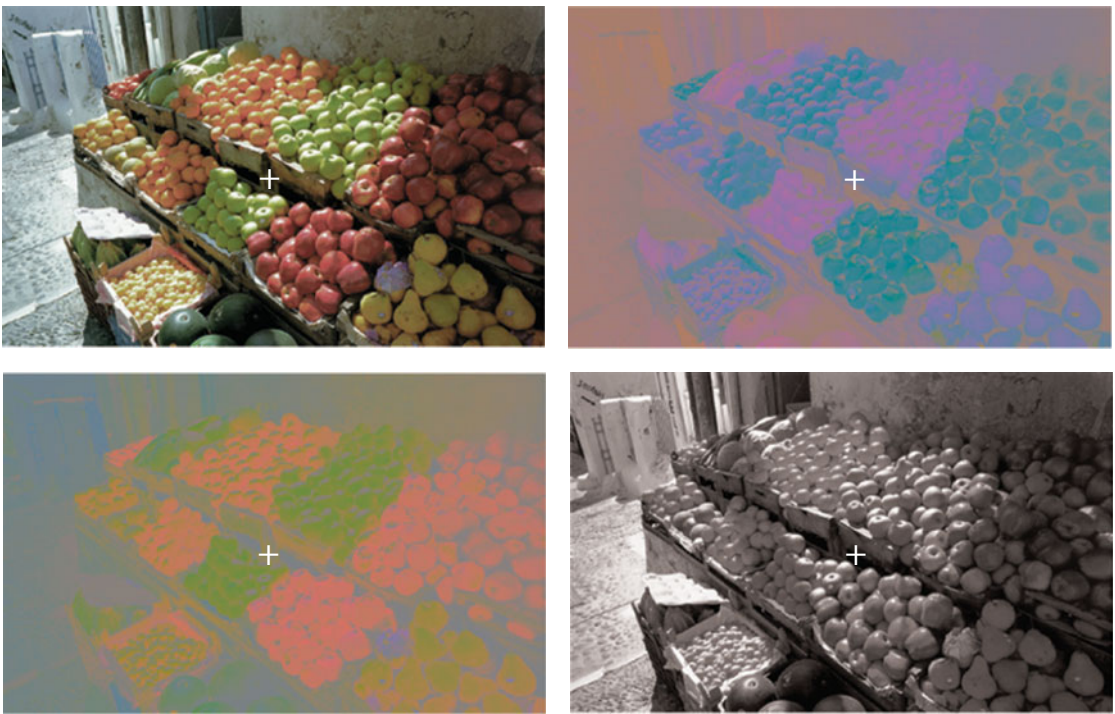
**Color Appearance and Stabilized Borders** The importance of contours on color appearance is made clear in experiments in which the retinal image is partially or wholly stabilized. When the retinal image is wholly stabilized, the color and brightness of the stimulus fades until it is no longer seen.<sup>397</sup> The color appearance of partially stabilized images in which stabilized and unstabilized borders are combined can be entirely independent of the local quantal absorptions. Krauskopf<sup>398</sup> found that when the border between a disc and a surrounding annulus is stabilized, and the outer border of the annulus unstabilized, the disc “disappears” and its color is filled in by the color of the annulus. When, for example, the border between a green, 529-nm annulus, and an orange, 640-nm disc, is stabilized, the disc appears green. These changes are not restricted to the stabilized disc. Unstabilized targets presented on stabilized discs also change their color appearance as the color of the disc fills in to the color of the annulus.<sup>399</sup> For example, if the annulus is yellow and the stabilized disc is either red or green, both discs fill in to take on the yellow color of the annulus. Unstabilized yellow targets presented in the center of the stabilized discs, however, take on the color appearance complementary to actual color of the disc, so that the yellow target on a green disc appears red, and the yellow target on a red disc appears green. Such changes are consistent with the local contrast signals at the target edges determining the filled-in appearance of the yellow target (see Experiment 4 of Ref. 400).

It has also been reported that stabilized boundaries can produce colors that appear reddish-green and yellowish-blue, and so violate the predictions of the opponent-colors theory that opposed colors (red and green, or yellow and blue) cannot be perceived simultaneously. These “forbidden colors” were produced by presenting two vertical stripes side-by-side, with their common border stabilized, but their outer borders unstabilized. When the juxtaposed stripes were red and green, most observers reported that the merged stripes appeared reddish-green or greenish-red, and when the stripes were blue and yellow most observers reported that they appeared bluish-yellow<sup>401</sup> (see Experiment 3 of Ref. 400). Forbidden colors were also reported by some subjects in another study, but only when the stripes were equiluminant.<sup>402</sup>

The results obtained using stabilized and unstabilized borders add to the view that color appearance need not be a local retinal phenomenon. A crucial question, then, is whether the changes in color appearance caused by image stabilization also influence phenomena that are ostensibly more low-level, and thus more likely to be local, such as detection. Two experiments have addressed this question. Neger, Piantanida, and Larimer<sup>403</sup> found that when a red disk was surrounded by a yellow annulus, stabilizing the edge between the two fields on the retina caused the yellow to fill in, making the disk appear yellow too. This filling-in affected the color appearance of small tests added to the disk, but it did not affect their increment threshold. The results for S-cone flicker detection, somewhat surprisingly, suggest that filling-in can change S-cone flicker sensitivity. Previously, Wisowaty and Boynton,<sup>404</sup> using flickering tritan pairs to isolate the S-cone response, had found that yellow background adaptation reduces the S-cone modulation sensitivity compared to no background adaptation. Piantanida<sup>405</sup> showed that the reduction in S-cone flicker sensitivity caused by a yellow field could also be caused by a dark field that only appeared yellow because its outer border with a yellow annulus was stabilized. In a related experiment, Piantanida and Larimer,<sup>400</sup> compared S-cone modulation sensitivities on yellow and green fields, and found that the S-cone modulation sensitivity was

lower on the yellow field. Again, this sensitivity difference depended upon the field appearance rather than upon its spectral content. Thus, the reduced sensitivity was found not only on the yellow field but also on a green field that appeared yellow because its border with a surrounding yellow annulus was stabilized. Similarly, the increased sensitivity was found not only on the green field but also on a yellow field that appeared green because its border with a surrounding green annulus was stabilized. The differences between studies may indicate that different sites of limiting noise act to limit detection in the different tasks.

**Contours and Aftereffects** Daw<sup>406</sup> showed that the saliency of colored after-images depended upon the presence of an aligned monochrome image. Thus, adaptation to a colored image produces a clear afterimage if the after-image is aligned with a monochrome image with the same contours, but produces a poor after-image if the two are misaligned. Figure 36 demonstrates this effect.



**FIGURE 36** Color and contour. Four different images of a fruit stall are shown. The full color image (top left) has been decomposed into its chromatic (bottom left) and luminance components (bottom right). The image on the top right is the complementary or inverse of the chromatic image. Notice that the details of the image are much better preserved in the luminance image than in the chromatic image. The chromatic information, although important for object discrimination and identification, is secondary to the perception of form. Color fills-in the picture delineated by the luminance information. This can be demonstrated by fixating the cross in the center of the complementary chromatic image (top right) for several seconds, and then diverting your gaze quickly to the cross in the center of the luminance image (bottom right). You should see a correctly colored version of the picture. Notice that if you shift your gaze slightly, so that the after-image no longer aligns precisely with the luminance image, the color disappears. You can also try adapting to the chromatic image (bottom left) and repeating the experiment. The effects are stronger with projected CRT versions of the images and successive presentation, which can be seen at <http://www.cvrl.org>. Color scene after Fig. 1.14A of Sharpe et al.<sup>477</sup> from an original by Minolta Corp. A version of this aftereffect demonstration was first described in 1962 by Nigel Daw.<sup>406</sup> Other examples of this illusion along with instructions about how to produce comparable images can be found at: [http://www.johnsadowski.com/color\\_illusion\\_tutorial.html](http://www.johnsadowski.com/color_illusion_tutorial.html).

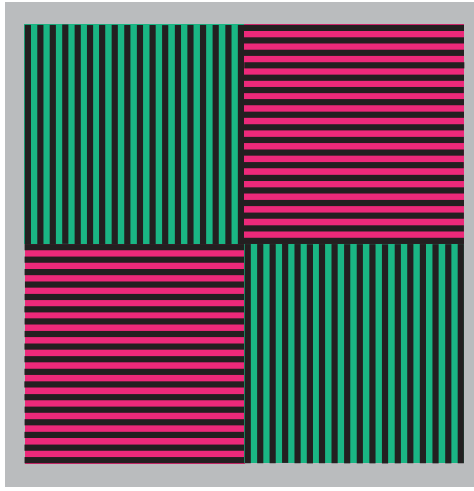
Impressive demonstrations of the influence of contours on after-images have recently been produced by van Lier and Vergeer<sup>407</sup> and van Lier, Vergeer, and Anstis.<sup>408</sup> Their demonstrations show that colored stimuli can produce different after-image colors at the same retinal location, depending on the positioning of contours presented with the after-image. In one version, observers adapt to a plaid pattern made up of blue, green, orange, and purple squares, and then view, together with the after-image, either horizontal or vertical contours aligned with the borders of the squares. The color of the after-image changes with the orientation of the contours, because the “mean” after-image along the horizontal rows and along the vertical columns is different, thanks to the arrangement of the adapting squares in the plaid. The demonstration can be seen at: <http://www-psy.ucsd.edu/~sanstis/SAai.html>.

**McCollough Effect** The McCollough effect is a well-known orientation-contingent color after-effect.<sup>409</sup> After prolonged viewing of colored “adapting” gratings of, say, vertical or horizontal orientation, neutral, black-and-white test gratings of the same orientation take on the complementary hue of the adapting grating. For example, following adaptation to red horizontal and green vertical gratings, horizontal and vertical black-and-white gratings appear, respectively, slightly greenish and reddish (see Fig. 37).

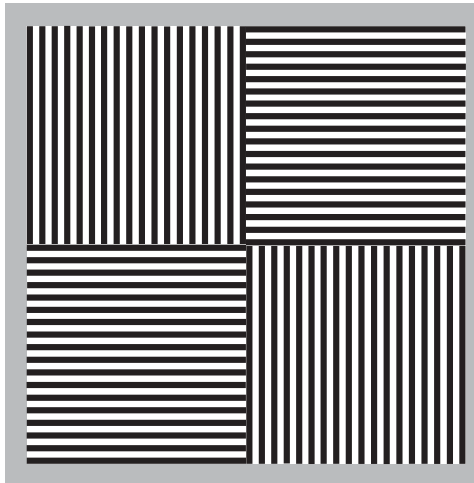
The McCollough effect can persist undiminished for many days provided that test gratings are not viewed.<sup>410</sup> This prolonged effect is distinct from colored afterimages, such as the effect demonstrated in the previous subsection, which decline relatively quickly. Vul, Krizay, and MacLeod<sup>411</sup> have recently identified two processes in the generation of the McCollough effect. The first process behaves like a leaky integrator, and produces an effect that declines exponentially with a time constant of roughly 30 s. The second process behaves like a perfect integrator, and produces an effect that shows no decay.<sup>411</sup> Several lines of evidence suggest that the McCollough effect is generated relatively early in the visual pathway. First, the McCollough effect exhibits little interocular transfer.<sup>409,412,413</sup> Second, it can be produced by alternating red-striped and green-striped adapting patterns at rates as high as 50 Hz, frequencies at which the colored patterns cannot be consciously perceived.<sup>414</sup> Third, it is dependent on the spectral content of the inducing stimuli, not their color appearance.<sup>415</sup> These findings suggest that the “adaptation” has an effect that occurs at a relatively early visual site that precedes binocularity, is able to respond high-frequency flicker, and retains some veridical information about the spectral content of lights. However, the contingency of the McCollough effect on grating orientation suggests that the site cannot be earlier than primary visual cortex, V1, where orientation selectivity is first clearly established.<sup>416,417</sup> There is an extensive literature on the McCollough effect (e.g., Refs. 418–423).

**Multiplexing Chromatic and Achromatic Signals** Information about color and contour may be transmitted—at least in part—by the same postreceptoral pathways. The ways in which these mixed signals are demultiplexed may give rise to some of the color and border phenomena described in the previous subsections. The evidence for the color-luminance multiplexing comes mainly from knowledge of the spatial properties of neuronal receptive fields in the early visual pathway. Cone-opponent, center-surround mechanisms that are also spatially opponent encode not only chromatic information, which is dependent upon the difference between the spectral sensitivities of their center and surround, but also “achromatic” information, which is dependent on the sum of their spectral sensitivities. For color, the center and surround behave synergistically producing a low-pass response to spatial variations in chromaticity, but for luminance they behave antagonistically producing a more band-pass response to spatial variations in luminance.<sup>149–151</sup> This type of multiplexing can, in principle, apply to any mechanism that is both chromatically and spatially opponent, but it is most often considered in the context of P-cells or midget ganglion cells, which make up as much as 80 percent of the primate ganglion cell population.<sup>424</sup> These cells are chromatically opponent with opposed L- and M-cone inputs,<sup>36</sup> but also respond to spatial differences in luminance.<sup>425,426</sup> In the fovea, they may be chromatically opponent simply by virtue of having single L- or M-cone centers<sup>427,428</sup> and mixed surrounds.<sup>429–431</sup> How segregated the L- and M-cone inputs are to P-cell surrounds and to P-cell centers in the periphery remains controversial.<sup>432–440</sup>

The multiplexing of color and luminance signals may just be an unwanted and unused artifact of having both spatial and chromatic opponency in P-cells. Indeed, Rodieck<sup>441</sup> has argued that the L–M



(a)



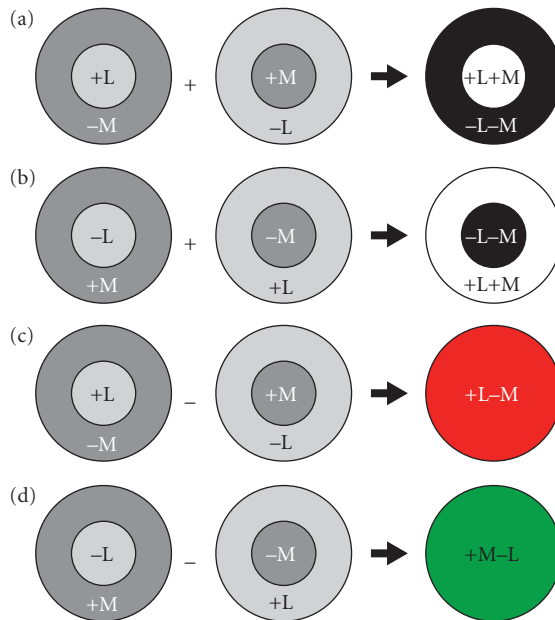
(b)

**FIGURE 37** McCollough effect. View the upper colored image (a) for several minutes letting your gaze fall on different colored areas for several seconds at a time. Look next at the lower monochrome image (b).

opponent system is mediated instead by a population of so-called Type II cells with coincident centers and surrounds, which are chromatically but not spatially opponent. For the multiplexing of color and luminance signals in cells with concentric center-surrounds to be unambiguously useful to later visual stages, the signals must be decoded. Several decoding strategies have been suggested, most of which depend on the chromatic signal being spatially low-pass filtered by the center-surround interaction, and the luminance signal being spatially band-pass filtered. Because of this filtering, signals between adjacent mechanisms will, in general, change slowly if the signals are chromatic and rapidly if they are achromatic. Thus, the chromatic and luminance signals can be decoded by spatially

low-pass or band-pass filtering, respectively, across adjacent mechanisms.<sup>149,150</sup> In such a scheme, low spatial-frequency luminance information and high spatial-frequency chromatic (edge) information is lost. The high-frequency luminance or edge information can be used both in “filling-in” the low spatial frequency luminance information, and to define missing chromatic edges. Several well-known visual illusions, such as the Craik-O’Brien-Cornsweet illusion (see Ref. 442), are consistent with filling-in, while others, such as the Boynton illusion (see Fig. 35a), are consistent with luminance edges defining chromatic borders.

Simple mechanisms for decoding the luminance and chromatic signals that difference or sum center-surround chromatically opponent neurons have been proposed.<sup>153,443–445</sup> Such mechanisms are, in fact, mathematically related to decoding using matched spatial filters.<sup>445</sup> In one version, spatially superimposed opponent cells with different cone inputs are either summed or differenced.<sup>153</sup> The band-pass luminance signal is decoded by summing  $+Lc-Ms$  and  $+Mc-Ls$  to give  $+(L+M)c-(L+M)s$  and by summing  $-Mc+Ls$  and  $-Lc+Ms$  to give  $-(L+M)c+(L+M)s$ , both of which are also bandpass and, in principle, achromatic (where  $c$  = center,  $s$  = surround). The low-pass chromatic signal is decoded by differencing  $+Lc-Ms$  and  $-Mc+Ls$  to give  $(+L-M)c,s$  and by differencing  $-Lc+Ms$  and  $+Mc-Ls$  to give  $(+M-L)c,s$ , both of which have spatially coincident centers and surrounds. These combination schemes are illustrated in Fig. 38. As Kingdom and Mullen<sup>446</sup> point out, superimposed centers with



**FIGURE 38** Double-duty L-M center-surround opponency. Achromatic and chromatic information can be demultiplexed from LM-cone-opponent center-surround mechanisms by summing or differencing different types of mechanisms. Spatially opponent On and Off center achromatic mechanisms can be produced by summing, respectively, either (a) L and M On-center mechanisms, or (b) L and M Off-center mechanisms. Spatially nonopponent L-M and M-L chromatic mechanisms can be produced by differencing, respectively, either (c) L and M On-center mechanisms, or (d) L and M Off-center mechanisms. See Fig. 24 of Lennie and D’Zmura<sup>478</sup> and Fig. 2 of Kingdom and Mullen.<sup>446</sup>



different cone inputs must be spatially offset, simply because two cones cannot occupy the same space. They found that such offsets produce significant crosstalk between the decoded luminance and chromatic outputs. However, this problem is mitigated somewhat by the blurring of the image by the eye's optics, and, in the case of chromatic detection, by neural blurring (see Fig. 7 of Ref. 200).

In another version of the decoding mechanism, De Valois and De Valois<sup>18</sup> added S-cone inputs to generate spectral sensitivities more consistent with opponent-colors theory (see subsection "Spectral Properties of Color-Opponent Mechanisms" in Sec. 11.5). Billock,<sup>445</sup> by contrast, opposed P-cells of the same sign in both the center and surround, thus directly generating double-opponent cells. This can be achieved by differencing (e.g., +L–M in the center differenced from +L–M in the surround) or by summing (e.g., +L–M in the center summed with –L+M in the surround).

The importance of multiplexing in spatially and chromatically opponent mechanisms, and, in particular, in P-cells, remains controversial. Clearly, the P-cells, which make up 80 percent of the ganglion cells and have small receptive fields, must be important for spatial, achromatic vision, but their importance for color vision, while likely, is not yet firmly established. Speculative arguments have been based on the phylogenetically recent duplication of the M/L-cone photopigment gene, which is thought to have occurred about 30 to 40 million years ago after the divergence of Old and New World primates.<sup>447</sup> The typical argument is that the color opponency in P-cells is parasitic on an "ancient" system that before the duplication detected only luminance contrast.<sup>448</sup> However, this raises the possibility that the "decoding" of color information might be done simply to remove it and so improve image fidelity rather than to provide a color signal.

Many psychophysical experiments have attempted to elucidate the properties of the P-cells by employing equiluminant stimuli.<sup>449</sup> The idea is that these stimuli silence the M-cells and thus reveal the properties of the P-cells. If we accept this logic, there is still, of course, an important limitation to these experiments: they do not, by their design, provide any information about how the P-cells respond to luminance modulation. Thus, these experiments can provide only a partial characterization of the P-cells' response.

## 11.6 CONCLUSIONS

In Sections 11.3 and 11.4, we introduced basic mechanistic models for color discrimination and color appearance. Section 11.5 reviewed evidence that both supports the basic model and indicates areas where it is deficient. Here we summarize what we see as the current state of mechanistic models of color vision.

### Low-Level and Higher-Order Color-Discrimination Mechanisms

Test methods provide clear evidence for the existence of three low-level "cardinal" postreceptoral mechanisms: L–M, S–(L+M), and L+M in detection and discrimination experiments (see subsection "Test Sensitivities" in Sec. 11.5). This simple picture is complicated by evidence that L–M and L+M may have small S-cone inputs (see subsection "Sensitivity to Different Directions of Color Space" in Sec. 11.5), but nevertheless the weight of the evidence suggests that just three bipolar cardinal mechanisms (or six unipolar mechanisms—see next section "Unipolar versus Bipolar Chromatic Mechanisms") are sufficient to understand the preponderance of the data. Consistent with the existence of cardinal mechanisms, near-threshold color-discrimination experiments, with one exception,<sup>241</sup> require only four or five unipolar detection mechanisms.<sup>179,197,260</sup> Field methods (see subsection "Field Sensitivities" in Sec. 11.5) also confirm the importance of the three cardinal mechanisms. This agreement across methods provides strong psychophysical evidence for the independent existence of these mechanisms.

Test methods provide little or no evidence for the existence of "higher-order" (or noncardinal) mechanisms. Evidence for higher-order mechanisms comes almost entirely from field methods carried



out mainly in the DKL equiluminance plane using detection and habituation,<sup>241</sup> detection and noise masking,<sup>245,248,450</sup> discrimination and habituation,<sup>241,255</sup> texture segmentation and noise,<sup>451</sup> and image segmentation and noise.<sup>452</sup> In contrast, three experiments using detection and noise carried out with stimuli specified in cone contrast space found little or no evidence for higher-order mechanisms.<sup>175,179,246</sup>

In principal, the choice of space should not influence the results. However, the choice of space used to represent the data does tend to affect the ensemble of stimuli studied, in the sense that stimuli that uniformly sample color directions when they are represented in one space do not necessarily do so when they are represented in another. Another concern is that experiments carried out at nominal equiluminance are liable to luminance artifacts, since the elimination of luminance signals is, in practice, difficult to achieve. Aside from limits on calibration precision or inaccuracies introduced by the assumption that the 1924 CIE  $V(\lambda)$  function (i.e., candelas/m<sup>2</sup>) defines equiluminance, a recurring difficulty in producing truly equiluminant stimuli is that the luminance spectral sensitivity varies not only with chromatic adaptation and experimental task, but also shows large individual differences (see subsection “Luminance” in Sec. 11.5). Moreover, the luminance channel almost certainly has cone-opponent inputs (see Fig. 25). Finally, even when the luminance mechanism is silenced, the low-level L–M mechanism may have a small S-cone input, which will rotate the mechanism axis away from the cardinal axis.

These concerns notwithstanding, the results of field methods also suffer from an inherent ambiguity. Just because noise masking or habituation alters the detection spectral sensitivity does not necessarily reveal transitions between multiple higher-order mechanisms. Such changes can also result if the simple model used to characterize mechanisms fails to capture, for example, low-level interactions and nonlinearities. Indeed, Zaidi and Shapiro<sup>243</sup> have suggested that such “failures” might actually be an adaptive orthogonalization among cardinal mechanisms that reduces sensitivity to the adapting stimulus and improves sensitivity to the orthogonal direction. Moreover, noise masking can produce rotated detection contours by changing the balance of noise at the first and second sites (see subsection “Sites of Limiting Noise” in Sec. 11.3) without the need to invoke high-order mechanisms.

Of course, it would be absurd to suppose that higher-order color mechanisms do not exist in some form. They are clearly found in more “cognitive” experiments such as Stroop or reverse-Stroop experiments, which show clearly the effects of color categorization (e.g., Ref. 453). The central question with respect to color discrimination, however, is how such mechanisms can be revealed and investigated psychophysically using conventional threshold techniques. For example, if the site of limiting noise precedes higher-order mechanisms for most stimulus conditions, this may not be possible. In that case, the study of higher-order mechanisms will require the use of appearance or other experimental tasks in addition to knowledge of the input characteristics.

On the other hand, some discrimination tasks do provide clear evidence for higher-order mechanisms. Zaidi and Halevy<sup>454</sup> showed that the discrimination thresholds for an excursion in color space away from a background, the chromaticity of which was modulated at 0.46 Hz along a circle in the DKL equiluminant-color plane, depended on the direction of the background color change. Discrimination was consistently worse if the excursion was in the same direction as the background color change, than if it was in the opposite direction. Since this effect was independent of the actual background chromaticity, the effect is inconsistent with there being a limited number of cardinal mechanisms. More generally, Flanagan, Cavanagh, and Favreau<sup>31</sup> used the tilt after-effect to investigate the spectral properties of the orientation-selective mechanisms thought to underlie the after-effect. They found that the orientation-selectivity occurs independently in each of the three cardinal mechanisms, but they also found evidence for secondary orientation-selective mechanisms at other directions in color space. In addition, they found evidence for another orientation-selective mechanism that was nonselective for color directions in any color direction. Other experiments that suggest high-order mechanisms include visual search,<sup>455</sup> color appearance,<sup>32,79,456</sup> and the motion coherence of plaid patterns.<sup>457</sup> A unified understanding of when and why higher-order mechanisms are revealed experimentally remains an important question for the field.

## Unipolar versus Bipolar Chromatic Mechanisms

The strong implication that cone-opponent and color-opponent mechanisms may be unipolar rather than bipolar came first from physiology. The inhibitory range of LGN opponent cells

(i.e., the extent of decreases in firing rate from the resting or spontaneous level) is seldom more than 10 to 20 impulses per second, compared with an excitatory range of several hundred (i.e., the extent of increases in firing rate from the resting level).<sup>458</sup> Consequently, cells of both polarities (i.e., L–M and M–L) are required in a push-pull relationship to encode fully both poles of a cone-opponent response. By the cortex, the low spontaneous firing rates of cortical neurons makes cells effectively half-wave rectifiers, because they cannot encode inhibitory responses.<sup>459,460</sup> Thus, the bipolar L–M, M–L, S–(L+M), and (L+M)–S cone-opponent mechanisms must become unipolar |L–M|, |M–L|, |S–(L+M)|, and |(L+M)–S| cone-opponent mechanisms, while the bipolar R/G and B/Y mechanism become unipolar R, G, B, and Y mechanisms.

Behavioral evidence for unipolar mechanisms at the cone-opponent level is sparse in the case of the L–M mechanism, because the opposite polarity L–M and M–L detection mechanisms have similar cone weights (see subsection “Detection contours in the L, M Plane” in Sec. 11.5), so in many ways they behave like a unitary bipolar mechanism.<sup>34</sup> Nevertheless, there is some evidence that the |L–M| and |M–L| poles behave differently. In particular, habituations to sawtooth chromaticity modulations can selectively elevate thresholds for |L–M| or |M–L| targets.<sup>14,461</sup> Sankeralli and Mullen<sup>462</sup> raised detection threshold using unipolar masking noise and found that masking was most evident when the test and noise had the same polarity (thus, |L–M| noise selectively masks an |L–M| test, and |M–L| noise selectively masks an |M–L| test). Other evidence suggests that |L–M| and |M–L| are separate mechanisms at the fovea.<sup>396,463</sup> Evidence for separate mechanisms is also apparent in the peripheral retina, where the |L–M| mechanism predominates.<sup>176,178,464</sup>

There is better evidence for the existence of unipolar S-cone mechanisms. Sawtooth chromaticity modulations selectively elevate thresholds for |S–(L+M)| or |(L+M)–S| targets,<sup>14,461</sup> as do unipolar |S–(L+M)| or |(L+M)–S| masks.<sup>462</sup> Other evidence shows that (1) the spatial summation for S-cone incremental and decremental flashes differs,<sup>465</sup> (2) the longer wavelength field sensitivity for transient tritanopia is different for incremental and decremental flashes, which indicates different L- and M-cone inputs for the two S cone polarities,<sup>466</sup> and (3) habituation or transient adapting flashes have differential effects on S increment and S decrement thresholds.<sup>467,468</sup> There is also ample physiological and anatomical evidence that S-cone ON and OFF pathways are distinct (e.g., Ref. 2). For a recent review, see Ref. 34.

Eskew<sup>34</sup> takes the more philosophical stance that a psychophysical mechanism should be defined as a univariant mechanism, the output of which is a “labeled line.”<sup>469,470</sup> This means that a bipolar cone-opponent mechanism cannot by definition be a unitary mechanism, since it can exist in two discriminable states (positive and negative).

## Discrepancies between Color-Discrimination and Color-Appearance Mechanisms

Much evidence indicates that color-discrimination and color-appearance mechanisms are distinct, at least given the linking hypotheses currently used to connect mechanism properties to performance.<sup>14,16–22</sup> As with color-discrimination mechanisms, much color appearance data may be accounted for by postulating three postreceptoral mechanisms. But the detailed properties of these mechanisms differ from those of the color-discrimination mechanisms.

In general, excitation of single cone-opponent mechanisms along cardinal directions does not give rise to the perceptions expected of isolated color-opponent mechanisms. Thus, the modulation of L–M around white produces a red/magenta to cyan color variation, whereas modulation of S around white produces a purple to yellow/green modulation.<sup>31,32</sup>

In specific tests of the relationship between color-discrimination and color-appearance mechanisms, unique hues, which correspond to the null vector of color-appearance mechanisms (see subsection “Spectral Properties of Color-Opponent Mechanisms” in Sec. 11.5), have been compared with the null vectors of color-discrimination mechanisms. Webster et al.<sup>298</sup> plotted pairs of unique colors of different saturation in DKL space. They found that only the unique red pair aligned with one of the cone-opponent axes, the L–M axis. The blue and yellow pair aligned roughly along an intermediate axis, but the green pair was not colinear with the red pair. Thus, unique blue and yellow, in particular,

must reflect the joint activity of L–M and S–(L+M). Chichilnisky and Wandell<sup>289</sup> and Wuerger, Atkinson, and Cropper<sup>22</sup> found that their four unique hue planes did not correspond to the null planes of color-discrimination mechanisms. Using a hue-naming technique, De Valois et al.<sup>21</sup> found that hue names are also not easily related to modulations of color-discrimination mechanisms (see Fig. 31).

The inconsistencies between color-discrimination and color-appearance mechanisms strongly suggest that the two are fundamentally different. But can they be considered to be mechanisms at different levels of a common serial processing stream or are they even more distinct? And if they do act serially, how might they be related? The simplified three-stage zone model described in the next section suggests that they could be parts of a common serial process and, at least for R/G, very simply related.

### Three-Stage Zone Models

As noted in the Introduction (see section “The Mechanistic Approach” in Sec. 11.2), linear three-stage Müller zone models,<sup>23</sup> in which the second stage is roughly cone-opponent and the third stage color-opponent, have been proposed by Judd<sup>24</sup> and more recently by Guth<sup>29</sup> and De Valois, and De Valois.<sup>18</sup>

The De Valois and De Valois<sup>18</sup> three-stage zone model is an interesting example based on physiological and anatomical assumptions about the relative numerosity of cone inputs to center-surround-opponent neurons. In their indiscriminate-surround model, neurons are assumed to have single cone inputs to their centers and mixed cone inputs to their surrounds (in the ratio of 10L:5M: S). Consequently, both the L–M and M–L cone-opponent stages have –S inputs, as a result of which the third color-opponent stage with an S input to R/G is arguably not needed (see Guth<sup>471</sup> and De Valois and De Valois<sup>472</sup>). [The dangers of assuming that cone weights can be simply related to relative cone numerosity were discussed before in the context of luminous efficiency (see subsection “Luminance” in Sec. 11.5).] The spectral sensitivities of the De Valois and De Valois<sup>18</sup> third-stage color-opponent mechanisms are shown in Fig. 32, as dashed black lines. Despite their physiologically based approach, the De Valois and De Valois<sup>18</sup> zone model inevitably derives from earlier psychophysical models, partly because there are only a limited number of ways in which the cone fundamentals can be linearly combined to produce plausible cone-opponent and color-opponent spectral sensitivities. As well as three-stage models, there are also many examples of two-stage models, in which the second stage is designed to account either for color-discrimination or for color-appearance data in isolation.<sup>8–13,15</sup>

As an exercise, we next provide an illustrative example of a linear three-stage zone model based on a few very simple assumptions about the signal transformations at each stage. Our goal was to see if we could derive plausible estimates of the zero crossings of opponent-colors theory without resorting to speculative physiological models or psychophysical data fitting.

*First zone* At the first stage, we assume the Stockman and Sharpe<sup>28</sup> cone fundamentals:  $\bar{l}(\lambda)$ ,  $\bar{m}(\lambda)$ , and  $\bar{s}(\lambda)$ , the spectral sensitivities of which are labelled Stage 1 in Fig. 4. These functions are normalized to unity peak.

*Second zone* At the second stage, we assume classical cone opponency. In the case of L–M (and M–L), we assign equal cone weights, thus yielding  $\bar{l}(\lambda) - \bar{m}(\lambda)$ , and its opposite-signed pair  $\bar{m}(\lambda) - \bar{l}(\lambda)$ . This is consistent with the evidence from psychophysical detection experiments for equal L- and M-cone contrast weights into L–M (see subsection “Sensitivity to Different Directions of Color Space” in Sec. 11.5). Note that the zero crossing of this cone-opponent L–M mechanism is 549 nm, which is far from unique yellow. For the zero crossing to be near the 580 nm—the unique yellow assumed at the next stage—the relative M:L cone weight would have to be increased from 1 to 1.55 (see, e.g., Fig. 7.4 of Ref. 15).

In the case of S–(L+M) and (L+M)–S, we assign half the weight to M as to L (in accordance with many other models) and then scale both by 0.69 to give equal weights (in terms of peak spectral sensitivity) to S and L+0.5M, thus yielding  $\bar{s}(\lambda) - 0.69[\bar{l}(\lambda) + 0.5\bar{m}(\lambda)]$  and its opposite signed pair  $0.69[\bar{l}(\lambda) + 0.5\bar{m}(\lambda)] - \bar{s}(\lambda)$ . The zero crossing of this mechanism is 486 nm, which is closer to unique blue (a zero crossing of R/G) than to unique green (a zero crossing of Y/B). For a linear Y/B to have a zero crossing near unique green, the B pole must have a contribution from M or L. The cone weights

into the cone-opponent mechanisms assumed at the second level are consistent with psychophysical measurements of color detection and discrimination estimates (e.g., Table 18.1 of Ref. 89).

The spectral sensitivities of the cone-opponent pairs are labelled Stage 2 in Fig. 4. (The L–M sensitivities have also been scaled by 2.55 in accordance with the proposals for the next stage.)

*Third zone* At the third stage, we sum the outputs of the second-stage mechanisms (or equivalently oppose ones of different signs), to give four color-opponent mechanisms (see also Refs. 29 and 18). The L–M cone-opponent input to the third stage is weighted by 2.55, so that R/G is zero for an equal-quantum white light, thus:

Red [L–M summed with S–(L+M)]

$$2.55[\bar{l}(\lambda) - \bar{m}(\lambda)] + (\bar{s}(\lambda) - 0.69[\bar{l}(\lambda) + 0.5\bar{m}(\lambda)]) = 1.86\bar{l}(\lambda) - 2.90\bar{m}(\lambda) + \bar{s}(\lambda)$$

(bipolar) or  $|1.86\bar{l}(\lambda) - 2.90\bar{m}(\lambda) + \bar{s}(\lambda)|$  (unipolar)

Green [M–L and (L+M)–S summed]

$$2.55[\bar{m}(\lambda) - \bar{l}(\lambda)] + (0.69[\bar{l}(\lambda) + 0.5\bar{m}(\lambda)] - \bar{s}(\lambda)) = -1.86\bar{l}(\lambda) + 2.90\bar{m}(\lambda) - \bar{s}(\lambda)$$

(bipolar) or  $|-1.86\bar{l}(\lambda) + 2.90\bar{m}(\lambda) - \bar{s}(\lambda)|$  (unipolar)

R/G is required to be zero for an equal-quantal white, because otherwise, according to opponent-colors theory, such whites would appear colored (see Ref. 283). Initially, we used the same weights for the cone-opponent inputs to B/Y and Y/B, thus:

Blue [M–L summed with S–(L+M)]

$$2.55[\bar{m}(\lambda) - \bar{l}(\lambda)] + (\bar{s}(\lambda) - 0.69[\bar{l}(\lambda) + 0.5\bar{m}(\lambda)]) = -3.24\bar{l}(\lambda) + 2.21\bar{m}(\lambda) + \bar{s}(\lambda)$$

(bipolar) or  $|-3.24\bar{l}(\lambda) + 2.21\bar{m}(\lambda) + \bar{s}(\lambda)|$  (unipolar)

Yellow [L–M summed with (L+M)–S]

$$2.55[\bar{l}(\lambda) - \bar{m}(\lambda)] + (0.69[\bar{l}(\lambda) + 0.5\bar{m}(\lambda)] - \bar{s}(\lambda)) = 3.24\bar{l}(\lambda) - 2.21\bar{m}(\lambda) - \bar{s}(\lambda)$$

(bipolar) or  $|3.24\bar{l}(\lambda) - 2.21\bar{m}(\lambda) - \bar{s}(\lambda)|$  (unipolar)

The spectral sensitivities of the color-opponent mechanisms are labelled Stage 3 in Fig. 4. The R/G mechanism yields reasonable estimates of unique blue (477 nm) and unique yellow (580 nm), and the B/Y mechanism yields a reasonable estimate of unique green (504 nm).

One potential problem, however, is that the opposing poles of B/Y and Y/B are unbalanced (as they also are in the De Valois and De Valois model), so that B/Y and Y/B will produce a nonzero response to an equal-quantum white. Given that  $Y > B$ , the white field would be expected to appear yellowish. This imbalance can be corrected by decreasing the relative contributions of Y perhaps after the half-wave rectification of B/Y into unipolar B and Y mechanisms. Alternatively, it can be corrected by decreasing the weights of the L- and M-cone inputs into B/Y and into Y/B. For this illustrative example, we choose the latter correction and accordingly scale the weights of the L- and M-inputs into Y/B by 0.34 to give a zero response to white, so that:

Blue [M–L summed with S–(L+M)] becomes

$$-1.10\bar{l}(\lambda) + 0.75\bar{m}(\lambda) + \bar{s}(\lambda) \text{ (bipolar) or } |-1.10\bar{l}(\lambda) + 0.75\bar{m}(\lambda) + \bar{s}(\lambda)| \text{ (unipolar)}$$

Yellow [L–M summed with (L+M)–S] becomes

$$1.10\bar{l}(\lambda) - 0.75\bar{m}(\lambda) - \bar{s}(\lambda) \text{ (bipolar) or } |1.10\bar{l}(\lambda) - 0.75\bar{m}(\lambda) - \bar{s}(\lambda)| \text{ (unipolar)}$$

The spectral sensitivities of the Y/B and B/Y color-opponent pair are shown in the lower right panel of Fig. 4 as dashed lines. The correction shifts unique green to 510 nm. Such post hoc adjustments are inevitably somewhat speculative, however. Moreover, given that B/Y is clearly nonlinear (see subsection “Linearity of Color-Opponent Mechanisms” in Sec. 11.5), any mechanism that adjusts the balance of B/Y is also likely to be nonlinear.

Nonetheless, this example demonstrates that plausible estimates of the zero crossings of third-stage color opponency can be generated by making a few very simple assumptions about the transformations at the cone-opponent and color-opponent stages. But, how close are the R/G and B/Y spectral sensitivities to color-appearance data at wavelengths removed from the zero crossings? Figure 32 compares the R/G and B/Y spectral sensitivities with color valence data. Figure 32*a* shows as red circles R/G valence data from Ingling et al.<sup>314</sup> scaled to best (least-squares) fit the R/G spectral sensitivity. As can be seen, the agreement between R/G and the valence data is remarkably good. Figure 32*b*, shows B/Y valence data for three subjects (AW, JF, and LK, dark blue triangles, purple inverted triangles, and light blue squares, respectively) from Werner and Wooten<sup>286</sup> each scaled to best (least-squares) fit B/Y. Consistent with their own failed attempts to find linear combinations of cone fundamentals to describe their B/Y valence data,<sup>286</sup> our B/Y spectral sensitivity agrees only approximately with the data for AW and LK, and poorly with the data for JF.

Also shown in Fig. 32 are the spectral sensitivities of R/G and B/Y color-opponent mechanisms proposed by De Valois and De Valois<sup>18</sup> scaled to best fit our functions. Their R/G and B/Y functions agree poorly with our functions and with the valence data. In contrast, our R/G spectral sensitivity agrees well with the proposal by Ingling et al.,<sup>314</sup> which was based on the valence data shown in Fig. 32*a* (see also Ref. 12).

In summary, simple assumptions about signal transformations at the second and third stages yield a reasonable estimate of the spectral sensitivity of the R/G color-opponent mechanism. Perhaps, then, the R/G mechanism does reflect a cortical summing of the chromatic responses of double-duty L–M center-surround neurons and S–(L+M) neurons as has been suggested before.<sup>18,29</sup> In contrast, the B/Y mechanism cannot be accounted for so simply, and may reflect much more complex nonlinear transformations (see subsection “Linearity of Color-Opponent Mechanisms” in Sec. 11.5). These differences between R/G and B/Y could be linked to the idea that the R/G mechanism represents the opportunistic use of an existing or perhaps slightly modified neural process following the relatively recent emergence of separate L- and M-cone photopigment genes,<sup>447</sup> whereas the B/Y mechanism is some more ancient neural process.<sup>448</sup>

Figure 39 shows a speculative “wiring” diagram of the three-stage Müller zone model.

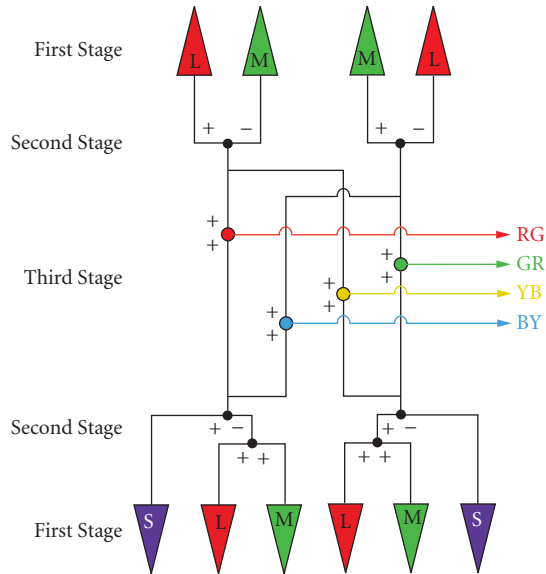
## Final Remarks

Although the three-stage zone model provides a way to integrate the separate two-stage models required for discrimination and for appearance, it is important to note several limitations.

First, as formulated, it does not account for the nonlinearities in unique hue and hue-scaling judgments described in “Linearity of Color-Opponent Mechanisms” in Sec. 11.5. It is possible that explicitly incorporating nonlinearities in the contrast-response functions of each stage could remedy this, but this remains an open question.

Second, the model as outlined just above is for a single state of adaptation. We know that adaptive processes act at both the first and second stages and, with the three-stage models, the effects of these would be expected to propagate to the third stage and thus affect both discrimination and appearance in a common fashion. Although some data, reviewed in “Color Appearance and Chromatic Detection and Discrimination” in Sec. 11.5, suggest that this is the case, further investigation on this point is required. In addition, test on pedestal data for crossed conditions suggest that somewhere in the processing chain there may be additional cross mechanism adaptive effects. Such effects have also been suggested to play a role in appearance data.<sup>381</sup>

Third, the model as formulated does not explicitly account for effects of the temporal and spatial structure of the stimuli. These components could certainly be added, but whether this can be done in a parsimonious fashion that does justice to the empirical phenomena is not clear. If our mechanistic understanding of color vision is to be applied to natural viewing, then its extension to handle the



**FIGURE 39** Three-stage Müller zone model. *First stage:* L-, M-, and S-cone photoreceptors (top and bottom). *Second stage:* L–M and M–L cone opponency (top) and S–(L+M) and (L+M)–S cone opponency (bottom). *Third stage:* Color opponency (center) is achieved by summing the various cone-opponent second-stage outputs.

complexity of natural retinal stimulation must be a high priority. At the same time, this is a daunting problem because of the explosion in stimulus parameters and the difficulties in controlling them adequately that occur when arbitrary spatiotemporal patterns are considered for both test and field.

Finally, we do not know whether models of this sort can provide a unified account of data across a wider range of tasks than simple threshold and appearance judgments.

Despite these unknowns and limitations, the type of three-stage model described here provides a framework for moving forward. It remains to be seen whether a model of this type will eventually provide a unified account of a wide range data, or whether what will be required, as was the case with Stiles'  $\pi$ -mechanism model which preceded it, is a reconceptualization of the nature of the psychophysical mechanisms and/or the linking hypotheses that connect them to behavioral data.

## 11.7 ACKNOWLEDGMENTS

The first author acknowledges a significant intellectual debt to Rhea Eskew not only for his writings on this subject, but also for the many discussions over many years about color detection and discrimination. He also acknowledges the support of Sabine Apitz, his wife, without whom this chapter could not have been written; and the financial support of the Wellcome Trust, the BBSRC, and Fight for Sight. The second author acknowledges the support of NIH R01 EY10016. The authors are grateful to Bruce Henning for helpful discussions, proofreading and help with Fig. 20, and Caterina Ripamonti for help with Fig. 16. They also thank Rhea Eskew, Bruce Henning, Ken Knoblauch, and Caterina Ripamonti for helpful comments on the manuscript.

## 11.8 REFERENCES

1. K. R. Gegenfurtner and D. C. Kiper, "Color Vision," *Annual Review of Neuroscience* **26**:181–206 (2003).
2. P. Lennie and J. A. Movshon, "Coding of Color and Form in the Geniculostriate Visual Pathway," *Journal of the Optical Society of America A* **10**:2013–2033 (2005).
3. S. G. Solomon and P. Lennie, "The Machinery of Colour Vision," *Nature Reviews Neuroscience* **8**:276–286 (2007).
4. T. Young, "On the Theory of Light and Colours," *Philosophical Transactions of the Royal Society of London* **92**:20–71 (1802).
5. H. von Helmholtz, *Handbuch der Physiologischen Optik*, Hamburg and Leipzig, Voss, 1867.
6. E. Hering, *Zur Lehre vom Lichtsinne. Sechs Mittheilungen an die Kaiserliche Akademie der Wissenschaften in Wien*, Carl Gerold's Sohn, Wien, 1878.
7. E. Hering, *Grundzüge der Lehre vom Lichtsinn*, Springer, Berlin, 1920.
8. L. M. Hurvich and D. Jameson, "An Opponent-Process Theory of Color Vision," *Psychological Review* **64**:384–404 (1957).
9. R. M. Boynton, "Theory of Color Vision," *Journal of the Optical Society of America* **50**:929–944 (1960).
10. P. L. Walraven, "On the Mechanisms of Colour Vision," Institute for Perception RVO-TNO, The Netherlands, 1962.
11. S. L. Guth and H. R. Lodge, "Heterochromatic Additivity, Foveal Spectral Sensitivity, and a New Color Model," *Journal of the Optical Society of America* **63**:450–462 (1973).
12. C. R. Ingling, Jr. and H. B.-P. Tsou, "Orthogonal Combination of the Three Visual Visual Channels," *Vision Research* **17**:1075–1082 (1977).
13. S. L. Guth, R. W. Massof, and T. Benzschawel, "Vector Model for Normal and Dichromatic color vision," *Journal of the Optical Society of America* **70**:197–212 (1980).
14. J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal Directions of Color Space," *Vision Research* **22**:1123–1131 (1982).
15. R. M. Boynton, *Human Color Vision*, Holt, Rinehart and Winston, New York, 1979.
16. S. A. Burns, A. E. Elsner, J. Pokorny, and V. C. Smith, "The Abney Effect: Chromaticity Coordinates of Unique and Other Constant Hues," *Vision Research* **24**:479–489 (1984).
17. M. A. Webster and J. D. Mollon, "Contrast Adaptation Dissociates Different Measures of Luminous Efficiency," *Journal of the Optical Society of America A* **10**:1332–1340 (1993).
18. R. L. De Valois and K. K. De Valois, "A Multi-Stage Color Model," *Vision Research* **33**:1053–1065 (1993).
19. I. Abramov and J. Gordon, "Color Appearance: On Seeing Red-or Yellow, or Green, or Blue," *Annual Review of Psychology* **45**:451–485 (1994).
20. R. T. Eskew, Jr. and P. M. Kortick, "Hue Equilibria Compared with Chromatic Detection in 3D Cone Contrast Space," *Investigative Ophthalmology and Visual Science (supplement)* **35**:1555 (1994).
21. R. L. De Valois, K. K. De Valois, E. Switkes, and L. Mahon, "Hue Scaling of Isoluminant and Cone-Specific Lights," *Vision Research* **37**:885–897 (1997).
22. S. M. Wuerger, P. Atkinson, and S. Cropper, "The Cone Inputs to the Unique-Hue Mechanisms," *Vision Research* **45**:25–26 (2005).
23. G. E. Müller, "Über die Farbenempfindungen," *Zeitschrift für Psychologie und Physiologie der Sinnesorgane, Ergänzungsband* **17**:1–430 (1930).
24. D. B. Judd, "Response Functions for Types of Vision According to the Müller Theory, Research Paper RP1946," *Journal of Research of the National Bureau of Standards* **42**:356–371 (1949).
25. D. B. Judd, "Fundamental Studies of Color Vision from 1860 to 1960," *Proceedings of the National Academy of Science of the United States of America* **55**:1313–1330 (1966).
26. G. Svaetichin and E. F. MacNichol, Jr., "Retinal Mechanisms for Chromatic and Achromatic Vision," *Annals of the New York Academy of Sciences* **74**:385–404 (1959).
27. R. L. De Valois, I. Abramov, and G. H. Jacobs, "Analysis of Response Patterns of LGN Cells," *Journal of the Optical Society of America* **56**:966–977 (1966).



28. A. Stockman and L. T. Sharpe, "Spectral Sensitivities of the Middle- and Long-Wavelength Sensitive Cones Derived from Measurements in Observers of Known Genotype," *Vision Research* **40**:1711–1737 (2000).
29. S. L. Guth, "A Model for Color and Light Adaptation," *Journal of the Optical Society of America A* **8**:976–993 (1991).
30. H. Hofer, J. Carroll, J. Neitz, M. Neitz, and D. R. Williams, "Organization of the Human Trichromatic Cone Mosaic," *Journal of Neuroscience* **25**:9669–9679 (2005).
31. P. Flanagan, P. Cavanagh, and O. E. Favreau, "Independent Orientation-Selective Mechanisms for the Cardinal Directions of Colour Space," *Vision Research* **30**:769–778 (1990).
32. M. A. Webster and J. D. Mollon, "The Influence of Contrast Adaptation on Color Appearance," *Vision Research* **34**:1993–2020 (1994).
33. W. S. Stiles, *Mechanisms of Colour Vision*, Academic Press, London, 1978.
34. R. T. Eskew, Jr, "Chromatic Detection and Discrimination," in *The Senses: A Comprehensive Reference, Volume 2: Vision II*, T. D. Albright and R. H. Masland, eds., Academic Press Inc., San Diego, 2008, pp. 101–117.
35. D. A. Baylor, B. J. Nunn, and J. L. Schnapf, "The Photocurrent, Noise and Spectral Sensitivity of Rods of the Monkey *Macaca Fascicularis*," *Journal of Physiology* **357**:575–607 (1984).
36. A. M. Derrington, J. Krauskopf, and P. Lennie, "Chromatic Mechanisms in Lateral Geniculate Nucleus of Macaque," *Journal of Physiology* **357**:241–265 (1984).
37. A. M. Derrington and P. Lennie, "Spatial and Temporal Contrast Sensitivities of Neurons in the Lateral Geniculate Nucleus of Macaque," *Journal of Physiology* **357**:219–240 (1984).
38. N. Graham and D. C. Hood, "Modeling the Dynamics of Adaptation: the Merging of Two Traditions," *Vision Research* **32**:1373–1393 (1992).
39. D. C. Hood, "Lower-Level Visual Processing and Models of Light Adaptation," *Annual Review of Psychology* **49**:503–535 (1998).
40. W. S. Stiles, "The Directional Sensitivity of the Retina and the Spectral Sensitivity of the Rods and Cones," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **B127**:64–105 (1939).
41. W. S. Stiles, "Incremental Thresholds and the Mechanisms of Colour Vision," *Documenta Ophthalmologica* **3**:138–163 (1949).
42. W. S. Stiles, "Further Studies of Visual Mechanisms by the Two-Colour Threshold Technique," *Coloquio sobre problemas opticos de la vision* **1**:65–103 (1953).
43. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, John Wiley & Sons, New York, 1966.
44. H. B. Barlow and W. R. Levick, "Threshold Setting by the Surround of Cat Retinal Ganglion Cells," *Journal of Physiology* **259**:737–757 (1976).
45. R. Shapley and C. Enroth-Cugell, "Visual Adaptation and Retinal Gain Controls," in *Progress in Retinal Research*, N. Osborne and G. Chader, eds., Pergamon Press, New York, 1984, pp. 263–346.
46. R. M. Boynton and D. N. Whitten, "Visual Adaptation in Monkey Cones: Recordings of Late Receptor Potentials," *Science* **170**:1423–1426 (1970).
47. J. M. Valetton and D. van Norren, "Light Adaptation of Primate Cones: An Analysis Based on Extracellular Data," *Vision Research* **23**:1539–1547 (1983).
48. D. A. Burkhardt, "Light Adaptation and Photopigment Bleaching in Cone Photoreceptors *in situ* in the Retina of the Turtle," *Journal of Neuroscience* **14**:1091–1105 (1994).
49. V. Y. Arshavsky, T. D. Lamb, and E. N. Pugh, Jr., "G Proteins and Phototransduction," *Annual Review of Physiology* **64**:153–187 (2002).
50. R. D. Hamer, S. C. Nicholas, D. Tranchina, T. D. Lamb, and J. L. P. Jarvinen, "Toward a Unified Model of Vertebrate Rod Phototransduction," *Visual Neuroscience* **22**:417–436 (2005).
51. E. N. Pugh, Jr., S. Nikonov, and T. D. Lamb, "Molecular Mechanisms of Vertebrate Photoreceptor Light Adaptation," *Current Opinion in Neurobiology* **9**:410–418 (1999).
52. A. Stockman, M. Langendörfer, H. E. Smithson, and L. T. Sharpe, "Human Cone Light Adaptation: From Behavioral Measurements to Molecular Mechanisms," *Journal of Vision* **6**:1194–1213 (2006).
53. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., Wiley, New York, 1982.



54. C. Sigel and E. N. Pugh, Jr., "Stiles's  $\pi_5$  Color Mechanism: Tests of Field Displacements and Field Additivity Properties," *Journal of the Optical Society of America* **70**:71–81 (1980).
55. B. A. Wandell and E. N. Pugh, Jr., "Detection of Long-Duration Incremental Flashes by a Chromatically Coded Pathway," *Vision Research* **20**:625–635 (1980).
56. E. H. Adelson, "Saturation and Adaptation in the Rod System," *Vision Research* **22**:1299–1312(1982).
57. W. S. Geisler, "Effects of Bleaching and Backgrounds on the Flash Response of the Cone System," *Journal of Physiology* **312**:413–434 (1981).
58. D. C. Hood and M. A. Finkelstein, "Sensitivity to Light," in *Handbook of Perception and Human Performance*, K. Boff, L. Kaufman and J. Thomas, eds., Wiley, New York, 1986, pp. 5-1–5-66.
59. M. M. Hayhoe, N. I. Benimof, and D. C. Hood, "The Time Course of Multiplicative and Subtractive Adaptation Processes," *Vision Research* **27**:1981–1996 (1987).
60. D. H. Kelly, "Effects of Sharp Edges in a Flickering Field," *Journal of the Optical Society of America* **49**:730–732 (1959).
61. C. F. Stromeyer, III, G. R. Cole, and R. E. Kronauer, "Second-Site Adaptation in the Red-Green Chromatic Pathways," *Vision Research* **25**:219–237 (1985).
62. A. Chaparro, C. F. Stromeyer, III, G. Chen, and R. E. Kronauer, "Human Cones Appear to Adapt at Low Light Levels: Measurements on the Red-Green Detection Mechanism," *Vision Research* **35**:3103–3118 (1995).
63. C. F. Stromeyer, III, G. R. Cole, and R. E. Kronauer, "Chromatic Suppression of Cone Inputs to the Luminance Flicker Mechanisms," *Vision Research* **27**:1113–1137 (1987).
64. A. Eisner and D. I. A. MacLeod, "Flicker Photometric Study of Chromatic Adaptation: Selective Suppression of Cone Inputs by Colored Backgrounds," *Journal of the Optical Society of America* **71**:705–718 (1981).
65. S. J. Ahn and D. I. A. MacLeod, "Link-Specific Adaptation in the Luminance and Chromatic Channels," *Vision Research* **33**:2271–2286 (1991).
66. Y. W. Lee, *Statistical Theory of Communication*, John Wiley & Sons, Inc., New York, 1968.
67. H. L. van Trees, *Detection, Estimation, and Modulation Theory*, Part I, Wiley, New York, 1968.
68. J. Nachmias, "Signal Detection Theory and its Applications to Problems in Vision," in *Visual Psychophysics, Handbook of Sensory Physiology*, Vol. VII/4, D. Jameson and L. H. Hurvich, eds., Springer-Verlag, Berlin, 1972, pp. 56–77.
69. C. H. Coombs, R. M. Dawes, and A. Tversky, *Mathematical Psychology*, Prentice-Hall, Englewood Cliff, New Jersey, 1970.
70. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification and Scene Analysis*, 2nd ed., John Wiley & Sons, New York, 2001.
71. H. B. Barlow, "Retinal and Central Factors in Human Vision Limited by Noise," in *Vertebrate Photoreception*, B. B. P. Fatt, ed., Academic Press, London, 1977, pp. 337–358.
72. A. B. Watson, H. B. Barlow, and J. G. Robson, "What Does the Eye See Best?," *Nature* **31**:419–422 (1983).
73. W. S. Geisler, "Physical Limits of Acuity and Hyperacuity," *Journal of the Optical Society of America A* **1**:775–782 (1984).
74. W. S. Geisler, "Sequential Ideal-Observer Analysis of Visual Discriminations," *Psychological Review* **96**:267–314 (1989).
75. D. H. Krantz, "Color Measurement and Color Theory: I. Representation Theorem for Grassmann Structures," *Journal of Mathematical Psychology* **12**:283–303 (1975).
76. D. H. Krantz, "Color Measurement and Color Theory: II Opponent-Colors Theory," *Journal of Mathematical Psychology* **12**:304–327 (1975).
77. D. L. MacAdam, "Chromatic Adaptation," *Journal of the Optical Society* **46**:500–513 (1956).
78. J. M. Hillis and D. H. Brainard, "Do Common Mechanisms of Adaptation Mediate Color Discrimination and Appearance? Uniform Backgrounds," *Journal of the Optical Society America A* **22**:2090–2106 (2005).
79. M. A. Webster and J. D. Mollon, "Changes in Colour Appearance Following Post-Receptor Adaptation," *Nature* **349**:235–238 (1991).
80. J. P. Moxley and S. L. Guth, "Hue Shifts Caused by Post-Receptor Adaptation," *Investigative Ophthalmology and Visual Science (supplement)* **20**:206 (1981).

81. S. L. Guth and J. P. Moxley, "Hue Shifts Following Differential Postreceptor Achromatic Adaptation," *Journal of the Optical Society of America* **72**:301–303 (1982).
82. J. M. Hillis and D. H. Brainard, "Do Common Mechanisms of Adaptation Mediate Color Discrimination and Appearance? Contrast Adaptation," *Journal of the Optical Society America A* **24**:2122–2133 (2007).
83. C. Noorlander, M. J. G. Heuts, and J. J. Koenderink, "Sensitivity to Spatiotemporal Combined Luminance and Chromaticity Contrast," *Journal of the Optical Society of America* **71**:453–459 (1981).
84. D. H. Kelly, "Flicker," in *Visual Psychophysics, Handbook of Sensory Physiology*, Vol. VII/4, D. Jameson and L. H. Hurvich, eds., Springer-Verlag, Berlin, 1972, pp. 273–302.
85. R. Luther, "Aus dem Gebiet der Farbreizmetrik," *Zeitschrift für technische Physik* **8**:540–558 (1927).
86. D. I. A. MacLeod and R. M. Boynton, "Chromaticity Diagram Showing Cone Excitation by Stimuli of Equal Luminance," *Journal of the Optical Society of America* **69**:1183–1186 (1979).
87. D. H. Brainard, "Cone Contrast and Opponent Modulation Color Spaces," in *Human Color Vision*, P. K. Kaiser, and R. M. Boynton, eds., Optical Society of America, Washington, D.C., 1996, pp. 563–579.
88. K. Knoblauch, "Dual Bases in Dichromatic Color Space," in *Colour Vision Deficiencies XII*, B. Drum, ed., Kluwer Academic Publishers, Dordrecht, 1995, pp. 165–176.
89. R. T. Eskew, Jr., J. S. McLellan, and F. Giulianini, "Chromatic Detection and Discrimination," in *Color Vision: From Genes to Perception*, K. Gegenfurtner and L. T. Sharpe, eds., Cambridge University Press, Cambridge, 1999, pp. 345–368.
90. A. Stockman and L. T. Sharpe, "Cone Spectral Sensitivities and Color Matching," in *Color Vision: From Genes to Perception*, K. Gegenfurtner and L. T. Sharpe, eds., Cambridge University Press, Cambridge, 1999, pp. 53–87.
91. B. Drum, "Short-Wavelength Cones Contribute to Achromatic Sensitivity," *Vision Research* **23**:1433–1439 (1983).
92. A. Stockman, D. I. A. MacLeod, and D. D. DePriest, "An Inverted S-cone Input to the Luminance Channel: Evidence for Two Processes in S-cone Flicker Detection," *Investigative Ophthalmology and Visual Science (supplement)* **28**:92 (1987).
93. J. Lee and C. F. Stromeyer, III, "Contribution of Human Short-Wave Cones to Luminance and Motion Detection," *Journal of Physiology* **413**:563–593 (1989).
94. A. Stockman, D. I. A. MacLeod, and D. D. DePriest, "The Temporal Properties of the Human Short-Wave Photoreceptors and Their Associated Pathways," *Vision Research* **31**:189–208 (1991).
95. W. S. Stiles, "Separation of the 'Blue' and 'Green' Mechanisms of Foveal Vision by Measurements of Increment Thresholds," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **B 133**:418–434 (1946).
96. W. S. Stiles, "The Physical Interpretation of the Spectral Sensitivity Curve of the Eye," in *Transactions of the Optical Convention of the Worshipful Company of Spectacle Makers*, Spectacle Maker's Company, London, 1948, pp. 97–107.
97. W. S. Stiles, "Color Vision: the Approach Through Increment Threshold Sensitivity," *Proceedings of the National Academy of Science of the United States of America* **45**:100–114 (1959).
98. W. S. Stiles, "Foveal Threshold Sensitivity on Fields of Different Colors," *Science* **145**:1016–1018 (1964).
99. J. M. Enoch, "The Two-Color Threshold Technique of Stiles and Derived Component Color Mechanisms," in *Handbook of Sensory Physiology*, Vol. VII/4, D. Jameson and L. H. Hurvich, eds., Springer-Verlag, Berlin, 1972 pp. 537–567.
100. R. M. Boynton, M. Ikeda, and W. S. Stiles, "Interactions Among Chromatic Mechanisms as Inferred from Positive and Negative Increment Thresholds," *Vision Research* **4**:87–117 (1964).
101. P. E. King-Smith, "Visual Detection Analysed in Terms of Luminance and Chromatic Signals," *Nature* **255**:69–70 (1975).
102. P. E. King-Smith and D. Carden, "Luminance and Opponent-Color Contributions to Visual Detection and Adaptation and to Temporal and Spatial Integration," *Journal of the Optical Society of America* **66**:709–717 (1976).
103. B. A. Wandell and E. N. Pugh, Jr., "A Field Additive Pathway Detects Brief-Duration, Long-Wavelength Incremental Flashes," *Vision Research* **20**:613–624 (1980).

104. W. S. Stiles and B. H. Crawford, "The Liminal Brightness Increment as a Function of Wavelength for Different Conditions of the Foveal and Parafoveal Retina," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **B113**:496–530 (1933).
105. H. G. Sperling and R. S. Harwerth, "Red-Green Cone Interactions in Increment-Threshold Spectral Sensitivity of Primates," *Science* **172**:180–184 (1971).
106. K. Kranda and P. E. King-Smith, "Detection of Colored Stimuli by Independent Linear Systems," *Vision Research* **19**:733–745 (1979).
107. J. E. Thornton and E. N. Pugh, Jr., "Red/Green Color Opponency at Detection Threshold," *Science* **219**:191–193 (1983).
108. M. Kalloniatis and H. G. Sperling, "The Spectral Sensitivity and Adaptation Characteristics of Cone Mechanisms Under White Light Adaptation," *Journal of the Optical Society of America A* **7**:1912–1928 (1990).
109. L. L. Sloan, "The Effect of Intensity of Light, State of Adaptation of the Eye, and Size of Photometric Field on the Visibility Curve," *Psychological Monographs* **38**:1–87 (1928).
110. C. R. Ingling, Jr., "A Tetrachromatic Hypothesis for Human Color Vision," *Vision Research* **9**:1131–1148 (1969).
111. D. J. Calkins, J. E. Thornton, and E. N. Pugh, Jr., "Monochromatism Determined at a Long-Wavelength/Middle-Wavelength Cone-Antagonistic Locus," *Vision Research* **13**:2349–2367 (1992).
112. W. d. W. Abney and E. R. Festing, "Colour Photometry," *Philosophical Transactions of the Royal Society of London* **177**:423–456 (1886).
113. W. d. W. Abney, *Researches in Colour Vision*, Longmans, Green, London, 1913.
114. H. E. Ives, "Studies in the Photometry of Lights of Different Colours. I. Spectral Luminosity Curves Obtained by the Equality of Brightness Photometer and Flicker Photometer under Similar Conditions," *Philosophical Magazine Series 6* **24**:149–188 (1912).
115. W. W. Coblenz and W. B. Emerson, "Relative Sensibility of the Average Eye to Light of Different Color and Some Practical Applications," *Bulletin of the Bureau of Standards* **14**:167–236 (1918).
116. E. P. Hyde, W. E. Forsythe, and F. E. Cady, "The Visibility of Radiation," *Astrophysical Journal*, **48**:65–83 (1918).
117. K. S. Gibson and E. P. T. Tyndall, "Visibility of Radiant Energy," *Scientific Papers of the Bureau of Standards* **19**:131–191 (1923).
118. A. Dresler, "The Non-Additivity of Heterochromatic Brightness," *Transactions of the Illuminating Engineering Society* **18**:141–165 (1953).
119. R. M. Boynton and P. Kaiser, "Vision: The Additivity Law Made to Work for Heterochromatic Photometry with Bipartite Fields," *Science* **161**:366–368 (1968).
120. S. L. Guth, N. V. Donley, and R. T. Marrocco, "On Luminance Additivity and Related Topics," *Vision Research* **9**:537–575 (1969).
121. Y. Le Grand, "Spectral Luminosity," in *Visual Psychophysics, Handbook of Sensory Physiology*, Vol. VII/4, D. Jameson and L. H. Hurvich, eds., Springer-Verlag, Berlin, 1972, pp. 413–433.
122. G. Wagner and R. M. Boynton, "Comparison of Four Methods of Heterochromatic Photometry," *Journal of the Optical Society of America* **62**:1508–1515 (1972).
123. J. Pokorny, V. C. Smith, and M. Lutze, "Heterochromatic Modulation Photometry," *Journal of the Optical Society of America A* **6**:1618–1623 (1989).
124. P. Lennie, J. Pokorny, and V. C. Smith, "Luminance," *Journal of the Optical Society of America A* **10**:1283–1293 (1993).
125. V. C. Smith, and J. Pokorny, "Spectral Sensitivity of the Foveal Cone Photopigments between 400 and 500 nm," *Vision Research* **15**:161–171 (1975).
126. D. B. Judd, "Report of U.S. Secretariat Committee on Colorimetry and Artificial Daylight," in *Proceedings of the Twelfth Session of the CIE, Stockholm*, Bureau Central de la CIE, Paris, 1951, pp. 1–60.
127. J. J. Vos, "Colorimetric and Photometric Properties of a 2-deg Fundamental Observer," *Color Research and Application* **3**:125–128 (1978).
128. L. T. Sharpe, A. Stockman, W. Jagla, and H. Jägle, "A Luminous Efficiency Function,  $V^*(\lambda)$ , for Daylight Adaptation," *Journal of Vision* **5**:948–968 (2005).

129. A. Stockman, H. Jägle, M. Pirzer, and L. T. Sharpe, "The Dependence of Luminous Efficiency on Chromatic Adaptation," *Journal of Vision* **8**(16):1, 1–26 (2008).
130. CIE, *Fundamental Chromaticity Diagram with Physiological Axes—Part 1. Technical Report 170-1*, Central Bureau of the Commission Internationale de l'Éclairage, Vienna, 2007.
131. H. De Vries, "Luminosity Curves of Trichromats," *Nature* **157**:736–737 (1946).
132. H. De Vries, "The Heredity of the Relative Numbers of Red and Green Receptors in the Human Eye," *Genetica* **24**:199–212 (1948).
133. R. A. Crone, "Spectral Sensitivity in Color-Defective Subjects and Heterozygous Carriers," *American Journal of Ophthalmology* **48**:231–238 (1959).
134. W. A. H. Rushton and H. D. Baker, "Red/Green Sensitivity in Normal Vision," *Vision Research* **4**:75–85 (1964).
135. A. Adam, "Foveal Red-Green Ratios of Normals, Colorblinds and Heterozygotes," *Proceedings Tel-Hashomer Hospital: Tel-Aviv* **8**:2–6 (1969).
136. J. J. Vos and P. L. Walraven, "On the Derivation of the Foveal Receptor Primaries," *Vision Research* **11**:799–818 (1971).
137. M. Lutze, N. J. Cox, V. C. Smith, and J. Pokorny, "Genetic Studies of Variation in Rayleigh and Photometric Matches in Normal Trichromats," *Vision Research* **30**:149–162(1990).
138. R. L. P. Vimal, V. C. Smith, J. Pokorny, and S. K. Shevell, "Foveal Cone Thresholds," *Vision Research* **29**:61–78 (1989).
139. J. Kremers, H. P. N. Scholl, H. Knau, T. T. J. M. Berendschot, and L. T. Sharpe, "L/M-Cone Ratios in Human Trichromats Assessed by Psychophysics, Electroretinography and Retinal Densitometry," *Journal of the Optical Society of America A* **17**:517–526 (2000).
140. K. R. Dobkins, A. Thiele, and T. D. Albright, "Comparisons of Red-Green Equiluminance Points in Humans and Macaques: Evidence for Different L:M Cone Ratios between Species," *Journal of the Optical Society of America A* **17**:545–556 (2000).
141. K. L. Gunther, and K. R. Dobkins, "Individual Differences in Chromatic (red/green) Contrast Sensitivity are Constrained by the Relative Numbers of L- versus M-cones in the Eye," *Vision Research* **42**:1367–1378 (2002).
142. A. Stockman, D. I. A. MacLeod, and J. A. Vivien, "Isolation of the Middle- and Long-Wavelength Sensitive Cones in Normal Trichromats," *Journal of the Optical Society of America A* **10**:2471–2490 (1993).
143. J. Carroll, C. McMahon, M. Neitz, and J. Neitz, "Flicker-Photometric Electroretinogram Estimates of L:M Cone Photoreceptor Ratio in Men with Photopigment Spectra Derived from Genetics," *Journal of the Optical Society of America A* **17**:499–509 (2000).
144. J. Carroll, J. Neitz, and M. Neitz, "Estimates of L:M Cone Ratio from ERG Flicker Photometry and Genetics," *Journal of Vision* **2**:531–542 (2002).
145. M. F. Wesner, J. Pokorny, S. K. Shevell, and V. C. Smith, "Foveal Cone Detection Statistics in Color-Normals and Dichromats," *Vision Research* **31**:1021–1037 (1991).
146. D. H. Brainard, A. Roorda, Y. Yamauchi, J. B. Calderone, A. Metha, M. Neitz, J. Neitz, D. R. Williams, and G. H. Jacobs, "Functional Consequences of the Relative Numbers of L and M Cones," *Journal of the Optical Society of America A* **17**:607–614 (2000).
147. J. Albrecht, H. Jägle, D. C. Hood, and L. T. Sharpe, "The Multifocal Electroretinogram (mfERG) and Cone Isolating Stimuli: Variation in L- and M-cone driven signals across the Retina," *Journal of Vision* **2**:543–558 (2002).
148. L. T. Sharpe, E. de Luca, T. Hansen, H. Jägle, and K. Gegenfurtner, "Advantages and Disadvantages of Human Dichromacy," *Journal of Vision* **6**:213–223 (2006).
149. C. R. Ingling, Jr., and E. Martinez-Uriegas, "The Relationship between Spectral Sensitivity and Spatial Sensitivity for the Primate r-g X-Channel," *Vision Research* **23**:1495–1500(1983).
150. C. R. Ingling, Jr. and E. Martinez, "The Spatio-Chromatic Signal of the r-g Channels," in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L. T. Sharpe, eds., Academic Press, London, 1983, pp. 433–444.
151. C. R. Ingling, Jr. and E. Martinez-Uriegas, "The Spatiotemporal Properties of the R-G X-cell Channel," *Vision Research* **25**:33–38 (1985).
152. C. R. Ingling, Jr. and H. B.-P. Tsou, "Spectral Sensitivity for Flicker and Acuity Criteria," *Journal of the Optical Society of America A* **5**:1374–1378 (1988).

153. P. Lennie and M. D’Zmura, “Mechanisms of Color Vision,” *CRC Critical Reviews in Neurobiology* **3**:333–400 (1988).
154. P. K. Kaiser, B. B. Lee, P. R. Martin, and A. Valberg, “The Physiological Basis of the Minimally Distinct Border Demonstrated in the Ganglion Cells of the Macaque Retina,” *Journal of Physiology* **422**:153–183 (1990).
155. P. Gouras and E. Zrenner, “Enhancement of Luminance Flicker by Color-Opponent Mechanisms,” *Science* **205**:587–589 (1979).
156. J. L. Brown, L. Phares, and D. E. Fletcher, “Spectral Energy Thresholds for the Resolution of Acuity Targets,” *Journal of the Optical Society of America* **50**:950–960 (1960).
157. J. Pokorny, C. H. Graham, and R. N. Lanson, “Effect of Wavelength on Foveal Grating Acuity,” *Journal of the Optical Society of America* **58**:1410–1414 (1968).
158. C. R. Ingling, S. S. Grigsby, and R. C. Long, “Comparison of Spectral Sensitivity Using Heterochromatic Flicker Photometry and an Acuity Criterion,” *Color Research & Application* **17**:187–196 (1992).
159. M. Ikeda, “Study of Interrelations between Mechanisms at Threshold,” *Journal of the Optical Society of America A* **53**:1305–1313 (1963).
160. S. L. Guth, “Luminance Addition: General Considerations and Some Results at Foveal Threshold,” *Journal of the Optical Society of America* **55**:718–722 (1965).
161. S. L. Guth, “Nonadditivity and Inhibition Among Chromatic Luminances at Threshold,” *Vision Research* **7**:319–328 (1967).
162. S. L. Guth, J. V. Alexander, J. I. Chumbly, C. B. Gillman, and M. M. Patterson, “Factors Affecting Luminance Additivity at Threshold Among Normal and Color-Blind Subjects and Elaborations of a Trichromatic-Opponent Color Theory,” *Vision Research* **8**:913–928 (1968).
163. M. Ikeda, T. Uetsuki, and W. S. Stiles, “Interrelations among Stiles  $\pi$  Mechanisms,” *Journal of the Optical Society of America* **60**:406–415 (1970).
164. J. E. Thornton and E. N. Pugh, Jr., “Relationship of Opponent-Colors Cancellation Measures to Cone Antagonistic Signals Deduced from Increment Threshold Data,” in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L. T. Sharpe, eds., Academic Press, London, 1983, pp. 361–373.
165. A. B. Poirson and B. A. Wandell, “The Ellipsoidal Representation of Spectral Sensitivity,” *Vision Research* **30**:647–652 (1990).
166. K. Knoblauch and L. T. Maloney, “Testing the Indeterminacy of Linear Color Mechanisms from Color Discrimination Data,” *Vision Research* **36**:295–306 (1996).
167. A. B. Poirson and B. A. Wandell, “Pattern-Color Separable Pathways Predict Sensitivity to Simple Colored Patterns” *Vision Research* **36**:515–526 (1996).
168. A. B. Poirson B. A. Wandell, D. C. Varner, and D. H. Brainard, “Surface Characterizations of Color Thresholds,” *Journal of the Optical Society of America A* **7**:783–789 (1990).
169. G. R. Cole, C. F. Stromeyer, III, and R. E. Kronauer, “Visual Interactions with Luminance and Chromatic Stimuli,” *Journal of the Optical Society of America A* **7**:128–140 (1990).
170. A. Chaparro, C. F. Stromeyer, III, E. P. Huang, R. E. Kronauer, and R. T. Eskew, Jr., “Colour is What the Eye Sees Best,” *Nature* **361**:348–350 (1993).
171. R. T. Eskew, Jr., C. F. Stromeyer, III, and R. E. Kronauer, “Temporal Properties of the Red-Green Chromatic Mechanism,” *Vision Research* **34**:3127–3137 (1994).
172. M. J. Sankeralli and K. T. Mullen, “Estimation of the L-, M-, and S-cone Weights of the Postreceptoral Detection Mechanisms,” *Journal of the Optical Society of America A* **13**:906–915 (1996).
173. A. Chaparro, C. F. Stromeyer, III, R. E. Kronauer, and R. T. Eskew, Jr., “Separable Red-Green and Luminance Detectors for Small Flashes,” *Vision Research* **34**:751–762 (1994).
174. G. R. Cole, T. Hine, and W. McIlhagga, “Detection Mechanisms in L-, M-, and S-cone Contrast Space,” *Journal of the Optical Society of America A* **10**:38–51 (1993).
175. F. Giulianini and R. T. Eskew, Jr., “Chromatic Masking in the ( $\Delta$ L/L,  $\Delta$ M/M) Plane of Cone-Contrast Space Reveals only Two Detection Mechanisms,” *Vision Research* **38**:3913–3926(1998).
176. J. R. Newton and R. T. Eskew, Jr., “Chromatic Detection and Discrimination in the Periphery: A Postreceptoral Loss of Color Sensitivity,” *Visual Neuroscience* **20**:511–521 (2003).
177. R. T. Eskew, Jr., C. F. Stromeyer, III, C. J. Picotte, and R. E. Kronauer, “Detection Uncertainty and the Facilitation of Chromatic Detection by Luminance Contours,” *Journal of the Optical Society of America A* **8**: 394–403 (1991).

178. C. F. Stromeyer, III, J. Lee, and R. T. Eskew, Jr., "Peripheral Chromatic Sensitivity for Flashes: a Post-Receptor Red-Green Asymmetry," *Vision Research* **32**:1865–1873 (1992).
179. R. T. Eskew, Jr., J. R. Newton, and F. Giulianini, "Chromatic Detection and Discrimination Analyzed by a Bayesian Classifier," *Vision Research* **41**:893–909 (2001).
180. R. M. Boynton, A. L. Nagy, and C. X. Olson, "A Flaw in Equations for Predicting Chromatic Differences," *Color Research and Application* **8**:69–74 (1983).
181. C. F. Stromeyer, III, A. Chaparro, C. Rodriguez, D. Chen, E. Hu, and R. E. Kronauer, "Short-Wave Cone Signal in the Red-Green Detection Mechanism," *Vision Research* **38**:813–826 (1998).
182. C. F. Stromeyer, III and J. Lee, "Adaptational Effects of Short Wave Cone Signals on Red-Green Chromatic Detection," *Vision Research* **28**:931–940 (1988).
183. J. D. Mollon and C. R. Cavonius, "The Chromatic Antagonisms of Opponent-Process Theory are not the Same as Those Revealed in Studies of Detection and Discrimination," in *Colour Deficiencies VIII, Documenta Ophthalmologica Proceedings Series, 46*, G. Verriest, ed., Nijhoff-Junk, Dordrecht, 1987, pp. 473–483.
184. H. De Lange, "Research into the Dynamic Nature of the Human Fovea-Cortex Systems with Intermittent and Modulated Light. II. Phase Shift in Brightness and Delay in Color Perception.," *Journal of the Optical Society of America* **48**:784–789 (1958).
185. D. Regan and C. W. Tyler, "Some Dynamic Features of Colour Vision," *Vision Research* **11**:1307–1324 (1971).
186. D. H. Kelly and D. van Norren, "Two-Band Model of Heterochromatic Flicker," *Journal of the Optical Society of America* **67**:1081–1091 (1977).
187. D. J. Tolhurst, "Colour-Coding Properties of Sustained and Transient Channels in Human Vision," *Nature* **266**:266–268 (1977).
188. C. E. Sternheim, C. F. Stromeyer, III, and M. C. K. Khoo, "Visibility of Chromatic Flicker upon Spectrally Mixed Adapting Fields," *Vision Research* **19**:175–183 (1979).
189. V. C. Smith, R. W. Bowen, and J. Pokorny, "Threshold Temporal Integration of Chromatic Stimuli," *Vision Research* **24**:653–660 (1984).
190. A. B. Metha and K. T. Mullen, "Temporal Mechanisms Underlying Flicker Detection and Identification for Red-Green and Achromatic Stimuli," *Journal of the Optical Society of America A* **13**:1969–1980 (1996).
191. A. Stockman, M. R. Williams, and H. E. Smithson, "Flicker-Clicker: Cross Modality Matching Experiments," *Journal of Vision* **4**:86a (2004).
192. G. J. C. van der Horst, C. M. M. de Weert, and M. A. Bouman, "Transfer of Spatial Chromaticity-Contrast at Threshold in the Human Eye," *Journal of the Optical Society of America* **57**:1260–1266 (1967).
193. G. J. C. van der Horst and M. A. Bouman, "Spatio-Temporal Chromaticity Discrimination," *Journal of the Optical Society of America* **59**:1482–1488 (1969).
194. R. Hilz and C. R. Cavonius, "Wavelength Discrimination Measured with Square-Wave Gratings," *Journal of the Optical Society of America* **60**:273–277 (1970).
195. E. M. Granger and J. C. Heurtley, "Visual Chromaticity-Modulation Transfer Function," *Journal of the Optical Society of America* **63**:1173–1174 (1973).
196. C. F. Stromeyer, III and C. E. Sternheim, "Visibility of Red and Green Spatial Patterns upon Spectrally Mixed Adapting Fields," *Vision Research* **21**:397–407 (1981).
197. K. T. Mullen and J. J. Kulikowski, "Wavelength Discrimination at Detection Threshold," *Journal of the Optical Society of America A* **7**:733–742 (1990).
198. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Efficiency in Detection of Isoluminant and Isochromatic Interference Fringes," *Journal of the Optical Society of America A* **10**:2118–2133 (1993).
199. N. Sekiguchi, D. R. Williams, and D. H. Brainard, "Aberration-Free Measurements of the Visibility of Isoluminant Gratings," *Journal of the Optical Society of America A* **10**:2105–2117 (1993).
200. D. R. Williams, N. Sekiguchi, and D. H. Brainard, "Color, Contrast Sensitivity, and the Cone Mosaic," *Proceedings of the National Academy of Science of the United States of America* **90**:9770–9777 (1993).
201. E. N. Pugh, Jr. and C. Sigel, "Evaluation of the Candidacy of the  $\pi$ -Mechanisms of Stiles for Color-Matching Fundamentals," *Vision Research* **18**:317–330 (1978).
202. O. Estévez, "On the Fundamental Database of Normal and Dichromatic Color Vision," PhD thesis, Amsterdam University, Amsterdam 1979.



203. H. J. A. Dartnall, J. K. Bowmaker, and J. D. Mollon, "Human Visual Pigments: Microspectrophotometric Results from the Eyes of Seven Persons," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **B 220**:115–130 (1983).
204. H. De Vries, "The Luminosity Curve of the Eye as Determined by Measurements with the Flicker Photometer," *Physica* **14**:319–348 (1948).
205. M. Ikeda and M. Urakubo, "Flicker HRTF as Test of Color Vision," *Journal of the Optical Society of America* **58**:27–31 (1968).
206. L. E. Marks and M. H. Bornstein, "Spectral Sensitivity by Constant CFF: Effect of Chromatic Adaptation," *Journal of the Optical Society of America* **63**:220–226 (1973).
207. P. E. King-Smith and J. R. Webb, "The Use of Photopic Saturation in Determining the Fundamental Spectral Sensitivity Curves," *Vision Research* **14**:421–429 (1974).
208. A. Eisner, "Comparison of Flicker-Photometric and Flicker-Threshold Spectral Sensitivities While the Eye is Adapted to Colored Backgrounds," *Journal of the Optical Society of America* **72**:517–518 (1982).
209. W. H. Swanson, "Chromatic Adaptation Alters Spectral Sensitivity at High Temporal Frequencies," *Journal of the Optical Society of America A* **10**:1294–1303 (1993).
210. C. F. Stromeyer, III, A. Chaparro, A. S. Tolia, and R. E. Kronauer, "Colour Adaptation Modifies the Long-Wave Versus Middle-Wave Cone Weights and Temporal Phases in Human Luminance (but not red-green) Mechanism," *Journal of Physiology* **499**:227–254 (1997).
211. W. B. Cushman and J. Z. Levinson, "Phase Shift in Red and Green Counter-Phase Flicker at High Frequencies," *Journal of the Optical Society of America* **73**:1557–1561 (1983).
212. P. L. Walraven and H. J. Leebeek, "Phase Shift of Sinusoidally Alternating Colored Stimuli," *Journal of the Optical Society of America* **54**:78–82 (1964).
213. D. T. Lindsey, J. Pokorny, and V. C. Smith, "Phase-Dependent Sensitivity to Heterochromatic Flicker," *Journal of the Optical Society of America A* **3**:921–927 (1986).
214. W. H. Swanson, J. Pokorny, and V. C. Smith, "Effects of Temporal Frequency on Phase-Dependent Sensitivity to Heterochromatic Flicker," *Journal of the Optical Society of America A* **4**:2266–2273 (1987).
215. V. C. Smith, B. B. Lee, J. Pokorny, P. R. Martin, and A. Valberg, "Responses of Macaque Ganglion Cells to the Relative Phase of Heterochromatically Modulated Lights," *Journal of Physiology* **458**:191–221 (1992).
216. A. Stockman, D. J. Plummer, and E. D. Montag, "Spectrally-Opponent Inputs to the Human Luminance Pathway: Slow +M and -L Cone Inputs Revealed by Intense Long-Wavelength Adaptation," *Journal of Physiology* **566**:61–76 (2005).
217. A. Stockman and D. J. Plummer, "Spectrally-Opponent Inputs to the Human Luminance Pathway: Slow +L and -M Cone Inputs Revealed by Low to Moderate Long-Wavelength Adaptation," *Journal of Physiology* **566**:77–91 (2005).
218. A. Stockman, E. D. Montag, and D. J. Plummer, "Paradoxical Shifts in Human Colour Sensitivity Caused by Constructive and Destructive Interference between Signals from the Same Cone Class," *Visual Neuroscience* **23**:471–478 (2006).
219. A. Stockman, E. D. Montag, and D. I. A. MacLeod, "Large Changes in Phase Delay on Intense Bleaching Backgrounds," *Investigative Ophthalmology and Visual Science (supplement)* **32**:841 (1991).
220. A. Stockman and D. J. Plummer, "The Luminance Channel Can be Opponent?," *Investigative Ophthalmology and Visual Science (supplement)* **35**:1572 (1994).
221. C. F. Stromeyer, III, P. D. Gowdy, A. Chaparro, S. Kladakis, J. D. Willen, and R. E. Kronauer, "Colour Adaptation Modifies the Temporal Properties of the Long- and Middle-Wave Cone Signals in the Human Luminance Mechanism," *Journal of Physiology* **526**:177–194 (2000).
222. A. Stockman, "Multiple Cone Inputs to Luminance," *Investigative Ophthalmology and Visual Science (supplement)* **42**:S320 (2001).
223. A. Stockman and D. J. Plummer, "Long-Wavelength Adaptation Reveals Slow, Spectrally Opponent Inputs to the Human Luminance Pathway," *Journal of Vision* **5**:702–716 (2005).
224. J. D. Mollon and P. G. Polden, "Saturation of a Retinal Cone Mechanism," *Nature* **259**:243–246 (1977).
225. C. F. Stromeyer, III, R. E. Kronauer, and J. C. Madsen, "Response Saturation of Short-Wavelength Cone Pathways Controlled by Color-Opponent Mechanisms," *Vision Research* **19**:1025–1040 (1979).
226. E. N. Pugh, Jr., "The Nature of the  $\pi_1$  Mechanism of W. S. Stiles," *Journal of Physiology* **257**:713–747 (1976).

227. P. G. Polden and J. D. Mollon, "Reversed Effect of Adapting Stimuli on Visual Sensitivity," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **B 210**:235–272 (1980).
228. E. N. Pugh, Jr. and J. Larimer, "Test of the Identity of the Site of Blue/Yellow Hue Cancellation and the Site of Chromatic Antagonism in the  $\pi_1$  Pathway," *Vision Research* **19**:779–788 (1980).
229. D. B. Kirk, "The Putative  $\pi_4$  Mechanism: Failure of Shape Invariance and Field Additivity," *Investigative Ophthalmology and Visual Science (supplement)* **26**:184 (1985).
230. A. Reeves, "Field Additivity of Stiles's Pi-4 Color Mechanism," *Journal of the Optical Society of America A* **4**:525–529 (1987).
231. E. N. Pugh, Jr. and J. D. Mollon, "A Theory of the  $\pi_1$  and  $\pi_3$  Color Mechanisms of Stiles," *Vision Research* **20**:293–312 (1979).
232. J. L. Schnapf, B. J. Nunn, M. Meister, and D. A. Baylor, "Visual Transduction in Cones of the Monkey *Macaca Fascicularis*," *Journal of Physiology* **427**:681–713 (1990).
233. D. C. Hood and D. G. Birch, "Human Cone Receptor Activity: The Leading Edge of the A-Wave and Models of Receptor Activity," *Visual Neuroscience* **10**:857–871 (1993).
234. B. B. Lee, D. M. Dacey, V. C. Smith, and J. Pokorny, "Horizontal Cells Reveal Cone Type-Specific Adaptation in Primate Retina," *Proceedings of the National Academy of Sciences of the United States of America* **96**:14611–14616 (1999).
235. B. B. Lee, D. M. Dacey, V. C. Smith, and J. Pokorny, "Dynamics of Sensitivity Regulation in Primate Outer Retina: The Horizontal Cell Network," *Journal of Vision* **3**:513–526 (2003).
236. G. J. Burton, "Evidence for Nonlinear Response Processes in the Human Visual System from Measurements on the Thresholds of Spatial Beat Frequencies," *Vision Research* **13**:1211–1225(1973).
237. D. I. A. MacLeod, and S. He "Visible Flicker from Invisible Patterns," *Nature* **361**:256–258 (1993).
238. D. I. A. MacLeod D. R. Williams, and W. Makous, "A Visual Nonlinearity Fed by Single Cones," *Vision Research* **32**:347–363 (1992).
239. O. Estévez and H. Spekreijse, "The 'Silent Substitution' Method in Visual Research," *Vision Research* **22**:681–691 (1982).
240. T. Benzschawel and S. L. Guth, "Post-Receptor Chromatic Mechanisms Revealed by Flickering vs Fused Adaptation," *Vision Research* **22**:69–75 (1982).
241. J. Krauskopf, D. R. Williams, M. B. Mandler, and A. M. Brown, "Higher Order Color Mechanisms," *Vision Research* **26**:23–32 (1986).
242. H. B. Barlow and P. Foldiak, "Adaptation and Decorrelation in the Cortex," in *The Computing Neuron*, R. Durbin, C. Miall, and G. J. Mitchison, eds., Addison-Wesley, Wokingham, 1989, pp. 54–72.
243. Q. Zaidi and A. G. Shapiro, "Adaptive Orthogonalization of Opponent-Color Signals," *Biological Cybernetics* **69**:415–428 (1993).
244. J. J. Atick, Z. Li, and A. N. Redlich, "What Does Post-Adaptation Color Appearance Reveal about Cortical Color Representation?," *Vision Research* **33**:123–129 (1993).
245. K. Gegenfurtner and D. C. Kiper, "Contrast Detection in Luminance and Chromatic Noise," *Journal of the Optical Society of America A* **9**:1880–1888 (1992).
246. M. J. Sankeralli and K. T. Mullen, "Postreceptor Chromatic Detection Mechanisms Revealed by Noise Masking in Three-Dimensional Cone Contrast Space," *Journal of the Optical Society of America A* **14**:2633–2646 (1997).
247. T. Hansen and K. Gegenfurtner, "Higher Level Chromatic Mechanisms for Image Segmentation," *Journal of Vision* **6**:239–259 (2006).
248. M. D'Zmura and K. Knoblauch, "Spectral Bandwidths for the Detection of Colour," *Vision Research* **38**:3117–3128 (1998).
249. G. Monaci, G. Menegaz, S. Süsstrunk, and K. Knoblauch, "Chromatic Contrast Detection in Spatial Chromatic Noise," *Visual Neuroscience* **21**:291–294 (2005).
250. F. Giulianini and R. T. Eskew, Jr., "Theory of Chromatic Noise Masking Applied to Testing Linearity of S-cone Detection Mechanisms," *Journal of the Optical Society of America A* **24**:2604–2021 (2007).
251. D. L. MacAdam, "Visual Sensitivities to Color Differences in Daylight," *Journal of the Optical Society of America* **32**:247–274 (1942).
252. Y. Le Grand, "Les Seuils Différentiels de Couleurs dans la Théorie de Young," *Revue d'Optique* **28**:261–278 (1949).



253. R. M. Boynton and N. Kambe, "Chromatic Difference Steps of Moderate Size Measured Along Theoretically Critical Axes," *Color Research and Application* **5**:13–23 (1980).
254. A. L. Nagy, R. T. Eskew, Jr., and R. M. Boynton, "Analysis of Color-Matching Ellipses in a Cone-Excitation Space," *Journal of the Optical Society of America A* **4**:756–768 (1987).
255. J. Krauskopf and K. Gegenfurtner, "Color Discrimination and Adaptation," *Vision Research* **11**:2165–2175 (1992).
256. J. Romero, J. A. Garcia, L. Jiménez del Barco, and E. Hita, "Evaluation of Color-Discrimination Ellipsoids in Two-color Spaces," *Journal of the Optical Society of America A* **10**:827–837 (1993).
257. R. W. Rodieck, *The Vertebrate Retina*, Freeman, San Francisco, 1973.
258. R. M. Boynton, R. T. Eskew, Jr., and A. L. Nagy, "Similarity of Normalized Discrimination Ellipses in the Constant-Luminance Chromaticity Plane," *Perception* **15**:755–763 (1986).
259. R. T. Eskew, Jr., Q. Wang, and D. P. Richters, "A Five-Mechanism Model of Hue Sensations," *Journal of Vision* **4**:315 (2004).
260. T. N. Cornsweet, and H. M. Pinsker, "Luminance Discrimination of Brief Flashes Under Various Conditions of Adaptation," *Journal of Physiology* **176**:294–310 (1965).
261. F. W. Campbell and J. J. Kulikowski, "Orientation Selectivity of the Human Visual System," *Journal of Physiology* **187**:437–445 (1966).
262. J. Nachmias and E. C. Kocher, "Discrimination of Luminance Increments," *Journal of the Optical Society of America* **60**:382–389 (1970).
263. J. Nachmias and R. V. Sansbury, "Grating Contrast: Discrimination May Be Better Than Detection," *Vision Research* **14**:1039–1042 (1974).
264. J. M. Foley and G. Legge, "Contrast Detection and Near-Threshold Discrimination in Human Vision," *Vision Research* **21**:1041–1053 (1981).
265. K. K. De Valois and E. Switkes, "Simultaneous Masking Interactions between Chromatic and Luminance Gratings," *Journal of the Optical Society of America* **73**:11–18 (1983).
266. C. F. Stromeyer, III and S. Klein, "Spatial Frequency Channels in Human Vision as Asymmetric (edge) Mechanisms," *Vision Research* **14**:1409–1420 (1974).
267. D. G. Pelli, "Uncertainty Explains Many Aspects of Visual Contrast Detection and Discrimination," *Journal of the Optical Society of America A* **2**:1508–1532 (1985).
268. E. Switkes, A. Bradley, and K. K. De Valois, "Contrast Dependence and Mechanisms of Masking Interactions among Chromatic and Luminance Gratings," *Journal of the Optical Society of America A* **5**:1149–1162 (1988).
269. K. T. Mullen and M. A. Losada, "Evidence for Separate Pathways for Color and Luminance Detection Mechanisms," *Journal of the Optical Society of America A* **11**:3136–3151 (1994).
270. R. W. Bowen and J. K. Cotten, "The Dipper and Bumper: Pattern Polarity Effects in Contrast Discrimination," *Investigative Ophthalmology and Visual Science (supplement)* **34**:708 (1993).
271. R. Hilz, G. Huppmann, and C. R. Cavonius, "Influence of Luminance Contrast on Hue Discrimination," *Journal of the Optical Society of America* **64**:763–766 (1974).
272. C.-C. Chen, J. M. Foley, and D. H. Brainard, "Detection of Chromoluminance Patterns on Chromoluminance Pedestals I: Threshold Measurements," *Vision Research* **40**:773–788 (2000).
273. C.-C. Chen, J. M. Foley, and D. H. Brainard, "Detection of Chromoluminance Patterns on Chromoluminance Pedestals II: Model," *Vision Research* **40**:789–803 (2000).
274. B. A. Wandell, "Measurement of Small Color Differences," *Psychological Review* **89**:281–302 (1982).
275. B. A. Wandell, "Color Measurement and Discrimination," *Journal of the Optical Society of America A* **2**:62–71 (1985).
276. A. C. Beare, "Color-Name as a Function of Wavelength," *The American Journal of Psychology* **76**:248–256 (1963).
277. G. S. Brindley, *Physiology of the Retina and the Visual Pathway*, Williams and Wilkins, Baltimore, 1970.
278. E. Schrödinger, "Über das Verhältnis der Vierfarben zur Dreifarben-theorie," *Sitzungsberichte. Abt. 2a, Mathematik, Astronomie, Physik, Meteorologie und Mechanik. Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse* **134**:471 (1925).
279. D. Jameson and L. M. Hurvich, "Some Quantitative Aspects of an Opponent-Colors Theory. I. Chromatic Responses and Spectral Saturation," *Journal of the Optical Society of America* **45**:546–552 (1955).

280. L. M. Hurvich and D. Jameson, "Some Quantitative Aspects of an Opponent-Colors Theory. II. Brightness, Saturation, and Hue in Normal and Dichromatic Vision," *Journal of the Optical Society of America* **45**:602–616 (1955).
281. D. Jameson and L. M. Hurvich, "Some Quantitative Aspects of an Opponent-Colors Theory. III. Changes in Brightness, Saturation, and Hue with Chromatic Adaptation," *Journal of the Optical Society of America* **46**:405–415 (1956).
282. L. M. Hurvich and D. Jameson, "Some Quantitative Aspects of an Opponent-Colors Theory. IV. A Psychological Color Specification System," *Journal of the Optical Society of America* **46**:1075–1089 (1956).
283. L. H. Hurvich, *Color Vision*, Sinauer, Sunderland, Massachusetts, 1981.
284. R. M. Boynton and J. Gordon, "Bezold-Brücke Hue Shift Measured by a Color-Naming Technique," *Journal of the Optical Society of America* **55**:78–86 (1965).
285. J. Gordon and I. Abramov, "Color Vision in the Peripheral Retina. II. Hue and Saturation," *Journal of the Optical Society of America* **67**:202–207 (1977).
286. J. S. Werner and B. R. Wooten, "Opponent Chromatic Mechanisms: Relation to Photopigments and Hue Naming," *Journal of the Optical Society of America* **69**:422–434 (1979).
287. C. R. Ingling, Jr., J. P. Barley, and N. Ghani, "Chromatic Content of Spectral Lights," *Vision Research* **36**:2537–2551 (1996).
288. A. Brückner, "Zur Frage der Eichung von Farbensystemen," *Zeitschrift für Sinnesphysiologie* **58**:322–362 (1927).
289. E. J. Chichilnisky and B. A. Wandell, "Trichromatic Opponent Color Classification," *Vision Research* **39**:3444–3458 (1999).
290. F. L. Dimmick and M. R. Hubbard, "The Spectral Location of Psychologically Unique Yellow, Green, and Blue," *American Journal of Psychology* **52**:242–254 (1939).
291. M. Ayama, T. Nakatsue, and P. E. Kaiser, "Constant Hue Loci of Unique and Binary Balanced Hues at 10, 100, and 1000 Td," *Journal of the Optical Society of America A* **4**:1136–1144 (1987).
292. B. E. Shefrin and J. S. Werner, "Loci of Spectral Unique Hues Throughout the Life-Span," *Journal of the Optical Society of America A* **7**:305–311 (1990).
293. G. Jordan and J. D. Mollon, "Rayleigh Matches and Unique Green," *Vision Research* **35**:613–620 (1995).
294. J. L. Nerger, V. J. Volbrecht, and C. J. Ayde, "Unique Hue Judgments as a Function of Test Size in the Fovea and at 20-deg Temporal Eccentricity," *Journal of the Optical Society of America A* **12**:1225–1232 (1995).
295. V. J. Volbrecht, J. L. Nerger, and C. E. Harlow, "The Bimodality of Unique Green Revisited," *Vision Research* **37**:404–416 (1997).
296. R. W. Pridmore, "Unique and Binary Hues as Functions of Luminance and Illuminant Color Temperature, and Relations with Invariant Hues," *Vision Research* **39**:3892–3908 (1999).
297. R. G. Kuehni, "Variability in Unique Hue Selection: A Surprising Phenomenon," *Color Research and Application* **29**:158–162 (2004).
298. M. A. Webster, E. Miyahara, G. Malkoc, and V. E. Raker, "Variations in Normal Color Vision. II. Unique Hues," *Journal of the Optical Society of America A* **17**:1545–1555 (2000).
299. R. G. Kuehni, "Determination of Unique Hues Using Munsell Color Chips," *Color Research and Application* **26**:61–66 (2001).
300. M. A. Webster, S. M. Webster, S. Bharadwaj, R. Verma, J. Jaikumar, G. Madan, and E. Vaithilingham, "Variations in Normal Color Vision. III. Unique Hues in Indian and United States Observers," *Journal of the Optical Society of America A* **19**:1951–1962 (2002).
301. W. d. W. Abney, "On the Change of Hue of Spectrum Colors by Dilution with White Light," *Proceedings of the Royal Society of London* **A83**:120–127 (1910).
302. K. Richter, "Antagonistische Signale beim Farbensehen und ihr Zusammenhang mit der empfindungsgemässen Farbordnung," PhD thesis, University of Basel, 1969.
303. J. D. Mollon and G. Jordan, "On the Nature of Unique Hues," in *John Dalton's Colour Vision Legacy*, I. M. D. C. C. Dickinson, ed., Taylor and Francis, London, 1997, pp. 381–392.
304. J. H. Parsons, *An Introduction to Colour Vision*, 2nd ed., Cambridge University Press, Cambridge, 1924.
305. A. Valberg, "A Method for the Precise Determination of Achromatic Colours Including White," *Vision Research* **11**:157–160 (1971).

306. D. Jameson and L. M. Hurvich, "Opponent-Response Functions Related to Measured Cone Photopigments," *Vision Research* **58**:429–430 (1968).
307. J. Larimer, D. H. Krantz, and C. M. Cicerone, "Opponent-Process Additivity—I: Red/Green Equilibria," *Vision Research* **14**:1127–1140 (1974).
308. C. R. Ingling, Jr., "The Spectral Sensitivity of the Opponent-Color Channels," *Vision Research* **17**:1083–1089 (1977).
309. J. Larimer, D. H. Krantz, and C. M. Cicerone, "Opponent-Process Additivity—II: Yellow/Blue Equilibria and Nonlinear Models," *Vision Research* **15**:723–731 (1975).
310. J. G. W. Raaijmakers and C. M. M. de Weert, "Linear and Nonlinear Opponent Color Coding," *Perception and Psychophysics* **18**:474–480 (1975).
311. Y. Ejima and Y. Takahashi, "Bezold-Brücke Hue Shift and Nonlinearity in Opponent-Color Process" *Vision Research* **24**:1897–1904 (1984).
312. S. Takahashi and Y. Ejima, "Spatial Properties of Red-Green and Yellow-Blue Perceptual Opponent-Color Response," *Vision Research* **24**:987–994 (1984).
313. M. Ayama, P. K. Kaiser, and T. Nakatsue, "Additivity of Red Chromatic Valence," *Vision Research* **25**:1885–1891 (1985).
314. C. R. Ingling, Jr., P. W. Russel, M. S. Rea, and B. H.-P. Tsou, "Red-Green Opponent Spectral Sensitivity: Disparity between Cancellation and Direct Matching Methods," *Science* **201**:1221–1223 (1978).
315. C. H. Elzinga and C. M. M. de Weert, "Nonlinear Codes for the Yellow/Blue Mechanism," *Vision Research* **24**:911–922 (1984).
316. M. Ayama and M. Ikeda, "Additivity of Yellow Chromatic Valence," *Vision Research* **26**:763–769 (1985).
317. K. Knoblauch and S. K. Shevell, "Relating Cone Signals to Color Appearance: Failure of Monotonicity in Yellow/Blue," *Visual Neuroscience* **18**:901–906 (2001).
318. M. Ikeda and M. Ayama, "Non-linear Nature of the Yellow Chromatic Valence," in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L. T. Sharpe, eds., Academic Press, London, 1983, pp. 345–352.
319. K. Knoblauch, L. Sirovich, and B. R. Wooten, "Linearity of Hue Cancellation in Sex-Linked Dichromacy," *Journal of the Optical Society America A* **2**:136–146 (1985).
320. W. von Bezold, "Über das Gesetz der Farbenmischung und die physiologischen Grundfarben," *Annalen der Physiologie und Chemie* **150**:221–247 (1873).
321. E. W. Brücke, "Über einige Empfindungen im Gebiete der Sehnerven," *Sitzungsberichte der Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse, Abteilung 3* **77**:39–71 (1878).
322. D. M. Purdy, "Spectral Hue as a Function of Intensity," *American Journal of Psychology* **63**:541–559 (1931).
323. J. J. Vos, "Are Unique and Invariant Hues Coupled?," *Vision Research* **26**:337–342 (1986).
324. P. L. Walraven, "On the Bezold-Brücke Phenomenon," *Journal of the Optical Society of America* **51**:1113–1116 (1961).
325. J. Walraven, "Discounting the Background—The Missing Link in the Explanation of Chromatic Induction," *Vision Research* **16**:289–295 (1976).
326. S. K. Shevell, "The Dual Role of Chromatic Backgrounds in Color Perception," *Vision Research* **18**:1649–1661 (1978).
327. S. K. Shevell, "Color Perception under Chromatic Adaptation: Equilibrium Yellow and Long-Wavelength Adaptation," *Vision Research* **22**:279–292 (1982).
328. P. Whittle, "The Brightness of Coloured Flashes on Backgrounds of Various Colours and Luminances," *Vision Research* **13**:621–638 (1973).
329. C. M. Cicerone, D. H. Krantz, and J. Larimer, "Opponent-Process Additivity—III: Effect of Moderate Chromatic Adaptation," *Vision Research* **15**:1125–1135 (1975).
330. J. von Kries, "Influence of Adaptation on the Effects Produced by Luminous Stimuli," in *Sources of Color Science (1970)*, D. L. MacAdam, ed., MIT Press, Cambridge, MA, 1905, pp. 120–1126.
331. R. W. Burnham, R. W. Evans, and S. M. Newhall, "Influence on Color Perception of Adaptation to Illumination," *Journal of the Optical Society of America* **42**:597–605 (1952).
332. H. V. Walters, "Some Experiments on the Trichromatic Theory of Vision," *Proceedings of the Royal Society of London* **131**:27–50 (1942).

333. D. L. MacAdam, "Influence of Chromatic Adaptation on Color Discrimination and Color Perception," *Die Farbe* **4**:133–143 (1955).
334. E. J. Chichilnisky and B. A. Wandell, "Photoreceptor Sensitivity Changes Explain Color Appearance Shifts Induced by Large Uniform Backgrounds in Dichoptic Matching," *Vision Research* **35**:239–254 (1995).
335. L. H. Hurvich and D. Jameson, "Further Developments of a Quantified Opponent Colors Theory," in *Visual Problems of Colour*, Volume 2, Her Majesty's Stationery Office, London, 1958, pp. 691–723.
336. D. Jameson and L. M. Hurvich, "Sensitivity, Contrast, and Afterimages," in *Visual Psychophysics*, Vol. VII/4, *Handbook of Sensory Physiology*, D. Jameson and L. H. Hurvich, eds., Springer-Verlag, Berlin, 1972, pp. 568–581.
337. S. K. Shevell, "Color Appearance," in *The Science of Color*, (2nd ed.), S. K. Shevell, ed., Elsevier, Oxford, 2003, pp. 149–190.
338. J. Walraven, "No Additive Effect of Backgrounds in Chromatic Induction," *Vision Research* **19**:1061–1063 (1979).
339. S. K. Shevell, "Unambiguous Evidence for the Additive Effect in Chromatic Adaptation," *Vision Research* **20**:637–639 (1980).
340. B. Drum, "Additive Effect of Backgrounds in Chromatic Induction," *Vision Research* **21**:959–961 (1981).
341. E. H. Adelson, "Looking at the World through a Rose-Colored Ganzfeld," *Vision Research* **21**:749–750 (1981).
342. J. Larimer, "Red/Green Opponent Colors Equilibria Measured on Chromatic Adapting Fields: Evidence for Gain Changes and Restoring Forces," *Vision Research* **21**:501–512 (1981).
343. J. Wei and S. K. Shevell, "Color Appearance under Chromatic Adaptation Varied Along Theoretically Significant Axes in Color Space," *Journal of the Optical Society of America A* **12**:36–46 (1995).
344. O. Rinner and K. Gegenfurtner, "Time Course of Chromatic Adaptation for Color Appearance and Discrimination," *Vision Research* **40**:1813–1826 (2000).
345. J. M. Hillis and D. H. Brainard, "Distinct Mechanisms Mediate Visual Detection and Identification," *Current Biology* **17**:1714–1719 (2007).
346. P. K. Kaiser, "Minimally Distinct Border as a Preferred Psychophysical Criterion in Heterochromatic Photometry," *Journal of the Optical Society of America* **61**:966–971 (1971).
347. A. Kohlrausch, "Theoretisches und Praktisches zur heterochromen Photometrie," *Pflügers Archiv für die gesamte Physiologie des Menschen und der Tiere* **200**:216–220 (1923).
348. H. Helson and V. B. Jeffers, "Fundamental Problems in Color Vision. II. Hue, Lightness, and Saturation of Selective Samples in Chromatic Illumination," *Journal of Experimental Psychology* **26**:1–27 (1940).
349. R. W. Burnham, R. M. Evans, and S. M. Newhall, "Prediction of Color Appearance with Different Adaptation Illuminations," *Journal of the Optical Society of America* **47**:35–42 (1957).
350. J. J. McCann, S. P. McKee, and T. H. Taylor, "Quantitative Studies in Retinex Theory: A Comparison between Theoretical Predictions and Observer Responses to the 'Color Mondrian' Experiments," *Vision Research* **16**:445–458 (1976).
351. L. E. Arend and A. Reeves, "Simultaneous Color Constancy," *Journal of the Optical Society of America A* **3**:1743–1751 (1986).
352. D. H. Brainard, W. A. Brunt, and J. M. Speigle, "Color Constancy in the Nearly Natural Image. 1. Asymmetric Matches," *Journal of the Optical Society of America A* **14**:2091–2110 (1997).
353. D. H. Brainard, "Color Constancy in the Nearly Natural Image. 2. Achromatic Loci," *Journal of the Optical Society of America A* **15**:307–325 (1998).
354. J. M. Kraft and D. H. Brainard, "Mechanisms of Color Constancy under Nearly Natural Viewing," *Proceedings of the National Academy of Sciences of the United States of America*, **96**:307–312 (1999).
355. D. H. Brainard, "Color Constancy," in *The Visual Neurosciences*, L. Chalupa and J. Werner, eds., MIT Press, Cambridge, MA, 2004, pp. 948–961.
356. H. E. Smithson, "Sensory, Computational, and Cognitive Components of Human Color Constancy," *Philosophical Transactions of the Royal Society of London B* **360**:1329–1346 (2005).
357. S. K. Shevell and F. A. A. Kingdom, "Color in Complex Scenes," *Annual Review of Psychology* **59**:143–166 (2008).

358. G. Buchsbaum, "A Spatial Processor Model for Object Colour Perception," *Journal of the Franklin Institute* **310**:1–26 (1980).
359. L. T. Maloney and B. A. Wandell, "Color Constancy: A Method for Recovering Surface Spectral Reflectances," *Journal of the Optical Society of America A* **3**:29–33 (1986).
360. D. H. Brainard and W. T. Freeman, "Bayesian Color Constancy," *Journal of the Optical Society of America A* **14**:1393–1411 (1997).
361. B. V. Funt, M. S. Drew, and J. Ho, "Color Constancy from Mutual Reflection," *International Journal of Computer Vision* **6**:5–24 (1991).
362. M. D'Zmura and G. Iverson, "Color Constancy. III. General Linear Recovery of Spectral Descriptions for Lights and Surfaces," *Journal of the Optical Society of America A* **11**:2389–2400 (1994).
363. G. D. Finlayson, P. H. Hubel, and S. Hordley, "Color by Correlation," in *Proceedings of the IS&T/SID Fifth Color Imaging Conference*, Scottsdale, AZ, 1997, pp. 6–11.
364. A. C. Hurlbert, "Computational Models of Color Constancy," in *Perceptual Constancy: Why Things Look As They Do*, V. Walsh and J. Kulikowski, eds., Cambridge University Press, Cambridge, 1998, pp. 283–322.
365. L. T. Maloney and J. N. Yang, "The Illuminant Estimation Hypothesis and Surface Color Perception," in *Colour Perception: From Light to Object*, R. Mausfeld and D. Heyer, eds., Oxford University Press, Oxford, 2001, pp. 335–358.
366. D. H. Brainard, J. M. Kraft, and P. Longère, "Color Constancy: Developing Empirical Tests of Computational Models," in *Colour Perception: Mind and the Physical World*, R. Mausfeld and D. Heyer, eds., Oxford University Press, Oxford, 2003, pp. 307–334.
367. D. H. Brainard, P. Longere, P. B. Delahunt, W. T. Freeman, J. M. Kraft, and B. Xiao, "Bayesian Model of Human Color Constancy," *Journal of Vision* **6**:1267–1281 (2006).
368. H. Boyaci, L. T. Maloney, and S. Hersh, "The Effect of Perceived Surface Orientation on Perceived Surface Albedo in Binocularly Viewed Scenes," *Journal of Vision* **3**:541–553 (2003).
369. H. Boyaci, K. Doerschner, and L. T. Maloney, "Perceived Surface Color in Binocularly Viewed Scenes with Two Light Sources Differing in Chromaticity," *Journal of Vision* **4**:664–679 (2004).
370. M. Bloj, C. Ripamonti, K. Mitha, S. Greenwald, R. Hauck, and D. H. Brainard, "An Equivalent Illuminant Model for the Effect of Surface Slant on Perceived Lightness," *Journal of Vision* **4**:735–746 (2004).
371. E. H. Land and J. J. McCann, "Lightness and Retinex Theory," *Journal of the Optical Society of America* **61**:1–11 (1971).
372. E. H. Land, "The Retinex Theory of Color Vision," *Scientific American* **237**:108–128 (1977).
373. E. H. Land, "Recent Advances in Retinex Theory," *Vision Research* **26**:7–21 (1986).
374. A. Hurlbert, "Formal Connections between Lightness Algorithms," *Journal of the Optical Society of America A* **3**:1684–1694 (1986).
375. G. West and M. H. Brill, "Necessary and Sufficient Conditions for von Kries Chromatic Adaptation to Give Color Constancy," *Journal of Mathematical Biology* **15**:249–250 (1982).
376. J. A. Worthey and M. H. Brill, "Heuristic Analysis of von Kries Color Constancy," *Journal of the Optical Society of America A* **3**:1708–1712 (1986).
377. D. H. Brainard and B. A. Wandell, "Analysis of the Retinex Theory of Color Vision," *Journal of the Optical Society of America A* **3**:1651–1661 (1986).
378. D. H. Foster and S. M. C. Nascimento, "Relational Colour Constancy from Invariant Cone-Excitation Ratios," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **257**:115–121 (1994).
379. M. A. Webster and J. D. Mollon, "Colour Constancy Influenced by Contrast Adaptation," *Nature* **373**:694–698 (1995).
380. Q. Zaidi, B. Spehar, and J. DeBonet, "Color Constancy in Variegated Scenes: Role of Low-level Mechanisms in Discounting Illumination Changes," *Journal of the Optical Society of America A* **14**:2608–2621 (1997).
381. M. D'Zmura and B. Singer, "Contast Gain Control," in *Color Vision: From Molecular Genetics to Perception*, K. Gegenfurtner and L. T. Sharpe, eds., Cambridge University Press, Cambridge, 1999, pp. 369–385.
382. W. S. Stiles, "Mechanism Concepts in Colour Theory," *Journal of the Colour Group* **11**:106–123 (1967).
383. D. Krantz, "A Theory of Context Effects Based on Cross-Context Matching," *Journal of Mathematical Psychology* **5**:1–48 (1968).

384. D. H. Brainard and B. A. Wandell, "Asymmetric Color-matching: How Color Appearance Depends on the Illuminant," *Journal of the Optical Society of America A* **9**:1433–1448 (1992).
385. Q. Zaidi, "Identification of Illuminant and Object Colors: Heuristic-Based Algorithms," *Journal of the Optical Society of America A* **15**:1767–1776 (1998).
386. J. Golz and D. I. A. MacLeod, "Influence of Scene Statistics on Colour Constancy," *Nature* **415**:637–640 (2002).
387. R. M. Boynton, M. M. Hayhoe, and D. I. A. MacLeod, "The Gap Effect: Chromatic and Achromatic Visual Discrimination as Affected by Field Separation," *Optica Acta* **24**:159–177 (1977).
388. R. T. Eskew, Jr. and R. M. Boynton, "Effects of Field Area and Configuration on Chromatic and Border Discriminations," *Vision Research* **27**:1835–1844 (1987).
389. B. W. Tansley and R. M. Boynton, "A Line, Not a Space, Represents Visual Distinctness of Borders Formed by Different Colors," *Science* **191**:954–957 (1976).
390. B. Pinna, "Un Effetto di Colorazione," in *Il laboratorio e la città. XXI Congresso degli Psicologi Italiani*, V. Majer, M. Maeran, and M. Santinello, eds., Società Italiana di Psicologia, Milano, 1987, p. 158.
391. B. Pinna, G. Brelstaff, and L. Spillmann, "Surface Color from Boundaries: A New 'Watercolor' Illusion," *Vision Research* **41**:2669–2676 (2001).
392. B. Pinna, J. S. Werner, and L. Spillmann, "The Watercolor Effect: A New Principle of Grouping and Figure-ground Organization," *Vision Research* **43**:43–52 (2003).
393. F. Devinck, P. B. Delahunt, J. L. Hardy, L. Spillmann, and J. S. Werner, "The Watercolor Effect: Quantitative Evidence for Luminance-Dependent Mechanisms of Long-Range Color Assimilation," *Vision Research* **45**:1413–1424 (2005).
394. E. D. Montag, "Influence of Boundary Information on the Perception of Color," *Journal of the Optical Society of America A* **14**:997–1006 (1997).
395. R. T. Eskew, Jr. "The Gap Effect Revisited: Slow Changes in Chromatic Sensitivity as Affected by Luminance and Chromatic Borders," *Vision Research* **29**:717–729 (1989).
396. P. D. Gowdy, C. F. Stromeyer, III, and R. E. Kronauer, "Facilitation between the Luminance and Red-Green Detection Mechanisms: Enhancing Contrast Differences Across Edges," *Vision Research* **39**:4098–4112 (1999).
397. L. A. Riggs, F. Ratliff, J. C. Cornsweet, and T. C. Cornsweet, "The Disappearance of Steadily Fixated Visual Test Objects," *Journal of the Optical Society of America* **43**:495–501 (1953).
398. J. Krauskopf, "Effect of Retinal Image Stabilization on the Appearance of Heterochromatic Targets," *Journal of the Optical Society of America* **53**:741–744 (1963).
399. A. L. Yarbus, *Eye Movements and Vision*, Plenum Press, New York, 1967.
400. T. P. Piantanida and J. Larimer, "The Impact of Boundaries on Color: Stabilized Image Studies," *Journal of Imaging Technology* **15**:58–63 (1989).
401. H. D. Crane and T. Piantanida, "On Seeing Reddish Green and Yellowish Blue," *Science* **221**:1078–1080 (1983).
402. V. A. Billock, G. A. Gleason, and B. H. Tsou, "Perception of Forbidden Colors in Retinally Stabilized Equiluminance Images: An Indication of Softwired Cortical Color Opponency," *Journal of the Optical Society of America A* **10**:2398–2403 (2001).
403. J. L. Nerger, T. P. Piantanida, and J. Larimer, "Color Appearance of Filled-in Backgrounds Affects Hue Cancellation, but not Detection Thresholds," *Vision Research* **33**:165–172 (1993).
404. J. J. Wisowaty and R. M. Boynton, "Temporal Modulation Sensitivity of the Blue Mechanism: Measurements Made Without Chromatic Adaptation," *Vision Research* **20**:895–909 (1980).
405. T. P. Piantanida, "Temporal Modulation Sensitivity of the Blue Mechanism: Measurements Made with Extraretinal Chromatic Adaptation," *Vision Research* **25**:1439–1444 (1985).
406. N. W. Daw, "Why After-Images are Not Seen in Normal Circumstances," *Nature* **196**:1143–1145 (1962).
407. R. van Lier and M. Vergeer, "Filling in the Afterimage after the Image," *Perception (ECVP Abstract Supplement)* **36**:200–201 (2007).
408. R. van Lier, M. Vergeer, and S. Anstis, "'Mixing-In' Afterimage Colors," *Perception (ECVP Abstract Supplement)* **37**:84 (2008).
409. C. McCollough, "Color Adaptation of Edge Detectors in the Human Visual System," *Science* **149**:1115–1116 (1965).



410. P. D. Jones and D. H. Holding, "Extremely Long-Term Persistence of the McCollough Effect," *Journal of Experimental Psychology: Human Perception & Performance* **1**:323–327 (1975).
411. E. Vul, E. Krizay, and D. I. A. MacLeod, "The McCollough Effect Reflects Permanent and Transient Adaptation in Early Visual Cortex," *Journal of Vision* **8**:4, 1–12 (2008).
412. G. M. Murch, "Binocular Relationships in a Size and Color Orientation Specific Aftereffect," *Journal of Experimental Psychology* **93**:30–34 (1972).
413. R. L. Savoy, "'Extinction' of the McCollough Effect does not Transfer Interocularly," *Perception & Psychophysics* **36**:571–576 (1984).
414. E. Vul and D. I. A. MacLeod, "Contingent Aftereffects Distinguish Conscious and Preconscious Color Processing," *Nature Neuroscience* **9**:873–874 (2006).
415. P. Thompson and G. Latchford, "Colour-Contingent After-Effects Are Really Wavelength-Contingent," *Nature* **320**:525–526 (1986).
416. D. H. Hubel and T. N. Wiesel, "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex," *Journal of Physiology* **160**:106–154 (1962).
417. D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *Journal of Physiology* **195**:215–243 (1968).
418. C. F. Stromeyer, III, "Form-Color Aftereffects in Human Vision," in *Perception, Handbook of Sensory Physiology*, Vol. VIII, R. Held, H. W. Leibowitz, and H. L. Teuber, eds., Springer-Verlag, Berlin, 1978, pp. 97–142.
419. D. Skowbo, B. N. Timney, T. A. Gentry, and R. B. Morant, "McCollough Effects: Experimental Findings and Theoretical Accounts," *Psychological Bulletin* **82**:497–510 (1975).
420. H. B. Barlow, "A Theory about the Functional Role and Synaptic Mechanism of Visual After-effects," in *Visual Coding and Efficiency*, C. B. Blakemore, ed., Cambridge University Press, Cambridge, 1990, pp. 363–375.
421. P. Dodwell and G. K. Humphrey, "A Functional Theory of the McCollough Effect," *Psychological Review* **97**:78–89 (1990).
422. G. K. Humphrey and M. A. Goodale, "Probing Unconscious Visual Processing with the McCollough Effect," *Consciousness and Cognition* **7**:494–519 (1998).
423. C. McCollough, "Do McCollough Effects Provide Evidence for Global Pattern Processing?," *Perception & Psychophysics* **62**:350–362 (2000).
424. V. H. Perry, R. Oehler, and A. Cowey, "Retinal Ganglion Cells that Project to the Dorsal Lateral Geniculate Nucleus in the Macaque Monkey," *Neuroscience* **12**:1101–1123 (1984).
425. T. N. Wiesel and D. Hubel, "Spatial and Chromatic Interactions in the Lateral Geniculate Body of the Rhesus Monkey," *Journal of Neurophysiology* **29**:1115–1156 (1966).
426. R. L. De Valois and P. L. Pease, "Contours and Contrast: Responses of Monkey Lateral Geniculate Nucleus Cells to Luminance and Color Figures," *Science* **171**:694–696 (1971).
427. H. Kolb and L. Dekorver, "Midget Ganglion Cells of the Parafovea of the Human Retina: A Study by Electron Microscopy and Serial Reconstructions," *Journal of Comparative Neurology* **303**:617–636 (1991).
428. D. J. Calkins, S. J. Schein, Y. Tsukamoto, and P. Sterling, "M and L Cones in Macaque Fovea Connect to Midget Ganglion Cells by Different Numbers of Excitatory Synapses," *Nature* **371**:70–72 (1994).
429. B. B. Boycott, J. M. Hopkins, and H. G. Sperling, "Cone Connections of the Horizontal Cells of the Rhesus Monkey's retina," *Proceedings of the Royal Society of London. Series B: Biological Sciences* **B229**:345–379 (1987).
430. W. Paulus and A. Kröger-Paulus, "A New Concept of Retinal Colour Coding," *Vision Research* **23**:529–540 (1983).
431. P. Lennie, P. W. Haake, and D. R. Williams, "The Design of Chromatically Opponent Receptive Fields," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds., MIT Press, Cambridge, MA, 1991, pp. 71–82.
432. R. C. Reid and R. M. Shapley, "Spatial Structure of Cone Inputs to the Receptive Fields in Primate Lateral Geniculate Nucleus," *Nature* **356**:716–718 (1992).
433. B. B. Lee, J. Kremers and T. Yeh, "Receptive Fields of Primate Retinal Ganglion Cells Studied with a Novel Technique," *Visual Neuroscience* **15**:161–175 (1998).
434. D. M. Dacey, "Parallel Pathways for Spectral Coding in Primate Retina," *Annual Review of Neuroscience* **23**:743–775 (2000).

435. K. T. Mullen and F. A. A. Kingdom, "Differential Distributions of Red-Green and Blue-Yellow Cone Opponency across the Visual Field," *Visual Neuroscience* **19**:109–118 (2002).
436. L. Diller, O. S. Packer, J. Verweij, M. J. McMahon, D. R. Williams, and D. M. Dacey, "L and M Cone Contributions to the Midget and Parasol Ganglion Cell Receptive Fields of Macaque Monkey Retina," *Journal of Neuroscience* **24**:1079–1088 (2004).
437. S. G. Solomon, B. B. Lee, A. J. White, L. Rüttiger, and P. R. Martin, "Chromatic Organization of Ganglion Cell Receptive Fields in the Peripheral Retina," *Journal of Neuroscience* **25**:4527–4539 (2005).
438. C. Vakrou, D. Whitaker, P. V. McGraw, and D. McKeefry, "Functional Evidence for Cone-Specific Connectivity in the Human Retina," *Journal of Physiology* **566**:93–102 (2005).
439. P. Buzas, E. M. Blessing, B. A. Szmajda, and P. R. Martin, "Specificity of M and L Cone Inputs to Receptive Fields in the Parvocellular Pathway: Random Wiring with Functional Bias," *Journal of Neuroscience* **26**:11148–11161 (2006).
440. P. R. Jusuf, P. R. Martin, and U. Grünert, "Random Wiring in the Midget Pathway of Primate Retina," *Journal of Neuroscience* **26**:3908–3917 (2006).
441. R. W. Rodieck, "What Cells Code for Color?," in *From Pigments to Perception. Advances in Understanding Visual Processes*, A. Valberg, and B. B. Lee, eds., Plenum, New York, 1991.
442. T. Cornsweet, *Visual Perception*, Academic Press, New York, 1970.
443. P. Lennie, "Recent Developments in the Physiology of Color Vision," *Trends in Neurosciences* **7**:243–248 (1984).
444. E. Martinez-Uriegas, "A Solution to the Color-Luminance Ambiguity in the Spatiotemporal Signal of Primate X Cells," *Investigative Ophthalmology and Visual Science (supplement)* **26**:183 (1985).
445. V. A. Billock, "The Relationship between Simple and Double Opponent Cells," *Vision Research* **31**:33–42 (1991).
446. F. A. A. Kingdom and K. T. Mullen, "Separating Colour and Luminance Information in the Visual System," *Spatial Vision* **9**:191–219 (1995).
447. J. Nathans, D. Thomas, and S. G. Hogness, "Molecular Genetics of Human Color Vision: The Genes Encoding Blue, Green and Red Pigments," *Science* **232**:193–202 (1986).
448. J. D. Mollon, "'Tho' She Kneel'd in That Place Where They Grew...' The Uses and Origins of Primate Colour Vision," *Journal of Experimental Biology* **146**:21–38 (1989).
449. M. S. Livingstone and D. H. Hubel, "Segregation of Form, Color, Movement, and Depth: Anatomy, Physiology, and Perception," *Science* **240**:740–749 (1988).
450. D. T. Lindsey and A. M. Brown, "Masking of Grating Detection in the Isoluminant Plane of DKL Color Space," *Visual Neuroscience* **21**:269–273 (2004).
451. A. Li and P. Lennie, "Mechanisms Underlying Segmentation of Colored Textures," *Vision Research* **37**:83–97 (1997).
452. T. Hansen and K. Gegenfurtner, "Classification Images for Chromatic Signal Detection," *Journal of the Optical Society of America A* **22**:2081–2089 (2005).
453. H. E. Smithson, S. Khan, L. T. Sharpe, and A. Stockman, "Transitions between Colour Categories Mapped with Reverse Stroop Interference and Facilitation," *Visual Neuroscience* **23**:453–460 (2006).
454. Q. Zaidi and D. Halevy, "Visual Mechanisms That Signal the Direction of Color Changes," *Vision Research* **33**:1037–1051 (1986).
455. M. D'Zmura, "Color in Visual Search," *Vision Research* **31**:951–966 (1991).
456. J. Krauskopf, Q. Zaidi, and M. B. Mandler, "Mechanisms of Simultaneous Color Induction," *Journal of the Optical Society of America A* **3**:1752–1757 (1986).
457. J. Krauskopf, H. J. Wu, and B. Farell, "Coherence, Cardinal Directions and Higher-order Mechanisms," *Vision Research* **36**:1235–1245 (1996).
458. A. Valberg, "Unique Hues: An Old Problem for a New Generation," *Vision Research* **41**:1645–1657 (2001).
459. J. A. Movshon, I. D. Thompson, and D. J. Tolhurst, "Spatial Summation in the Receptive Fields of Simple Cells in the Cat's Striate Cortex," *Journal of Physiology* **283**:53–77 (1978).
460. H. Spitzer and S. Hochstein, "Complex-Cell Receptive Field Models," *Progress in Neurobiology* **31**:285–309 (1988).



461. J. Krauskopf and Q. Zaidi, "Induced Desensitization," *Vision Research* **26**:759–762 (1986).
462. M. J. Sankeralli and K. T. Mullen, "Bipolar or Rectified Chromatic Detection Mechanisms?," *Visual Neuroscience* **18**:127–135 (2001).
463. P. D. Gowdy, C. F. Stromeyer, III, and R. E. Kronauer, "Detection of Flickering Edges: Absence of a Red-Green Edge Detector," *Vision Research* **39**:4186–4191 (1999).
464. M. Sakurai and K. T. Mullen, "Cone Weights for the Two Cone-Opponent Systems in Peripheral Vision and Asymmetries of Cone Contrast Sensitivity," *Vision Research* **46**:4346–4354 (2006).
465. A. Vassilev, M. S. Mihaylova, K. Rachevaa, M. Zlatkovab, and R. S. Anderson, "Spatial Summation of S-cone ON and OFF Signals: Effects of Retinal Eccentricity," *Vision Research* **43**:2875–2884 (2003).
466. J. S. McClellan and R. T. Eskew, Jr., "ON and OFF S-cone Pathways Have Different Long-Wave Cone Inputs" *Vision Research* **40**:2449–2465 (2000).
467. A. G. Shapiro and Q. Zaidi, "The Effects of Prolonged Temporal Modulation on the Differential Response of Color Mechanisms," *Vision Research* **32**:2065–2075 (1992).
468. Q. Zaidi, A. G. Shapiro, and D. C. Hood, "The Effect of Adaptation on the Differential Sensitivity of the S-cone Color System," *Vision Research* **32**:1297–1318 (1992).
469. N. V. S. Graham, *Visual Pattern Analyzers*, Oxford University Press, New York, 1989.
470. A. B. Watson and J. G. Robson, "Discrimination at Threshold: Labelled Detectors in Human Vision," *Vision Research* **21**:1115–1122 (1981).
471. S. L. Guth, "Comments on 'A Multi-Stage Color Model'" *Vision Research* **36**:831–833 (1996).
472. R. L. De Valois and K. K. De Valois, "On 'A Three-Stage Color Model,'" *Vision Research* **36**:833–836 (1996).
473. M. E. Chevreul, *De la loi du Contraste Simultané des Couleurs*, Pitois-Levreault, Paris, 1839.
474. A. Kitaoka, "Illusion and Color Perception," *Journal of the Color Science Association of Japan* **29**:150–151 (2005).
475. K. Sakai, "Color Representation by Land's Retinex Theory and Belsey's Hypothesis," Gradual thesis, *Department of Psychology*, Ritsumeikan University, Japan, 2003.
476. W. von Bezold, *Die Farbenlehre in Hinblick auf Kunst und Kunstgewerbe*, Westermann, Braunschweig, 1874.
477. L. T. Sharpe, A. Stockman, H. Jägle, and J. Nathans, "Opsin Genes, Cone Photopigments, Color Vision and Colorblindness," in *Color Vision: From Genes to Perception*, K. Gegenfurtner and L. T. Sharpe, eds., Cambridge University Press, Cambridge, 1999, pp. 3–51.
478. P. Lennie, "Parallel Visual Pathways: A Review," *Vision Research* **20**:561–594 (1980).

---

# ASSESSMENT OF REFRACTION AND REFRACTIVE ERRORS AND THEIR INFLUENCE ON OPTICAL DESIGN

---

**B. Ralph Chou**

*School of Optometry  
University of Waterloo  
Waterloo, Ontario, Canada*

---

## 12.1 GLOSSARY

---

### Definitions

**Accommodation.** The increase of refractive power of the eye by changing the shape of the crystalline lens that enables focusing of the eye on a near object.

**Amblyopia.** Reduced visual acuity that is not improved with corrective lenses and occurs in the absence of anatomical or pathological anomalies in the eye.

**Ametropia.** A refractive condition of the unaccommodated eye in which light from optical infinity does not focus on the retina.

**Aniseikonia.** A relative difference in the perceived size and/or shape of the images in the two eyes.

**Anisometropia.** Unequal refractive state in the two eyes, which may result in aniseikonia.

**Aphakia.** Absence of the crystalline lens from the eye.

**Astigmatism.** Refractive state of the eye in which rays from a point object form two line images at different distances from the retina.

**Back vertex power.** Reciprocal of the distance in meters between the pole of the back surface of a lens and the axial position of the image formed of an infinitely distant object. The unit of BVP is the **diopter** (or reciprocal meter).

**Base curve of a contact lens.** The radius of curvature of the ocular surface of a contact lens.

**Base curve of a spectacle lens.** The power of the flattest meridian on the front surface of a spectacle lens. Alternatively, for a spectacle lens design, the reference surface power for a series of lenses of different back vertex powers.

**Emmetropia.** Refractive state of an unaccommodated eye in which light from an infinitely distant object is focused on the retina.

**Far point.** A point in the object space of an ametropic eye which is conjugate to the retina.

**Presbyopia.** Reduced ability to accommodate, usually due to age-related changes in the crystalline lens.

**Prismatic effect.** The deviation of a ray of light as it passes through an optical system, considered as if due to a prism of known deviating power that replaces the optical system. The unit of prismatic effect is the **prism diopter**, which is 1 cm of displacement per meter of travel of the ray along the optical axis.

**Rimless mounting.** A system of mounting spectacle lenses in which there is no frame; all parts for resting the spectacles on the nose and ears are attached directly to the lenses.

**Visual acuity.** Clinical measure of the minimum angle of resolution of the eye, with or without corrective lenses.

**Working distance.** Distance of the object of regard in front of the eye, usually considered 40 cm for reading and other near tasks.

## Symbols

$f$	spectacle lens focal length
$F$	spectacle lens power
$F_v$	back vertex power of a spectacle lens
$F_x$	effective back vertex power of a spectacle lens displaced a distance $x$
$K$	keratometry reading of corneal curvature
$L$	axial length of the eye
$P$	power of intraocular implant lens
$x$	displacement distance of a spectacle lens
$P_H$	prismatic effect in a lens of power $F$ at a point $x$ cm from the optical center

## Equations

**Equation (1)** is the lens effectivity equation which is used to adjust the power of a spectacle lens when it is moved a distance  $x$  meter from its original position.

$F_x$	effective back vertex power of a spectacle lens displaced a distance $x$
$F_v$	original back vertex power of the spectacle lens
$x$	displacement distance of the spectacle lens

**Equation (2)** estimates the spectacle correction needed after cataract extraction.

$F$	back vertex power of the postoperative spectacle lens
$F_{\text{old}}$	back vertex power of the preoperative spectacle lens

**Equation (3)** is the SRKII formula for power of an intraocular lens needed to produce emmetropia in an eye after cataract removal.

$P$	IOL power
$A1$	a constant
$K$	keratometer reading of the central corneal curvature
$L$	axial length of the eye in mm

**Equation (4)** is Prentice's Rule.

$P_H$	prismatic effect
$x$	distance in centimeters of the point on the lens through which the line of sight passes from the optical center of the lens
$F$	the back vertex power of the lens

**Equation (5)** is the exact expression for spectacle magnification of a spectacle lens.

$M$	magnification
$d_v$	distance between the back vertex of the lens and the cornea

- $F$  back vertex power of the spectacle lens  
 $t$  axial thickness of the spectacle lens in meters  
 $n$  index of refraction of the lens  
 $F_1$  front surface power of the lens

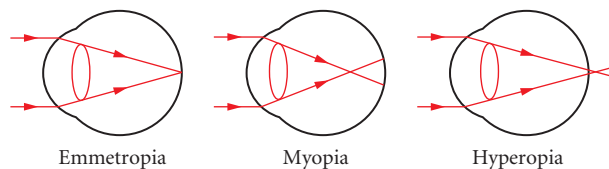
Equation (6) is the approximate formula for spectacle magnification.

## 12.2 INTRODUCTION

At the beginning of the twenty-first century, much of the information upon which we rely comes to us through vision. Camera view finders, computer displays, liquid crystal displays at vehicle controls, and optical instruments are only a few examples of devices in which refractive error and its correction can have important implications for both their users and optical designers. The options for correction of refractive error are many and varied; each has its advantages and disadvantages when the use of optical devices is considered. Most observers are binocular, and whether one or both eyes are fully or optimally corrected may affect one's ability to use a given optical device or instrument. In addition, the optical and physiological changes associated with aging of the eye may also have a profound effect on visual performance.

## 12.3 REFRACTIVE ERRORS

As described by Charman<sup>1</sup> in Chap. 1, the eye consists of two refractive elements, the cornea and the crystalline lens, which are separated by the watery aqueous humor, and the gel-like vitreous humor, which fills the rest of the eyeball between the crystalline lens and the retina (Fig. 1). The cornea provides approximately one-third of the optical power of the eye and the crystalline lens the remainder. Light entering the cornea from optical infinity ideally is focused through the dioptrics of the eye onto the retina, where it stimulates the photoreceptors and triggers the cascade of neurophysiological processes leading to the visual percept. An eye in which the retinal image is in sharp focus is described as emmetropic. The shape of the crystalline lens can be changed through the action of intraocular muscles in the process called *accommodation*. Accommodation increases the power of the eye so that light from a “near” object at a finite distance in front of the eye is sharply focused on the *retina*. The neural mechanisms governing accommodation affect the neural control of convergent and divergent eye movement. This ensures that as the eyes fixate on a near object, both eyes form images that overlap, giving rise to a single binocular percept. The crystalline lens continues to grow throughout life and physical changes within it gradually cause a loss of flexibility which reduces the amplitude of accommodation. The clinical onset of *presbyopia* (old eye) is marked by the inability to maintain focus on a target viewed at distance of 40 cm in front of the eye. This normally occurs at the age of 40 to 50 years.



**FIGURE 1** Focusing collimated light in emmetropic, myopic, and hyperopic eyes.

The failure to accurately focus light from a remote object onto the retina is referred to as *ametropia*.<sup>1</sup> Ametropia or refractive error is present in many eyes. Uncorrected ametropia has been identified as one of the leading treatable causes of blindness around the world.<sup>2</sup> The degree of ametropia is quantified by the optical power  $F$  of the spectacle lens that “corrects” the focus of the eye by bringing light from the distant object to focus at the retina. The dioptric power of the lens is given by the reciprocal of its focal length  $f$ :  $F = 1/f$ . The unit of focal power is the *diopter*, abbreviated as D, which has dimensions of reciprocal meters. Most spectacle lenses can be regarded as thin lenses,<sup>3</sup> and the power of superimposed thin lenses can be found as the algebraic sum of the powers of the individual lenses. This is convenient in ophthalmic applications such as measurement of refractive error and the use of “trial” corrections made with combinations of loose or trial case lenses.

## Types of Refractive Error

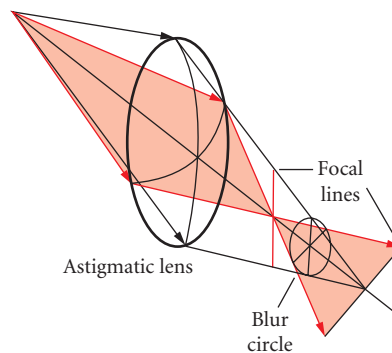
An unaccommodated eye which focuses light from a distant object onto the retina is *emmetropic* and requires no corrective lenses (Fig. 1).

A “near sighted” or *myopic* eye focuses light in front of the retina (Fig. 1). The myope sees closer objects in focus, but distant objects are blurred. Myopia is corrected by lenses of negative power that diverge light.

A “far sighted” or *hyperopic* eye focuses light behind the retina (Fig. 1). The hyperope can exercise accommodation to increase the power of the eye and focus the image of a distant object clearly on the retina. Nearer objects can also be seen clearly by exercising more accommodation, but the increased effort to maintain accommodation and a single binocular percept may result in symptoms of “eye-strain” and blurred near point vision. If accommodative fatigue occurs, the observer’s ability to maintain a clearly focused image of the distant object is lost, and blur at both distance and near results. Measurement of the refractive error of a hyperope often results in a finding of apparent emmetropia when accommodation is being exercised (latent hyperopia) and varying degrees of manifest hyperopia when control of accommodation is lost. Hyperopia is corrected by lenses of positive power that converge light.

The foregoing discussion assumes that the dioptics of the eye bring collimated light to a point focus. However, most eyes are *astigmatic*, that is, the optical components form two line foci that are usually perpendicular to one another. *Astigmatism* results in blur for all object distances, and the greater the astigmatism, the greater the blur. If the orientation of the line foci is in a direction away from the horizontal and vertical, the effects of ocular astigmatism on vision may be very serious, hindering development of the visual system.

Astigmatism is corrected with a *spherocylindrical lens* (Fig. 2), which usually has one spherical and one toric surface. Optically, the cylinder component brings the two line foci of the eye together,



**FIGURE 2** The astigmatic pencil formed by a spherocylindrical lens.

and the sphere component places the superimposed line foci on the retina. In clinical practice, the lens prescription is written in terms of *sphere*, *cylinder*, and *axis*. The position of the cylinder axis is specified using the trigonometric coordinate system centered on the eye as seen by the examiner, with 0° to the examiner's right or the patient's left, in what is referred to as the TABO or Standard Notation.<sup>3</sup> For example, the prescription  $+1.00 - 2.50 \times 030$  corresponds to a +1.00 D spherical lens (sometimes designated +1.00 DS) superimposed on a -2.50 D cylindrical lens with its axis at 30° (sometimes designated -2.00 DC  $\times$  030). Occasionally, for clinical reasons, astigmatism may be corrected by the *spherical equivalent lens*, a spherical lens that causes the line foci to straddle the retina. The power of the spherical equivalent is calculated as the sum of the sphere plus one-half of the cylinder in the spectacle prescription. The spherical equivalent of the above prescription would be -0.25 D.

Spectacle lenses are usually manufactured in quarter diopter steps. By ophthalmic convention, spectacle prescriptions are written with a plus or minus sign and to two decimal places. When the power is less than  $\pm 1.00$  D, a 0 is written before the decimal point, for example, +0.25 D.

Refractive errors are usually one of these simple types because the curvatures of the various refractive surfaces of the eye vary slowly over the area of the entrance pupil.<sup>4,5</sup> In some instances, however, *irregular astigmatism* may occur and no lens of any type can produce sharp vision. These are usually cases of eye injury or disease, such as distortion of the cornea in advanced keratoconus, corneal scarring after severe trauma, corneal dystrophy, disruption of the optics of the crystalline lens due to cataract, and traumatic damage to the lens. Tilting or displacement of the crystalline lens and distortion of the cornea following rigid contact lens wear may also lead to irregular refraction.<sup>6</sup> Williams<sup>7</sup> has described a case of irregular refraction due to tilting of the retina.

## 12.4 ASSESSMENT OF REFRACTIVE ERROR

In ophthalmic practice, the measurement of refractive error or ocular refraction is an iterative process in which a series of tests is employed to refine the results until a satisfactory end point is reached. The usual starting point is an objective test where no patient interaction is required, followed by subjective tests that require patient responses. Clinical tests are generally accurate to within  $\pm 0.25$  D, which is comparable to the depth of field of the human eye.<sup>8</sup>

### Objective Tests

Objective tests require only that the patients fix their gaze upon a designated target, normally located at a distance of 6 m from the eye so that with accurate fixation there is a negligible amount of accommodation being exercised. Three common tests are keratometry, direct ophthalmoscopy, and retinoscopy.

*Keratometry* measures the curvature of the front surface of the cornea. A bright object, the *keratometer mire*, is reflected by the cornea (1st Purkinje image) and its magnification determined from the lateral displacement of a doubled image of the mire.<sup>9</sup> The curvature of the corneal surface (*K* reading) in its principal meridians is read from a calibrated dial along with the axis of any corneal astigmatism. The total power of the cornea may be estimated from the *K* reading, and the corneal astigmatism may be used to predict the total amount and orientation of the astigmatism of the whole eye. Various rules of thumb have been developed (e.g., Javal's rule) to estimate the total astigmatism from the keratometric readings of normal eyes. This is due to the observation that large degrees of ocular astigmatism are almost always due to the cornea. In an *aphakic* eye, from which the crystalline lens has been removed, the corneal astigmatism is the ocular astigmatism.

Although primarily designed to view the retina of the living eye, the *direct ophthalmoscope* can also be used to estimate refractive error. Light is directed from a source in the ophthalmoscope through the pupil to illuminate the patient's retina. The examiner views the resulting image through lenses of varying power until a clear view of the retina is obtained. The algebraic sum of the lens power

and refractive error of the examiner is the patient's refractive error. When astigmatism is taken into account, this result is a crude estimate of the spherical equivalent of the patient's ocular refraction and is not sufficiently accurate to prescribe a corrective lens power. It may be helpful in assessing the refractive error of incommunicative patients when other methods cannot be employed.

The most common objective measurement of ocular refraction is retinoscopy. The retinoscope is designed to view the red reflex in the pupil as light is reflected from the retina. This is the same red reflex that often is seen in flash photographs of faces. The retinoscope is held at a known *working distance* from the eye, and the examiner observes the red reflex through the peephole in the retinoscope mirror as the retinoscope beam is moved across the patient's eye. Vignetting of the beam by the pupils of the patient and the examiner results in a movement of the reflex across the pupil. The direction and speed of the reflex motion depends on the patient's refractive error.<sup>10</sup>

Most refractionists use either a slightly divergent or collimated retinoscope beam at a working distance of 50 or 66 cm from the eye. If the red reflex moves in the same direction as the beam, this is a "with" movement and indicates that the retina is conjugate to a point closer than the working distance. As lenses with increasing plus power are interposed in 0.25 D increments, the with movement becomes faster until the end point is reached where a lens neutralizes the refractive error and the reflex moves infinitely fast or instantly fills the pupil when the beam touches the edge of the pupil. The next lens in the sequence should cause the reflex motion to reverse direction. The patient's refractive error is the value of the end point lens minus the working distance correction (+2.00 D for 50 cm working distance and +1.50 D for 66 cm). When the red reflex shows an "against" movement (it moves in the opposite direction to the retinoscope beam), the retina is conjugate to a point farther than the working distance.

Neutralization is achieved by interposing lenses of increasing minus power until the reflex shows a with movement. The end point lens is one incremental step back in power; with this lens, the reflex instantly fills the pupil when the beam touches the edge of the pupil. The patient's refractive error is calculated by correcting for the examiner's working distance.

The shape and motion of the reflex can be used to determine the axis of astigmatism. Once the two principal meridians of the eye are identified, the refractive error is measured in each separately. The sphere, cylinder, and axis components of the ocular refraction can then be calculated.

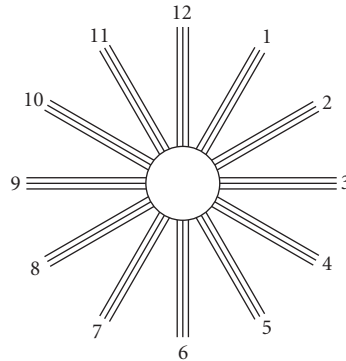
Retinoscopy provides an estimate of the refractive error very quickly, but requires some skill. A streak-shaped beam can facilitate the process in many patients, but retinoscopic findings may still vary from the true ocular refraction in some patients.

Automated equipment has been developed to determine the refractive error and corneal curvature without depending on the examiner's skill or judgment. Autokeratometers and autorefractors are frequently used to provide a starting point for the subjective refraction. More recently, aberrometers, instruments that measure ocular aberrations as well as the paraxial ocular refraction, have entered the ophthalmic market.

## Subjective Techniques

The final refinement of the refractive findings in an oculo-visual examination is usually obtained through subjective refraction. Although many techniques have been developed over the years, in practice only two or three are used. These methods rely on patient interaction to determine the sphere, cylinder, and axis components of the refractive error.

Subjective methods rely on the use of a visual acuity chart and trial lenses that can be placed in a trial frame worn by the patient, or a phoropter, a mechanical device that contains a range of powers of spherical and cylindrical lenses, prisms and accessory filters and lenses. The visual acuity chart may be either printed or projected and contains letters, numbers, or pictographs arranged in a sequence of sizes from large at the top to small at the bottom. The target size is rated according to the distance at which the finest detail within it would subtend 1 min of arc. Both recognition and resolution of detail within the target are required to read the smallest characters on the chart. The *visual acuity* is a representation of the minimum angle of resolution of the eye viewing the chart with or without corrective lenses. For example, an eye with a 1.5 min minimum angle of resolution should just read



**FIGURE 3** The astigmatic dial.

a target rated for 9 m or 30 ft when viewed at a distance of 6 m or 20 ft. The visual acuity would be recorded as 6/9 in metric, or 20/30 in imperial units. The refractive error can be estimated as the quotient of the denominator in imperial units divided by 100—in this example the refractive error would be approximately  $\pm 0.30$  D spherical equivalent. We usually expect ametropic patients with healthy eyes to have corrected visual acuity of 6/4.5 (20/15) or 6/6 (20/20). Rabbetts<sup>11</sup> has an excellent review of the design and use of visual acuity charts.

Subjective refraction usually begins with the patient's retinoscopic finding or previous spectacle correction in the phoropter or trial frame. Sometimes it may be preferable or necessary to add plus lenses in 0.25 D steps before the unaided eye to blur or fog the patient, then reduce lens power or add minus power incrementally until the best possible visual acuity is achieved with a spherical lens power (unfogging).

The astigmatism is measured next. If the patient has an equivalent sphere lens in front of the eye, vision is fogged while viewing the astigmatic dial, which is a chart with radial lines arranged in 30° intervals (Fig. 3). The patient identifies the line that appears clearest or darkest; the line corresponds to the line focus nearest to the retina and its orientation allows for calculation of the axis of the astigmatism. Since the chart looks like an analog clock dial, the hour value times 30 gives the approximate axis value, where 12 o'clock has the value 0 for the TABO axis notation. Cylinder power is added with this axis position until all lines appear equally dark or clear.

Refinement of the astigmatic correction is achieved with the *Jackson cross-cylinder* check test. This uses a spherocylindrical lens with equal but opposite powers in its principal meridians; normally a  $\pm 0.25$  D or  $\pm 0.50$  D lens is used. The Jackson cross-cylinder lens is designed to rotate around one of its principal meridians to check cylinder power and around an axis 45° from that meridian to check cylinder axis orientation. The patient looks through the spherocylinder correction and Jackson cross-cylinder lens and notes changes in clarity of small letters on the acuity chart as the cross cylinder is flipped and adjustments are made first to the axis position, then the cylinder power until both cross-cylinder orientations give the same degree of clarity. The sphere is then adjusted for best corrected visual acuity. Further adjustments may be made to equalize the visual acuity between the two eyes before the final corrective power is confirmed with trial lenses and frame.

## Presbyopia

Presbyopia or “old eye” is the refractive condition in which the ability to accommodate or focus the eyes on a near target is reduced. This is an age-related phenomenon due to the gradually increasing rigidity of the crystalline lens. It generally becomes clinically significant in the mid to late 40s; absolute presbyopia occurs by the late 50s when there is no accommodation. The onset of presbyopia is



earlier among hyperopes whose first symptom is the inability to focus on distant objects that were previously seen clearly without corrective lenses. Myopes can compensate for early presbyopia by removing their corrective lenses to view near objects. Presbyopia is corrected by adding positive power (the reading addition or add) to the distance correction until the patient is able to achieve the desired near visual acuity for reading. Single vision reading glasses or bifocal lenses may be prescribed for near work.

The added power may be determined either objectively using techniques such as dynamic retinoscopy or subjectively using a variety of techniques.<sup>12</sup>

## 12.5 CORRECTION OF REFRACTIVE ERROR

### The Art of Prescribing

Ametropia is usually corrected by placing a lens in front of the eye that will focus light from a distant object onto the axial position that is conjugate to the retina of the ametropic eye. This position is the *far point* or *punctum remotum* of the eye. Near objects can then be seen clearly through the lens by exercising accommodation.

The fact that humans usually have two functioning eyes and the frequent occurrence of astigmatism in ametropia require eye care practitioners to consider a given patient's binocular status and visual needs in addition to the refractive errors of the eyes. Prescribing corrective lenses is therefore as much an art as a science.

The spatial distortion inherent in astigmatic corrections can cause adaptation problems, whether or not the patient is well adapted to spectacle correction. As a result, most practitioners will try to minimize the amount of cylindrical correction that is needed for clear comfortable vision. In some cases, particularly in first corrections, the spherical equivalent may be preferred over even a partial correction of the astigmatism to facilitate adaptation to the lenses.

Younger ametropes may find vision slightly more crisp when they exercise a small degree of accommodation when looking at a distant target. Beginning refractionists often err in "over minus-ing" patients who accommodate during the refraction. The methods of subjective refraction are designed to minimize this tendency. While absolute presbyopes may be expected to require a full distance correction, and one expects myopic individuals will normally require full correction of their ametropia, younger hyperopes may often show a manifest error that is considerably lower than what can be found in cycloplegic refraction. They may require a partial correction until they learn to relax their accommodation when looking at distant objects.

Anatomical and neuromuscular anomalies of extraocular muscles that control eye movement may result in ocular deviations, both latent and manifest, that affect the ability of the visual system to provide a clear binocular visual percept for all target distances in all directions of gaze through corrective lenses. Differential *prismatic effects* may lead to double vision when the patient looks in certain directions or at certain distances in front of the eyes. The practitioner may need to modify the lens prescription in order to provide comfortable vision.

As presbyopia develops, a reading addition is required to maintain single clear binocular vision during near visual tasks like reading. The power of the add usually is chosen so that the patient uses about one-half of the available accommodation when reading. Addition power is determined using a standardized testing distance of 40 cm, then adjusted using trial lenses and the patient's preferred working distance for near vision tasks. The design of the lenses (bifocal, multifocal, invisible bifocal, single vision readers) to be prescribed largely depends on the wearer's visual tasks, intended use of the lenses, and need for eye protection.

It should be noted that presbyopes are not the only ones who can benefit from using bifocals and reading glasses. Since these corrections can reduce the demand for accommodation, they may be prescribed for younger patients who have anomalous accommodation and complain of eyestrain, visual discomfort, and even double vision. Finally, some individuals with low degrees of myopia often find that they are most comfortable when reading without any spectacle correction at all.

## Spectacle Correction

**Optical Considerations** Spectacles have been in use for at least 700 years, although it was only after the late nineteenth century that they were generally used to correct distance vision. Prior to that time, the lenses were self-selected mostly to aid in reading, and rarely for distance correction. The now familiar *bent* shape of modern spectacle lenses was not generally adopted until the early 1900s.

Spectacle lenses are often treated as thin lenses and it is often assumed that their exact position in front of the eyes is not critical. These assumptions apply to the majority of patients who have low-to-moderate refractive errors. It is only when we consider lenses for correction of high refractive error that thick lens optics must be used.

Spectacle lenses are specified by their *back vertex power*, which is the reciprocal of the distance from the pole of the ocular or back surface of the lens to its second focal point. The powers of phoropter and trial case lenses are also given as back vertex powers. The *focimeter* or *lensometer* is an instrument used to measure the back vertex power of spectacle lenses.<sup>13</sup>

The back vertex power of a corrective lens depends on the *vertex distance*, which is the distance between the back surface of the lens and the cornea. The vertex distance is typically between 12 and 15 mm. When the vertex distance of the phoropter or trial frame used in determining a patient's refractive error differs from the actual vertex distance of the corrective lenses placed in the spectacle frame, the back vertex power of the lenses can be modified using the formula

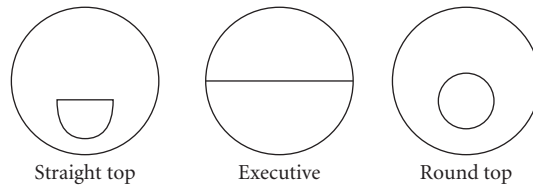
$$F_x = \frac{F_v}{1 - xF_v} \quad (1)$$

where  $x$  is the distance in meters between the vertex distance of the refraction and the vertex distance of the spectacles ( $x > 0$  when the spectacles are moved closer to the eye). It can be shown that changes in vertex distance of 1 or 2 mm are insignificant unless the lens power is greater than 8.00 D.<sup>14</sup>

Of all the parameters that are considered in designing spectacle lenses, the most significant is the *base curve*, which defines the lens form. The base curve is adjusted so that the lens forms images of distant objects near the far point of the wearer's eye in all directions of gaze. Ideally the image surface coincides with the far point sphere of the eye, so that the same spectacle prescription is required in all fields of gaze.<sup>15</sup> In practice, this means that the lens design will minimize unwanted oblique astigmatism and control curvature of field. Many corrected curve or best form lens designs have been developed, each with its own assumptions about vertex distance, field size, thickness, material, and relative importance of the Seidel aberrations, using spherical curves in the principal meridians of the lens. High plus power lenses must be made with aspheric curves to achieve adequate performance. Astigmatic lenses are made in minus cylinder form, with a spherical front surface and toric back.

The introduction of computer-aided design and computer numerically controlled (CNC) surfacing machines has led to the free form surfacing technology of today's ophthalmic lens industry. By specifying the deviation of both lens surfaces from reference spherical surfaces at many points, a diamond cutter can be used to form any desired surface shape. Double aspheric lenses which may provide optimal visual performance for wearers of any spectacle correction can be made readily with this technology.

**Spectacle Lens Materials** Spectacle lenses of all types are available in many glass and plastic materials. The final choice of material for a given pair of lenses depends on the diameter and thickness of the lenses (which determine their weight), the need for impact and scratch resistance, and the sensitivity of the wearer to color fringes caused by transverse chromatic aberration. The glass lenses of the twentieth century have been largely replaced by plastics, the most common being CR39 with a refractive index of 1.498. Polycarbonate (index 1.586) is frequently used because of its high impact resistance for protective lenses for occupational and sports use and for patients with high refractive errors. This material requires scratch-resistant coatings because of its soft surface and it suffers from chromatic aberration. Trivex (index 1.53) is a recently introduced plastic material that is more impact resistant than CR39 and can be used in *rimless mountings* and high power prescriptions where polycarbonate shows poorer performance. High index urethane lenses have largely replaced the high index glass lenses of the late 1900s with refractive indices of 1.6 to 1.74 available. Many



**FIGURE 4** Common bifocal styles.

lenses are supplied with antireflection coatings to enhance visual performance and cosmetic appearance; the coating can significantly reduce impact resistance.<sup>16–18</sup> Tints can be applied for cosmetic, protective, or vision enhancement purposes.<sup>19</sup>

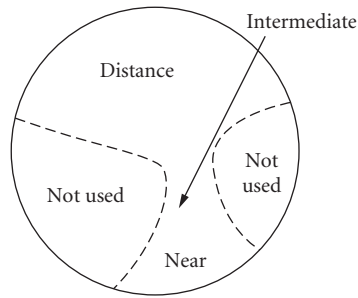
**Presbyopic Corrections** Most presbyopic patients prefer a bifocal correction in which the top part of the spectacle lens provides the distance correction and an area in the lower part, the *segment*, contains the correction for near vision. The difference between the segment and distance power is the add of the bifocal. Adds typically range from +1.00 D to +2.50 D, with the power increasing with age. A small minority of patients prefer separate pairs of glasses for distance use and reading.

The most common bifocal segments are the *straight top*, *Executive*, and *round top* designs shown in Fig. 4. Although many segment designs were developed in the early 1900s, only a few remain commercially available. All plastic bifocals are one piece, that is, the distance and near powers are achieved by changing the surface curvature on one side of the lens. Glass Executive bifocals are also one piece, but the straight top and round top styles are made by fusing a high index glass button into a countersink in the front surface of a lower index glass lens. They are referred to as fused bifocals. Bifocal segments are normally placed so that the top of the segment (the segment line) is at the edge of the lower eyelid. This allows comfortable near vision with minimal downgaze and keeps the segment as inconspicuous as possible. The style and width of the bifocal segment used is mainly determined by the wearer's visual needs at near, particularly with regard to width of the near visual field, as well as cosmetic appearance.

Patients with adds greater than +1.75 D often find that objects at a distance of 50 to 60 cm appear blurred through both the distance and segment portions of their bifocals. A trifocal lens that contains an intermediate segment between the distance and near parts of the lens allows for clear vision at arm's-length distance. Trifocals are positioned with the top line at or just below the lower edge of the pupil.

Special occupational multifocal lenses can be prescribed for presbyopes who need to see intermediate or near objects overhead. These lenses contain an intermediate or near add in the upper part of the lens in addition to the usual bifocal or trifocal segment. They are often referred to as occupational trifocal or quadrifocal lenses, respectively. Such lenses are recommended for presbyopic mechanics, electricians, plumbers, carpenters, painters, and librarians, among others. The *progressive addition lens* or *PAL* is an increasingly popular alternative to bifocal and trifocal lenses. The PAL design (Fig. 5) features a continuously varying power in the lens from the distance visual point, the pupil position when looking at distance, to the near visual point that is used to read. Originally a single aspheric progressive surface on the front of the lens was used, but more recent designs that take advantage of freeform surfacing technology have incorporated the power progression on either or both surfaces. Clear vision is achieved not only at the distance and near visual points, but also through a narrow optical corridor or umbilicus that joins them.

Outside of the distance and near vision zones and the corridor, the changes in surface shape result in distorted blurred images. The myriad of PAL designs in the ophthalmic lens market represents the many solutions to designing the progressive surface(s) to optimize single clear binocular vision through as much of the lens as possible. Some manufacturers claim to include preference for head or eye movements to view objects to the side, differences in prismatic effect between the eyes, and other factors in their designs.<sup>20</sup> Although marketed as lineless bifocals, PALs have proven popular because they hide the fact the wearer is old enough to need a reading prescription and provide clear vision at



**FIGURE 5** A progressive addition lens (PAL) has a large area of distance prescription power and an area of addition power connected by a narrow corridor. Areas to either side of the corridor and reading area are less usable due to blur induced by the progressive power surface.

all distances once the patient has learned to use them. The lens must be properly positioned in front of the eyes and the patient instructed on how to use it. Most patients adapt quickly to their PALs and prefer them to conventional bifocals despite the extra cost.

Recently PAL technology has been adapted for both patients approaching presbyopia as well as advanced presbyopes, who work with computer displays. These office lenses focus on arm's length and near working distances to provide a wider intermediate field of view than can be achieved in the corridor of a PAL. Prepresbyopes with complaints of eyestrain and presbyopes suffering "bifocal neck" while working at a computer display may find that these lens designs relieve their symptoms. Since the lenses have the intermediate portion in front of the pupil in the straight ahead gaze position, the head need not be tilted backward to enable the wearer to see the screen. Musicians, hairdressers, and other workers who require wide clear intermediate fields may also find these lenses useful.

## Contact Lenses

A contact lens can be considered as a spectacle lens with a zero vertex distance. The spectacle lens power must be adjusted for the change in vertex distance from the spectacle plane; in practice lens powers under 5 D are not adjusted for vertex distance since the change in power is less than 0.25 D. The contact lens power will be more positive or less negative than the spectacle lens power. Since the contact lens is supported by the cornea, the fit of the contact lens must be assessed to ensure that the back surface of the contact lens is supported without significantly affecting the shape of the cornea. The tear layer between the cornea and the contact lens may affect the fit as well as the power of the contact lens.

Most contact lenses are prescribed for cosmetic reasons and many patients ask for them in sports activity. Patients with very high refractive errors often find that contact lenses provide better quality of vision, particularly with regard to magnification effects and extent of the visual field, than spectacles. They are also much more comfortable than when wearing thick, heavy spectacle lenses.

The main contraindications for wearing contact lenses are dry eye and allergies.

**Rigid Contact Lenses** The first contact lenses were made from polymethylmethacrylate (PMMA). This plastic has very good optical and physical properties, but is impermeable to oxygen. Contact lenses had to be designed so that the overall diameter and the curvature of the back surface (the base curve of the contact lens) allowed a lens to move over the eye with the blink and permit tear fluid to be exchanged under the lens. This allowed enough oxygen to reach the cornea to maintain its physiology. Poorly fitting PMMA lenses could cause many problems related to disrupted corneal physiology due to hypoxia, including hypoesthesia, edema, surface irregularities, and overwear syndrome.

Deformation of the cornea could also occur, especially in corneal astigmatism, if the base curve did not provide a good mechanical fit to the toric corneal surface.

Modern rigid lenses made of gas permeable materials<sup>21</sup> have greatly reduced many of the problems caused by PMMA. These biocompatible materials transmit oxygen and can be made larger and thinner than PMMA lenses without adversely affecting corneal physiology. With less movement, the lenses are more comfortable, although initially there is a period of adaptation. Adaptation usually occurs over a few weeks as the patient starts wearing the lenses for a few hours each day, then gradually increases the wearing time. Eventually most patients wear their contact lenses between 10 and 14 hours a day; some patients may attempt extended wear, when the lenses are worn continuously for several days. Once adapted, the patient must continue to wear the lenses on a schedule that maintains the eyes' adaptation.

Rigid lenses do not conform to the shape of the underlying cornea. Thus, the tears filling the space between the ocular surface of the contact lens and the cornea form a *tear lens*. The sum of the power of the tear lens plus the power of the contact lens itself must equal the vertex adjusted power of the spectacle correction. Very thin gas permeable lenses tend to flex and change shape on the eye because of lid pressure and the tear lens. Most contact lens fitters will use trial lenses to estimate how the lens behaves on the eye and thus choose a base curve and lens power that will provide both a good physical fit and optimal vision correction.

Because the tears and cornea have almost the same index of refraction, the optical power of the cornea is replaced by that of the front surface of the tear layer beneath the lens. A spherical base curve therefore effectively masks the corneal astigmatism that often contributes most of the refractive astigmatism in ametropia. As a result, many patients with moderate astigmatism can be managed satisfactorily with spherical contact lenses, provided that an acceptable fit of the base curve to the cornea can be achieved. If the astigmatism is too great, or if a suitable physical fit cannot be achieved, rigid contact lenses with toric front and/or back surfaces can be used. In these cases, the lenses must be oriented using prism ballast in which prism power is ground into the lens so that the thicker base edge is at the bottom of the lens—its orientation is maintained by gravity.

Rigid lenses, particularly those made of PMMA, will change the shape of the cornea.<sup>22</sup> The curvature change may vary from day to day, making it difficult to determine a spectacle prescription for use when the patient is not wearing the contact lenses. This is described as spectacle blur. In extreme cases, the warped cornea may have a permanent irregular astigmatism. *Orthokeratology* is a clinical approach to reducing or eliminating refractive error by corneal molding with rigid contact lenses.<sup>23</sup> The procedure involves a series of contact lenses with successively flatter base curves to reduce corneal curvature. After several months, the patient experiences a mild reduction of 1 or 2 D of myopia. This is a temporary (i.e., several hours) effect. Retainer lenses must be worn for several hours each day to prevent the cornea from rebounding to its original shape.

Most rigid lenses are made with a handling tint, often a very light blue, brown, or green, to facilitate locating a lens dropped on the floor or countertop. The tint should not affect the wearer's colour perception. The right lens of a pair is normally identified with a dot engraved on the front surface. Lenses must be cleaned and disinfected with special care solutions after each wearing. The fit of the lens should be checked every year. The expected service life of a rigid gas permeable contact lens is 2 to 3 years.

**Hydrogel Lenses** Soft or hydrogel contact lenses were introduced in the 1970s and today they dominate the contact lens market. A hydrogel is a polymer that is able to absorb a significant amount of water.<sup>21</sup> Although brittle and fragile when dry, a hydrogel becomes soft and flexible when hydrated. The first commercially successful lens, the Bausch & Lomb Soflens®, was made of hydroxyethylmethacrylate (HEMA), which contains 38.6 percent water by weight when fully hydrated. Manufacturers have since developed HEMA derivatives with water content up to 77 percent.<sup>24</sup> A hydrogel lens that is saturated with ophthalmic saline solution buffered to the pH of the tears will smoothly drape over the cornea if an appropriate base curve has been selected. The base curve of the lens is determined from the value of  $K$ , the corneal curvature measured by a keratometer. A trial lens of known power with this base curve is placed on the eye and an over refraction is performed to determine the final contact lens power. A properly sized hydrogel lens will cover the entire cornea

and overlap the sclera by about 1 mm; the peripheral flattening of the cornea is accommodated by flattening of the edge zone of the ocular surface of the contact lens. The optic zone of the lens, where the base curve is found, covers most of the cornea.

Most soft contact lenses are spherical with small amounts of astigmatism being corrected with the equivalent sphere of the prescription. Because a hydrogel lens conforms more to the shape of the cornea than a rigid lens, higher amounts of astigmatism must be corrected with a toric lens. Maintaining the orientation of a toric lens on the eye is a challenge, since interaction of the lens with the underlying cornea, the tear layer, and the action of the lids will influence how it sits on the eye. Various methods have been used by manufacturers to stabilize lens position.

Soft lenses can be worn on a daily, disposable/frequent replacement basis. In daily wear, the lenses are worn up to 12 to 14 hours a day and removed each night. The lenses must be cleaned and disinfected before they are worn again. Cleaning with digital rubbing of the lens surfaces removes mucous deposits while disinfection eliminates infectious agents. These lenses can be worn for about 1 year. Disposable or frequent replacement lenses are worn between 1 day and several weeks on a daily basis, then discarded. They are prescribed for patients who experience rapid build-up of protein deposits or who have limited access to cleaning and sterilizing solutions. Extended wear lenses are worn up to a week at a time before being removed for cleaning and disinfection. Their higher water content and reduced thickness increase oxygen permeability but make them more susceptible to damage when handled. Extended wear lenses are usually replaced monthly.

One important advantage of hydrogel over rigid lens materials is that the lens is comfortable to wear with little or no adaptation time. Tear exchange is not as important since hydrogel transmits oxygen to the cornea. Thin lenses with high water content transmit significantly more oxygen. The lenses do not dislodge easily, making them ideal for sports activity; however, it has been reported that hydrogel lenses worn in chlorinated pool water exhibit significantly more microbial colonization than lenses that were never worn in the pool. This may increase the risk of bacterial keratitis if the lenses are worn in water sports.<sup>25</sup> Lenses worn in water sports should be discarded as soon as possible after the activity is finished.<sup>26</sup>

In the last few years, silicone hydrogel lenses have entered the market. Silicone hydrogel has very high oxygen transmissivity, but is stiffer than HEMA-based hydrogel. Initially there is more awareness of the lens in situ; however, the oxygen permeability of these lenses has largely eliminated many of the complications of earlier hydrogel lenses.

Most soft contact lenses are either clear or have a slight handling tint. Cosmetic lenses with an overall tint can be used to make light irides appear in different colors. Lenses with arrays of tinted dots can be used to change the color appearance of dark irides with varying degrees of success. Novelty tints can be used to change the appearance of the eye (e.g., slit pupils, “unnatural” pupil color); however, there is a controversy over whether such lenses should only be supplied through eye care practitioners because of serious risks to eye health.<sup>27</sup> Similar types of lens can be used to disguise scarred or disfigured eyes due to damage or disease.<sup>28</sup>

**Contact Lenses for Presbyopia** Presbyopia has become an important clinical problem as patients who began wearing their contact lenses in the 1970s and 1980s enter their 40s or 50s. The quality of vision and convenience of a largely spectacle-free life are compromised. Reading glasses worn over the contact lenses is a simple approach, but defeats the purpose of having contacts in the first place.

Monovision is the practice of fitting one eye for distance correction and the other for near. This approach reduces the need for spectacles, but relies on the patient’s ability to suppress the vision of one or the other eye when looking at a given target. The resulting reduction of binocular vision may create more serious problems involving depth and space perception.

Bifocal contact lenses have met with varying degrees of success. Alternating vision rigid contact lenses have a segment ground into the lens. When the patient looks at a near object the lower eyelid moves the lens over the cornea to bring the segment in front of the pupil. The theory of this design is seldom found in practice.

The simultaneous vision lenses are more successful. Both the distance portion and the segment are within the pupil so that superimposed clear and blurred images of distant and near objects are seen. There is a resultant loss of clarity and contrast. A diffractive contact lens design uses concentric



rings cut into the base curve of the lens to provide a near correction while the front surface curvature provides the distance correction simultaneously.<sup>29</sup>

For a more detailed discussion of contact lens optics and technology, see Chap. 20.

## Refractive Surgery

The desire of many ametropic patients to be free of both spectacles and contact lenses has driven the development of refractive surgery to reduce or neutralize ametropia. Earlier procedures that used incisions in the cornea to alter its shape (radial keratotomy) have since been largely abandoned due to complications including corneal perforation<sup>30</sup> and the possibility of rupture of the compromised cornea.<sup>31</sup> More recently, the excimer laser has been used in *photorefractive keratectomy* (PRK) and *laser in-situ keratomileusis* (LASIK) to reshape the cornea. These procedures are not without risks of compromised vision and complications include stromal haze, regression of refractive effect, infection, and optical and/or mechanical instability of the corneal flap in LASIK.<sup>30</sup> Almost all patients achieve visual acuity of at least 20/40 (6/12) and most achieve postoperative visual acuity of at least 20/20 (6/6). LASIK results in increased higher-order ocular aberrations; a new wavefront-guided LASIK procedure improves on this result.<sup>32</sup> A more detailed description of PRK and LASIK can be found in Chap. 16.

Extremely high myopia (over 12.00 D) may be treated by clear lens exchange in which the crystalline lens is extracted and an intraocular lens of suitable power is implanted. Phakic lens implants have also been used in either the anterior chamber or posterior chamber.

Low myopes can achieve visual acuity of better than 20/40 (6/12) when treated with intrastromal corneal rings (ICR). Circular arcs are implanted concentrically to the optical axis of the cornea in the stroma of the peripheral cornea to flatten the corneal curvature mechanically. The procedure is at least partly reversed by removal of the rings.

None of these procedures is guaranteed to result in perfect vision. All have some deleterious effect on quality of vision (glare, loss of contrast, increased higher-order aberrations<sup>30</sup>) and none eliminates the need for glasses for certain visually demanding tasks. As these patients become presbyopic, they will still require some form of reading prescription.

## Aphakia and Pseudophakia

*Cataract* is the general term for a crystalline lens that loses transparency because of aging or injury. Cortical cataract is seen in the outer layers of the crystalline lens, the cortex, and may comprise general haze, punctuate or wedge shaped opacities, and bubbles. In earlier stages, cortical cataract does not greatly affect visual acuity that is measured with high contrast targets, but there may be a significant reduction of contrast sensitivity. Patients may complain of glare in bright sunlight and difficulty reading. Nuclear cataract often appears as a gradual yellowing of the core of the crystalline lens. This is thought to arise from photochemical changes in the lens crystalline proteins triggered by long-term chronic exposure to ultraviolet radiation.<sup>33</sup> Visual consequences include reduced color discrimination and visual acuity.

Treatment of cataract is by surgical removal of the crystalline lens. *Intracapsular cataract extraction* (ICCE) is the removal of the lens and its surrounding lens capsule. ICCE has been largely replaced by *extracapsular cataract extraction* (ECCE) in which the lens is removed, but the capsule remains in the eye. Phacoemulsification is a procedure for breaking up the lens with an ultrasonic probe to facilitate ECCE. A very small incision near the edge of the cornea is required for ECCE, whereas a much larger circumferential incision is required for ICCE. Consequently, the potential for optical and physical postoperative complications in ECCE is much smaller than for ICCE. Approximately one in three patients who undergo ECCE will experience postoperative opacification of the lens capsule. This is remedied by an in-office YAG laser capsulotomy to open a hole in the posterior capsule to restore a clear optical path to the retina.<sup>34</sup> An eye which has had its lens removed is described as *aphakic*.

**Postsurgical Correction** Since the crystalline lens contributes about one-third of the optical power of the eye (see Chap. 1), the aphakic eye requires a high plus optical correction whose approximate power is

$$F = +11.25 + 0.62 F_{\text{old}} \quad (2)$$

where  $F_{\text{old}}$  is the preoperative equivalent spectacle correction.<sup>35</sup> This can be provided in the form of spectacles, contact lenses, or intraocular lens implants.

Aphakic spectacle corrections, sometimes referred to as cataract lenses, are heavy, thick, and cosmetically unattractive. They also have significant optical problems including distortion, field curvature, and oblique astigmatism that cannot be minimized using best-form principles, high magnification, and ring scotoma arising from prismatic effect in the periphery.<sup>36</sup> Aspheric lenses and lenticular designs can minimize some of these effects, but aphakic spectacle corrections remain problematic.

Contact lenses provide much better visual performance for aphakic patients. Cosmetic appearance is improved and most of the problems due to the aberrations and prismatic effects of spectacle lenses are eliminated. The magnification is closer to that of the phakic eye, making spatial adaptation and hand-eye coordination easier.<sup>37</sup> The optical advantages are somewhat offset by the difficulties of fitting these lenses due to the thick optical center and the challenge of providing sufficient oxygen to maintain corneal physiology.

Except in rare cases, most patients with clinically significant cataracts now have their natural crystalline lenses replaced with *intraocular lenses* (IOLs). They are described as being *pseudophakic*. In general, posterior chamber IOLs that are implanted in the lens capsule after ECCE, provide better optical performance than anterior chamber lenses that are fixed in the anterior chamber angle or clipped to the iris following ICCE. The power of the IOL is determined from the keratometric measurement of corneal curvature and the axial length of the eye, as determined by ultrasound A-scan. There are many formulas for estimating the power of the IOL. One of the most widely used, the SRK II formula is

$$P = A1 - 0.9 K - 2.5 L \quad (3)$$

where  $P$  is the IOL power needed for emmetropia,  $A1$  is a constant that varies according to the axial length of the eye  $L$  in millimeters, and  $K$  is the keratometer reading of central corneal curvature.<sup>38</sup> IOLs for astigmatism and bifocal IOLs have also been produced and recently designs incorporating wavefront correction have been introduced. IOL formulas have also been developed for use with patients who have undergone LASIK.<sup>39</sup> A more detailed discussion of the optics and clinical considerations of IOLs is found in Chap. 21.

## 12.6 BINOCULAR FACTORS

### Convergence and Accommodation

In most individuals the two eyes have similar refractive errors. When looking at a distant object, the amount of defocus is therefore similar in both eyes and accommodation to view a near object results in retinal images that appear similar in size and clarity. (It is assumed that both eyes will accommodate by the same amount.) The visual system has an easier task to fuse the two retinal images into a single binocular percept.

Ideally, corresponding points in the two retinal images would be fused into a single perceived image. However, binocular fusion does not require exactly matching images. When a single point in the retina of one eye is stimulated, an area of retina in the other eye surrounding the corresponding point will give rise to a single fused percept when any point in this area is stimulated. This is called *Panum's area*. The size of Panum's area varies with location in the retina. It is very small in the macula and increases toward the periphery of the retina. Disparities between the two points are used by the visual system to determine stereoscopic depth.



Control of the extraocular muscles is neurologically linked to the process of accommodation. As accommodation increases, the eyes converge so that near objects can be seen as a single binocular percept. When the patient accommodates while wearing corrective lenses, the lines of sight pass through off-axis points of the lens and prismatic effect will alter the amount of convergence needed to maintain fusion. Myopes experience less convergence demand while hyperopes experience more; the amount of this difference from the emmetropic convergence demand can be estimated for each eye using Prentice's rule:

$$P_H = xF \quad (4)$$

where  $P_H$  is the prismatic effect in *prism diopters* along the horizontal direction,  $x$  is the distance in centimeters along the horizontal meridian of the line of sight from the optical center of the lens, and  $F$  is the power of the corrective lens in the horizontal meridian.

The foregoing discussion assumes that the ametropia is spherical. If the ametropia is astigmatic, it is open to question whether accommodation is driven to focus on one or the other principal meridian or on the circle of least confusion (spherical equivalent). Most writers assume that focus is to the spherical equivalent in the interval of Sturm.

A more extensive discussion of binocular factors is found in Chap. 13.

## Anisometropia and Aniseikonia

*Anisometropia* is considered clinically significant when the spherical equivalent refractive error of the two eyes differs by 1.00 D or more.<sup>40</sup> Whether corrected or uncorrected, anisometropia can lead to significant problems with binocular vision for the patient.

**Uncorrected Ametropia** In uncorrected ametropia, depending on the refractive errors of the two eyes and the distance of the object of regard, one retinal image or the other may be in focus, or neither, since both eyes accommodate by the same amount. If the patient is a young child, the eye that more frequently has a clear retinal image will develop normally; however, the eye that tends to have a blurred image more of the time is likely to develop *amblyopia*. This is often the case where one eye is hyperopic and the other emmetropic, or if both eyes are hyperopic but to different degrees. If one eye is myopic and the other emmetropic, the child may learn to use the myopic eye for near objects and the emmetropic one for distant objects, in which case there is no amblyopia, but binocular fusion and stereopsis may not develop normally.

**Corrected Ametropia** When anisometropia is corrected, accommodation brings both retinal images of a near object into focus; however, differences in prismatic effect in off-axis gaze and in retinal image size may lead to problems with binocular vision, particularly fusion. This is particularly a problem when the patient looks down to read. Although the visual system has the capacity to fuse images with substantial horizontal disparities, it is extremely limited when disparities in the vertical meridian are encountered. In downward gaze the lines of sight pass through a spectacle lens approximately 1.0 cm below the optical center. Many patients experience visual discomfort when the vertical prismatic imbalance between the eyes is about 1 prism diopter (this unit is often written as  $\Delta$ ) and vertical diplopia (double vision) when it is over  $2\Delta$ . While a younger patient can avoid this problem by tipping the chin down so that the lines of sight pass through the paraxial zone of the lens in the vertical meridian, presbyopic anisometropes would need either bifocals or PALs with slab-off prism in the near portion of the lens with more minus or less plus distance power or a separate pair of single vision reading glasses.<sup>40</sup> Bifocal contact lenses, monovision, or reading glasses with contact lenses may avoid the problem altogether, but may cause other complications with the patient's vision.

**Aniseikonia** Anisometropia corrected with spectacles almost invariably results in *aniseikonia*, a condition in which binocular fusion is difficult or impossible because of the disparity in size of the retinal or cortical images from the two eyes.

The spectacle magnification  $M$  of a spectacle lens is given by the formula

$$M = \frac{1}{1 - d_v F} \frac{1}{1 - \frac{t}{n} F_1} \quad (5)$$

where  $d_v$  is the distance between the back vertex of the spectacle lens and the cornea,  $F$  is the back vertex power of the lens,  $t$  is the axial thickness of the lens,  $n$  its refractive index, and  $F_1$  its front surface power. The first factor is the power factor, the magnification due to a thin lens of power  $F$  at distance  $d_v$  in front of the cornea and the second is the shape factor, the magnification due to an afocal lens with front surface power  $F_1$ , index  $n$ , and thickness  $t$ . For thin lenses the equation can be written as

$$M = 1 + d_v F + \frac{t}{n} F_1 \quad (6)$$

using a thin lens approximation.<sup>41</sup>

Aniseikonia can be measured with a space eikonometer or estimated from the refractive error. The percentage difference in image size is between 1 and 1.5 percent per diopter of anisometropia and can be attributed to differences in refractive power of the eyes rather than axial length. When relative spectacle magnification is calculated on the basis of axial versus refractive ametropia, aniseikonia must arise from differences between the eyes in refractive ametropia. Size lenses, afocal lenses with magnification, can be used to determine the amount of change in spectacle magnification of the corrective lenses needed to neutralize the aniseikonia. The spectacle magnification formula can then be used to calculate the new lens parameters needed to modify the shape factor of the lens.<sup>41</sup>

Aniseikonia is relatively rare; however, cataract surgery and corneal refractive surgery often induces it particularly when only one eye is operated. Clinicians require more understanding of this clinical phenomenon than is assumed in their training. The reader is referred to Remole and Robertson's book on this subject<sup>42</sup> as well as the Web site [www.opticaldiagnostics.com/info/aniseikonia.html](http://www.opticaldiagnostics.com/info/aniseikonia.html)<sup>43</sup> for further information.

## 12.7 CONSEQUENCES FOR OPTICAL DESIGN

Designers of optical equipment must take into account that many potential users of their devices will have ametropia, either corrected or uncorrected, and perhaps presbyopia as well. While each refractive condition has its challenges, the combination can make for a great deal of discomfort and inconvenience when using optical equipment. For example, digital point-and-shoot cameras have LCD panels to display images as well as control menus and exposure information. The panel is also used to aim the camera. Most users will hold the device approximately 40 cm from the eyes—viewing the display is uncomfortable and difficult, particularly if the user is wearing bifocal or PAL spectacles. It is also important to determine whether the device is to be used monocularly or binocularly. In the latter case, the optical design must allow for convergence of the eyes and the possible change in convergence demand if the user has to accommodate. A further complication is the size of the exit pupil of the device: as patients age, their pupil size decreases and results in vignetting of the exit beam. Human visual performance factors must be considered carefully in the design of future optical devices.

## 12.8 REFERENCES

1. W. N. Charman, "Optics of the Eye," Chap. 1, in V. Lakshminarayanan and Enoch J. M. (eds.), *OSA Handbook of Optics*, Vol. 3, McGraw-Hill, New York, 2009.
2. World Health Organization. *Elimination of Avoidable Visual Disability due to Refractive Errors*, WHO/PBL/00.79, WHO, Geneva, 2001.

3. M. Jalie, *Ophthalmic Lenses and Dispensing*, 3rd ed., Butterworth Heinemann, Boston, 2008.
4. W. Long, "Why Is Ocular Astigmatism Regular?" *Am. J. Optom. Physiol. Opt.* **59**(6):20–522 (1982).
5. W. Charman and G. Walsh, "Variations in the Local Refractive Correction of the Eye Across its Entrance Pupil," *Opt. Vis. Sci.* **66**(1):34–40 (1989).
6. C. Sheard, "A Case of Refraction Demanding Cylinders Crossed at Oblique Axes, Together with the Theory and Practice Involved," *Ophth. Rec.* **25**:558–567 (1916).
7. T. D. Williams, "Malformation of the Optic Nerve Head," *Am. J. Optom. Physiol. Opt.* **55**(10):706–718 (1978).
8. R. B. Rabbetts, *Clinical Visual Optics*, 3rd ed., Butterworth Heinemann, Boston, 1998, pp. 288–289.
9. R. B. Rabbetts, *Clinical Visual Optics*, 3rd ed., Butterworth Heinemann, Boston, 1998, pp. 380–389.
10. R. B. Rabbetts, *Clinical Visual Optics*, 3rd ed., Butterworth Heinemann, Boston, 1998, pp. 330–350.
11. R. B. Rabbetts, *Clinical Visual Optics*, 3rd ed., Butterworth Heinemann, Boston, 1998, pp. 19–61.
12. J. M. Newman, "Analysis, Interpretation and Prescription for the Ametropias and Heterophorias," Chap. 22, in W. J. Benjamin (ed.), *Borish's Clinical Refraction*, 2nd ed., Butterworth Heinemann, Boston, 2006, pp. 1002–1009.
13. M. Jalie, *Ophthalmic Lenses and Dispensing*, 3rd ed., Butterworth Heinemann, Boston, 2008, pp. 24–27.
14. W. F. Long, "Paraxial Optics of Vision Correction," Chap. 4, in W. N. Charman (ed.), *Vision and Visual Dysfunction*, Vol. 1, *Visual Optics and Instrumentation*, MacMillan, London, 1991, p. 55.
15. M. Jalie, *The Principles of Ophthalmic Lenses*, 3rd ed., Association of Dispensing Opticians, London, 1984, pp. 413–468.
16. B. R. Chou and J. K. Hovis, "Durability of Coated CR-39 Industrial Lenses," *Optom. Vis. Sci.* **80**(10):703–707 (2003).
17. B. R. Chou, A. Gupta, and J. K. Hovis, "The Effect of Multiple Antireflective Coatings and Centre Thickness on Resistance of Polycarbonate Spectacle Lenses to Penetration by Pointed Missiles," *Optom. Vis. Sci.* **82**(11):964–969 (2005).
18. B. R. Chou and J. K. Hovis, "Effect of Multiple Antireflection Coatings on Impact Resistance of Hoya Phoenix Spectacle Lenses," *Clin. Exp. Optom.* **89**(2):86–89 (2006).
19. D. G. Pitts and B. R. Chou, "Prescription of Absorptive Lenses," Chap. 25, in W. J. Benjamin and I. M. Borish (eds.), *Borish's Clinical Refraction*, 2nd ed., W.B. Saunders Company: Cambridge, MA, 2006, pp. 1153–1187.
20. M. Jalie, *Ophthalmic Lenses and Dispensing*, 3rd ed., Butterworth Heinemann, Boston, 2008, pp. 169–195.
21. M. F. Refojo, "Chemical Composition and Properties," Chap. 2, in M. Guillon and M. Ruben (eds.), *Contact Lens Practice*, Chapman & Hall, London, 1994, pp. 29–36.
22. G. E. Lowther, "Induced Refractive Changes," Chap. 44, in M. Guillon and M. Ruben (eds.), *Contact Lens Practice*, Chapman & Hall, London, 1994.
23. L. G. Carney, "Orthokeratology," Chap 37, in M. Guillon and M. Ruben (eds.), *Contact Lens Practice*, Chapman & Hall, London, 1994.
24. M. Callender, B. R. Chou, and B. E. Robinson (eds.), *Contact Lenses and Solutions Available in Canada Univ. Waterloo Contact Lens J.* **34**(1):5–70 (2007).
25. J. Choo, K. Vuu, P. Bergenske, K. Burnham, J. Smythe, and P. Caroline, "Bacterial Populations on Silicone Hydrogel and Hydrogel Contact Lenses after Swimming in a Chlorinated Pool," *Optom. Vis. Sci.* **82**(2):134–137 (2005).
26. D. Lam and T. B. Edrington, "Contact Lenses and Aquatics," *Contact Lens Spectrum*, **21**(5):2–32 (2006).
27. Food and Drug Administration. FDA Reminds Consumers of Serious Risks of Using Decorative Contact Lenses without Consulting Eye Care Professionals. *FDA News* (Online) October 27, 2006. Accessed June 15, 2008, [www.fda.gov/bbs/topics/NEWS/2006/NEW01499.html](http://www.fda.gov/bbs/topics/NEWS/2006/NEW01499.html).
28. M. J. A. Port, "Cosmetic and Prosthetic Contact Lenses," Chap 20.6, in A. J. Phillips and L. Speedwell (eds.), *Contact Lenses*, 4th ed., Butterworth Heinemann, Oxford, 1997, pp. 752–754.
29. A. L. Cohen, "Diffractive Bifocal Lens Designs," *Optom. Vis. Sci.*, **70**(6):461–468 (1993).
30. J. P. G. Bergmanson and E. J. Farmer, "A Return to Primitive Practice? Radial Keratotomy Revisited," *Contact Lens Anterior Eye* **22**(1):2–10(1999).
31. P. Vinger, W. Mieler, J. Oestreicher, and M. Easterbrook, "Ruptured Globe Following Radial and Hexagonal Keratotomy," *Arch. Ophthalmol.* **114**:129–134 (1996).

32. M. J. Lee, S. M. Lee, H. J. Lee, W. R. Wee, J. H. Lee, and M. K. Kim, "The Changes of Posterior Corneal Surface and High-order Aberrations after Refractive Surgery in Moderate Myopia," *Korean J. Ophthalmol*, **21**(3): 131–136 (2007).
33. D. G. Pitts, L. L. Cameron, J. G. Jose, S. Lerman, E. Moss, S. D. Varma, S. Zigler, S. Zigman, and J. Zuclich, "Optical Radiation and Cataracts," Chap. 2, in M. Waxler and V. M. Hitchins (eds.), *Optical Radiation and Visual Health*. CRC Press, Boca Raton, FL, 1986, pp. 23–25.
34. B. Noble and I. Simmons, *Complications of Cataract Surgery: A Manual*, Butterworth Heinemann, Boston, 2001, pp. 83–85.
35. H. H. Emsley, "Optics of Vision," Vol. 1, *Visual Optics*, 5th ed. Butterworths, London, 1952, p. 111.
36. T. E. Fannin and T. P. Grosvenor, *Clinical Optics*, 2nd ed, Butterworth Heinemann, Boston, 1996, pp. 328–345.
37. T. E. Fannin and T. P. Grosvenor, *Clinical Optics*, 2nd ed., Butterworth Heinemann, Boston, 1996, p. 396.
38. D. R. Sanders, J. Retzlaff, and M. C. Kraff, "Comparison of the SRK II Formula and other Second Generation Formulas," *J. Cataract Refract Surg*. **14**:136–141 (1988).
39. L. Wang, M. A. Booth, and D. D. Koch, "Comparison of Intraocular Lens Power Calculation Methods in Eyes that Have Undergone Laser-assisted in-situ Keratomileusis," *Trans. Am. Ophthalmol Soc.*, **102**:189–197 (2004).
40. T. E. Fannin and T. P. Grosvenor, *Clinical Optics*, 2nd ed., Butterworth Heinemann, Boston, 1996, pp. 294–299.
41. T. E. Fannin and T. P. Grosvenor, *Clinical Optics*, 2nd ed., Butterworth Heinemann, Boston, 1996, pp. 300–323.
42. A. Remole and K. M. Robertson, *Aniseikonia and Anisophoria*, Runestone Publishing, Waterloo, Ontario, Canada, 1996.
43. About Aniseikonia. *Optical Diagnostics*. (Online) Accessed June 25, 2008, [www.opticaldiagnostics.com/info/aniseikonia.html](http://www.opticaldiagnostics.com/info/aniseikonia.html).

*This page intentionally left blank*

---

# BINOCULAR VISION FACTORS THAT INFLUENCE OPTICAL DESIGN

---

Clifton Schor

*School of Optometry  
University of California  
Berkeley, California*

---

## 13.1 GLOSSARY

---

**Accommodation.** Change in focal length or optical power of the eye produced by change in power of the crystalline lens as a result of contraction of the ciliary muscle. This capability decreases with age.

**Ametrope.** An eye with a refractive error.

**Aniseikonia.** Unequal perceived image sizes from the two eyes.

**Baseline.** Line intersecting the entrance pupils of the two eyes.

**Binocular disparity.** Differences in the perspective views of the two eyes.

**Binocular fusion.** Act or process of integrating percepts of retinal images formed in the two eyes into a single combined percept.

**Binocular parallax.** Angle subtended by an object point at the nodal points of the two eyes.

**Binocular rivalry.** Temporal alternation of perception of portions of each eye's visual field when the eyes are stimulated simultaneously with targets composed of dissimilar colors or different contour orientations.

**Center of rotation.** A pivot point within the eye about which the eye rotates to change direction of gaze.

**Concomitant.** Equal amplitude synchronous motion or rotation of the two eyes in the same direction.

**Conjugate.** Simultaneous motion or rotation of the two eyes in the same direction.

**Convergence.** Inward rotation of the two eyes.

**Corresponding retinal points.** Regions of the two eyes, which when stimulated result in identical perceived visual directions.

**Cyclopean eye.** A descriptive term used to symbolize the combined view points of the two eyes into a single location midway between them.

**Cyclovergence.** Unequal torsion of the two eyes.

**Disconjugate.** Simultaneous motion or rotation of the two eyes in opposite directions.

**Divergence.** Outward rotation of the two eyes.

**Egocenter.** Directional reference point for judging direction relative to the head from a point midway between the two eyes.

**Emmetrope.** An eye with no refractive error.

**Entrance pupil.** The image of the aperture stop formed by the portion of an optical system on the object side of the stop. The eye pupil is the aperture stop of the eye.

**Extraretinal cues.** Nonvisual information used in space perception.

**Eye movement.** A rotation of the eye about its center of rotation.

**Fixation.** The alignment of the fovea with an object of interest.

**Focus of Expansion.** The origin of velocity vectors in the optic flow field.

**Frontoparallel plane.** Plane that is parallel to the face and orthogonal to the primary position of gaze.

**Haploopia.** Perception of a single target by the two eyes.

**Heterophoria.** Synonymous with phoria.

**Horofter.** Locus of points in space whose images are formed on corresponding points of the two retinas.

**Hyperopia.** An optical error of the eye in which objects at infinity are imaged behind the retinal surface while the accommodative response is zero.

**Midsagittal plane.** Plane that is perpendicular to and bisects the baseline. The plane vertically bisecting the midline of the body.

**Motion parallax.** Apparent relative displacement or motion of one object or texture with respect to another that is usually produced by two successive views by a moving observer of stationary objects at different distances.

**Myopia.** An optical error of the eye in which objects at infinity are imaged in front of the retinal surface while the accommodative response is zero.

**Nonconcomitant.** Unequal amplitudes of synchronous motion or rotation of the two eyes in the same direction.

**Optic flow.** Pattern of retinal image movement.

**Percept.** That which is perceived.

**Perception.** The act of awareness through the senses, such as vision.

**Perspective.** Variations of perceived size, separation, and orientation of objects in 3-D space from a particular viewing distance and vantage point.

**Phoria.** An error of binocular alignment revealed when one eye is occluded.

**Primary position of gaze.** Position of the eye when the visual axis is directed straight ahead and is perpendicular to the frontoparallel plane.

**Retinal disparity.** Difference in the angles formed by two targets with the entrance pupils of the eyes.

**Shear.** Differential displacement during motion parallax of texture elements along an axis perpendicular to meridian of motion.

**Skew movement.** A vertical vergence or vertical movement of the two eyes in opposite directions.

**Stereopsis.** The perception of depth stimulated by binocular disparity.

**Strabismus.** An eye turn or misalignment of the two eyes during attempted binocular fixation.

**Tilt.** Amount of angular rotation in depth about an axis in the frontoparallel plane—synonymous with orientation.

**Torsion.** Rotation of the eye around the visual axis.

**Vantage point.** Position in space of the entrance pupil through which 3-D space is transformed by an optical system with a 2-D image or view plane.

**Visual field.** The angular region of space or field of view limited by the entrance pupil of the eye, the zone of functional retina, and occlusion structures such as the nose and orbit of the eye.

**Visual plane.** Any plane containing the fixation point and entrance pupils of the two eyes.

**Yoked movements.** Simultaneous movement or rotation of the two eyes in the same direction.

## 13.2 COMBINING THE IMAGES IN THE TWO EYES INTO ONE PERCEPTION OF THE VISUAL FIELD

### The Visual Field

Optical designs that are intended to enhance vision usually need to consider limitations imposed by having two eyes and the visual functions that are made possible by binocular vision. When developing an optical design it is important to tailor it to specific binocular functions that you wish to enhance or at least not limit binocular vision. Binocular vision enhances visual perception in several ways. First and foremost, it expands the visual field.<sup>1</sup> The horizontal extent of the visual field depends largely on the placement and orientation of eyes in the head. Animals with laterally placed eyes have panoramic vision that gives them a full 360° of viewing angle. The forward placement of our eyes reduces the visual field to 190°. Eye movements let us expand our field of view. The forward placement of the eyes adds a large region of binocular overlap so that we can achieve another binocular function, stereoscopic depth perception. The region of binocular overlap is 114° and the remaining monocular portion is 37° for each eye. Each eye sees slightly more of the temporal than nasal-visual field and they also see more of the ipsilateral than the contralateral side of a binocularly viewed object.

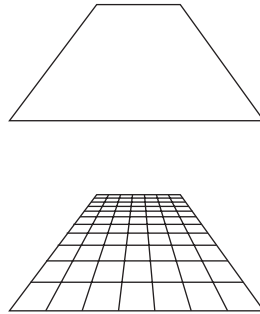
Visual perception is not uniform throughout this region. The fovea or central region of the retina is specialized to allow us to resolve fine detail in the central 5° of the visual field. Eye movements expand the regions of available space. They allow expansion of the zone of high resolution, and accommodation expands the range of distances within which we can have high acuity and clear vision from viewing distances ranging from as near as 5 cm in the very young to optical infinity.<sup>2</sup> The peripheral retina is specialized to aid us in locomotor tasks such as walking so that we can navigate safely without colliding with obstacles. It is important that optical designs allow users to retain the necessary visual to perform tasks aided by the optical device.

### Perceived Space

Space perception includes our perception of direction, distance, orientation, and shape of objects; object trajectories with respect to our location; sensing body orientation, location, and motion in space; and heading. These percepts can be derived from a 2-D image projection on the two retinas of a 3-D object space. Reconstruction of a 3-D percept from a 2-D image requires the use of information within the retinal image that is geometrically constrained by the 3-D nature of space. Three primary sources of visual or vision-related information are used for this purpose and they include monocular information, binocular information, and extraretinal or motor information. Monocular visual or retinal cues include familiarity with size and shape of objects, linear perspective and shape distortions, texture density, shading, partial image occlusion or overlap, size expansion and optic flow patterns, and motion parallax.<sup>3</sup> Binocular information is also available, including stereoscopic depth sensed from combinations of horizontal and vertical disparity, and motion in depth.<sup>4</sup> Extraretinal cues include accommodation, convergence, and gaze direction of the eyes. Many of these cues are redundant and provide ways to check their consistency, and to sense their errors that can be corrected by adaptably reweighting their emphasis or contribution to the final 3-D percept.

**Monocular Cues** Static monocular cues include familiarity with size and shape of objects, linear perspective and shape distortions, texture density, partial image occlusion or overlap.<sup>3</sup> Linear perspective refers to the distortions of the retinal image that result from the projection or imaging of a 3-D object located at a finite viewing distance onto a 2-D surface such as the retina. Texture refers to a repetitive pattern of uniform size and shape such as a gravel bed. Familiarity with the size and shape of a target allows us to sense its distance and orientation. As targets approach us and change their orientation, even though the retinal image expands and changes shape, the target appears to maintain a constant perceived size and rigid shape. Retinal image shape changes are interpreted as changes in object orientation produced by rotations about three axes. Texture density can indicate which axes the object is rotated about. Rotation about the vertical axis produces tilt and causes a compression of texture along





**FIGURE 1** When the edges of an object are parallel, they meet at a point in the retinal image plane called the vanishing point.

the horizontal axis. Rotation about the horizontal axis produces slant and it causes a compression of texture along the vertical axis. Rotation about the  $z$  axis produces roll or torsion and causes a constant change in orientation of all texture elements. Combinations of these rotations cause distortions or foreshortening of images; however, if the distortions are decomposed into three rotations, the orientation of rigid objects can be computed. The gradient of compressed texture varies inversely with target distance. If viewing distance is known, the texture density gradient is a strong cue to the amount of slant about a given axis.

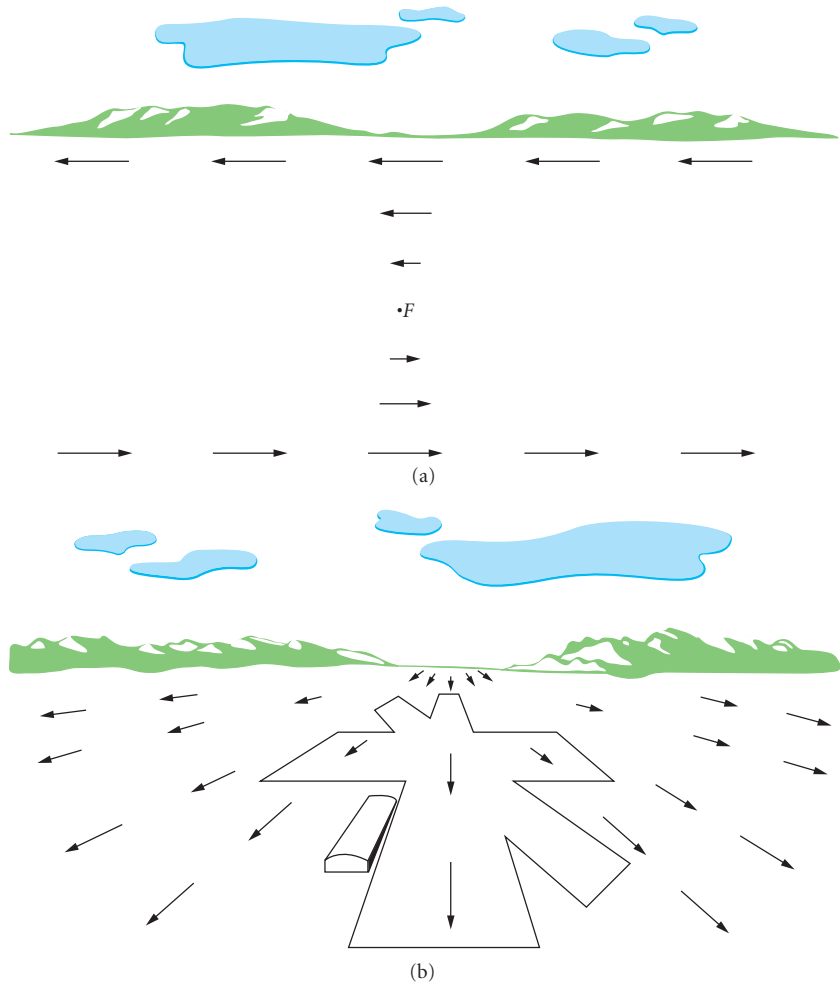
Perspective cues arising from combinations of observer's view point, object distance, and orientation can contribute to perception of distance and orientation. Perspective cues utilize edge information that is extrapolated until it intersects another extrapolated edge of the same target (Fig. 1).<sup>5</sup> When the edges of an object are parallel, they meet at a point in the retinal image plane called the vanishing point. Normally we do not perceive the vanishing point directly, and it needs to be derived by extrapolation. It is assumed that the line of sight directed at the vanishing point is parallel to the edges of the surface such as a roadway that extrapolates to the same vanishing point. As with texture density gradients, perspective distortion of the retinal image increases with object proximity. If the viewing distance is known, the amount of perspective distortion is an indicator of the amount of slant or tilt about a given axis.

Depth ordering can be computed from image overlap. This is a very powerful cue and can override any of the other depth cues, including binocular disparity. A binocular version of form overlap is called DaVinci stereopsis where the overlap is compared between the two eyes' views.<sup>6</sup> Because of their lateral separation, each eye sees a slightly different view of 3-D objects. Each eye perceives more of the ipsilateral side of the object than the contralateral side. Thus for a given eye, less of the background is occluded by a near object on the ipsilateral than the contralateral side. This binocular discrepancy is sufficient to provide a strong sense of depth in the absence of binocular parallax.

**Kinetic Cues** Kinetic monocular motion cues include size expansion, optic flow patterns, and motion parallax. Distance of targets that we approach becomes apparent from the increased velocity of radial flow of the retina (loom).<sup>7</sup> Motion of texture elements across the whole image contributes to our perception of shape and location of objects in space. The relative motion of texture elements in the visual field is referred to as optic flow and this flow field can be used to segregate the image into objects at different depth planes as well as to perceive the shape or form of an object. Motion parallax is the shear of optic flow in opposite directions resulting from the translation of our view point.<sup>8</sup> Movies that pan the horizon yield a strong sense of 3-D depth from the motion parallax they produce between near and far parts of 3-D objects. Two sequential views resulting from a lateral shift in the viewpoint (motion parallax) are analogous to two separate and simultaneous views from two eyes with separate viewpoints (binocular disparity). Unlike horizontal disparity, the shear

between sequential views by itself is ambiguous, but once we know which way our head is translating, we can correctly identify the direction or sign of depth.<sup>9</sup> The advantage of motion parallax over binocular disparity is that we can translate in any meridian, horizontal or vertical, and get depth information, while stereo only works for horizontal disparities.<sup>8</sup>

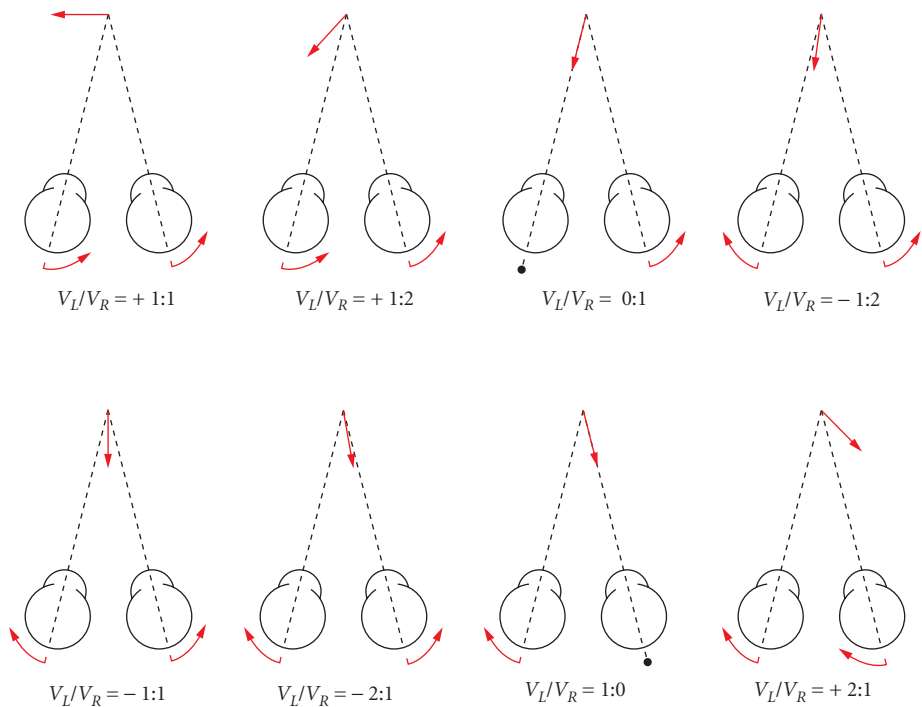
Motion perception helps us to determine where we are in space over time. Probably the most obvious application is to navigation and heading judgments (Fig. 2). Where are we headed, and can



**FIGURE 2** (a) Motion Parallax. Assume that an observer moving toward the left fixates a point at  $F$ . Objects nearer than  $F$  will appear to move in a direction opposite to that of the movement of the observer; objects farther away than  $F$  will appear to move in the same direction as the observer. The length of the arrows signifies that the apparent velocity of the optic flow is directly related to the distance of objects from the fixation point. (b) Motion perspective. The optical flow in the visual field as an observer moves forward. The view is that as seen from an airplane in level flight. The direction of apparent movement in the terrain below is signified by the direction of the motion vectors (arrows); speed of apparent motion is indicated by the length of the motion vectors. The expansion point in the distance from which motion vectors originate is the heading direction.

we navigate to a certain point in space? Generally we perceive a focus of expansion in the direction we are heading as long as our eyes and head remain fixed.<sup>7,10-13</sup> These types of judgments occur whenever we walk, ride a bicycle, or operate a car. Navigation also involves tasks to intercept or avoid moving objects. For example, you might want to steer your car around another car or avoid a pedestrian walking across the crosswalk. Or you may want to intercept something such as catching or striking a ball. Another navigation task is anticipation of time of arrival or contact. This activity is essential when pulling up to a stop sign or traffic light. If you stop too late you enter the intersection and risk a collision. These judgments are also made while walking down or running up a flight of stairs. Precise knowledge of time to contact is essential to keep from stumbling. Time to contact can be estimated from the distance to contact divided by the speed of approach. It can also be estimated from the angular subtense of an object divided by its rate of expansion.<sup>14</sup> The task is optimal when looking at a surface that is in the frontoparallel plane, and generally we are very good at it. It looks like we use the rate of expansion to predict time of contact since we make unreliable judgments of our own velocity.<sup>15</sup>

Another class of motion perception is self- or ego-motion.<sup>16</sup> It is related to the vestibular sense of motion and balance and it tells us when we are moving. Self-motion responds to vestibular stimulation as well as visual motion produced by translational and rotational optic flow of large fields. Finally, we can use binocular motion information to judge the trajectory of a moving object and determine if it will strike us in the head or not. This motion in depth results from a comparison of the motion in one eye to the other. The visual system computes the direction of motion from the amplitude ratio and relative direction of horizontal retinal image motion (Fig. 3).<sup>17</sup> If horizontal motion in the two



**FIGURE 3** Motion in depth. Relative velocities of left and right retinal images for different target trajectories. When the target moves along a line passing between the eyes, its retinal images move in opposite directions in the two eyes; when the target moves along a line passing wide of the head, the retinal images move in the same direction, but with different velocities. The ratio ( $V_L/V_R$ ) of left- and right-eye image velocities provides an unequivocal indication of the direction of motion in depth relative to the head.

eyes is equal in amplitude and direction, it corresponds to an object moving in the frontoparallel plane. If the horizontal retinal image motion is equal and opposite, the object is moving toward us in the midsagittal plane and will strike us between the eyes. If the motion has any disconjugacy at all, it will strike us. A miss is indicated by yoked, albeit noncomitant, motion.

***Binocular Cues and Extraretinal Information from Eye Movements*** Under binocular viewing conditions, we perceive a single view of the world as though seen by a single cyclopean eye, even though the two eyes receive slightly different retinal images. The binocular percept is the average of monocularly sensed shapes and directions (allelotropia). The benefit of this binocular combination is to allow us to sense objects as single with small amounts of binocular disparity so that we can interpret depth from the stereoscopic sense. Interpretation of the effects of prism and magnification upon perceived direction through an instrument needs to consider how the visual directions of the two eyes are combined. If we only had one eye, direction could be judged from the nodal point of the eye, a site where viewing angle in space equals visual angle in the eye, assuming the nodal point is close to the radial center of the retina. However, two eyes present a problem for a system that operates as though it has only a single cyclopean eye. The two eyes have view points separated by approximately 6.5 cm. When the two eyes converge accurately on a near target placed along the midsagittal plane, the target appears straight ahead of the nose, even when one eye is occluded. In order for perceived egocentric direction to be the same when either eye views the near target monocularly, there needs to be a common reference point for judging direction. This reference point is called the cyclopean locus or egocenter, and is located midway on the interocular axis. The egocenter is the percept of a reference point for judging visual direction with either eye alone or under binocular viewing conditions.

## Perceived Direction

Direction and distance can be described in polar coordinates as the angle and magnitude of a vector originating at the egocenter. For targets imaged on corresponding retinal points, this vector is determined by the location of the retinal image and by the direction of gaze that is determined by the average position of the two eyes (conjugate eye position). The angle the two retinal images form with the visual axes is added to the conjugate rotational vector component of binocular eye position (the average of right- and left-eye position). This combination yields the perceived egocentric direction. Convergence of the eyes, which results from disconjugate eye movements, has no influence on perceived egocentric direction. Thus, when the two eyes fixate near objects to the left or right of the midline in asymmetric convergence, only the conjugate component of the two eyes' positions contributes to perceived direction. These facets of egocentric direction were summarized by Hering<sup>18</sup> as five laws of visual direction, and they have been restated by Howard.<sup>19</sup> The laws are mainly concerned with targets imaged on corresponding retinal regions (i.e., targets on the horopter).

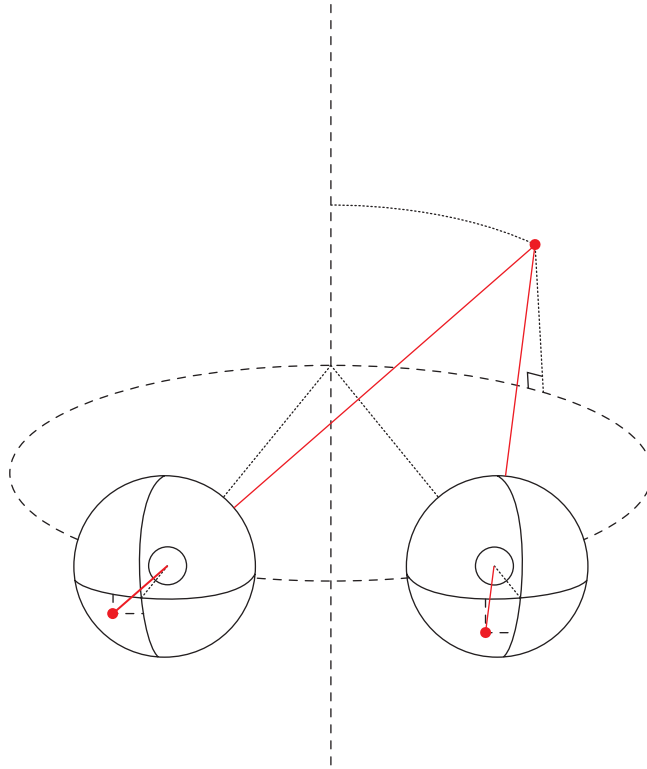
We perceive space with two eyes as though they were merged into a single cyclopean eye. This merger is made possible by a sensory linkage between the two eyes that is facilitated by the anatomical superposition or combination of homologous regions of the two retinas in the visual cortex. The binocular overlap of the visual fields is very specific. Unique pairs of retinal regions in the two eyes (corresponding points) must receive images of the same object so that these objects can be perceived as single and at the same depth as the point of fixation. This requires that retinal images must be precisely aligned by the oculomotor system with corresponding retinal regions in the two eyes. As described in Sec. 13.8, binocular alignment is achieved by yoked movements of the eyes in the same direction (version) and by movements of the eyes in opposite direction (vergence). Slight misalignment of similar images from corresponding points is interpreted as depth. Three-dimensional space can be derived geometrically by comparing the small differences between the two retinal images that result from the slightly different vantage points of the two eyes caused by their 6.5-cm separation. These disparities are described as horizontal, vertical, and torsional as well as distortion or shear differences between the two images. The disparities result from surface shape,

depth, and orientation with respect to the observer as well as direction and orientation (torsion) of the observer's eyes.<sup>20</sup> These disparities are used to judge the layout of 3-D space and to sense the solidness or curvature of surfaces. Disparities are also used to break through camouflage in images such as seen in tree foliage.

**Binocular Visual Direction—Corresponding Retinal Points and the Horopter** Hering<sup>18</sup> defined binocular correspondence by retinal locations in the two eyes, which when stimulated, resulted in a percept in identical visual directions. For a fixed angle of convergence, projections of corresponding points along the equator and midline of the eye converge upon real points in space. In other cases, such as oblique eccentric locations of the retina, corresponding points have visual directions that do not intersect in real space. The horopter is the locus in space of real objects or points whose images can be formed on corresponding retinal points. It serves as a reference throughout the visual field for the same depth or disparity as at the fixation point. To appreciate the shape of the horopter, consider a theoretical case in which corresponding points are defined as homologous locations on the two retinas. Thus corresponding points are equidistant from their respective foveas. Consider binocular matches between the horizontal meridians or equators of the two retinas. Under this circumstance, the visual directions of corresponding points intersect in space at real points that define the longitudinal horopter. This theoretical horopter is a circle whose points will be imaged at equal eccentricities from the two foveas on corresponding points except for the small arc of the circle that lies between the two eyes.<sup>22</sup> While the theoretical horopter is always a circle, its radius of curvature increases with viewing distance. This means that its curvature decreases as viewing distance increases. In the limit, the theoretical horopter is a straight line at infinity that is parallel to the interocular axis. Thus a surface representing zero disparity has many different shapes depending on the viewing distance. The consequence of this spatial variation in horopter curvature is that the spatial pattern of horizontal disparities is insufficient information to specify depth magnitude or even depth ordering or surface shape. It is essential to know viewing distance to interpret surface shape and orientation from depth related retinal image disparity. Viewing distance could be obtained from extraretinal information such as the convergence state of the eyes, or from retinal information in the form of vertical disparity as described further.

The empirical or measured horopter differs from the theoretical horopter in two ways.<sup>22</sup> It can be tilted about a vertical axis and its curvature can be flatter or steeper than the Vieth-Muller circle. This is a circle that passes through the fixation point in the midsagittal plane and the entrance pupils of the two eyes. Points along this circle are imaged at equal retinal eccentricities from the foveas of the two eyes. The tilt of the empirical horopter can be the result of a horizontal magnification of the retinal image in one eye. Image points of a frontoparallel surface subtend larger angles in the magnified images. A similar effect occurs when no magnifier is worn and the fixation plane is tilted toward one eye. The tilt causes a larger image to be formed in one eye than the other. Thus magnification of a horizontal row of vertical rods in the frontoparallel plane causes the plane of the rods to appear tilted to face the eye with the magnifier.

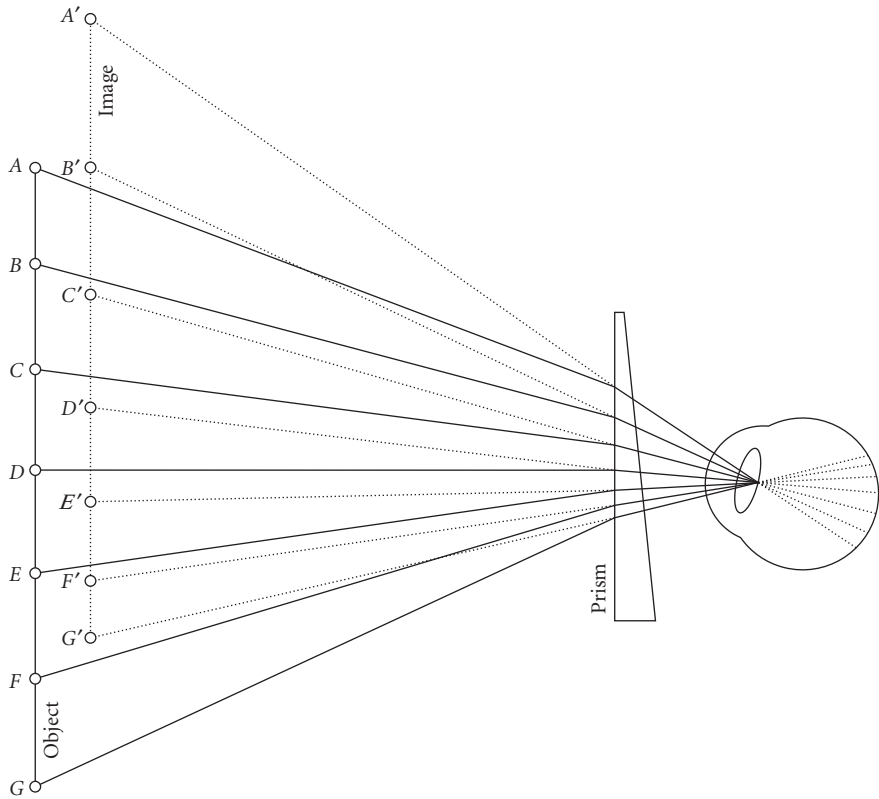
How does the visual system distinguish between a surface fixated in eccentric gaze that is in the frontoparallel plane (parallel to the face), but tilted with respect to the Vieth-Muller circle, and a plane fixated in forward gaze, that is tilted toward one eye? Both of these planes project identical patterns of horizontal retinal image disparity, but they have very different physical tilts in space. In order to distinguish between them, the visual system needs to know the horizontal gaze eccentricity. One way is to register the extraretinal eye position signal, and the other is to compute gaze eccentricity from the vertical disparity gradient associated with eccentrically located targets. It appears that the latter method is possible, since tilt of a frontoparallel plane can be induced by a vertical magnifier before one eye (the induced effect). The effect is opposite to the tilt produced by placing a horizontal magnifier before the same eye (the geometric effect). If both a horizontal and vertical magnifier are placed before one eye, in the form of an overall magnifier, no tilt of the plane occurs. Interestingly, in eccentric gaze, the proximity to one eye causes an overall magnification of the image in the nearer eye, and this could be sufficient to allow observers to make accurate frontoparallel settings in eccentric gaze (Fig. 4). Comparison of vertical and horizontal magnification seems to be sufficient to disambiguate tilt from eccentric viewing.



**FIGURE 4** The horopter. The theoretical horopter or Vieth-Muller circle passes through the fixation point and two entrance pupils of the eyes. The theoretical vertical horopter is a vertical line passing through the fixation point in the midsagittal plane. Fixation on a tertiary point in space produces vertical disparities because of greater proximity to the ipsilateral eye. (Reprinted by permission from Tyler and Scott.<sup>21</sup>)

Flattening of the horopter from a circle to an ellipse results from nonuniform magnification of one retinal image. If the empirical horopter is flatter than the theoretical horopter, corresponding retinal points are more distant from the fovea on the nasal than temporal hemiretina. Curvature changes in the horopter can be produced with nonuniform magnifiers such as prisms (Fig. 5). A prism magnifies more at its apex than its base. If base-out prism is placed before the two eyes, the right half of the image is magnified more in the left eye than in the right eye and the left half of the image is magnified more in the right eye than in the left eye. This causes flat surfaces to appear more concave and the horopter to be less curved or more convex.

**The Vertical Horopter** The theoretical vertical point horopter for a finite viewing distance is limited by the locus of points in space where visual directions from corresponding points will intersect real objects.<sup>21</sup> These points are described by a vertical line in the midsagittal plane that passes through the Vieth-Muller circle. Eccentric object points in tertiary gaze (points with both azimuth and elevation) lie closer to one eye than the other eye. Because they are imaged at different vertical eccentricities from the two foveas, tertiary object points cannot be imaged on theoretically corresponding retinal points. However, all object points at an infinite viewing distance can be imaged on corresponding retinal regions and at this infinite viewing distance, the vertical horopter becomes a plane.



**FIGURE 5** Figure showing nonuniform magnification of a prism. Disparateness of retinal images, producing stereopsis.

The empirical vertical horopter is declinated (top slanted away from the observer) in comparison to the theoretical horopter. Helmholtz<sup>23</sup> reasoned that this was because of a horizontal shear of the two retinal images which causes a real vertical plane to appear inclined toward the observer. Optical infinity is the only viewing distance that the vertical horopter becomes a plane. It is always a vertical line at finite viewing distances. Targets that lie away from the midsagittal plane always subtend a vertical disparity due to their unequal proximity and retinal image magnification in the two eyes. The pattern of vertical disparity varies systematically with viewing distance and eccentricity from the midsagittal plane. For a given target height, vertical disparity increases with horizontal eccentricity from the midsagittal plane and decreases with viewing distance. Thus a horizontally extended target of constant height will produce a vertical disparity gradient that increases with eccentricity. The gradient will be greater at near than far viewing distances. It is possible to estimate viewing distance from the vertical disparity gradient or from vertical disparity at a given point if target eccentricity is known. This could provide a retinal source of information about viewing distance to allow the visual system to scale disparity to depth and to compute depth ordering. Several investigators have shown that modifying vertical disparity can influence the magnitude of depth (scaling) and surface slant, such that vertical disparity is a useful source of information about viewing distance that can be used to scale disparity and determine depth ordering.

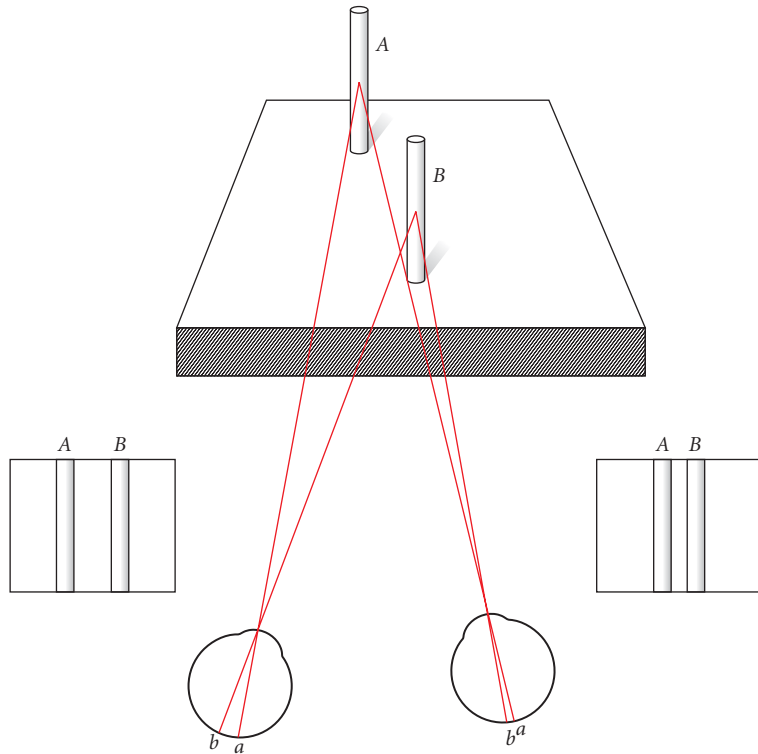
## Stereopsis

Three independent variables involved in the calculation of stereodepth are retinal image disparity, viewing distance, and the separation in space of the two viewpoints (i.e., the baseline or interpupillary distance) (Fig. 6). In stereopsis, the relationship between the linear depth interval between two objects and the retinal image disparity that they subtend is approximated by the following expression

$$\Delta d = \frac{\eta \times d^2}{2a}$$

where  $\eta$  is retinal image disparity in radians,  $d$  is viewing distance,  $2a$  is the interpupillary distance and  $\Delta d$  is the linear depth interval.  $2a$ ,  $d$ , and  $\Delta d$  are all expressed in the same units (e.g., meters). The formula implies that in order to perceive depth in units of absolute distance (e.g., meters), the visual system utilizes information about the interpupillary distance and the viewing distance. Viewing distance could be sensed from the angle of convergence<sup>24</sup> or from other retinal cues such as oblique or vertical disparities. These disparities occur naturally with targets in tertiary directions from the point of fixation.<sup>25–30</sup>

The equation illustrates that for a fixed retinal image disparity, the corresponding linear depth interval increases with the square of viewing distance and that viewing distance is used to scale the



**FIGURE 6** The differences in the perspective views of the two eyes produce binocular that can be used to perceive depth.



horizontal disparity into a linear depth interval. When objects are viewed through base-out prisms that stimulate additional convergence, perceived depth should be reduced by underestimates of viewing distance. Furthermore, the pattern of zero retinal image disparities described by the curvature of the longitudinal horopter varies with viewing distance. It can be concave at near distances and convex at far distances in the same observer.<sup>22</sup> Thus, without distance information, the pattern of retinal image disparities across the visual field is insufficient to sense either depth ordering (surface curvature) or depth magnitude.<sup>25</sup> Similarly, the same pattern of horizontal disparity can correspond to different slants about a vertical axis presented at various horizontal gaze eccentricities.<sup>22</sup> Convergence distance and direction of gaze are important sources of information used to interpret slant from disparity fields associated with slanting surfaces.<sup>31</sup> Clearly, stereodepth perception is much more than a disparity map of the visual field.

## Binocular Fusion and Suppression

Even though there is a fairly precise point-to-point correspondence between the two eyes for determining depth in the fixation plane, images in the two eyes can be combined into a single percept when they are anywhere within a small range or retinal area around corresponding retinal points. Thus a point in one eye can be combined perceptually with a point imaged within a small area around its corresponding retinal location in the other eye. These ranges are referred to as Panum's fusional area (PFA) and they serve as a buffer zone to eliminate diplopia for small disparities near the horopter. PFA allows for the persistence of single binocular vision in the presence of constant changes in retinal image disparity caused by various oculomotor disturbances. For example, considerable errors of binocular alignment ( $>15$  arc min) may occur during eye tracking of dynamic depth produced either by object motion or by head and body movements.<sup>32</sup> Stereopsis could exist without singleness but the double images near the fixation plane would be a distraction. The depth of focus of the human eye serves a similar function. Objects that are nearly conjugate to the retina appear as clear as objects focused precisely on the retina. The buffer for the optics of the eye is much larger than the buffer for binocular fusion. The depth of focus of the eye is approximately 0.75 D. Panum's area, expressed in equivalent units, is only 0.08 meter angles or approximately one tenth the magnitude of the depth of focus. Thus we are more tolerant of focus errors than we are of convergence errors.

Fusion is only possible when images have similar size, shape, and contrast polarity. This similarity ensures that we only combine images that belong to the same object in space. Natural scenes contain many objects that are at a wide range of distances from the plane of fixation. Some information is coherent, such as images formed within Panum's fusional areas. Some information is fragmented, such as partially occluded regions of space resulting in visibility to only one eye. Finally, some information is uncorrelated because it is either ambiguous or in conflict with other information, such as the superposition of separate diplopic images arising from objects seen by both eyes behind or in front of the plane of fixation. One objective of the visual system is to preserve as much information from all three sources as possible to make inferences about objective space without introducing ambiguity or confusion of space perception. In some circumstances conflicts between the two eyes are so great that conflicting percepts are seen alternately every 4 s (binocular rivalry suppression) or in some cases one image is permanently suppressed such as when you look through a telescope with one eye while the other remains open. The view through the telescope tends to dominate over the view of the background seen by the other eye. For example, the two ocular images may have unequal clarity or blur such as in asymmetric convergence, or large unfusible disparities originating from targets behind or in front of the fixation plane may appear overlapped with other large diplopic images. Fortunately, when dissimilar images are formed within a fusible range of disparity that perception of the conflicting images is suppressed.

Four classes of stimuli evoke what appear to be different mechanisms of interocular suppression. The first is unequal contrast or blur of the two retinal images which causes interocular blur suppression. The second is physiologically diplopic images of targets in front or behind the singleness horopter which result in suspension of one of the redundant images.<sup>33</sup> The third is targets of different shape

presented in identical visual directions. Different size and shape result an alternating appearance of the two images referred to as either binocular retinal rivalry or percept rivalry suppression. The fourth is partial occlusion that obstructs the view of one eye such that the portions of the background are only seen by the unoccluded eye and the overlapping region of the occluder that is imaged in the other eye is permanently suppressed.

### 13.3 DISTORTION OF SPACE BY MONOCULAR MAGNIFICATION

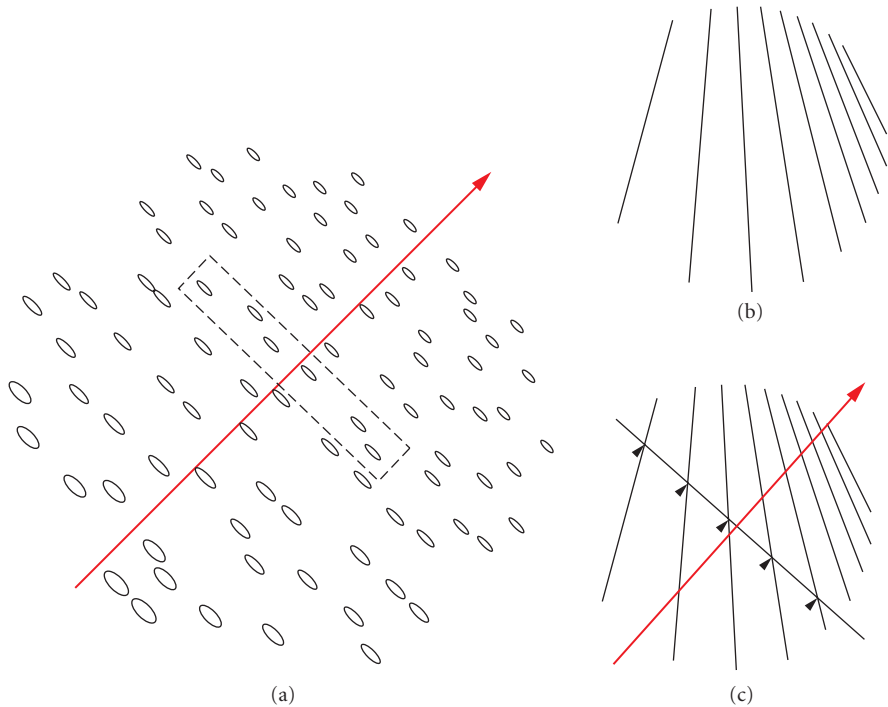
The magnification and translation of images caused by optical aids will distort certain cues used for space perception while leaving other cues unaffected. This will result in cue misrepresentation of space as well as cue conflicts that will produce errors of several features of space perception. These include percepts of direction, distance, trajectory or path of external moving targets, heading and self-motion estimates, as well as shape and orientation (curvature, slant, and tilt) of external objects. Fortunately, if the optical distortions are long standing, the visual system can adapt and make corrections to their contributions to space perception.

#### Magnification and Perspective Distortion

Most optical systems magnify or minify images, and this can produce errors in perceived direction. When objects are viewed through a point other than the optical center of a lens, they appear displaced from their true direction. This is referred to as a prismatic displacement. Magnification will influence most of the monocular cues mentioned above except overlap. Magnification will produce errors in perceived distance and produce conflicts with cues of texture density and linear perspective. For example, the retinal image of a circle lying on a horizontal ground plane has an elliptical shape when viewed at a remote distance, and a more circular shape when viewed from a shorter distance. When a remote circle is viewed through a telescope, the uniformly magnified image appears to be closer and squashed in the vertical direction to form an ellipse (shape distortion). Uniform magnification also distorts texture-spacing gradients such that the ground plane containing the texture is perceived as inclined or tilted upward toward vertical. Normally, the change in texture gradient is greatest for near objects (Fig. 7). When a distant object is magnified, it appears nearer but its low texture density gradient is consistent with a plane inclined toward the vertical. Magnification also affects slant derived from perspective cues in the same way. For this reason, a tiled rooftop appears to have a steeper pitch when viewed through binoculars or a telescope, the pitcher-to-batter distance appears reduced in telephoto view from the outfield, and the spectators may appear larger than the players in front of them in telephoto lens shots.

#### Magnification and Stereopsis

Magnification increases both retinal image disparity and image size. The increased disparity stimulates greater stereo depth while increased size stimulates reduced perceived viewing distance.<sup>34</sup> Greater disparity will increase perceived depth while reduced perceived viewing distance will scale depth to a smaller value. Because stereoscopic depth from disparity varies with the square of viewing distance, the perceptual reduction of viewing distance will dominate the influence of magnification on stereoscopic depth. Thus, perceived depth intervals sensed stereoscopically from retinal image disparity appears smaller with magnified images because depth derived from binocular retinal image disparity is scaled by the square of perceived viewing distance. Objects appear as flat surfaces with relative depth but without thickness. This perceptual distortion is referred to as cardboarding. Most binocular optical systems compensate for this reduction in stereopsis by increasing the baseline or separation of the two ocular objectives. For example, the prism design of binoculars folds the optical path into



**FIGURE 7** Texture gradients. The tilt of a surface is the direction in which it is slanted away from the gaze normal of the observer. (a) If the surface bears a uniform texture, the projection of the axis of tilt in the image indicates the direction in which the local density of the textures varies most, or, equivalently, it is perpendicular to the direction that the texture elements are most uniformly distributed (dotted rectangle). Either technique can be used to recover the tilt axis, as illustrated. Interestingly, however, the tilt axis in situations like (b) can probably be recovered most accurately using the second method, i.e., searching for the line that is intersected by the perspective lines at equal intervals. This method is illustrated in (c). (Reprinted by permission from Stevens.<sup>35</sup>)

a compact package and also expands the distance between the objective lenses creating an expanded interpupillary distance (telestereoscope). This increase of effective separation between the two eyes' views exaggerates disparity and objects can appear in hyperstereoscopic depth if it exceeds the depth reduction caused by perceived reductions in viewing distance. The depth can be predicted from the formula in the preceding section that computes disparity from viewing distance, interpupillary distance, and physical linear depth interval.

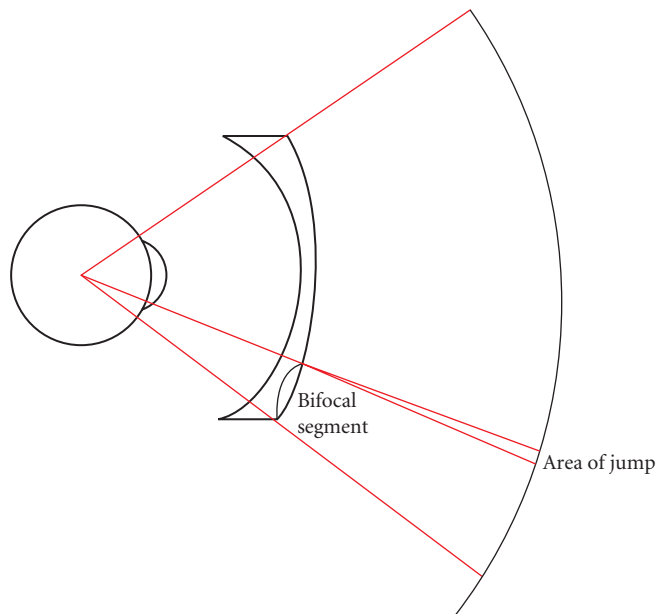
### Magnification and Motion Parallax—Heading

Magnification will produce errors in percepts derived from motion cues. Motion will be exaggerated by magnification and perceived distance will be shortened. As with binocular disparity, depth associated with motion parallax will be underestimated by the influence of magnification on perceived distance. Heading judgments in which gaze is directed away from the heading direction will be in error by the amount the angle of eccentric gaze is magnified. Thus if your car is headed north and your gaze is shifted  $5^\circ$  to the left, heading errors will be the percentage the  $5^\circ$  gaze shift is magnified by the optical system and you will sense a path that is northeast. Heading judgments that are constantly

changing, such as along a curved path, might be affected in a similar way, if the gaze lags behind the changing path of the vehicle. In this case, changes of perceived heading along the curved path would be exaggerated by the magnifier. Unequal magnification of the two retinal images in anisometropia will cause errors in sensing the direction of motion in depth by changing the amplitude ratio of the two moving images.

## Bifocal Jump

Most bifocal corrections have separate optical centers for the distance lens and the near addition. One consequence is that objects appear to change their direction or jump as the visual axis crosses the top line of the bifocal segment (Fig. 8). This bifocal jump occurs because prism is introduced as the eye rotates downward behind the lens away from the center of the distance correction and enters the bifocal segment. When the visual axis enters the optical zone of the bifocal, a new prismatic displacement is produced by the displacement of the visual axis from the optical center of the bifocal addition. All bifocal additions are positive so that the prismatic effect is always the same; objects in the upper part of the bifocal appear displaced upward compared to views just above the bifocal. This results in bifocal jump. A consequence of bifocal jump is that a region of space seen near the upper edge of the bifocal is not visible. It has been displaced out of the field of view by the vertical prismatic effect of the bifocal. The invisible or missing part of the visual field equals the prismatic jump which can be calculated from Prentice's rule [distance (cm) between the upper edge and center of the bifocal times the power of the add]. Thus for a 2.5-D bifocal add with a top to center distance of 0.8 cm, approximately  $1^\circ$  of the visual field is obscured by bifocal jump. This loss may seem minor; however, it is very conspicuous when paying attention to the ground plane while walking. There is a clear horizontal line of discontinuity that is exaggerated by the unequal speed of optic flow above and below the bifocal boundary. Optic flow is exaggerated by the magnification of the positive bifocal



**FIGURE 8** Jump figure. The missing region of the visual field occurs at the top edge of the bifocal segment as a result of prism jump.

segment. It can produce errors of estimated walking speed and some confusion if you attend to the ground plane in the lower field while navigating toward objects in the upper field. This can be a difficult problem when performing such tasks as walking down steps. It is best dealt with by maintaining the gaze well above the bifocal segment while walking, which means that the user is not looking at the ground plane in the immediate vicinity of the feet. Some bifocal designs eliminate jump by making the pole of the bifocal and distance portion of the lens concentric. The bifocal line starts at the optical centers of both lenses. This correction always produces some vertical displacement of the visual field since the eyes rarely look through the optical center of the lens; however, there is no jump.

## Discrepant Views of Objects and Their Images

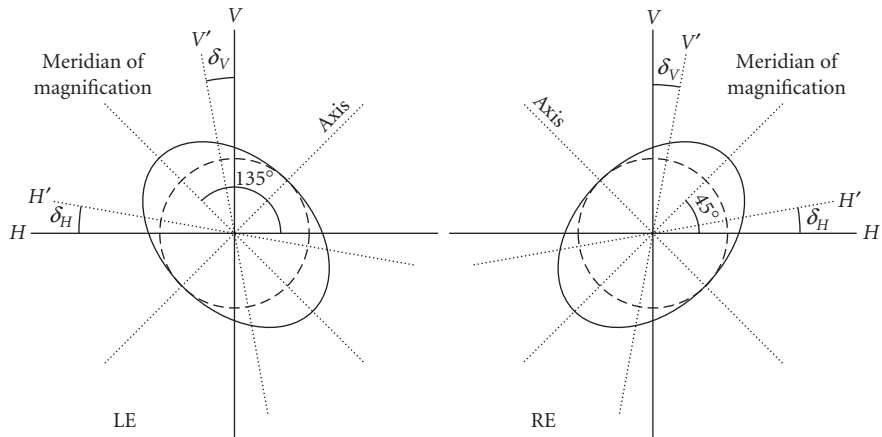
Translation errors in viewing a scene can lead to errors in space perception. A common problem related to this topic is the interpretation of distance and orientation of objects seen on video displays or in photographs. The problem arises from discrepancies between the camera's vantage point with respect to a 3-D scene, and the station point of the observer who is viewing the 2-D projected image after it has been transformed by the optics of the camera lens. Errors of observer distance and tilt of the picture plane contribute to perceptual errors of perspective distortion, magnification, and texture density gradients expected from the station point. The station point represents the position of the observer with respect to the view plane or projected images. To see the picture in the correct perspective, the station point should be the same as the location of the camera lens with respect to the film plane. Two violations of these requirements are errors in viewing distance and translation errors (i.e., horizontal or vertical displacements). Tilt of the view screen can be decomposed into errors of distance and translation errors. In the case of the photograph, the correct distance for the station point is the camera lens focal length multiplied by the enlargement scaling factor. Thus, when a 35-mm negative taken through a 55-mm lens is enlarged 7 times to a 10-in print, it should be viewed at a distance of 38.5 cm or 15 in ( $7 \times 5.5$  cm).

## 13.4 DISTORTION OF SPACE PERCEPTION FROM INTEROCULAR ANISO-MAGNIFICATION (UNEQUAL BINOCULAR MAGNIFICATION)

---

Optical devices that magnify images unequally or translate images unequally in the two eyes produce errors of stereoscopic depth and perceived direction as well as eye alignment errors that the oculomotor system is not accustomed to correcting. If these sensory and motor errors are longstanding, the visual system adapts perceptual and motor responses to restore normal visual performance. Patients who receive optical corrections that have unequal binocular magnification will adapt space percepts in 10 days to 2 weeks.

**Lenses and Prisms** Magnification produced by lenses is uniform compared to the nonuniform magnification produced by prisms, where the magnification increases toward the apex of the prism. Uniform magnification differences between the horizontal meridians of the two eyes produce errors in perceived surface tilt about a vertical axis (geometric effect). A frontoparallel surface appears to face the eye with the more magnified image. This effect is offset by vertical magnification of one ocular image that causes a frontoparallel surface to appear to face the eye with the less magnified image. Thus perceptual errors of surface tilt occur mainly when magnification is greater in the horizontal or vertical meridian. These meridional magnifiers are produced by cylindrical corrections for astigmatism. Nonuniform magnification caused by horizontal prism such as base-out or base-in (Fig. 5), cause surfaces to appear more concave or convex, respectively. Finally, cylindrical corrections for astigmatism cause scissor or rotational distortion of line orientations (Fig. 9). When the axes of astigmatism are not parallel in the two eyes, cyclodisparities are introduced. These disparities



**FIGURE 9** Magnification ellipse. Meridional magnification of images in the two eyes in oblique meridians cause “scissors” or rotary deviations of images of vertical (and horizontal) lines and affect stereoscopic spatial localization in a characteristic manner. (Reprinted from Ogle and Boder.<sup>36</sup>)

produce slant errors and cause surfaces to appear inclined or declined. These spatial depth distortions are adapted to readily so that space appears veridical; however, some people are unable to cope with the size differences of the two eyes and suffer from an anomaly referred to as aniseikonia. While aniseikonic errors are tolerated by many individuals, they are subject of spatial distortions which, in given situations, may cause meaningful problems.

## Aniseikonia

Unequal size of the two retinal images in anisometropia can precipitate a pathological mismatch in perceived size of the two ocular images (aniseikonia). Aniseikonia is defined as a relative difference in perceived size and/or shape of the two ocular images. Most difficulty with aniseikonia occurs when anisometropia is corrected optically with spectacles. Patients experience a broad range of symptoms including spatial disorientation, depth distortion, diplopia, vertigo, and asthenopia. These symptoms result from several disturbances including spatial distortion and interference with binocular sensory and motor fusion.

Two optical distortions produced by spectacle corrections of anisometropia are called axial and lateral aniseikonia. Axial aniseikonia describes size distortions produced by differences in the magnification at the optical centers of ophthalmic spectacle corrections for anisometropia. The magnification difference produced by the spectacle correction depends in part on the origin of the refractive error. Retinal images in the uncorrected axial and refractive ametropias are not the same. In axial myopia, the uncorrected retinal image of the elongated eye is larger than that of an emmetropic eye of similar refractive power. An uncorrected refractive myope has a blurred image of similar size to an emmetropic eye with the same axial length. Knapp’s law states that image size in an axial ametropic eye can be changed to equal the image size in an emmetropic eye having the same refractive power by placing the ophthalmic correction at the anterior focal plane of the eye. The anterior focal plane of the eye is usually considered to correspond approximately to the spectacle plane. However, this is only a very general approximation. The anterior focal point will vary with power of the eye and in anisometropia, it will be unequal for the two eyes. Assuming the anterior focal point is the spectacle plane, then the abnormally large image of an uncorrected axial-myopic eye can be reduced to near normal with a spectacle correction, whereas a contact lens correction would leave the retinal image of the myopic eye enlarged. If the ametropia is refractive in nature (i.e., not axial), then the uncorrected

retinal images are equal in size and a contact lens correction will produce less change in image size than a spectacle correction. If the refractive myope is corrected with spectacles, the retinal image becomes smaller than that in an emmetropic eye of similar refractive power. This change in image size of refractive anisometropes corrected with spectacle lenses is one of several factors leading to ocular discomfort in aniseikonia.

Paradoxically, spectacle correction of axial myopes often produces too drastic a reduction in ocular image size. The ocular image of the axially myopic eye, corrected with spectacle lenses, appears smaller than that of the less ametropic eye in anisometropia.<sup>37-42</sup> As mentioned earlier, ocular image size depends not only upon retinal image size but also upon the neural factor.<sup>22</sup> For example, in cases of axial myopia, the elongation of the eye produces stretching of the retina and pigment epithelium. Accordingly, retinal images of normal size occupy a smaller proportion of the stretched retina than they would in a retina with normal retinal element density. Indeed, anisometropes with axial myopia report minification of the ocular image in their more myopic eye after the retinal images have been matched in size.<sup>41,43,44</sup> Thus, in the uncorrected state, myopic anisometropic individuals have optical magnification and neural minification. As a result, a spectacle correction would overcompensate for differential size of ocular images.

Surprisingly, not all patients with a significant amount of aniseikonia (4 to 6%) complain of symptoms. This could be due in part to suppression of one ocular image. However, normally we are able to fuse large size differences up to 15 percent of extended targets.<sup>45</sup> Sinusoidal gratings in the range from 1 to 5 cpd can be fused with grating differing in spatial frequency by 20 to 40 percent<sup>46-48</sup> and narrow band pass filtered bars can be fused with 100- to 400-percent difference in interocular size.<sup>49,50</sup> These enormous size differences demonstrate the remarkable ability of the visual system to fuse large size differences in axial aniseikonia. However, in practice, it is difficult to sustain fusion and stereopsis with greater than 8- to 10-percent binocular image size difference.

With extended targets, magnification produces an additional lateral displacement due to cumulative effects such that nonaxial images are enlarged and imaged on noncorresponding points. Given our tolerances for axial disparity gradients in size, it appears that most of the difficulty encountered in aniseikonia is lateral. Von Rohr<sup>51</sup> and Erggelet<sup>52</sup> considered this prismatic effect as the most important cause of symptomatic discomfort in optically induced aniseikonia.

Differential magnification of the two eyes images can be described as a single constant; however, the magnitude of binocular-position disparity increases proportionally with the eccentricity that an object is viewed from the center of a spectacle lens. Prentice's rule approximates this prismatic displacement from the product of distance from the optical center of the lens in centimeters and the dioptric power of the lens. In cases of anisometropia corrected with spectacle lenses, this prismatic effect produces binocular disparity along the meridian of eccentric gaze. When these disparities are horizontal they result in distortions of stereoscopic depth localization, and when they are vertical, they can produce diplopia and eye strain. Fortunately, there is an increasing tolerance for disparity in the periphery provided by an increase in the binocular sensory fusion range with retinal eccentricity.<sup>22</sup> In cases of high anisometropia, such as encountered in monocular aphakia, spectacle corrections can cause diplopia near the central optical zone of the correction lens. In bilateral aphakia, wearing glasses or contact lenses does not interfere with the normal motor or sensory fusion range to horizontal disparity; however, stereothresholds are elevated.<sup>53</sup> In unilateral aphakia, contact lenses and intraocular lens implants support a normal motor fusion range; however, stereopsis is far superior with the interocular lens implants.<sup>53</sup> Tolerance of size difference errors varies greatly with individuals and their work tasks. Some may tolerate 6 to 8 percent errors while others may complain with only 0.25- to 0.5-percent errors.

## Interocular Blur Suppression with Anisometropia

There is a wide variety of natural conditions that present the eyes with unequal image contrast. These conditions include naturally occurring anisometropia, unequal amplitudes of accommodation, and asymmetric convergence on targets that are closer to one eye than the other. This blur can be eliminated in part by a limited degree of differential accommodation of the two eyes<sup>54</sup> and by



interocular suppression of the blur. The latter mechanism is particularly helpful for a type of contact lens patient who can no longer accommodate (presbyopes) and prefer to wear a near contact lens correction over one eye and a far correction over the other (monovision) rather than wearing bifocal spectacles. For most people (66%), all of these conditions result in clear, nonblurred, binocular percepts with a retention of stereopsis<sup>55</sup> albeit with the stereothreshold elevated by approximately a factor of two. Interocular blur suppression is reduced for high contrast targets composed of high spatial frequencies.<sup>55</sup> There is an interaction between interocular blur suppression and binocular rivalry suppression. Measures of binocular rivalry reveal a form of eye dominance defined as the eye that is suppressed least when viewing dichoptic forms of different shape. When the dominant eye for rivalry and aiming or sighting is the same, interocular suppression is more effective than when dominance for sighting and rivalry are crossed (i.e., in different eyes).<sup>56</sup>

When the contrast of the two eyes' stimuli is reduced unequally, stereoacuity is reduced more than if the contrast of both targets is reduced equally. Lowering the contrast in one eye reduces stereo acuity twice as much as when contrast is lowered in both eyes. This contrast paradox only occurs for stereopsis and is not observed for other disparity-based phenomenon including binocular sensory fusion. The contrast paradox occurs mainly with low spatial frequency stimuli (<2.0 cycles/degree). Contrast reduction caused by lens defocus is greater at high spatial frequencies, and perhaps this is why stereopsis is able to persist in the presence of interocular blur caused by anisometropia.

### 13.5 DISTORTIONS OF SPACE FROM CONVERGENCE RESPONSES TO PRISM

Horizontal vergence responses to prism influence estimates of viewing distance.<sup>24,57–59</sup> The resulting errors in perceived distance will influence depth percepts that are scaled by perceived viewing distance such as stereopsis and motion parallax. Time of arrival or contact with external objects can be estimated from the distance to contact divided by the speed of approach. If distance is underestimated while wearing convergent or base-out (temporal) prism, then time to contact, computed from velocity and distance, will also be underestimated. In addition, prism stimuli to convergence will produce conflicts in distance estimates from vergence and retinal image size. When vergence responds to prism, retinal image size remains constant instead of increasing as it would in response to real changes in viewing distance. The conflict is resolved by a perceived decrease in perceived size. This illusion is referred to as convergence micropsia. An average magnitude of shrinkage is 25 percent for an accommodation and convergence increase equivalent to 5 D or 5 meter angles.<sup>60</sup> Very large prism stimuli to vergence present major conflicts between the fixed stimulus to accommodation and the new vergence stimulus. The oculomotor system compromises by partially responding to the vergence stimulus. If physical size of a target in a virtual display has been adjusted to correspond to an expected full vergence response, (increase retinal image size with convergence stimuli), but a smaller vergence response actually occurs, the perceptual compensation for size is based on the smaller vergence response. As a result, a paradoxical increase of perceived size accompanies convergence (convergence macropsia).<sup>61–63</sup> The incompletely fused images are perceived at the distance of convergence alignment of the eyes rather than at the intended distance specified by physical size and disparity. Thus, aniseikonic and anisophoric differences in the two eyes, which are made worse by anisometropia, can cause difficulty when performing binocular tasks associated with the use of optical instruments.

### 13.6 EYE MOVEMENTS

There are two general categories of eye movements. One class stabilizes the motion of the retinal image generated by our own body and head movements and the other class stabilizes the retinal image of an object moving in space. In the former case, the whole retinal image is moving in unison



whereas in the latter case only the retinal image of the moving object moves relative to the stationary background and when that image is tracked it remains stabilized on the fovea during target inspection. In all, there are five types of eye movements (VOR, OKN, saccades, pursuits, and vergence) and they are classified as either stabilizing or tracking movements.

Two classes of reflex eye movements perform the image stabilization task. The first class of stabilizing eye movement compensates for brief head and body rotation and is called the vestibulo-ocular reflex (VOR). During head movements in any direction, the semicircular canals of the vestibular labyrinth signal how fast the head is rotating and the oculomotor system responds reflexively to this signal by rotating the eyes in an equal and opposite rotation. During head rotation, this reflex keeps the line of sight fixed in space and visual images stationary on the retina. Body rotation in darkness causes a repetitious sequence of eye movements which move the line of sight slowly in the direction opposite to head rotation (slow phase) followed by a rapid (saccadic) movement in the direction the head moves (fast phase). This reflex is almost always active and without it we would be unable to see much of anything due to constant smear of the retinal image. The vestibulo-ocular reflex is extremely accurate for active head motion because it has an adaptive plasticity that makes it easily calibrated in response to altered visual feedback.<sup>64,65</sup> The VOR may be inhibited or suppressed by the visual fixation mechanism<sup>66</sup> as a result of the system's attempt to recalibrate to the head-stationary image.

Optokinetic reflex is an ocular following response that stabilizes whole field motion when we move about in a visual scene. Unlike the vestibulo-ocular reflex, this optokinetic reflex requires a visible retinal image whereas the vestibulo-ocular reflex operates in total darkness. Rotation or translation of the visual field causes the eyes to execute a slow following and fast reset phase sequence of eye movements that have the same pattern as the VOR. The optokinetic reflex supplements the vestibulo-ocular reflex in several ways. The vestibulo-ocular reflex responds to acceleration and deceleration of the head but not to constant velocity. In contrast, optokinetic reflex responds to constant retinal image velocity caused by constant body rotation or translation. The vestibulo-ocular reflex controls initial image stabilization and optokinetic reflex maintains the stabilization.

Tracking eye movements place the retinal image of the object of regard on the fovea and keep it there, even if the object moves in space. The class of eye movements that places the image on the fovea is called a saccade. This is a very fast eye movement that shifts the image in a step-like motion. Saccades can be made voluntarily, in response to visual, tactile, and auditory stimuli, and even in darkness to willed directions of gaze. Their chief characteristic is they are fast, reaching velocities of nearly 1000 deg/s. Once the target is on the fovea, slow following eye movements called pursuits maintain the image of the target on the fovea. They can keep images stabilized that are formed of objects moving as fast as 30 deg/s in humans. Generally, pursuits cannot be made voluntarily in the absence of a visual moving stimulus. However, they can respond to a moving sound or tactile sensation in darkness.

---

## 13.7 COORDINATION AND ALIGNMENT OF THE TWO EYES

---

If you observe the motion of the two eyes, you will notice that they either move in the same or opposite direction. When they move in the same direction they are said to be yoked, or conjugate, like a team of horses. These are also called version movements. At other times the eyes move in opposite directions, especially when we converge from far to near. These are called disjunctive or vergence movements. Both conjugate and disjunctive movements can be either fast or slow depending if the eyes are shifting fixation or maintaining foveation (stable fixation) of a target. A century ago, Ewald Hering described the movements of the two eyes as equal and symmetrical (Hering's law). Our eyes are yoked when we are born but the yoking can be modified in response to anisometric spectacles that produce unequal image sizes and require us to make unequal eye movements to see bifoveally. This image inequality leads to the binocular anomalies of aniseikonia (unequal perceived image size) and anisophoria (noncomitant phoria or a phoria that varies with direction of gaze).

Targets are brought into alignment with corresponding points by movements of the eyes in opposite directions. These vergence movements have 3° of freedom (horizontal, vertical, and torsional). Horizontal vergence can rotate the eyes temporalward (divergence) or nasalward (convergence). Since near targets only require convergence of the eyes, there is a limited need for divergence other than to overcome postural errors of the eyes. Consequently, the range of divergence is less than half the range of convergence. The eyes are able to diverge approximately 4° and converge approximately 8 to 12° in response to disparity while focusing accurately on a target at a fixed viewing distance. Vergence movements serve the same function as conjugate movements of the eyes which is to preserve the alignment of the two retinal images on corresponding retinal points. Vergence is required when changing viewing distance, whereas versions accomplish the task when changing direction of gaze.

The stimulus to vergence depends upon both the viewing distance and the distance between the two eyes (interpupillary distance). It is equal to the angle subtended by a target in space at the entrance pupils of the two eyes (binocular parallax). As will be described in Sec. 13.9, "Prism-Induced Errors of Eye Alignment," the stimulus to vergence can be changed by stimulating accommodation with lenses, or changing disparity (binocular parallax) with prisms. The exact stimulus value of lenses and prisms to vergence depends upon the placement of the prism relative to the eye and this relationship will be described in Sec. 13.8, "Effects of Lenses and Prism on Vergence and Phoria." Vergence is measured in degrees of ocular rotation and prism diopters. The meter angle is also used in order to get a comparable metric for accommodation and convergence. Meter angles equal the reciprocal of the viewing distance in meters, and the value is approximately equal to the accommodative stimulus in diopters.

$$MA = \frac{1}{\text{target distance in meters}}$$

Meter angles (MA) can be converted into prism diopters simply by multiplying MA by the interpupillary distance (IPD) in cm. Prism diopters equal the 100 tan vergence angle in degrees. Vergence eye movements respond to four classes of stimuli described nearly a century ago by Maddox.

## Intrinsic Stimuli to Vergence

There are four classes of stimuli to align the eyes with horizontal vergence at various viewing distances. The first is an intrinsic innervation that develops postnatally. During the first 6 weeks of life, our eyes diverge during sleep. This divergence posture is known as the anatomical position of rest and it represents the vergence angle in the absence of any innervation. As we develop during the first 1.5 months of life, the resting vergence angle is reduced to nearly zero by the development of tonic vergence which reduces the resting position from the anatomical position of rest to the physiological position of rest. The anatomical position of rest refers to eye alignment in the absence of all innervation, such as in deep anesthesia. The physiological position of rest refers to the angle that vergence assumes when we are awake and alert, in the absence of any stimulus to accommodation or binocular fusion or perceived distance. Usually, this physiological position of rest differs from the vergence demand, (i.e., the amount of vergence needed to fuse a distance target). The discrepancy between the vergence demand at a far viewing distance and the physiological position of rest is called the phoria. The phoria is an error of binocular alignment, but the error is only manifest during monocular viewing or any condition that disrupts binocular vision. If binocular vision is allowed (not disrupted) other components of vergence reduce the error nearly to zero. Adaptation of the phoria continually calibrates eye alignment and keeps the phoria constant throughout life.

There are two components of phoria adaptation that compensate for eye misalignment. One is a conjugate process that is called prism adaptation and it refers to our ability to adjust our phoria in response to a uniform prism. This prism is constant throughout the visual field and it produces a conjugate change in phoria in all directions of gaze. The conjugate adaptation is very rapid and can be completed in several minutes in a normal patient. The eyes are able to adapt horizontal vergence several degrees over a 1 h time span and larger amounts for unlimited adaptation periods. The slow-anomalous fusional movements reported by Bagolini<sup>67</sup> for strabismus patients are similar in time course to adaptive responses to prism and they indicate that the potential range of adaptation must exceed 10°.

The second component of phoria adaptation is a nonconjugate process that adjusts the phoria separately for each direction and distance of gaze. For example, anisometric spectacle corrections disrupt the normal alignment pattern needed for fusion. A person wearing a plus (converging) lens over the right eye needs to over converge in left gaze and over diverge in right gaze. Our oculomotor system is able to make these noncomitant adjustments such that the phoria varies with direction and distance of gaze consistent to compensate for the problems encountered with anisometric spectacle corrections. This disconjugate adaptation occurs in a matter of hours and it persists when the spectacle lenses are removed. It provides a means to align the eyes without retinal cues of binocular disparity. The adapted variations of phoria are guided by a synergistic link with the eye positions in which they were learned. For example when an anisometric person first wears a plus lens spectacle over the right eye, the right eye is hypophoric in upgaze and hyperphoric in downgaze. After wearing this spectacle correction for several hours the vergence errors are adapted to so that there is no longer a phoria or vergence error while wearing the spectacles. However, when the spectacles are removed, the right eye is hyperphoric in upgaze and hypophoric in downgaze. This demonstrates a synergistic link between vertical eye position and vertical phoria where eye position guides the vertical vergence alignment of the two eyes, even when only one eye is open and the other is occluded.

## Binocular Parallax

Horizontal vergence also responds to misalignment of similar images from corresponding retinal points (retinal image disparity). It reduces the disparity to a value within Panum's fusional area and causes images to appear fused or single. This fusional or disparity vergence is analogous to optical reflex accommodation. It serves as a rapid fine adjustment mechanism that holds or maintains fixation of a given target and corrects for minor changes of disparity caused by small movements of the head or the target being inspected. It has a range of approximately  $3^\circ$  of convergence and divergence. Rapid adaptations of the phoria account for increased ranges of fusion in response to disparity.

Vertical and cyclovergence eye movements also respond to binocular disparity. Vertical vergence can rotate one eye up or down relative to the other eye (skew movements) over a range of  $2^\circ$ . These movements respond to vertical disparity that is always associated with near targets in tertiary directions of gaze (targets at oblique eccentricities). Vertical disparity is a consequence of spatial geometry. Tertiary targets are closer to one eye than the other so that they subtend unequal image sizes in the two eyes. Unequal vertical eccentricities resulting from unequal image magnification requires a vertical vergence to align tertiary targets onto corresponding retinal regions. At near distances greater than 10 in, vertical disparities are very small and only require a limited range of movement.

Torsional or cyclovergence can twist one eye about its visual axis relative to the orientation of the other eye. Normally, torsional vergence keeps the horizontal meridians of the retinas aligned when the eyes are converged on a near target in the upper or lower visual field. Torsional vergence stimulated by cyclodisparities is limited to a range of approximately  $10^\circ$ . Like vertical vergence, torsional vergence responds to unequal orientations or torsional disparities that result from the perspective distortions of near targets viewed by the two eyes. Ocular torsion is mainly stimulated by lateral head roll and amounts to approximately 10 percent of the amount of head roll, up to a maximum of  $6^\circ$  ( $60^\circ$  head roll). The head roll can produce vertical disparities of the near fixation targets if the eyes are not accurately converged. With the head tilted toward the right shoulder, a divergence fixation error can cause the right eye to be depressed from the target and a convergence fixation error will cause the right eye to be elevated with respect to the fixation target. However, no vertical fixation errors result from head roll if the eyes are accurately converged on the fixation target.

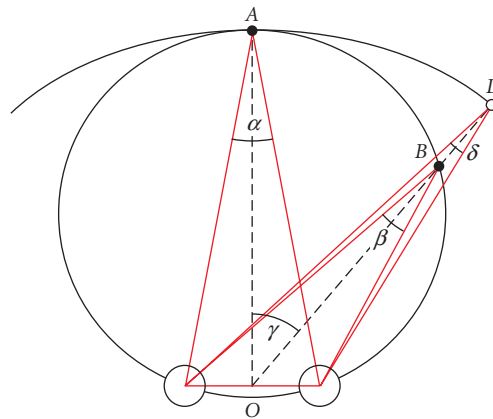
## Defocus and Efforts to Clear Vision

Stimulation of either convergence or accommodation causes synergistic changes in convergence, accommodation, and pupil size. This synergy is called the near triad. During monocular occlusion, when the eyes focus from optical infinity to 1 m (1 D) they converge approximately  $2$  to  $3^\circ$  without

disparity cues. This accommodative/vergence response brings the eyes into approximate binocular alignment, and any residual alignment errors produce binocular disparity that is corrected by fusional vergence. The activation of convergence when accommodation is stimulated while one eye is covered is called accommodative vergence. The gain or strength of this interaction is called the accommodative convergence/accommodation (AC/A) ratio which describes the amount of convergence activated per diopter of accommodation. Similarly, if convergence is stimulated while wearing pinhole pupils, accommodation also responds without blur cues. This reflex is called convergence accommodation and the gain or strength of this interaction is the convergence-accommodation/convergence (CA/C) ratio. Normally these ratios are tuned appropriately for the geometry of our eye separation so that efforts of either accommodation or convergence bring the eyes into focus and alignment at nearly the same viewing distance. Occasionally, because of uncorrected refractive errors, or large abnormal phorias, or because of optical distortions caused by instruments, there is an imbalance of accommodation and vergence causing them to respond to different distances. The mismatch may result in fatigue and associated symptoms if the visual task is extended. The resulting mismatch needs to be corrected by fusional vergence and optical reflex accommodation. If the mismatch is beyond the range of the compensation, then vision remains blurred or double.

### Variations of Cross-Coupling with Direction and Distance of Gaze

In straight-ahead gaze (zero azimuth), targets in the midsagittal plane produce equal magnitudes of disparity and defocus when they are expressed in meter angles [ $MA = 1/\text{viewing distance (m)}$ ] and diopters ( $D = 1/\text{viewing distance (m)}$ ) respectively. However, when targets are viewed in asymmetric vergence [i.e., with some combination of horizontal version (azimuth) and vergence], conflicts arise between stimuli for accommodation and convergence as a consequence of 3-D spatial geometry. Figure 10 compares a plan view of the isoaccommodation circle to the isovergence circle. The isoaccommodation circle describes the locus of points that subtend a constant average accommodation



**FIGURE 10** Diagram showing isovergence circle (large complete circle), isoaccommodation circle (shown as an arc) and various azimuth angles. The two small circles represent the right and the left eye. The point A in the figure represents the spatial location that corresponds to matched stimuli for accommodation and convergence.  $\alpha$  = vergence angle at point A (angle made by the intersection of two lines of sight),  $\beta$  = vergence angle at point B, and  $\alpha = \beta$  along the isovergence circle.  $\gamma$  = lateral gaze angle. Note that at point D on the isoaccommodation circle, the convergence and accommodation demand is lower than at point B.

stimulus to the two eyes with increasing azimuth. This circle has a radius equal to the viewing distance from the cyclopean eye to the object of regard in any horizontal direction of gaze lying in a common visual plane. The isovergence circle passes through the fixation point and two centers of eye rotation.<sup>68</sup> Its distance to the cyclopean eye varies with both viewing distance and direction of gaze (azimuth). The circle has a radius equal to the interpupillary distance (PD) divided by twice the sine of the angle of convergence ( $\theta$ ).

$$\text{Isovergence radius} = \frac{\text{PD}}{2 * \sin \theta}$$

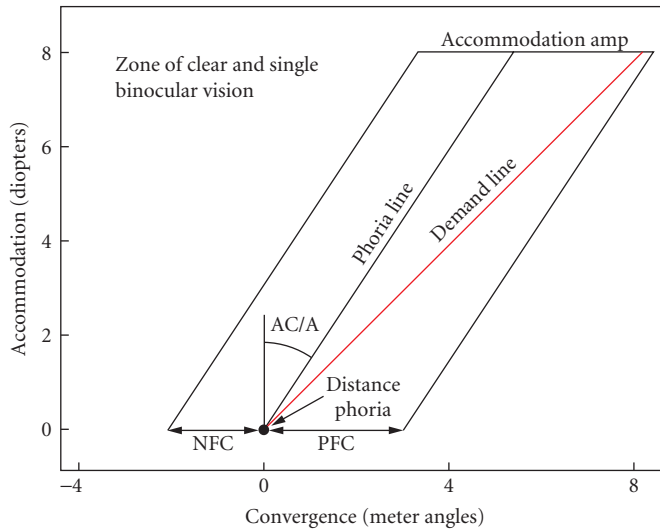
The only spatial location that corresponds to matched stimuli for accommodation and convergence is straight ahead where these two circles intersect (Fig. 10). For points lying to the left or right, the stimulus to accommodation is always greater than the stimulus to convergence. This conflict is resolved by having the normal cross-link interaction between accommodative-convergence and accommodation (AC/A ratio) set for an ideal value for 50° azimuth.<sup>69</sup> This low ratio results in a small divergence error of vergence at lesser azimuths that is easily overcome with fusional convergence. The normal convergence accommodation/convergence interaction (CA/C ratio) is ideal for straight ahead. This produces a lag of accommodation in small lag of accommodation that increases with gaze eccentricity.

## Perceived Distance

The eyes align and focus in response to perceived distance. The term proximal refers to perceived distance. Perceived distance is classified as a spatiotopic cue because it results in a conscious percept of distance and direction. Our perception of space is made up of an infinite number of different targets in different directions and distances. Our eyes can only select and fuse one of these target configurations at a time. One of the main tasks of the oculomotor system is to select one fixation target from many alternatives. This selection process usually results in a shift of binocular fixation from one distance and direction to another. Proximal vergence serves as a course adjustment mechanism and accounts for nearly 100 percent of the eye movement when gaze shifts are too large for us to sense them from blur and disparity. We perform these large gaze shifts without needing to see the new intended target clearly or even with both eyes.<sup>70</sup> We only need to have an estimate of target location using monocular cues for distance and direction. The majority of binocular shifts in fixation are stimulated by proximal vergence. Once the shift has been made another component of vergence (fusional vergence) takes over to maintain the response at the new target location. Errors in perceived distance can result in errors of eye alignment and focus. Mislocalization of large print or symbols in a head mounted display, or of the distance of a microscope slide can cause the eyes to accommodate excessively and produce a blurred image of a target that is imaged at optical infinity.

## The Zone of Clear and Single Binocular Vision

These components of vergence add linearly to form a range of accommodative and disparity stimuli that the observer is able to respond to with single and clear binocular vision (Fig. 11).<sup>2</sup> The range is determined by the convergence and divergence fusional vergence range from the resting phoria position, plus the linkage between accommodation and vergence described in the section on “Defocus and Efforts to Clear Vision” that increases with accommodative effort until the amplitude of accommodation is reached. Geometrically matched stimuli for convergence and accommodation that correspond to common viewing distances normally lie within this range of clear single binocular vision unless there is an alignment error such as an eye turn or strabismus. Mismatches between stimuli for accommodation and convergence produced by lenses or prisms or optical instruments can be compensated for as long as the combined stimuli for accommodation and convergence lie within this stimulus operating zone for binocular vision. The ability to respond to unmatched stimuli for convergence and accommodation does not result from decoupled independent responses of the



**FIGURE 11** Figure of the ZCSBV. Schematic representation of the zone of clear single binocular vision in terms of five fundamental variables (amplitude of accommodation, distance phoria, AC/A ratio positive fusional convergence, negative fusional convergence).

two systems. Accommodation and convergence always remain cross-coupled, but fusional vergence and optical reflex accommodation are able to make responses to overcome differences in phorias produced by accommodative vergence and prisms.

## 13.8 EFFECTS OF LENSES AND PRISM ON VERGENCE AND PHORIA

Errors of horizontal and vertical vergence can result from differences in the two ocular images caused by differential magnification and translation. Optical power differences will cause magnification differences. Translation of the image is caused by prism or by decentration of the line of sight from the optical center of the lens system.

### Magnification-Induced Errors of Eye Alignment

Magnification of the ocular image is influenced by several factors including index of refraction, lens thickness, front surface power, overall power of the lens, and the ocular distance from the eyepiece. For a spectacle lens the magnification is described by the product of a shape factor and a power factor. Magnification described by the shape factor ( $M_s$ ) equals

$$M_s = \frac{1}{1 - F_1 t / n}$$

and by the power factor ( $M_p$ ) equals

$$M_p = \frac{1}{1 - P_v h}$$

where  $F_1$  is the front surface power,  $t$  is the lens thickness in centimeters,  $n$  is the index of refraction,  $h$  is the distance of the spectacle plane from the cornea in centimeters, and  $P_v$  is the thin lens power. Percent angular magnification caused by the power factor is approximately 1.5 percent per diopter power. The percent magnification ( $M\%$ ) can be approximated by the following formula:<sup>22</sup>

$$M\% = \frac{F_1 t}{n + P_v h}$$

Differences between the lenses before the two eyes will produce image size differences, depending on the source of refractive error. Two basic sources are an abnormal axial length (axial ametropia) or abnormal corneal and lenticular power (refractive ametropia). Magnification differences between the two eyes caused by interocular differences in axial length can be eliminated largely by correcting axial errors with spectacle lenses located in the primary focal plane (the spectacle plane) and placing the lenses not too near to the eyes. Magnification differences between the two eyes caused by unequal powers of the two corneas can be prevented by correcting refractive errors with contact lenses instead of spectacle lenses (Knapp's law). If Knapp's law is violated by correcting a refractive myope with spectacle lenses placed in the primary focal plane of the eye (16.67 mm), the myopic eye will have a smaller image than an emmetropic eye with a normal corneal power. Similarly, if an axial myope is corrected with a contact lens, the myopic eye will have a larger image than an emmetropic eye with a normal axial length. However, the perceived image size in axial refractive errors also depends on whether the retina is stretched in proportion to the axial length. If it is, then the contact lens correction for an axial ametropia would produce a perceived image size equal to that of an emmetropic eye with a normal axial length. This analysis only applies to perceived image size and it is an attempt to correct perceived image size differences between the two eyes (aniseikonia). However, it does not address the unequal ocular movements stimulated by anisometropic spectacle corrections (anisophoria) resulting, as described in Sec. 13.10 on gaze, from the separation of the entrance pupil and center of rotation.

Anisophoria is defined by Friedenwald<sup>71</sup> as a heterophoria which varies in magnitude with direction of gaze. It is produced by spectacle corrections for anisometropia independent of whether refractive errors are axial or refractive in nature and only concerns the refractive power of the spectacle lens correction. The lateral prismatic effect occurs even when size distortions of an axial refractive error have been properly corrected with a spectacle lens. A prismatic deviation will always exist whenever the visual axis of the eye and optical axis of a lens do not coincide. As a result, the initial movement of an ametropic eye that is corrected with spectacle lenses will be too small in hypermetropia and too large in myopia. These optical factors produce differential prism in anisometropia corrected with spectacles. Initially, anisometropes corrected with spectacles will make vergence errors during versional movements to eccentric targets because the eyes follow Hering's law of equal innervation to the extraocular muscles. Interocular differences in power of spectacle lenses can stimulate vergence eye movements in all meridians of anisometropia (vertical, horizontal, and oblique). This optically induced anisophoria can be compensated for easily in the horizontal meridian by disparity vergence eye movements. The vertical errors will be the most troublesome because of the limited range and slow speed of vertical vergence.<sup>72,73</sup> Difficulty in compensating for optically induced anisophoria is believed to contribute to the symptoms of aniseikonia corrected with spectacle lenses.<sup>71,74</sup> Anisophoria may explain why both refractive and axial anisometropes prefer contact lens corrections that do not produce anisophoria. Contact lenses move with the eye, and the line of sight always passes through the optical center such that fixation errors of targets at infinity always equal the correct motor error to guide the oculomotor system during gaze shifts. Optically induced anisophoria can be very large. For example, Ogle<sup>22</sup> computes a 4.2 prism diopter phoria produced by a +4-D anisometropia spectacle correction during a 20° eye movement. This constitutes a 12-percent error. Normally the oculomotor system is capable of recalibration of Hering's law to compensate for optically induced anisophoria. For example, anisometropic patients wearing bifocal spectacle corrections do not exhibit the predicted change in vertical phoria while viewing in downgaze.<sup>75,76</sup> Adjustments of heterophoria in all directions of gaze occur in approximately 2.5 h for differential afocal magnification of up to 10 percent.<sup>77,78</sup> Similar oculomotor adjustments occur normally in asymmetric convergence whereby the abducted eye makes larger saccades than the adducted eye in order to compensate for geometrically induced aniseikonia.<sup>79</sup> Unequal prism before the two eyes produced by a 10-percent magnification difference,



such as observed in monocular aphakia corrected with spectacles or spectacle and contact lens combinations,<sup>80</sup> is extremely difficult to adapt to. Presumably a 10-percent magnification difference sets an upper limit to the degree of anisometropia that will not disrupt binocular versional eye movements and binocular alignment. It is possible that individuals who have symptoms of aniseikonia when corrected with spectacle lenses, are simply unable to make the necessary adjustments in conjugacy to overcome their optically induced anisophoria. Indeed Charnwood<sup>81</sup> observed that aniseikonia patients preferred prism incorporated into their spectacle corrections to reduce anisophoria and Field<sup>82</sup> reported that nearly all of 25 patients with clinically significant aniseikonia had oculomotor imbalance. Clearly anisophoria is an important by-product of anisometropia corrected with spectacles.

Torsional vergence errors can also result from optical corrections that have cylindrical corrections for astigmatism. The power of a cylindrical lens varies with the square of the cosine of the angular separation from the major power meridian. This power variation produces a meridian-dependent magnification of the ocular image described as a magnification ellipse. The ellipse produces rotational deviations of lines at different orientations toward the meridian of maximum magnification. The rotary deviation between a line and its image meridian ( $\delta$ ) depends on the angle between the orientation of the line ( $\gamma$ ) and the major axis of the magnification ellipse ( $\theta$ ). The rotary deviation is approximated by the following equation:

$$100 \tan \delta = 1/2 f \sin 2(\gamma\theta)$$

where  $f$  equals the percent of magnification produced by the cylinder.<sup>22</sup> This equation demonstrates that most of the shear or scissor distortion of lines will occur in meridians between the major and minor axes of the magnification ellipse. These rotary deviations produce lines imaged on noncorresponding retinal meridians or torsional disparities because the axis of astigmatism is normally not parallel in the two eyes. Rather, astigmatism tends to be mirror symmetric about oblique meridians such that an axis of  $45^\circ$  in one eye is often paired with an axis of  $135^\circ$  in the other eye. As a result, the shear or rotary deviations of vertical meridians tend to be in opposite directions. The resulting cyclodisparities produce perceptual declination errors of stereo surface slant (tilt of a surface about a vertical or horizontal axis). Cyclodisparities about horizontal axes stimulate cyclovergence movements of the two eyes (torsion of the two eyes in opposite directions). If the disparities are long lasting, the cyclovergence response adapts and persists even in the absence of binocular stimulation.<sup>83</sup>

### 13.9 PRISM-INDUCED ERRORS OF EYE ALIGNMENT

Errors of binocular eye alignment also result from translation of one ocular image with respect to the other. The translation can be caused by prism or by displacement of the optical center of the lens from the line of sight. When an object is viewed through a prism, it appears in a new direction that is deviated toward the prism apex and the eye must move in that direction to obtain foveal fixation. The amount the eye must rotate to compensate for the prismatic displacement of targets at finite viewing distances decreases as the distance between the eye and prism increases. This is because the angle that the displaced image subtends at the center of eye rotation decreases with eye-prism separation. This reduction is referred to as prism effectivity.<sup>84</sup> It does not occur for targets at infinity since the target distance is effectively the same to both the prism and center of rotation. The effective prism magnitude at the eye ( $Z_e$ ) is described by the following equation:

$$Z_e = \frac{Z}{1 - C_{\text{rot}} \times V}$$

where  $Z$  is prism value,  $C$  is center of rotation to prism separation in meters, and  $V$  is the reciprocal of the object to lens separation in meters which corresponds to the dioptric vergence leaving the prism.<sup>84</sup>



Binocular disparities induced by a prism or lens determines the direction of the phoria or vergence error. For example, base-out prism produces a crossed disparity and divergence error (exophoria) which is corrected under binocular conditions by fusional convergence. Minus lenses stimulate accommodation and accommodative convergence, which under binocular conditions produce a convergence error (esophoria) that is corrected by fusional divergence. Similarly, base-in prism produces an uncrossed disparity and convergence error (esophoria) which is corrected under binocular conditions by fusional divergence, and plus lenses stimulate a relaxation of both accommodation and accommodative convergence, which produces a divergence error (exophoria) that is corrected by fusional convergence.

The prismatic translation caused by optical displacement is described by Prentice's rule, where the prism displacement ( $Z$ ) equals the power of the lens in diopters ( $P$ ) times the displacement of the lens center from the line of sight ( $d$ ) in centimeters ( $Z = dP$ ). The angular deviation of any paraxial ray that passes through a lens at a distance from the optical center is described by this rule. Prentice's rule can also be applied to cylindrical lenses where the magnification occurs in a power meridian. The axis of a cylindrical lens is perpendicular to its power meridian. The prism induced can be calculated from the vector component of the displacement that is perpendicular to the axis of the cylinder. This distance ( $Hp$ ) equals the sine of the angle ( $\theta$ ) between the displacement direction and the axis of the cylinder times the magnitude of the displacement ( $h$ )

$$Hp = h \sin \theta$$

Thus, according to Prentice's rule, prism induced by displacing a cylindrical lens equals

$$Z = Ph \sin \theta$$

The magnitude of prism induced by lens displacement also depends on the separation between the lens and the center of rotation of the eye. Effective prism displacement ( $Z_e$ ) decreases as the distance of the lens from the eye increases according to the following equation:<sup>84</sup>

$$Z_e = \frac{(Zc + Z)}{1 - C_{\text{rot}} \times V}$$

where  $Zc$  is  $100 \tan c$ ,  $c$  is the angle formed between the object point and the displaced point from the optical center (major reference point) and the straight ahead line,  $V$  is the reciprocal of the image distance to the lens in meters which corresponds to the dioptric vergence leaving the prism,  $Z$  is the prism calculated from Prentice's rule,  $C_{\text{rot}}$  is separation between the lens and center of rotation of the eye. The reduction of prismatic displacement with eye to lens separation is referred to as prism effectivity with lenses.

When the two eyes view targets through lenses that are displaced from the optical axes, a disparity stimulus is introduced. Most typically, when the optical centers of a binocular optical system are separated by an amount greater or less than the interpupillary distance, disparate horizontal prism is introduced that requires the eyes to make a horizontal vergence adjustment to achieve binocular alignment. For example, when looking through a pair of plus lenses that are farther apart than the interpupillary distance, each eye is viewing through a point that is nasalward from the optical axis. This introduces base-out prism that stimulates convergence of the two eyes. This prism vergence stimulus may produce conflicts with the accommodative stimulus that is not affected by lens displacement so that the eyes need to focus for one distance and converge for another. Accommodation and convergence are linked synergistically so that their responses are closely matched to similar target distances. There is a limited degree of disparity vergence compensation available to overcome any discrepancy between the two systems, and these efforts can result in headaches and other ocular discomfort.

Anisometropic ophthalmic lens prescriptions with bifocal corrections in the lower segment of the lens can also introduce vertical prism while viewing near objects through the reading addition. The amount of vertical prism is predicted by Prentice's rule applied to the amount of anisometropia. Thus, for a 2-D anisometropic individual, a patient looking downward 1 cm through their bifocal has a vertical phoria of two prism diopters. The vertical phoria can be corrected with a slab-off

prism (a base-down prism that is cemented onto the bifocal over the lens with the most plus power). However, this procedure is rarely necessary, since the patient usually is able to adapt the vertical phoria to vary nonconjugately with vertical eye position as described above.

### 13.10 HEAD AND EYE RESPONSES TO DIRECTION (GAZE CONTROL)

When objects attract our attention in the peripheral visual field, we are able to fixate or shift our central gaze toward them using head and eye movements. The eyes move comfortably over a range of  $15^\circ$  starting from primary position. Larger gaze shifts require combinations of head and eye movements. Normally, the eyes begin the large movements, then the head moves toward the target while the eye movements are reduced by the VOR, and when the head stops the eyes refine their alignment. The reduction of eye motion during head rotation results from a gaze stabilization reflex that allows us to keep our eyes fixed on a stationary object as we move our body or head. This reflex operates with head motion sensors called the vestibular organs.

The manner in which we shift our gaze influences optical designs. Optical systems usually are mounted on the head. They need to provide a large enough viewing angle to accommodate the normal range of eye movements, and to remain stable on the head as it moves. They also need to provide peak resolution in all directions that the oculomotor system is able to shift gaze while the head remains stationary. Gaze shifts can be inaccurate because of interactions between spectacle refractive corrections and movements of the eyes.

The accuracy of eye movements is affected by stationary optical systems such as spectacles that are mounted in front of the eyes as opposed to optical systems such as contact lenses that move with the eye. The problem occurs when viewing distant targets because the entrance pupil of the eye and its center of rotation are at different distances from the spectacle plane. The entrance pupil is at a distance of about 15 mm from the spectacle lens and the ocular center of rotation lies about 27 mm behind the spectacle plane. Thus the center of rotation and entrance pupil are separated by about 12 mm and a target in space subtends different angles at these two points. The angle at the entrance pupil determines the visual error and the angle at the center of rotation determines the motor error. Whether a spectacle lens is worn or not, this difference alters retinal image feedback during eye rotations for any target imaged at a finite viewing distance from the eye. Images at infinity don't have this problem because they subtend equal angles at both the entrance pupil and center of rotation. Thus eye movements cause motion parallax between targets viewed at near and far viewing distances and some animals like the chameleon use this parallax to judge distance.<sup>85</sup>

Problems occur when the spectacle lens forms an image of an object at infinity at some finite distance from the eye. This image of the distant target subtends different angles at the entrance pupil and center of rotation because of their 12-mm separation. Thus, perceived visual angle and motor error are unequal. Depending on whether the image is formed in front (myopic error) or behind (hyperopic error) the eye, this discrepancy requires that a fixational eye movement must be smaller or larger, respectively, than the perceived eccentricity or retinal angular error of the target sensed prior to the eye movement. The amount of eye rotation can be computed from the equation describing prism effectivity with lenses ( $Z_e$ ):<sup>84</sup>

$$Z_e = \frac{Zc}{1 - C_{\text{rot}} \times V}$$

where  $Zc$  equals  $100 \tan c$  ( $c$  equals angular eccentricity of the target measured at the lens plane), and  $V$  equals the dioptric vergence of the image distance from the lens plane, and  $C_{\text{rot}}$  equals the lens to center of rotation separation in meters. When different power spectacle lenses are worn before the two eyes to correct unequal refractive errors (anisometropia), a target that would normally require equal movements of the two eyes now requires unequal binocular movements. The oculomotor system learns to make these unequal movements and it retains this ability even when one eye is occluded after wearing anisometric spectacle corrections for only 1 hour.

During head movements, magnification of the retinal image by spectacles increases retinal image motion during head rotation and causes the stationary world to appear to move opposite to the direction the head rotates. This can produce vertigo as well as stimulate a change in the vestibular stabilization reflex. When a magnifier is worn for only 1 h, the reflex can be exaggerated to compensate for the excessive retinal image motion. Following this adaptation, when the optical aid is removed, the exaggerated VOR causes the world to appear to move in the same direction that the head rotates. If you wear glasses, you can demonstrate this adaptation to yourself by removing your glasses and shaking your head laterally. If you wear negative lenses to correct for myopia, the minification they produce has caused a reduction in image stabilization and without your glasses, the world will appear to move opposite to the direction of head shake.

### 13.11 FOCUS AND RESPONSES TO DISTANCE

Our visual sense of space requires that we are able to accurately perceive objects at different distances. Single and clear views of targets at a finite viewing distance require that the visual system align the similar images of the target on corresponding retinal regions and that the optics of the eye adjust the focus to accommodate the near target (accommodation). The range of accommodation varies with age, starting at birth at about 18.5 D and decreasing 1 D every 3 years until approximately age 55 when it is no longer available (absolute presbyopia). The amplitude of accommodation available to an individual of a given age is remarkably stable across the population if extremes of refractive are not considered.

Spectacle corrections and other optical systems can influence both aspects of the near response. Accommodation declines with age until eventually, the near point is beyond the near working distance (functional presbyopia). In this situation, near targets can be imaged farther away from the eye with plus lenses so that they lie within the available range of remaining accommodation. The accommodative system is only able to maintain 2/3 of its full amplitude for long periods of time and when this range is exceeded by the optical demands of working distance, bifocals are typically prescribed. The magnitude of the bifocal correction depends upon the accommodative amplitude and the near working distance. For a typical 16-in or 40-cm near working distance that requires 2.5 D of accommodation, a bifocal correction is needed when the amplitude of accommodation is less than 3.75 D. Typically this amplitude is reached at about 45 years of age.

If spectacles are worn, the amplitude of accommodation is influenced by the sign (plus or minus) of the spectacle correction. Accommodation demand equals the difference in the dioptric value of the far point and the near point of the eye. These dioptric values can be described relative to the front surface of the eye or to the spectacle plane. The reciprocal of a specific viewing distance to the spectacle plane describes the spectacle accommodative demand, and this value is independent of the lens correction. However, the ocular accommodative demand (amount the eye needs to change power) required to focus targets at this distance does depend on the power and sign of the spectacle correction. Ocular accommodation is computed by the difference between the dioptric vergence subtended by the near point target at the cornea, after it is imaged by the spectacle lens, and the vergence subtended at the cornea by the far point of the eye. For myopes corrected with minus spectacle lenses, ocular accommodation through spectacles is greater than the spectacle accommodative demand and the reverse is true for hyperopes. For example, the ocular accommodative response to a target place 20 cm in front of the spectacle plane that is 14 mm from the eye is 4.67 D from an emmetrope, 4 D for a myope wearing a  $-6$ -D corrective spectacle lens, and 5.56 D for a hyperope wearing a  $+6$  D corrective spectacle lens. Thus, near targets present unequal accommodative stimuli to the two eyes of anisometropes corrected with spectacles. Although the accommodative response is consensual (equal in the two eyes) the difference in the two stimuli will go unnoticed as long as it lies within the depth of focus of the eye (approximately 0.25 D). However, in the example given above, a 6-D anisometrope would have about a 1-D difference in the stimulus to accommodation presented by a target at a 20-cm viewing distance. Finally, this lens effectivity causes spectacle lens wearers who are hyperopic to become presbyopic before myopes.

## 13.12 VIDEO HEAD SETS, HEAD'S UP DISPLAYS AND VIRTUAL REALITY: IMPACT ON BINOCULAR VISION

A variety of head mounted visual display systems (HMD) have been developed to assist performance of a variety of tasks. Telepresence is an environment in which images from remote cameras are viewed on video screens to perform tasks at distal sites. In other applications, information in the form of symbology is added to a natural image to augment reality, such as in a heads up aviation application, and in still another application, virtual environments are simulated and displayed for entertainment or instructional value. The head mounted video systems usually display information in one of three configurations. These are monocular or monoscopic that present information to one eye, biocular where both eyes view an identical image, and binocular where slight differences in the two eyes' images stimulate stereoscopic depth perception. The biocular view avoids image conflicts between the two eyes that arise in the monoscopic systems, and stereopsis in the binocular systems helps to sort out or distinguish elevation and distance information in the perspective scene. Stereopsis also provides the user with fine depth discrimination and shape sensitivity for objects located within an arm's length. Each of these systems attempts to present coherent cues to space perception, however, due to physical limitations this is not always possible and cue conflicts result.

### Distance Conflicts

The current designs of HMD devices introduce three stimulus conflicts. They present conflicting information about target distance to accommodation and convergence, conflicting visual-vestibular information, and conflicts concerning the spatial location of information to the two eyes. For example, the symbology presented to aviators in heads-up displays is set for infinity but large symbols appear to be near and this can initiate an inappropriate accommodative response. However, given sufficient time, subjects will accommodate correctly.<sup>86</sup> Currently all HMD units present a fixed optical viewing distance that is not always matched to the convergence stimulus in biocular units or the many states of convergence stimulated in binocular units. This situation is exacerbated by the presence of uncorrected refractive errors that can cause images to be blurred in the case of myopia or to force extra accommodation in the case of hyperopia, or to present unequal blur and accommodative stimuli in the case of anisometropia. Normally, observers with uncorrected refractive errors are able to adjust target distance to clear images but the HMD has a fixed target distance that is not adjustable. Units that provide self-adjusting focus run the risk of users introducing larger errors or mismatching the clarity of the two ocular images.

As described above, accommodation and convergence are tightly coupled to respond to approximately the same viewing distance. There is some flexibility to compensate for small differences between stimuli to accommodation and convergence but long periods of usage, particularly those requiring fusional divergence, to overcome an esophoria caused by excessive accommodation can produce eye strain. Tolerance for accommodative errors can be increased by extending the depth of focus with small exit pupils or with screens that have reduced resolution. It is also helpful to set the viewing distance at an intermediate value of approximately 2 m<sup>87</sup> to allow uncorrected mild myopes to see clearly while allowing a 3° range of stereo depth disparities. In biocular units where disparity is not varied, the optical distance can be set closer to 1 m to further compensate for uncorrected myopia and to place images near arm's length. Prolonged viewing at close distances can cause instrument myopia, a form of excessive accommodation, and postimmersion short-term aftereffects such as accommodative spasm.<sup>88</sup> Accommodative fatigue also can result after 30 min of stereo display use.<sup>89</sup> If refractive corrections are worn, the optimal image distance will depend on the tasks performed. Simulated computer screens should be at arm's length whereas views for driving simulators should be remote to provide more realism in the display.

Errors of perceived distance can occur when vergence distance is not matched to image size or perceptual task. These errors can result in perceptual image expansion (macropsia) or constriction (micropsia), particularly in binocular displays that stimulate a range of convergence values. These effects can be minimized by limiting the range of binocular disparities to a few degrees for near

objects and to use a binocular display when presenting views of distant objects simultaneously with foreground detail. Perspective cues along with motion parallax, overlap or superposition, and texture density gradients provide powerful depth cues for remote or complex scenes containing a wide range of depth information and binocular disparity adds little to these displays. Stereopsis is most useful for facilitating manual dexterity tasks with fine depth discrimination between objects and their 3-D shape located within an arm's length of the user. Static displays that don't utilize motion parallax will benefit more from the addition of stereoscopic cues.

Some HMDs and other virtual environments combine real and virtual information with a see-through visor or with remote rear projection systems that surround the viewer panoramically as in the Computer-Automatic Virtual Environment.<sup>90</sup> These mixed environments can introduce multiple and sometimes conflicting accommodative stimuli. When real variable focus and virtual fixed focus images are at different image distances, both cannot be cleared at the same time, though they may appear to be in similar locations in space. Conflicting near virtual targets can interfere with recognition of real distant targets.<sup>91</sup> Several optical solutions have been suggested to minimize these focus conflicts.<sup>54</sup> The depth of focus can be extended with pinhole pupils. A pinhole exit pupil projected to the eye pupil in a Maxwellian view system will preserve field of view. The virtual image is viewed through the Maxwellian pupil imaged via a beam splitter while the real environment is viewed through the natural pupil. Eye tracking might be necessary in Maxwellian systems to shift the exit pupil to follow eye movements.<sup>92</sup> Monovision is another possible solution in which a near lens is worn over one eye and a far lens over the other. A target that is in focus for one eye will be blurred for the other. The visual system is able to suppress the blurred image and retain the focused images of both eyes and also retain a large functional range of stereopsis.<sup>93</sup> Monovision has been used successfully in presbyopes with contact lens corrections, and it has been shown to work on prepresbyopes in most individuals.<sup>55</sup> The limitation of monovision is that it provides clear vision only for the two focal depths, plus and minus the depth of focus. The range of the two clear vision distance might be extended by aniso-accommodation; the ability of the two eyes to focus independently of each other while remaining binocular.<sup>54</sup> Another solution is a chromatic bifocal that takes advantage of the longitudinal chromatic aberration of the eye. Wavelengths of light at opposite ends of the visual spectrum have a 2-D difference in dioptric power for a given object distance. Given that nearer objects tend to be in the lower visual field, different monochromatic wavelengths combined with field segregation could be used in the virtual image control panel. Users viewing near objects could view long-wavelength images in the lower half of the virtual control panel. Thus, both real and red virtual images would stimulate an increase in accommodation. For distant real objects, the user could attend to the upper short wavelength images in the virtual display panel and real and blue virtual targets would stimulate a relaxation of accommodation. A simpler solution is to place a bifocal lens segment in the lower portion of the see-through window, bringing near real images to the same focal distance as the images seen on the virtual display.

## Visual-Vestibular Conflicts

Virtual systems that are worn on the head or remain fixed in space both present conflicting visual-vestibular information. Normally, when the head rotates or translates, the eyes rotate in the opposite direction to keep the images of objects in space stabilized on the retina. The relationship between retinal image motion and head and eye motion is disrupted by the virtual display. In the case of an external virtual projection system like CAVE (Computer-Augmented Virtual Environment),<sup>90</sup> simulated movements of the visual scene that correspond to body movements are not accompanied by the appropriate vestibular signals. For example, a driving simulator displays changes in the road view while steering the virtual car. When making turns, the visual field appears to rotate but there is no corresponding vestibular signal. As a result, drivers of virtual systems tend to oversteer the turns because they underestimate the curved path of the automobile. Perceptual disturbances during similar phenomenon in flight simulators are referred to as simulator sickness. In contrast head mounted video displays can present a fixed visual stimulus while the head rotates such that vestibular signals are not accompanied by the appropriate visual motion stimuli. These conflicts can also

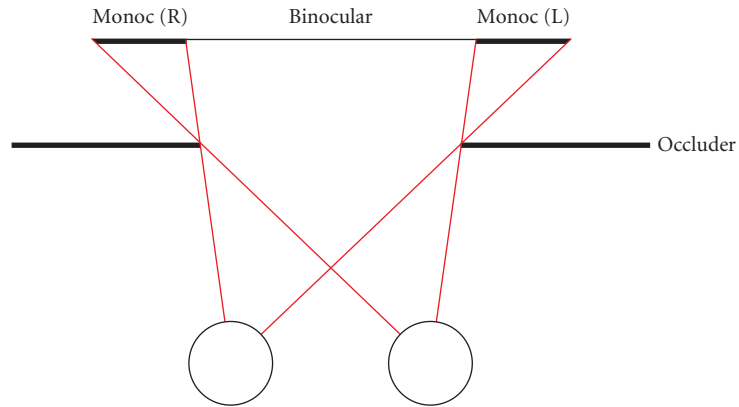
result in symptoms of nausea, vertigo, and general malaise even when very small head movements occur.<sup>94-96</sup> Visual-vestibular conflicts also occur with telepresence systems that present images that are either magnified or minified as a result of differences between the camera distance to objects and the viewer distance to the viewscreen. As a result, head movements that cause camera movements will stimulate excessive retinal image motion with magnification or too little retinal image motion with image minification. The mismatch will cause adaptive adjustments of the VOR and possible simulator sickness.

The oculomotor system responds to this visual-vestibular conflict during steady fixation by suppressing the VOR and adaptively reducing its gain. Thus, a user viewing a 2-h motion picture on an HMD while making a normal range of head movements during conversation with other persons will adapt their VOR to be less responsive to head motion. This adaptation can produce postimmersion symptoms where the adapted VOR remains suppressed such that head movements without the HMD will not stimulate a sufficient vestibular ocular movement to stabilize the retinal image. The user will perceive the world to move with his head and this can cause nausea, vertigo, and malaise. Attempts to move images with the head using a head tracking system that updates image position during head rotation have been unsuccessful<sup>97</sup> because of small delays in image processing that desynchronize expected image motion and head movements. It is possible that an image could be blanked during large head movements and stabilized with a head tracker during small head movements. The blanking would be similar to the normal suppression of vision during saccadic eye movements. Other solutions are to use see-through and see-around virtual displays so that real objects in the periphery are always visible during normal head movements so that some visual and vestibular stimuli remain coherent.

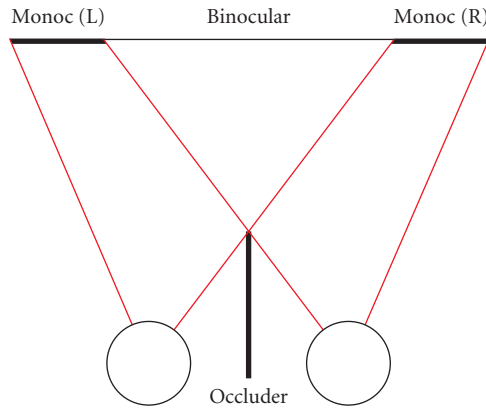
## Spatial Location Conflicts

No existing HMD has the wide field of view of the human visual system in a natural environment. The horizontal field of view in currently available HMDs ranges from 22.5° to 155°.<sup>98</sup> The larger fields of view in binocular systems are expanded by partially overlapping the two monocular images. Note that the entire monocular and binocular portions of the field of view lie within the normal 114 to 120° binocular overlap of the normal visual system. Partial overlap in HMDs can be produced by displacing the right field to the right and left field to the left. Each eye has a monocular flanking component in its temporal visual field. It is also possible to reverse the displacement so that the monocular components are in the nasal-visual field of each eye (Fig. 12). These two organizations are consistent with two different occlusion aperture configurations. The former has the same organization as a binocular field seen on a near screen in front of a background. Thus the total image is seen as a discontinuous surface in depth with the central binocular portion at a proximal distance compared to the monocular portion that is seen farther away.<sup>6</sup> The latter case has the same organization as a single continuous surface viewed through an aperture (porthole effect). The former system is more likely to cause binocular rivalry in the monocular regions of the field than is the latter approach. This suppression can cause a moonlike crescent shape (luning) in the flanking monocular region that corresponds to the border of the image seen by the contralateral eye and this blocks visibility of the monocular field crescent seen by the ipsilateral eye. The luning can be reduced by blurring the edges of each eye's image borders or by using the crossed view configuration. Interestingly, the latter crossed-view approach is preferable in head mounted displays.<sup>99</sup>

Visual suppression and binocular rivalry suppression are also stimulated in monocular image displays. Rivalry occurs between the image presented to one eye and the natural scene presented to the other. Even if one eye is occluded, the dark image of the patched eye will rival with the visible image of the open eye. This rivalry can be minimized by presenting the patched eye with a uniform field with the same mean luminance and color as the open eye image. This can be accomplished easily by placing a diffusing filter over the nonseeing eye. In cases where the two eyes view independent images, it might be possible to control which eye is suppressed by moving the image in the seeing eye and keeping it stationary in the nonseeing eye. This technique is based on the observation that motion is a dominant feature for binocular rivalry.



Porthole occlusion configuration  
 Contralateral monocular crescents



Septum occlusion configuration  
 Ipsilateral monocular crescents

**FIGURE 12** Occlusion figure with port hole and septum. Partial overlap of the visual field simulates obstructed views by a port-hole aperture or septum in the mid-sagittal plane.

## Optical Errors

The success of any binocular viewing instrument depends in large part on the precision of alignment and quality of distortion free optics. Alignment errors can produce vertical, horizontal, and torsional disparities. Of these, the vertical errors are the most troublesome. There is limited range for vertical vergence and slow response times.<sup>77</sup> The oculomotor system can adapt its vertical alignment posture (vertical phoria) if the vertical disparity is constantly present; however, if the HMD is worn intermittently then no adaptation will occur. Horizontal errors, especially in the convergence direction, can be overcome easily by vergence eye movements. Torsional errors cause overall torsional disparities which can cause perception of slant about the horizontal axis if the virtual scene is superimposed



by see-through optics on the natural visual environment. However, if an opaque visor is used so that only the virtual display is visible, slant distortion will not be perceived with cyclotorsion disparity.<sup>100</sup> This is because the visual system only interprets slant when there is horizontal shear disparity between vertical lines but no vertical shear disparity between horizontal lines. When both forms of shear are present, this is interpreted as resulting from eye rotation rather than real object slant. Cyclodisparity in large field displays will stimulate cyclovergence responses that can have postimmersion aftereffects in the form of cyclophoria.<sup>83</sup> Errors in the baseline separation of the entrance pupils of the optical system can produce divergent disparities if it is set wider than the interpupillary distance. The oculomotor system has limited divergence ability and this stimulus can cause headaches and visual discomfort. Finally, low quality optics that have distortions such as barrel, pincushion, and keystone will introduce horizontal, vertical, and torsional disparities in both biocular and binocular displays if the distortions are not identical for the two eyes or if the visual fields are displaced slightly in the two eyes. Displacements could result from alignment errors or from partial field overlap used to expand the field of view. The disparities produced by these distortions can cause idiosyncratic warpage of the virtual image and vertical disparities that stress the oculomotor system.

### 13.13 REFERENCES

1. G. L. Walls, *The Vertebrate Eye and Its Adaptive Radiation*, Hafner, New York, 1967.
2. H. W. Hofstetter, "Graphical Analysis," Chapter 13 in *Vergence Eye Movements: Clinical and Basic Aspects*, C. M. Schor and K. Ciuffreda (eds.), Butterworth, Boston, 1983.
3. W. Dember and J. Warm, *Psychology of Perception*, Holt, Rinehart, and Winston, New York, 1979.
4. B. Cumming, "Motion—In Depth," Chapter 12 in *Visual Detection of Motion*, A. T. Smith and R. J. (eds.), Snowden Academic Press, San Diego, CA, 1994.
5. W. Epstein, (ed.), *Stability and Constancy in Visual Perception: Mechanisms and Processes*, John Wiley and Sons, New York, 1977.
6. K. Nakayama and S. Shimojo, "DaVinci Stereopsis: Depth and Subjective Occluding Contours from Unpiped Image Points," *Vis. Res.* **30**:1811–1825 (1990).
7. M. G. Harris, "Optic Retinal Flow," Chapter 2 in *Visual Detection of Motion*, A. T. Smith and R. J. Snowden (eds.), Academic Press, San Diego, CA, 1994.
8. B. Rogers and B. J. Graham, "Similarities between Motion Parallax and Stereopsis in Human Depth Perception," *Vis. Res.* **22**:261–270 (1982).
9. M. E. Ono, J. Rivest, and H. Ono, "Depth Perception as a Function of Motion Parallax and Absolute-Distance Information," *J. Exp. Psychol: Human Perception and Performance.* **12**:331–337 (1986).
10. L. R. Harris, "Visual Motion Caused by Movements of the Eye, Head and Body," Chapter 14 in *Visual Detection of Motion*, A. T. Smith and R. J. Snowden (eds.), Academic Press, San Diego, CA, 1994.
11. D. Regan and K. I. Beverley, "Visually Guided Locomotion: Psychophysical Evidence for a Neural Mechanism Sensitive to Flow Patterns," *Science* **205**:311–313 (1979).
12. W. H. Warren and D. J. Hannon, "Direction of Self-Motion is Perceived from Optical Flow," *Nature* **336**:162–163 (1988).
13. J. A. Crowell and M. S. Banks, "Ideal Observer for Heading Judgments," *Vis. Res.* **36**:471–490 (1996).
14. D. N. Lee, "A Theory of Visual Control of Braking Based on Information about Time to Collision," *Perception* **5**:437–459 (1976).
15. S. P. McKee and S. Watamaniuk, "The Psychophysics of Motion Perception," Chapter 4 in *Visual Detection of Motion*, A. T. Smith and R. J. Snowden (eds.), Academic Press, San Diego, CA, 1994.
16. T. Brandt, J. Dichgans, and W. Koenig, "Differential Central and Peripheral Contributions to Self-Motion Perception," *Exp. Brain Res.* **16**:476–91 (1973).
17. K. I. Beverley and D. Regan, "Evidence for the Existence of Neural Mechanisms Selectively Sensitive to the Direction of Movement in Space," *J. Physiol. Lond.* **235**:17–29 (1973).
18. E. Hering, *Spatial Sense and Movements of the Eye* (in German), C. A. Radde (trans.), American Academy of Optometry, Baltimore, 1879.



19. I. P. Howard, *Human Visual Orientation*, John Wiley and Sons, New York, 1982.
20. R. van Ee and C. J. Erkelens, "Stability of Binocular Depth Perception with Moving Head and Eyes," *Vis. Res.* **36**:3827–3842 (1996).
21. C. W. Tyler and A. B. Scott, "Binocular Vision," in *Physiology of the Human Eye and Visual System*, R. Records (ed.), Harper and Row, Hagerstown, 1979.
22. K. N. Ogle, *Researches in Binocular Vision*, Hafner, New York, 1964.
23. H. V. Helmholtz, *Handbuch der Physiologischen Optik*, 3d German ed. (1962) English translation by J. P. C. Southall, (trans.), 1909.
24. J. Foley, "Binocular Distance Perception," *Psychol. Rev.* **87**:411–434 (1980).
25. J. Garding, J. Porrill, J. E. W. Mayhew, and J. P. Frisby, "Stereopsis, Vertical Disparity and Relief Transformations," *Vis. Res.* **35**(5):703–722 (1995).
26. J. E. W. Mayhew and H. C. Longuet-Higgins, "A Computational Model of Binocular Depth Perception," *Nature* **297**:376–378 (1982).
27. B. Gillam and B. Lawergren, "The Induced Effect, Vertical Disparity, and Stereoscopic Theory," *Perception and Psychophysics* **34**:121–130 (1983).
28. L. Liu, S. B. Stevenson, and C. M. Schor, "A Polar Coordinate System for Describing Binocular Disparity," *Vis. Res.* **34**(9):1205–1222 (1994).
29. B. J. Rogers and M. F. Bradshaw, "Vertical Disparities, Differential Perspective and Binocular Stereopsis," *Nature* **361**:253–255 (1993).
30. G. Westheimer and M. W. Pettet, "Detection and Processing of Vertical Disparity by the Human Observer," *Proc. R. Soc. Lond. B Biol. Sci.* **250**(1329):243–247 (1992).
31. B. T. Backus, M. S. Banks, R. van Ee, and J. A. Crowell, "Horizontal and Vertical Disparity, Eye Position, and Stereoscopic Slant Perception," *Vis. Res.* **39**:1143–70 (1999).
32. R. M. Steinman and H. Collewijn "Binocular Retinal Image Motion during Active Head Rotation," *Vis. Res.* **20**:415–429 (1980).
33. D. Cline, H. W. Hofstetter, and J. R. Griffin, *Dictionary of Visual Science*, 4th ed. Chilton, Radnor, PA, 1989.
34. G. Westheimer, "Effect of Binocular Magnification Devices on Stereoscopic Depth Resolution," *J. Opt. Soc. Am.* **46**:278–280 (1956).
35. K. Stevens, "Surface Perception from Local Analysis of Texture and Contour," Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 1979.
36. K. N. Ogle and P. Boder, "Distortion of Stereoscopic Spatial Localization," *J. Opt. Soc. Am.* **38**:723–733 (1948).
37. C. Bourdy, "Aniseikonia and Dioptric Elements of the Eye," *J. Am. Opt. Assn.* **39**:1085–1093 (1968).
38. R. S. Arner, "Eikonometer Measurements in Anisometropes with Spectacles and Contact Lenses," *J. Am. Optom. Assn.* **40**:712–715 (1969).
39. L. Rose and A. Levenson, "Anisometropia and Aniseikonia," *Am. J. Optom.* **49**:480–484 (1972).
40. S. Awaya and G. K. von Noorden, "Aniseikonia Measurement by Phase Difference Haploscope in Myopia, Anisometropia, and Unilateral Aphakia (with Special Reference to Knapp's Law and Comparison Between Correction with Spectacle Lenses and Contact Lenses)," *Jpn. Contact Lens Soc.* **13**:131 (1971).
41. A. Bradley, J. Rabin, and R. D. Freeman, "Nonoptical Determinants of Aniseikonia," *Invest. Ophthalmol. Vis. Sci.* **24**:507–512 (1983).
42. B. Winn, C. A. Ackerley, F. K. Brown, J. Murray, J. Prars, and M. F. St. John, "Reduced Aniseikonia in Axial Anisometropia with Contact Lens Correction," *Ophthalmic Physiol. Opt.* **8**:341–344 (1987).
43. R. M. Burnside and C. Langley, "Anisometropia and the Fundus Camera," *Am. J. Ophthalmol.* **58**:588–594 (1964).
44. E. Engle, "Incidence and Magnitude of Structurally Imposed Retinal Image Size Differences," *Percept. Motor Skills* **16**:377–384 (1963).
45. B. Julesz, *Foundations of Cyclopean Perception*, Chicago University of Chicago Press, Chicago, 1971.
46. C. B. Blakemore, "A New Kind of Stereoscopic Vision," *Vis. Res.* **10**:1181–1199 (1970).
47. A. Fiorentini and L. Maffei, "Binocular Depth Perception without Geometrical Cues," *Vis. Res.* **11**:1299–1305 (1971).

48. C. W. Tyler and E. Sutter, "Depth from Spatial Frequency Difference: an Old Kind of Stereopsis?" *Vis. Res.* **19**:858–865 (1979).
49. C. M. Schor, I. C. Wood, and J. Ogawa, "Spatial Tuning of Static and Dynamic Local Stereopsis," *Vis. Res.* **24**:573–578 (1984).
50. C. M. Schor and T. Heckmen, "Interocular Differences in Contrast and Spatial Frequency: Effects on Stereopsis and Fusion," *Vis. Res.* **29**:837–847 (1989).
51. M. von Rohr, *Die Brille als Optisches Instrument*, Wilhelm Englemann, Leipzig, pp. 172, 1911.
52. H. Erggelet, "Über Brillenwirkungen," *Ztschr. F. Ophth. Optik* **3**:170–183 (1916).
53. J. Nolan, A. Hawkswell and S. Becket, "Fusion and Stereopsis in Aphakia," in *Orthoptics: Past Present and Future*, S. Moore, J. Mein, and L. Stockbridge (eds.), Shatten Int. Cont. Med. Book Co., New York, pp. 523–529 1975.
54. L. Marran, and C. M. Schor, "Lens Induced Aniso-Accommodation," *Vis. Res.* **38**(22):3601–3619 (1998).
55. C. M. Schor, L. Landsman, and P. Erickson, "Ocular Dominance and the Interocular Suppression of Blur in Monovision," *Am. J. Optom. and Physiol. Optics* **64**:723–736 (1987).
56. M. J. Collins and A. Goode, "Interocular Blur Suppression and Monovision," *Acta Ophthalmologica* **72**: 376–380 (1994).
57. H. Wallach and K. J. Frey "Adaptation in Distance Perception Based on Oculomotor Cues," *Perception and Psychophysics* **11**:77–83 (1972).
58. D. A. Owens and H. W. Leibowitz, "Accommodation, Convergence and Distance Perception in Low Illumination," *Am. J. Optom. Physiol. Opt.* **57**:540–50 (1980).
59. S. M. Ebenholtz, "Hysteresis Effects in the Vergence Control System: Perceptual Implications," in D. F. Fisher, R. A. Monty, and J. W. Senders (eds.), *Eye Movements: Visual Perception and Cognition*, Erlbaum, Hillsdale, NJ, 1981.
60. D. W. McCready, "Size, Distance Perception and Accommodation Convergence Micropsia. A Critique," *Vis. Res.* **5**:189–206 (1965).
61. T. Morita and N. Hiruma, "Effect of Vergence Eye Movements for Size Perception of Binocular Stereoscopic Images," in *Proc. Third International Display Workshops*, Kobe, Japan, SID, 1996.
62. S. R. Ellis, U. J. Bucher, and B. M. Menges, "The Relationship of Binocular Convergence and Errors in Judged Distance to virtual Objects," in *Proc. International Federation of Automatic Control*, Boston, MA (1995).
63. Y. Y. Yeh, "Visual and Perceptual Issues in Stereoscopic Color Displays," Chapter 4, in *Stereo Computer Graphics and Other True 3D Technologies*, D. F. McAllister (ed.), Princeton University Press, Princeton, NJ, 1993.
64. A. Gonshor and G. Melvill-Jones, "Extreme Vestibulo-Ocular Adaptation Induced by Prolonged Optical Reversal of Vision," *J. Physiol. (Lond.)* **256**:381–414 (1976).
65. G. M. Gauthier and D. A. Robinson, "Adaptation of the Human Vestibular Ocular Reflex to Magnifying Lenses," *Brain Res.* **92**:331–335 (1975).
66. R. M. Burde, "The Extraocular Muscles," in *Adler's Physiology of the Eye: Clinical Applications*, R. A. Moses (ed.), St. Louis, CV Mosby, pp 84–183 (1981).
67. B. Bagolini, "Part II Sensorial–Motor Anomalies in Strabismus (Abnormal Movements)," *Doc. Ophthalmol.* **41**:23–41 (1976).
68. R. K. Luneburg, *Mathematical Analysis of Binocular Vision*, Princeton University Press, Princeton, NJ, (1948).
69. D. Nguyen, I. Vedamurthy, and C. M. Schor, "Cross-Coupling Between Accommodation and Convergence Is Optimized for a Broad Range of Directions and Distances of Gaze," *Vis. Res.* **48**(7):893–903 (2008).
70. C. M. Schor, J. Alexander, L. Cormack, and S. Stevenson, "A Negative Feedback Control Model of Proximal Convergence and Accommodation," *Ophth. and Physiol. Optics.* **12**:307–318 (1992).
71. J. S. Friedenwald, "Diagnosis and Treatment of Anisophoria," *Arch. Ophth.* **15**:283–304 (1936).
72. A. L. Perlmutter and A. Kertesz, "Measurement of Human Vertical Fusional Response," *Vis. Res.* **18**:219–223 (1978).
73. W. A. Houtman, T. H. Roze, and W. Scheper, "Vertical Motor Fusion," *Doc. Ophthalmol.* **44**:179–185 (1977).
74. A. Romole "Dynamic versus Static Aniseikonia," *Aust. J. Optom.* **67**:108–113 (1984).
75. V. Ellerbrock and G. Fry, "Effects Induced by Anisometric Corrections," *Am. J. Optom.* **19**:444–459 (1942).

76. P. L. Cusick and H. W. Hawn, "Prism Compensation in Cases of Anisometropia," *Arch Ophthalmol.* **25**:651–654 (1941).
77. D. B. Henson and B. Dharamsi, "Oculomotor Adaptation to Induced Heterophoria and Anisometropia," *Invest. Ophthalmol. Vis. Sci.* **22**:234–240 (1982).
78. C. M. Schor, J. Gleason, and D. Horner, "Selective Nonconjugate Binocular Adaptation of Vertical Saccades and Pursuits," *Vis. Res.* **30**(11):1827–1845 (1990).
79. L. C. Morrison, "Stereoscopic Localization with the Eyes Asymmetrically Converged," *Am. J. Optom.* **54**:556–566 (1977).
80. J. M. Enoch, "A Spectacle-Contact Lens Combination used as a Reverse Galilean Telescope in Unilateral Aphakia," *Am. J. Optom.* **45**:231–240 (1968).
81. J. R. B. Charnwood, *An Essay on Binocular Vision*, London, Hatton, pp. 73–88, 1950.
82. H. B. Field, "A Comparison of Ocular Imagery," *Arch. Ophthalmol.* **29**:981–988 (1943).
83. J. S. Maxwell and C. M. Schor, "Adaptation of Ocular Torsion in Relation to Head Position," *Vis. Res.* **36**(8):1195–1205 (1999).
84. M. P. Keating, *Geometric, Physical and Visual Optics*, Butterworths, Boston, MA, 1988.
85. M. F. Land, "Fast-Focus Telephoto Eye," *Nature* **373**:658 (1995).
86. C. M. Schor and L. Task, "Effects of Overlay Symbolism in Night Vision Goggles on Accommodation and Attention Shift Reaction Time," *Aviation, Space and Environmental Medicine* **67**:1039–1047 (1996).
87. D. A. Southard, "Transformations for Stereoscopic Visual Stimuli," *Computers and Graphics* **16**:401–410 (1992).
88. M. Rosenfield and K. J. Ciuffreda, "Cognitive Demand and Transient Nearwork-Induced Myopia," *Optom. Vis. Sci.* **71**:381–385 (1994).
89. T. Inoue and H. Ohzu, "Accommodation and Convergence When Looking at Binocular 3D Images," in *Human Factors in Organizational Design and Management III*, K. Noro and O. Brown, Jr. (eds), Elsevier Science Publishers, Amsterdam, pp. 249–252, 1990.
90. R. V. Kenyon, T. A. DeFanti, and J. T. Sandin, "Visual Requirements for Virtual-Environment Generation," *Society for Information Display Digest* **3**:357–360 (1995).
91. J. Norman and S. Ehrlich, "Visual Accommodation and Virtual Image Displays: Target Detection and Recognition," *Hum. Factors* **28**:135–151 (1986).
92. G. C. de Wit and R. A. E. W. Beek, "Effects of a Small Exit Pupil in a Virtual Reality Display System," *Opt. Eng.* **36**:2158–2162 (1997).
93. P. Erickson and C. M. Schor, "Visual Functions with Presbyopic Contact Lens Corrections," *Optometry and Visual Science* **67**:22–28 (1990).
94. R. L. Hughes, L. R. Chason, and J. C. H. Schwank, *Psychological Considerations in the Design of Helmet-Mounted Displays and Sights: Overview and Annotated Bibliography*, U.S. Government Printing Office, AMRL-TR-73-16, National Technical Information Service and U.S. Dept. of Commerce, Washington D.C., 1973.
95. P. A. Howarth, "Empirical Studies of Accommodation, Convergence and HMD Use," in *Proc. Hoso-Bunka Foundation Symposium: The Human Factors in 3D Imaging*, Hoso-Bunka Foundation, Tokyo, Japan.
96. P. A. Howarth and P. J. Costello, "The Occurrence of Virtual Simulation Sickness Symptoms When an HMD was Used as a Personal Viewing System," *Displays* **18**:107–116 (1997).
97. T. Piantanida, D. K. Bowman, J. Larimer, J. Gille, and C. Reed, "Studies of the Field of View/Resolution Trade-Off in Virtual-Reality Systems," *Hum. Vis. Visual Proc. Digital Display III, Proc. SPIE* **1666**:448–456 (1992).
98. E. T. Davis, "Visual Requirements in HMDs: What Can We See and What Do We Need to See," Chapter 8 in *Head Mounted Displays*, J. E. Melzer and K. Moffett, (eds.) Optical and Electro-Optical Engineering Series, McGraw-Hill, New York, 1997.
99. J. E. Melzer and K. W. Moffitt, "Ecological Approach to Partial Binocular Overlap," *Large Screen Projection, Avionic and Helmet-Mounted Displays, Proc. SPIE* **1456**:124 (1991).
100. I. P. Howard and H. Kaneko, "Relative Shear Disparities and the Perception of Surface Inclination," *Vis. Res.* **34**:2505–2517 (1994).

---

# OPTICS AND VISION OF THE AGING EYE

---

## John S. Werner

*Department of Ophthalmology & Vision Science  
University of California, Davis  
Sacramento, California*

## Brooke E. Scheffrin

*Department of Psychology  
University of Colorado  
Boulder, Colorado*

## Arthur Bradley

*School of Optometry  
Indiana University  
Bloomington, Indiana*

---

## 14.1 GLOSSARY

---

**Aphakia.** An eye lacking a lens, usually as a result of cataract surgery.

**Age-related macular degeneration (AMD).** A disease that results in loss of vision in the central retina or macula due to damage to the photoreceptors and retinal pigment epithelium of the retina, often secondary to changes in the choroid (blood vessel layer) and Bruch's membrane. There are two major forms, called "wet," or exudative, and "dry," or atrophic. The dry form is more common, but the wet form is responsible for the majority of cases involving severe vision loss.

**Cataract.** An opacity or clouding of the lens resulting in a reduction in the amount of light transmitted to the retinal image and an increase in the light scatter. This is the leading cause of blindness worldwide, but can be treated successfully by removal of the lens.

**Diabetic retinopathy.** A disease of the eye caused by diabetes that leads to a proliferation of blood vessels. These blood vessels may swell and leak. When the resultant scar tissue shrinks the surrounding vitreous humor, it may cause the retina to detach, leading to loss of vision in the affected region of retina.

**Glaucoma.** A group of diseases, often but not always, associated with high intraocular pressure, resulting in death of ganglion cells.

**Intraocular lens (IOL).** A prosthetic lens usually implanted in the eye to replace the natural lens following its removal. IOLs may be placed in front of the pupillary plane (anterior chamber lenses) or

behind the pupil (posterior chamber lenses). They are manufactured from a variety of materials, the most common of which are polymethylmethacrylate, silicone, and hydrogels.

**Phakia.** An eye with a natural lens, from the Greek *phakos* or lens.

**Presbyopia.** Greek for “old sight,” is a condition in which an individual has lost the ability to accommodate sufficiently for near work. This results in the need for “reading glasses” around the age of 40 or 50 years.

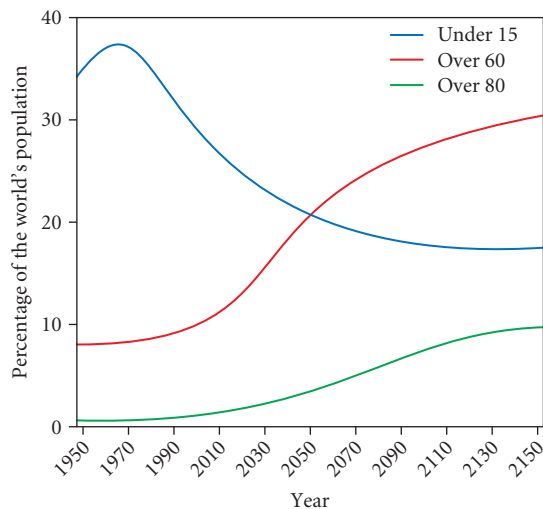
**Pseudophakia.** An eye with an artificial or intraocular lens implanted to replace the natural lens following its removal.

## 14.2 INTRODUCTION

Senescence occurs in the optics of all human eyes and the resulting changes in the spatial and spectral distribution of light comprising the retinal image can have a profound impact on vision. These optical changes have enormous economic consequences for health care and society, and although some effective treatments exist, optical quality in the senescent eye remains inferior to that of younger eyes. Before describing the optical and visual changes that characterize normal aging, it is useful to consider the scope of the problem in light of demographic changes in the world’s aging population.

## 14.3 THE GRAYING OF THE PLANET

The proportion of the world’s population that is elderly is higher now than at any other point in recorded history. In the last two decades, the number of individuals above age 65 years has increased at more than twice the rate as that of individuals below 65 years.<sup>1</sup> At present, every month, another 1 million of the world’s population enters its sixth decade of life. As a result, the number of individuals age 60 and above is likely to change from one in ten today to one in five by 2050, and one in three by the year 2150. As illustrated by Fig. 1, the global growth rate for individuals over age 80 is occurring



**FIGURE 1** Percentage of the global population projected from 1950 to 2150 for three age groups based on statistics provided by the United Nations.<sup>2</sup>

at an even faster rate. Indeed, the number of centenarians in the United States has doubled every 10 years since 1960. At the same time, there is a corresponding decline in the proportion of the population below 15 years of age.

Many factors have contributed to this remarkable demographic shift in the age distribution, including a change in the fertility rate, improved hygiene, nutrition, disease control, and education. These trends are generally more apparent in less developed countries, and the demographic shift in the elderly population is now more rapid in these parts of the world. For example, in 1995 there was a global increase of 12 million persons at or above age 60 and nearly 80 percent of it was due to a change in less developed countries. The countries with the fewest resources at present will thus have less time to cope with the social and economic consequences of an elderly population. It is not clear how long these present demographic shifts will continue into the future and when they will reach equilibrium.

## Some Implications of Living Longer

Some theoretical calculations place the maximum human life span at about 120 years, but there are reasons to think that this may not apply to the eye. The demographic shifts in longevity of the eye are apparently not keeping pace with that of the rest of the body. Research to Prevent Blindness estimates that the number of blind individuals will double between 1999 and 2030. If true, the social, health, and economic consequences will be profound. One eminent authority put it this way: “Mathematically speaking, it is permissible to say that those who prolong life promote blindness” (Ref. 3, p. 306).

**Health Implications and the Problem of Defining Normal Aging** Current projections for the future are that as the baby boomers (the cohort of 76 million Americans born between 1946 and 1964) start to turn age 65 in the year 2011, they will be better educated, have better health, and more financial resources than did their parents when they retired. All indications are that retiring baby boomers will read more and want to spend more time in outdoor leisure activities than their parents did at comparable ages. For the age group over 65 years of age, there are also some indications that the rates of disability and disease are decreasing compared to previous generations, but this is not so for vision. Older adults experience a higher rate of visual disability and ocular disease than younger individuals, and the rate accelerates after about 40 years of age. For example, glaucoma is about 8 times more common for individuals above age 65 than in the general population. In the Framingham study,<sup>4</sup> 50 percent of participants age 75 and over had at least one of four visual “diseases:” age-related macular degeneration, cataract, diabetic retinopathy, or glaucoma.

Many studies of “normal” aging eyes fail to include significant numbers of eyes older than 80 years (a notable exception being Haegerstrom-Portnoy<sup>5</sup>). Thus, most of the aging effects described in this chapter fail to describe very old (80 to 100 years) eyes. In addition, most studies of aging attempt to characterize the visual system with many factors controlled, such as by testing with optimal refraction rather than habitual refraction. These controls are essential for separating the factors underlying normal aging, but they necessarily overestimate the quality of vision experienced outside the laboratory.

There is little agreement about where to draw the line between senescent degradation in vision and some forms of ocular pathology. For many researchers, normal aging is defined by exclusion—the changes in vision that occur when overt ocular disorder and disease are not detected. An alternative distinction between aging and disease is based on the amount of tissue changed or when visual loss exceeds a criterion. Both of these definitions are used to define cataract, a loss of transparency of the lens resulting from cumulative changes in products of oxidation that can dramatically elevate intraocular light scatter. There are several different types of cataract that can degrade the optical quality of images formed on the retina and consequently undermine the quality of the “neural” image transmitted by the retina. The biochemical events leading up to these lenticular changes occur continuously across the life span so the presence of this disorder is largely a matter of where the criterion is set on the continuum of normal aging.

Most individuals will still be able to approach their golden years without frank visual pathology. However, no individual should expect to have the quality of vision experienced in youth. Indeed, perhaps 20 years prior to retirement, most individuals already will have experienced the need for

“reading glasses” to correct near vision due to the senescent reduction in or loss of accommodation. Some problems of optical origin can be corrected or compensated by optical means, but additional losses in vision may be expected due to neural changes for which there are no known remedies. Some current research has focused on factors that might contribute to normal aging of the eye that could be controlled by life style choices such as diet<sup>6</sup> or exposure to ultraviolet<sup>7</sup> and short-wave visible<sup>8</sup> radiation. Whatever the cause, any impairment of our dominant sensory system will inevitably have behavioral consequences. In some cases, this may be an inconvenience such as slower reading or the need for better lighting, while in other cases it may lead to more serious problems such as accidents, falls, and greater dependence on others for daily activities. For many, senescent changes in vision will become the dominant factor limiting their quality of life.<sup>9</sup>

***Economic and Social Implications of Aging Eyes*** There is a gradual reduction in accommodative amplitude throughout adulthood resulting in a complete loss of accommodation during the sixth decade of life. Combining this information with the fact that each day 10,000 people in the United States reach the age of 50, there is little wonder that ophthalmic corrections for reading and near work will account for approximately half the revenues in the multibillion dollar ophthalmic lens industry.

Visual impairment secondary to diseases of the eye also creates a sizeable economic impact on our society. For example, visual impairment and blindness for individuals above age 40 have an estimated cost of \$51 billion per annum in the United States alone.<sup>10</sup> In addition, approximately 2.8 million cataract surgeries are performed each year in the United States at an annualized cost of about \$25 billion. The costs of presbyopic spectacle correction may be even higher. These numbers are almost certain to increase due to the increasing proportion of the population that is elderly.

Additional economic consequences of our aging population will be created because the elderly generally require more special services, housing needs, and medical care, often as a result of visual impairment or blindness. For example, vision impairment increases the risks of falls by sixfold.<sup>11</sup> Senescent changes in vision are also likely to impose new burdens on public transportation systems in view of the fact that about half the world’s elderly population today live in cities, and that proportion is expected to increase to about three-fourths, at least in more developed countries. New demands will be made for housing these elderly individuals in order to promote independent living.<sup>12</sup> Being able to care for oneself is not only important for self-esteem, but it also has enormous economic impact. Architects and lighting designers will have to take into account age-related changes to the visual system in planning for new environments inhabited by the elderly. Current standards for lighting are almost entirely based on college-age students, although there have been new calls for explicit attention to the visual needs of aged individuals.<sup>13</sup> The goal of such standards would be to find conditions that can compensate for sensory losses, to be described later, while not introducing conditions that promote “disability glare.” Such lighting considerations might promote greater safety in the home to facilitate independent mobility and support communication through computers and reading and other visual tasks involving hobbies and work.

## 14.4 SENESCENCE OF THE EYE’S OPTICS

Senescent changes influencing image formation are prevalent in all optical components of the eye. The following sections describe some of these changes in detail.

### Senescent Changes in the Optical Components of the Eye

***Tears*** The anterior surface of the eye is bathed by tears that produce an optically smooth air-tear interface mirroring the global shape of the cornea. The refractive index of the tears approaches that of the underlying corneal epithelium, therefore, the optical impact of the rough corneal epithelium surface is essentially eliminated by index matching. Because the tear film is so thin ( $3\ \mu\text{m}$ )<sup>14</sup> it is vulnerable to evaporation effects, and it is inherently unstable. Stability and uniformity are enhanced by a



surface monolayer of lipids and blinking. Insufficient blinking and/or inadequate tear volume or lipid coating can cause the uniform tears to “break-up” into a patchwork of tears leaving regions of dehydrated mucins and the optically rough epithelium exposed, resulting in increased optical scatter and aberrations<sup>15</sup> and reduced visual performance.<sup>16</sup> The presence of dry eye is an indicator of inadequate tears, and the prevalence of dry eye syndrome increases with age.<sup>17</sup> With increasing age, tear production by the accessory and main lacrimal glands decreases,<sup>3</sup> the lipid layer gets thinner,<sup>18</sup> corneal innervation and sensitivity decline,<sup>19</sup> and in combination, these factors lead to reduced tear production<sup>20,21</sup> and blink rates. All of these factors may contribute to reduced optical quality of tears in older eyes, which may be most pronounced during activities such as reading which lowers blink rates.

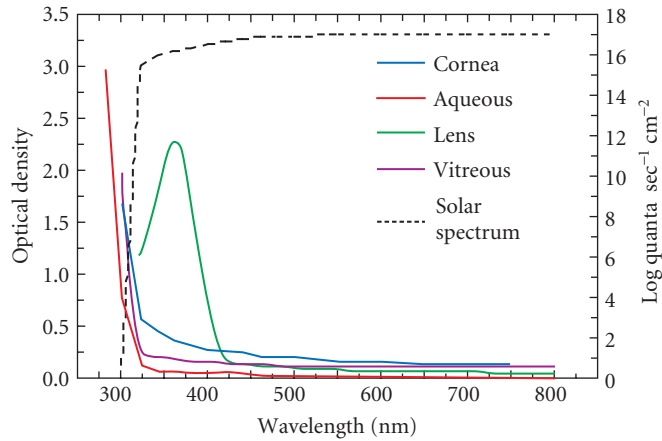
**Cornea** The human cornea consists of three discrete layers separated by two basement membranes. The outermost epithelial layer, composed of basal, wing, and surface cells, acts as a barrier to the external environment and is subject to a rapid rate of renewal, with the result that its cells are replaced every 5 to 7 days.<sup>22</sup> The middle layer, or stroma, is a matrix of collagen fibers as well as fibroblast-like cells called keratocytes and accounts for approximately 90 percent of the corneal thickness. The stroma is sandwiched between two membranes. Anterior and posterior to the stroma are Bowman’s and Decemet’s membrane, respectively. The transparency of the stroma is due to the uniform density and subwavelength spacing of collagen fibers, and any disruption of this order caused by such factors as increased hydration will cause light scatter and loss of transparency. Because of the tendency of the stroma to imbibe water there must be a way to regulate corneal hydration. The extent of corneal hydration is largely controlled by energy-dependent-ion coupled fluid transport mechanisms located at the innermost layer of the cornea, the corneal endothelium. The endothelium is made up of a single layer of cells. At birth, the corneal endothelium has about 4500 cells/mm<sup>2</sup>, but the number is reduced to about 2000 to 2500 cells/mm<sup>2</sup> by age 70 or 80. Although the surviving cells become larger to maintain approximately constant total tissue volume, a smaller total number of fluid transport mechanisms remain in the corneal endothelium. As a result, proper corneal hydration is sometimes a problem in the elderly. The rate of decline in endothelial cell density is accelerated by cataract surgeries,<sup>23</sup> but for the typical eye sufficient endothelial cells remain throughout life.

The cornea is the most powerful refractive element in the human eye, accounting for about 40 to 45 D of optical power. Although it has a lower refractive index than the lens, the difference of refractive indices, which is directly proportional to dioptric power, is highest at the air-tear/cornea interface. The cornea is not spherical; its reduced curvature in the periphery helps reduce the spherical aberration of the cornea. The refractive index of the cornea<sup>24</sup> and curvature of its anterior surface<sup>25</sup> appear to undergo little change with age, but with aging the axis of corneal astigmatism loses the “with the rule” bias seen in younger eyes.<sup>26,27</sup>

As can be seen from Fig. 2, the cornea absorbs most of the ultraviolet (UV) radiation below about 320 nm. Absorption of this radiation in sufficient intensity can damage the cornea (e.g., actinic keratitis), but long-term effects are rare because damage is typically confined to the corneal epithelium which is capable of rapid renewal. Stem cells in the corneal margins continue to generate epithelial cells, which then migrate across the cornea to generate the basal layer of the epithelium. This may be why the absorption spectrum of the cornea undergoes only small changes with advancing age.<sup>28</sup> Cumulative absorption of UV light may, however, lead to corneal abnormalities such as pterygium, a condition that involves the advancing growth of abnormal tissue from the border between the cornea and sclera, the limbus, onto the clear cornea. This increases in prevalence with increasing age.<sup>29</sup>

**Anterior and Posterior Chambers** The cornea and lens are separated by a fluid-filled space. This fluid (the aqueous humor) maintains the metabolism of the cornea and lens, as the transparency requirements of these structures necessitate that they have no blood supply. The much larger volume between the cornea and iris is the “anterior” chamber, and the smaller volume between the iris and the lens is the “posterior” chamber. The aqueous humor is produced by the ciliary body around the margins of the posterior chamber, flows to the anterior chamber, and is drained through the trabecular meshwork and into Schlemm’s canal around the radial margins of the anterior chamber. The optical role of the aqueous, therefore, is twofold. It enables the avascular lens and corneal endothelium to function (and thus retain lens and corneal transparency) while maintaining its own

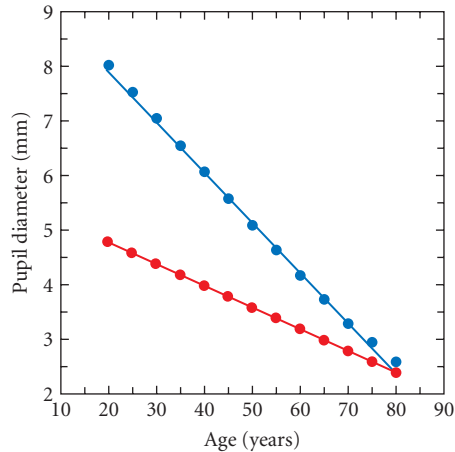




**FIGURE 2** Optical density (decadic logarithm of the reciprocal of transmittance) of human ocular media is plotted as function of wavelength (left ordinates; from Boettner and Wolter<sup>30</sup>). Log quanta in sunlight reaching the surface of the earth from an angle of  $30^\circ$  from the horizon is plotted as a function of wavelength (right ordinates; from Brandhorst et al.<sup>31</sup>). (After Werner.<sup>32</sup>)

transparency. The optical density of the aqueous changes negligibly with age<sup>30</sup> and any damage sustained by the aqueous resulting from absorption of light is likely to be unimportant for senescence because it is rapidly replaced.<sup>22</sup> Although there are pathological conditions that directly affect aqueous transparency (e.g., uveitis in response to infection), they are primarily transient and not related directly to age. There is a senescent reduction in both the production and drainage of aqueous,<sup>33</sup> and an increased likelihood of failure of the homeostatic production and drainage system. As drainage is unable to keep up with production, intraocular pressure rises, and glaucoma can ensue (see section “Glaucomas”). One optical side effect of glaucoma can occur when fluid in the corneal epithelial layer builds up between the basal epithelial cells,<sup>34</sup> causing colored diffraction halos similar to those generated by nocturnal eye closure, contact lens wear, or prolonged exposure to fresh water during swimming. A healthy epithelium maintains its high transparency by tight junctions between adjacent cells; however, as fluid builds up between the regular array of cells, classic wavelength-dependent diffraction occurs which is visible as colored halos. Glaucoma is generally associated with corneal epithelial edema only when pressures become very high, which is more typical of acute angle closure glaucoma. Interestingly, a much rarer condition in older eyes, ocular hypotony, can generate loss of corneal transparency because of stromal edema resulting from IOP being too low.<sup>35</sup> Both cases emphasize the critical role played by intraocular pressure in maintaining corneal hydration and thus corneal transparency.

**Pupil** The dilator and sphincter muscles of the iris control the size of the pupillary aperture that, for young adults, varies in diameter with the level of the prevailing illumination from approximately 2 to 8 mm. As in any optical system, the size of the aperture has important consequences for image formation. The smallest pupillary diameters can decrease optical quality through diffraction effects. The optimal pupil diameter is about 2.4 mm from the point of view of classic image quality.<sup>36</sup> Traditionally, larger pupils at lower light levels are considered to represent a compromise between improved sensitivity (more light entering the eye) and diminished image quality (more aberrations in peripheral optics). However, this view ignores the very significant role of photon noise on image quality.<sup>37</sup> Including photon noise in image calculations, Laughlin<sup>38</sup> argued that pupil dilation as light levels drop ensures optimal image quality. Employing a similar hypothesis (pupil size is adjusted to optimize retinal image quality), pupil miosis that accompanies senescence may also be considered as an adaptation to poorer optical quality in older eyes.



**FIGURE 3** Mean diameter of the human eye pupil is plotted as a function of age. Blue and red symbols show dark- and light-adapted values, respectively. (Data from Kornzweig.<sup>39</sup>)

Maximal pupil diameter occurs in adolescence and then progressively decreases with increasing age. As shown in Fig. 3, age-related change in diameter differs for dark-adapted and light-adapted conditions. Changes in pupil diameter with light level provide a mechanism of adaptation through adjustment of visual sensitivity with illumination. However, the change from maximum to minimum size influences retinal illuminance by only about a factor of 10, so the pupil plays only a small role in adaptation mechanisms that allow us to function over a range of ambient illumination covering about 12 log units.

Age-related changes in pupil size are noticeable in the elderly but actually begin in the early adult years. For example, when fully dark adapted, there is essentially a monotonic decrease in pupil diameter beyond 10 years of age.<sup>40</sup> Spatial vision depends on light level, and although a smaller pupillary diameter allows less incident light to fall onto the retina, it is unlikely that a smaller pupil accounts for significant age-related changes in spatial vision. For example, it has been demonstrated that age-related differences in contrast sensitivity under photopic and mesopic ambient light levels cannot be explained solely by differences in pupillary diameter.<sup>41</sup> In fact, there actually are some benefits associated with an age-related reduction in pupil diameter. The smaller pupil of older persons diminishes age-related changes in higher-order aberrations<sup>42</sup> and the optical transfer function of the eye at low luminance levels.<sup>43</sup> In addition, the smaller pupil in the elderly reduces the diameter of the retinal blur circle, thereby increasing the depth of focus by about 0.5 D between the ages of 30 and 70 years.<sup>44</sup> More significantly, since blur circle size (radians) = defocus (Diopters) \* pupil diameter (m),<sup>45</sup> as pupil size is halved, blur circle size is also halved. For example, if defocus was 4 D, then the impact of halving the pupil would be the same as halving the defocus, or a 2-D change in image quality. Thus halving of the photopic pupil and the more than factor of 3 reduction in mesopic pupil diameters in older eyes will greatly reduce the level of blur. Senescent pupil miosis may be nature's attempt at minimizing the impact of presbyopia.

**The Lens** The lens consists of only one cell type, but its thickness increases throughout the entire life span due to the accumulation of cells to form a series of concentric layers.<sup>46</sup> This results in a continuous axial thickening and reduction in the depth of the anterior chamber.<sup>47</sup> The oldest cells are thus found in the center or nucleus (indeed these are the oldest cells in the body in that they are born early in gestation and are never replaced), while the youngest cells comprise the outer or cortical layers. As the cells of the lens age, they lose their organelles and become essentially metabolically inert.

Molecules that absorb UV and short-wave visible light are present in the newborn lens and their numbers increase continuously over the life span.<sup>48</sup> Thus, the curve for the lens shown in Fig. 2 can be, to a first approximation, scaled multiplicatively with age.<sup>49</sup> A number of studies have measured the optical density of the human lens as a function of age. Some have described these changes in density by a linear function of age<sup>49,50</sup> while other evidence demonstrates a bilinear function due to an acceleration in the aging function after about 60 years.<sup>51</sup> The precise form of the aging function is difficult to determine because of substantial individual differences at every age (on the order of 1.0 log unit at 400 nm). However, it should be noted that the average infant lens transmits about 25 times more light at 400 nm to the retina than the average 70-year-old eye!

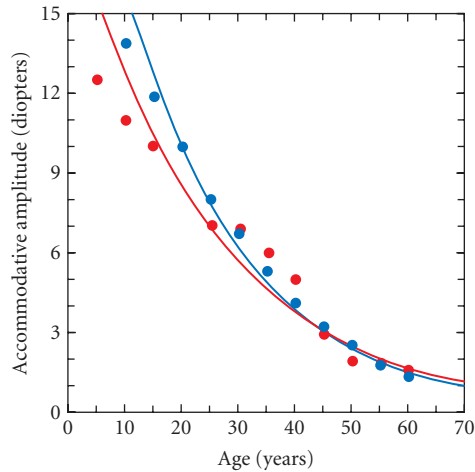
*Transparency/cataract* Cataract refers to an opacity of the lens that decreases vision by a somewhat arbitrary criterion amount. An acceleration in the population mean rate of change in lens optical density beyond approximately 60 years of age could be due to the inclusion of subjects having cataract.<sup>52</sup> However, it is not clear where normal aging ends and cataract begins because this increase in optical density continues throughout life. In his comprehensive monograph on this subject, Young<sup>53</sup> summarized a substantial literature which lead him to conclude that: “No discontinuity between senescent and cataractous changes can be detected at the molecular level in the human lens....The most sophisticated techniques of modern biophysical and biochemical analysis have so far failed to uncover any feature of the cataractous lens that suggests cataract is anything more than an advanced stage of normal aging” (Ref. 53, p. 56).

It should be added that there are several forms of cataract depending on the part of the lens affected (nuclear, cortical, posterior subcapsular) or etiology (congenital, age-related, chemically induced, trauma induced, disease induced). Losses of lenticular transparency can sometimes occur through occupational hazards associated with infrared light exposure such as in “glassblowers cataract” due to the conduction of radiant heat that is absorbed by the cornea and, to a lesser extent, the iris.<sup>54</sup> Most often, changes in the transparency of the lens are associated with cumulative exposure to ultraviolet radiation. Weale<sup>55</sup> has proposed that the specific form of cataract depends on genetic factors while the common etiology is most often cumulative exposure to solar radiation (see section “Life-Span Environmental Radiation Damage”).

Currently, the only effective treatment for cataract requires surgical removal of the lens and in more than 90 percent of such surgeries performed in the United States an intraocular lens (IOL) is implanted in its place. A properly selected IOL obviates the need to wear thick spectacle corrections or contact lenses for the hyperopia that would otherwise ensue. Most of these lenses now contain UV-absorbing chromophores, but this is a relatively recent design improvement.

*Accommodation and presbyopia* Accommodation refers to the ability of the eye to adjust its power to focus the images of objects located at varying distances in the environment. Contraction of the ciliary muscle located within the ciliary body of the eye allows the anterior, and to a lesser degree the posterior, surface of the lens to change shape thereby altering the dioptric power of the lens (see recent review by Charman<sup>56</sup>). The amplitude of accommodation refers to the *dioptric* difference between the farthest and nearest distances in which an object remains in focus. Age-related loss in accommodative amplitude (Fig. 4) has been attributed to various causes including reduced ciliary body function, changes in the geometric relationships of the lens and the surrounding zonules and ciliary body based on progressive circumferential enlargement of the crystalline lens, but primarily it is due to changes in the elastic properties of the lens and to a lesser extent the capsule.<sup>56,57</sup> The red symbols in Fig. 4 are from the Basel longitudinal study of individuals followed for 20 years, with different subjects starting the study at different ages.<sup>58</sup> These data agree well with the classic data of Donders<sup>59</sup> shown by blue symbols.

As the amplitude of accommodation decreases, a near correction for tasks such as reading may be required. This typically occurs within the fifth decade of life and is known as presbyopia (meaning “old sight”). Because the accommodative amplitude changes continuously, the definition of presbyopia and its onset are somewhat arbitrary. The two sets of data in Fig. 4, although separated in time by about a century, are from similar geographic locations. Age of onset of presbyopia depends on many factors including geographical latitude and average ambient temperature.<sup>3</sup>



**FIGURE 4** Amplitude of accommodation (diopters) is plotted as a function of age. Note that the tasks used for measurement involved both accommodation and depth of focus (which the clinical literature often calls pseudoaccommodation). This depends on both optical and neural factors. Red symbols (curve) from Brückner et al.<sup>58</sup> and blue symbols (curve) from Donders.<sup>59</sup> Curves are based on best-fitting exponential functions.

**Vitreous** Separating the lens from the retina is the vitreous chamber. This space, which occupies about 80 percent of the eye's volume, is filled with a colorless gelatinous substance called the vitreous humor. It is made up of about 99-percent water mixed into a gel-like mesh of collagen fibers and hyaluronic acid. Its role is primarily structural and not metabolic. In order to function it must retain its volume and remain transparent. As the vitreous ages, the gel is gradually replaced by unbound water. This liquefaction of the vitreous is responsible for the increased movement of the vitreous during and following eye movements in the aging eye.<sup>61</sup> Local shrinkage of the vitreous gel is responsible for the increased prevalence of posterior vitreous detachment which can be present in up to 65 percent of eyes older than 65 years.<sup>60</sup> The liquefaction and shrinkage can lead to increased numbers and movement of floaters (visible shadows and diffraction phenomena associated with clusters of fibers and other optical inhomogeneities in the vitreous) consistent with increased movement of the vitreous in old eyes.<sup>61</sup>

**Retina** The optical role of the retina is complex. Its primary optical role is to absorb photons in the photopigment. However, the secondary optical role reflects needs for transparency, opacity, and waveguiding. Because the retina is "inverted," light must pass through the neural retina prior to reaching the photopigment. Thus, the neural retina must remain highly transparent. For example, it is for this reason that retinal ganglion cell axons are generally unmyelinated until they have exited the eye. Even though most of the retinal tissue is transparent, the blood supply to the postreceptoral neurons remains opaque. The lateral displacement of the foveal postreceptoral neurons removes the need for any retinal blood supply in the central fovea. Thus, most eyes have a foveal avascular zone with an accompanying increase in transparency and presumably image quality in the central fovea.

External to the layer of photopigment (which resides in the outer segments of the photoreceptors) lies a second layer of pigment, melanin, in the pigmented epithelium cells (melanin and hemoglobin in the choroid are also effective pigments). It has been suggested that melanin in the RPE has a primary antioxidant role protecting the retina from age-related oxidative damage.<sup>62</sup> Optically, this pigment plays the same role as the black paint used to coat the inside of a camera; it reduces the

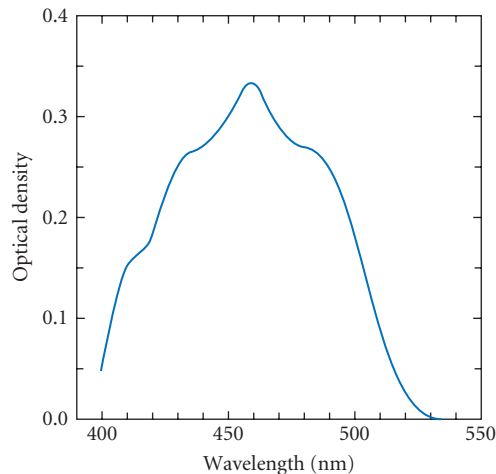
reflection of photons that are not absorbed by the photopigment. For example, at the peak sensitivity of the photopigment, about 2/3 of the photons entering the outer segment are absorbed by the photopigment (0.5 optical density), and the outer segments cover about 70 percent of the retinal area, and thus, slightly less than 50 percent of photons arriving at the retina will be absorbed. This percentage is lower in the peripheral retina where the outer segments are shorter, and will generally be lower due to photopigment bleaching (which never exceeds 50% under natural conditions), and of course will be lower at wavelengths other than the peak. The effectiveness of the photopigment and the pigment epithelium combine to improve image quality of the eye but make it difficult for clinicians to obtain high intensity images of the fundus as little light is reflected back out of the eye.

A third major pigment in the retina, most dense around the fovea, is a yellow pigment called the macular pigment (MP). It is located in the receptor fiber and inner plexiform layers<sup>63</sup> where it selectively absorbs short-wavelength light en route to the photoreceptors. Figure 5 shows the spectral absorbance of the foveal macular pigment for an average observer. The peak density is at 460 nm, near the peak sensitivity of the short-wave-sensitive cones (440 nm).

Several different methods (fundus photography, entoptic visualization, and psychophysics) indicate that the peak density of the MP is at the fovea and follows an exponential decay to negligible levels at approximately 5 to 10° retinal eccentricity.<sup>65</sup> There are large individual differences in the absorption of the macular pigment (from about 0 to 1.0 log unit at 460 nm), but there is little age dependency after early childhood.<sup>66,67</sup>

While the function of the MP is still uncertain, it has been suggested that the presence of this pigment improves the optical image quality at the retina and may aid in maintaining the health of the macula. Because MP selectively absorbs short-wave light, it has been proposed that it helps to minimize the effects of chromatic aberration<sup>68</sup> thereby improving polychromatic image quality in the central retina. With regard to the health of the eye, MP is made up of carotenoids that tend to neutralize photosensitized molecules<sup>69</sup> and inhibit free radical reactions that are toxic to retinal cells. Thus, it has been suggested that the MP may protect the retina from actinic damage that may lead to age-related changes in visual sensitivity and to the development of age-related macular degeneration.<sup>8</sup>

A fourth, rhodopsin-based, visual pigment has recently been observed in some retinal ganglion cells. It has peak sensitivity around 460 nm, and it has been suggested that it may be responsible for light-induced melatonin suppression and thus maintenance of the circadian rhythm. In a recent



**FIGURE 5** Optical density of the human macular pigment is plotted as a function of wavelength. (Data from Vos.<sup>64</sup>)

analysis, Charman<sup>56</sup> showed that the combined effects of senescent miosis and lens yellowing would reduce the photon catch of this pigment in old eyes to only 10 percent of that of younger eyes and thus may contribute to age-related changes in circadian rhythms.

In addition to the transparency and pigmentation properties described above, the retinal photoreceptors have important optical waveguide properties. Photoreceptors must contain large numbers of photopigment molecules because large numbers of photons need to be captured, and photopigment regeneration is slow (half-life of 2 and 7 min for cones and rods, respectively). Also, in order to adequately sample the retinal image spatially, photoreceptors must be small. Typical photoreceptors contain long tubes filled with pigment with their axis aligned approximately perpendicular to the image plane (e.g., outer segments of foveal cones can be 60  $\mu\text{m}$  long and only 2  $\mu\text{m}$  in diameter). Such geometry creates fiber-optic or waveguide properties in these cells. These waveguide properties play an important role in reducing the impact of highly aberrated and scattered light so that they maximize the probability of photon catch for light entering in and around the pupil center [the Stiles-Crawford effect (SCE) see Chap. 8], and in so doing, enhance the optical quality of the retinal image.<sup>70</sup> A recent report<sup>71</sup> has demonstrated that retinal Müller cells may also act as optical fibers funneling light through the retina to the photoreceptors to avoid scattering by the neural and vascular cells of the retina. No significant senescent changes occur in the SCE implying stability in photoreceptor orientation.<sup>72,73</sup>

There are several senescent changes in the optics of the retina. The depth of the foveal pit decreases with increasing age,<sup>74</sup> while the pigmented epithelial cells accumulate lipofuscin throughout life as a by-product of phagocytosis of photoreceptor outer segments. Over age 70, lipofuscin and melanolipofuscin granules may occupy up to 20 percent to 33 percent of the free cytoplasmic space of the cell.<sup>75</sup> Also, between the ages of 10 and 90 years, there is a 2.5-fold decrease in melanin concentration within the retinal pigmented epithelial cells.<sup>76</sup> All of these changes can alter the final destination and distribution of photons within the retina. In addition, these cellular alterations are manifest in an age-related increase in lipofuscin autofluorescence,<sup>77</sup> which may serve as a possible marker for cellular aging, oxidative damage, and dysfunction associated with retinal disease.<sup>78</sup> The combined impact of cellular senescence on the retinal optics is unknown, but it likely contributes to the increased scatter observed in older eyes.<sup>79</sup>

**Eye Size** The globe of the eye grows rapidly in the first year of life and then more slowly until about 5 years of age when it reaches an asymptotic sagittal length of about 23.5 mm. While the globe of the eye changes relatively little over the adult years, the lens continues to grow. These and other structural changes alter the refractive state of the eye throughout the life span. The coordinated growth of eye size and optical power in many eyes enables a focused retinal image without need for accommodation or refractive treatments. Such eyes are said to be emmetropic, but emmetropia is not the norm during all stages of development and senescence. Infant eyes are almost always hyperopic but during the first 2 years of life they shift toward emmetropia. For many older children and adults, a further slow shift toward myopia relegates many to an adult life of near sightedness (e.g., recent estimates indicate that myopia rates can be as high as 71 and 84 percent in Hong Kong<sup>80</sup> and Taiwan,<sup>81</sup> respectively), whereas in later life there is an overall tendency for the refractive error of the eye to shift in a hyperopic direction. Although the prevalence of emmetropia can be high in modern populations, both eyes tend to maintain highly correlated growth resulting in similar refractive states in the two eyes. Recent reports, however, indicate that this correlated growth pattern begins to fail as the eyes age,<sup>82</sup> and can result in significant interocular differences in refractive error (anisometropia) among the population older than 70 years,<sup>83</sup> compared to children and younger adults.

## Senescent Changes in Retinal Image Quality

Many factors act to alter the retinal image throughout the life span. In this section, their combined effects are described along with one aspect of the retinal image (chromatic aberration) that does not seem to be age dependent.

**Intraocular Scatter** The spatial distribution of light on the retina from a point source is described by the point spread function (PSF). The fraction of light transmitted from a point source in the forward direction is the total transmittance, while the fraction transmitted within a cone of a particular angle  $a$ , is the direct transmittance. The value of  $a$  is arbitrary, but typically is about  $1^\circ$ . The difference between the total and direct transmittances produces a “tail” in the PSF and is called stray light. However, because the photoreceptors are directionally selective (see Chap. 9), the optical PSF is not necessarily equivalent to the functional PSF. Westheimer and Liang<sup>84</sup> have used psychophysical methods to measure the PSF of younger and older observers for a broadband source. The total scattered light from  $7'$  to  $90^\circ$  is about 5 times larger in a 69-year-old observer compared to a young adult. Put another way, 90 percent of the light from a point source falls within a region of  $7'$  for a young subject, but only 50 percent of the light falls within  $7'$  for the older subject. (This analysis of the PSF does not take into account differences in the density of the ocular media which would further increase the difference between young and old.)

A number of studies have shown that large particle scatter by the human ocular media increases across the life span (e.g., Wolf<sup>85</sup>). The relation between light scatter and age has been described as nonlinear due to an acceleration in the function in middle age.<sup>86</sup> A more recent study<sup>87</sup> measured the integral of the scatter function of the eye from 2.2 to  $90^\circ$ . The results show little or no increase in scatter from 16 to 40 years of age but a rapid increase afterward (i.e., a factor of 5 increase at 65 years compared to young adults).

Stray light in the human eye is not strongly related to wavelength,<sup>88,89</sup> implying that it is not due to scatter by small particles (smaller than the wavelength of light). A small wavelength dependence, greater at long than middle wavelengths (opposite to Rayleigh scatter, where scatter =  $k\lambda^{-4}$ ), has been attributed to fundal reflections, especially in lightly pigmented eyes.<sup>90</sup> A recent study<sup>91</sup> has proposed that intraocular scatter is a combination of small particle Rayleigh scatter and a red dominated scatter produced by light entering through the sclera. This latter source must pass through the choroidal blood and thus is red biased. The combination of these effects in lightly pigmented eyes produces scatter levels that vary little with wavelength. This can be important in evaluating stray light in the elderly because pigmentation levels in the RPE decrease with age. The lens seems to be the main source of age-related change in intraocular stray light.<sup>85</sup> The cornea produces light scatter as well, but the amount does not change significantly with age.<sup>30</sup>

Scattered light in the elderly may also be due to the age-related accumulation of two fluorophores in the lens.<sup>92</sup> These are molecules that emit light at one wavelength (e.g., 520 nm) when irradiated by another (e.g., 410 nm). As these emissions are scattered randomly, they would be expected to degrade visual resolution. Elliott et al.<sup>93</sup> demonstrated a small effect of fluorescence on resolution of low contrast targets.

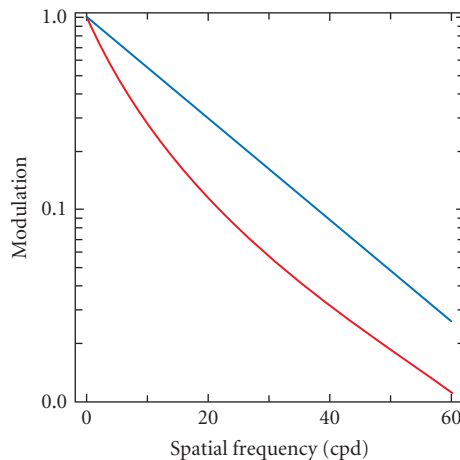
Intraocular stray light reduces the contrast of the retinal image. Normal levels of scattered light do not affect color detection thresholds significantly, but high levels can affect chromatic saturation and estimates of color constancy.<sup>94</sup> The primary effect of intraocular stray light is to impair visual performance when lighting varies dramatically within a visual scene, such as when driving at night, because even small amounts of scatter from a high intensity source can dramatically reduce contrast of the retinal images of nearby low intensity targets. The effects of light scatter are sometimes referred to as disability glare. Discomfort glare, on the other hand, describes a subject's visual discomfort when high intensity glare sources are present. The level of discomfort glare shows large intersubject variability, but it is generally correlated with the amount of disability glare. The older eye is not necessarily more sensitive to glare, but it experiences more of it due to greater scatter.<sup>79,95</sup> Low luminance and low contrast visual acuity exhibit marked deficits with aging, which is further exacerbated by glare (veiling auto headlights, and so on). With the exception of cataract surgery, there are no treatments for increased ocular scatter, and therefore older people suffer the visual consequences of scatter or modify their behavior to avoid its most serious consequences, for example, changing night driving behavior.

**Monochromatic Aberrations** Almost 95 percent of the wavefront error measured monochromatically is caused by the lower-order aberrations of defocus and astigmatism,<sup>96</sup> and even in well-refracted eyes these lower-order aberrations continue to dominate.<sup>97</sup> However, there has been considerable interest generated recently in the higher-order aberrations of the eye.



Although the typical older eye exhibits appreciable levels of higher-order aberration, this results less from a specific ocular surface than from a decreased ability of lenticular aberrations to compensate for corneal aberrations—in contrast to younger eyes. With aging there is a change in meridian of the corneal astigmatism, the cornea develops negative spherical aberration, and the corneal coefficients for coma change as well. The changes are not accompanied by compensatory lenticular changes, resulting in greater whole-eye aberrations.<sup>27,98</sup> Thus, senescent changes in sign and not the magnitude of both the cornea and lens aberrations are the primary contributor to increased ocular aberrations seen in older eyes. The age dependence of this trend, however, is not completely clear for two reasons. First, population data show large variability between eyes, and most studies clearly show that the distribution of aberrations in older eyes is mostly overlapping with the distribution of younger eyes.<sup>26,99</sup> Thus, many older eyes have young eye levels of higher-order aberrations. Second, although most studies generally indicate best optical quality in young adult eyes (20 to 30 age group), one shows that optical quality is maximum at around age 40.<sup>99</sup> Finally, there is a shortage of data from eyes above 65 years.

A quantitative description of the quality of an optical system is provided by its modulation transfer function (MTF), the relation between contrast attenuation and spatial frequency. The eye's MTF can be derived from *in vivo* measurements of the optical point spread function on the retina. Guirao et al.<sup>43</sup> derived the MTFs from carefully collected point spread functions of 60 observers ranging in age from 20 to 70 years while controlling for such important factors as pupil size and the refractive state of the eye. In Fig. 6, the blue and red curves represent MTFs for observers with mean ages of 24 and 63 years, respectively. Modulation is scaled logarithmically to highlight differences at middle and high spatial frequencies that are obscured on a linear plot. Optical performance decreases with increasing spatial frequency for both age groups, and there is an age-related loss in performance at all spatial frequencies. Results similar to these have also been reported by Artal et al.<sup>100</sup> The results in Fig. 6 may be attributed to optical aberrations but not ocular scatter because it is factored out in the measurement procedure. Inclusion of intraocular scatter would tend to produce a more or less uniform “halo” in the retinal image that further reduces the contrast of stimuli in the elderly eye.



**FIGURE 6** Modulation transfer functions for average individuals at 23 (blue curve) and 63 (red curve) years of age for a 3-mm pupil, 543-nm stimulus, calculated from equation and parameters from Guirao et al.<sup>43</sup>



Perhaps the most important finding is that a smaller pupil can compensate for the age-related increase in monochromatic aberrations. Indeed senescent miosis may provide a complete compensation for the failed correction of corneal aberrations by lenticular aberrations. That is, one compensation mechanism evolves as the other disappears, and this does not include additional compensation mechanisms at a neural level.<sup>101</sup> The bottom line is that the major loss in *optimal* retinal image quality with age will not be due to increased monochromatic or chromatic aberrations.

**Chromatic Aberration** The refractive index of the ocular media is wavelength dependent so that the spectral components of a broadband stimulus will not be imaged in the same plane within the eye. This is an example of chromatic aberration. Chromatic aberration can be appreciated from the colored fringes that can be observed in many telescopes and microscopes. Refractive indices of the optical elements of the eye decline as wavelength increases moving the long wavelength images farther from the cornea. The degree of chromatic aberration can be quantified in terms of the lens power required to image a particular wavelength on the retina when the eye is focused for a reference wavelength (change in refractive error as a function of wavelength). With a 578-nm reference, the aberration varies from  $-1.70$  D at 405 nm to 0.40 D at 691 nm.<sup>102</sup> Howarth et al.<sup>103</sup> (1988) measured the longitudinal chromatic aberration for four younger (27 to 33 years) and six older (48 to 72 years) participants using three different methods. All methods failed to show a change in axial chromatic aberration with age.

Interestingly, one suggestion for treating presbyopia employs a hyperchromatic lens design which can double the effective chromatic aberration of the eye. Experimental studies have shown that such a lens can increase the depth of field of the presbyopic eye.<sup>104</sup> Conversely, diffractive IOL designs have been proposed that can cancel the refractive chromatic aberration of the human pseudophakic eye.<sup>105,106</sup>

## Summary of Optical Senescence

Senescent changes in the eye's optics are widespread and varied, but the overall effects are dominated by three important changes. First, the reflexive autofocus feature of the eye (accommodation), which allows objects at any distance to be well focused on the retina, slowly declines with age and disappears altogether during the sixth decade of life. From the fifth decade onward, the inability to focus objects in our three-dimensional world is a major source of retinal image degradation (due to spherical defocus). Second, due to senescent changes in lens and other ocular media transparency,<sup>107</sup> and pupil miosis, retinal image illuminance is chronically lower in older eyes, and this is particularly true for short wavelength light sources. Interestingly, the widely available low transparency yellow filters used in many sunglass prescriptions that mimic the older eye transmission characteristics indicate that this aging effect will have little impact in high light environments. The third significant and widespread optical change with age is the increased optical scatter (forward scatter by the eye's optics, and increased scatter by the retina and through the sclera due to reduced pigmentation). Scatter will always have the greatest impact at night when environmental light levels can vary by huge amounts from one location to another. In combination, the reduced transparency and the increased scatter will compromise vision at night. Also, the loss of accommodation and increased scatter will compromise retinal image quality when reading at near, particularly with high gloss materials that reflect light specularly, and thus can create local high intensity regions on the page.

## 14.5 SENESCENT CHANGES IN VISION

Age-related changes in visual performance are due to both optical and neural factors. Changes in performance that cannot be explained solely by optical factors are assumed to be due to senescent changes in neural pathways. An understanding of neuro-visual senescence thus depends on an understanding of the optics of the aging eye.

## Senescent Changes in Sensitivity under Scotopic and Photopic Conditions

Visible radiation in natural viewing conditions is normally limited to the band between 400 and 700 nm. As can be inferred from Fig. 2, the limit at short wavelengths is due to the absorption characteristics of the ocular media; the limit at long wavelengths is due to the absorption spectrum of the photopigments contained in the rod and cone receptors.

Various studies using psychophysical procedures have demonstrated age-related losses in the sensitivity of scotopic and photopic mechanisms. Dark adaptation functions (absolute threshold versus time in the dark) have two scalloped-shaped branches, the first branch being mediated by cone photoreceptors and the second mediated by rods. A number of early studies have shown that the asymptotes of each of these branches become elevated as a function of age. In addition, it has been noted that the rate of photopic and scotopic dark adaptation decreases with increasing age.<sup>108</sup> This latter result has been confirmed by more recent studies.<sup>109,110</sup> It was not clear from some of the early studies whether these effects were entirely due to age-related changes in retinal illuminance caused by, among other factors, smaller pupils (see subsection “Pupils” in Sec. 14.4). Most, but not all,<sup>111</sup> of the studies that have taken age-related changes in pupil size and ocular media density into account have demonstrated age-related losses in scotopic sensitivity.<sup>112,113</sup> Based on those studies that have shown losses in sensitivity that cannot be explained solely by preretinal changes, it can be inferred that losses in sensitivity reflect neural changes that take place at or beyond the level of the photoreceptors.

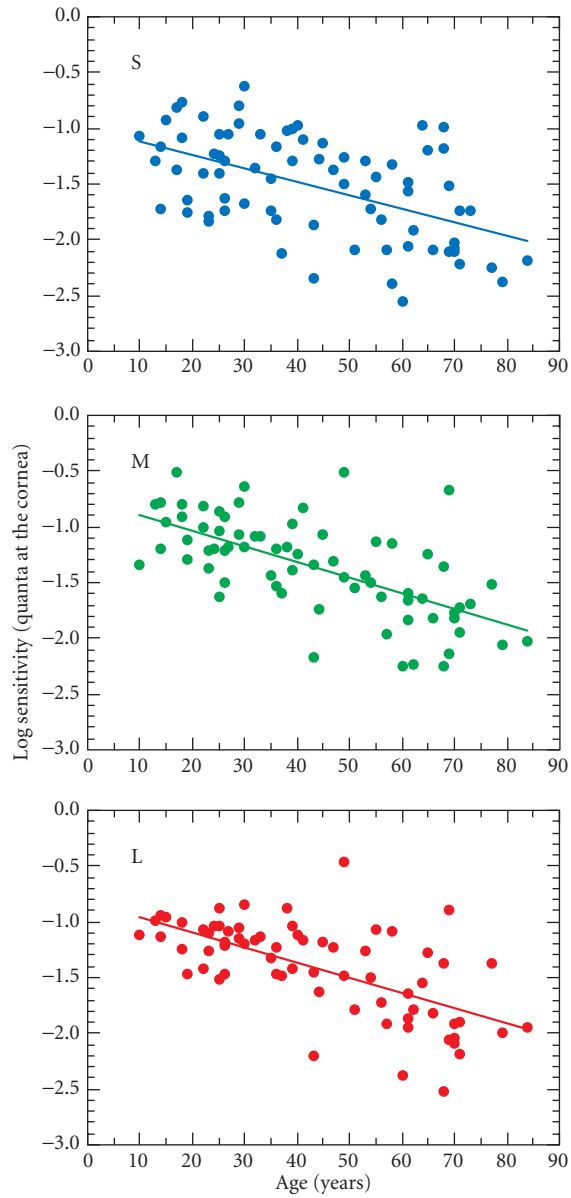
Several studies indicate that changes at the receptor level may partially account for age-related changes in light and dark adaptation and for overall sensitivity under scotopic and photopic conditions. *In vivo* measurements of the photopigment kinetics indicate that a change in the rate of dark adaptation is due, at least in part, to a decrease in the rate of photopigment regeneration.<sup>109,114</sup> Other changes at the receptor level involve the relative ability of “aging” rods to efficiently capture light quanta. Although there are substantial losses in numbers of rod photoreceptors with age,<sup>115</sup> the amount of rhodopsin photopigment in the retina decreases only slightly, if at all, with increasing age.<sup>116</sup> This implies that individual surviving rods must contain more photopigment in the older retina. However, some outer segments of these aged photoreceptors undergo morphological changes such that they appear to be more convoluted in shape.<sup>117</sup> These relatively deformed outer segments may contribute to reductions in scotopic sensitivity because rhodopsin molecules now possess less than optimal orientations for absorbing light.

Photopic vision is mediated by three different classes of cone photoreceptor, with each class of receptor having peak sensitivity at either short- (S), middle- (M), or long- (L) wavelengths. Several studies have shown that the sensitivity of the S cones decreases with age.<sup>118–120</sup> In addition, the results of Werner and Steele (Fig. 7) indicate that M and L cones also sustain significant sensitivity losses with age, and that the rate of loss is similar for all three cone types.

One complication in specifying the rates of age-related sensitivity loss for individual cone types is that the magnitudes of these losses depend on the level of light adaptation. For example, it has been shown for an S-cone mechanism that age-related differences are greatest near the absolute threshold and gradually diminish at higher light levels.<sup>121</sup> For this reason, Werner et al.<sup>122</sup> have measured the absolute threshold of cone mechanisms at the fovea, and at locations 4 and 8° eccentric to the fovea, along the horizontal meridian in the temporal retina. Their study revealed linear declines in sensitivity with age for all three cone mechanisms that could not be accounted for by age-related changes in preretinal structures of the eye. Thus, losses in sensitivity that occur during our life span reflect, in part, neural changes taking place in both scotopic and photopic mechanisms.

## Color Vision

The ability to discriminate hue declines with age beginning in early adulthood, as assessed by color matching<sup>123,124</sup> and arrangement tasks.<sup>125–127</sup> These losses in discrimination may be due to age-related changes in retinal illuminance caused by smaller pupil size and increased short-wave absorption by the lens. Results from the Farnsworth-Munsell 100-hue test indicate that losses in discrimination



**FIGURE 7** Log relative sensitivity (on a quantal basis) of S-, M-, and L-cone mechanisms is plotted as a function of age. (Data from Werner and Steele.<sup>120</sup>)

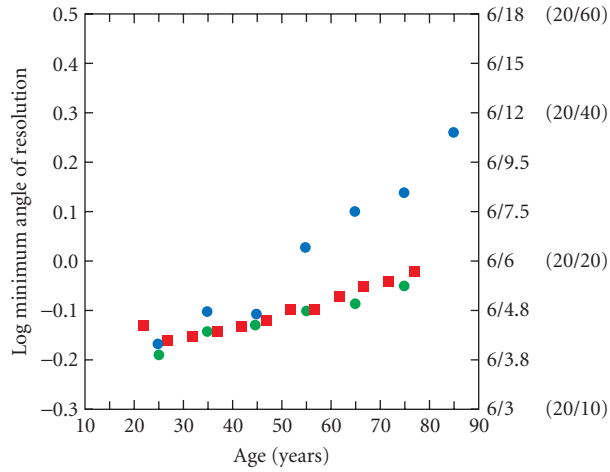
mediated by S cones occur earlier in life than do discriminations mediated by M and L cones. However, this effect is due, at least in part, to the construction of the FM-100 test.<sup>128</sup> When chromatic discriminations are measured under conditions that equate the stimuli individually for retinal illuminance, losses occur not only for S-cone mediated discriminations,<sup>129</sup> but throughout color space for both spectral<sup>130</sup> and nonspectral stimuli.<sup>131</sup> These latter results indicate that age-related losses in chromatic discrimination are due, in part, to neural processes.

Senescence of the lens results in a reduction in retinal illuminance and a change in the spectral distribution of stimuli arriving at the retina. Concurrently, there is a reduction in sensitivity of the photoreceptors and numerous morphological changes in the visual pathways. Nevertheless, there is a remarkable degree of stability in color perception across the life span. Color naming of broadband reflective samples, the wavelengths of unique blue and yellow, saturation, and the achromatic locus are all relatively invariant with normal aging.<sup>132</sup> The proportion of different hue responses (e.g., blue or green) to a large set of simulated Munsell samples is independent of individual ocular media density.<sup>133</sup> Stability of color perception despite many changes in the retinal stimulus implies that the visual system continuously renormalizes itself to maintain constancy of perception. Thus, an elderly person may call the same stimulus “white” as he or she did 70 years ago, even though it must be based upon a markedly different retinal stimulus and strength of retinal signals.<sup>134</sup> To characterize the renormalization implicit in these findings, the chromaticity of the achromatic point was measured before and after cataract surgery.<sup>135</sup> There was a shift following cataract surgery (removal of a brunescient lens) that was initially toward yellow in color space, but over the course of months, it drifted back in the direction of the achromatic point before surgery. This long-term renormalization is as it should be; otherwise, the white of the young would be the yellow of the old. Such adjustments may occur by calibrating the average responses in color mechanisms according to the average color in scenes.

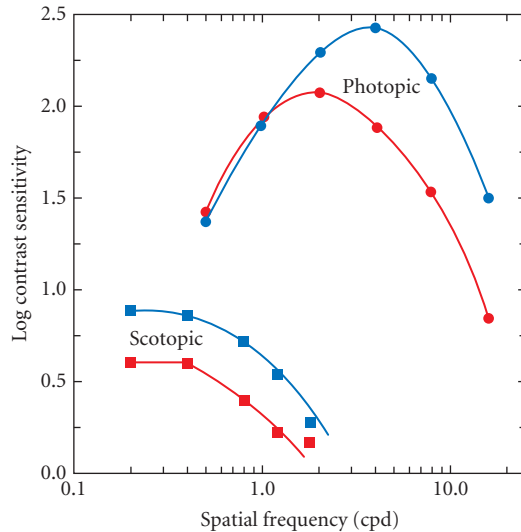
## Spatial Vision

Visual acuity, the minimum angle of resolution, is the most time-honored measure of human visual performance, and in many clinical settings it is the only measured visual ability. It can be measured quickly and reliably with a variety of test patterns such as Snellen letters or Landolt C's. Traditionally, visual acuity is measured with letters having high contrast and moderate luminances. On average, such data suggest some stability in acuity until about age 50 and then a linear decrease with increasing age.<sup>136</sup> The results have been quite variable between studies, and the failures to screen subjects for disease and to optimize their refractions for the test distance make it difficult to know the true course of age-related change in visual acuity. Figure 8 shows results from several recent studies plotted in terms of log minimum angle of resolution (left axis) or equivalent Snellen and metric acuities (right axis). Data from Owsley et al.<sup>137</sup> resemble earlier studies, although participants were screened for visual disease and refracted for the test distance. However, with more stringent screening and optimized refraction, the change in visual acuity follows a somewhat different course according to more recent studies.<sup>138,139</sup> As shown in Fig. 8, these latter studies show better overall visual acuity with a progressive decline after about age 25 years. Conversely, when subjects are not screened for visual disease and are tested with their habitual refractive correction, their acuity is considerably worse than shown in Fig. 8.<sup>5</sup> Which data sets represent “normal” aging is unclear, but the better acuity values indicate that individuals can often reliably resolve finer details, or higher spatial frequencies, than suggested by earlier research, for example, 80 year olds can have 20/20 acuity.

A more general characterization of spatial vision is provided by the contrast sensitivity function (CSF), defined by the reciprocal of the minimum contrast required to detect sinusoidal gratings that vary in spatial frequency (the number of cycles per degree of visual angle, cpd). Figure 9 shows CSFs for nonflickering stimuli interpolated from a large data set by linear regression for hypothetical 20 (blue symbols) and 75-year-old (red symbols) observers. These functions could be computed in this way because the age-related changes in contrast sensitivity appear to be linear with age. A comparison of the photopic CSFs (circles) show little change in sensitivity with age at low spatial frequencies, but notable declines in sensitivity are found at middle and high spatial frequencies. These results are consistent with other studies that measured contrast sensitivity in subjects refracted for the test



**FIGURE 8** Log minimum angle of resolution (left) and Snellen equivalent acuity (right) plotted as a function of age. Blue circles are from Owsley et al.;<sup>137</sup> red squares and green circles are from Frisén & Frisén<sup>138</sup> and Elliott et al.,<sup>139</sup> respectively.



**FIGURE 9** Log contrast sensitivity is plotted as a function of spatial frequency for static sinusoidal gratings. Blue and red circles represent average observers age 20 and 75 years, respectively, determined from regression equations for stimuli centered on the fovea. (From Owsley et al.<sup>137</sup>) Blue and red squares represent average observers age 20 and 75 years, respectively, determined from regression equations for stimuli centered at 8° nasal. (From Scheffrin et al.<sup>144</sup>)

distance and free from ocular disease.<sup>140–143</sup> Currently it is unclear whether age-related differences in contrast sensitivity, especially at high spatial frequencies, become greater if the temporal modulation of luminance varying sinusoidal gratings is increased.<sup>137,142,143</sup>

Evidence from a number of studies suggests that prerenal and neural factors are required to explain the decline in contrast sensitivity with age. Preretinal factors such as smaller pupils, increases in ocular media density and intraocular scatter can partially account for these results. However, losses in contrast sensitivity are still demonstrated by studies that have controlled for pupil size<sup>141</sup> and have attempted to mimic age-related reductions in retinal illuminance due to smaller pupils and increases in lens density.<sup>143</sup> In addition, age-related losses of contrast sensitivity at high spatial frequencies are found when measured with laser interferometry,<sup>145</sup> or when ocular aberrations, including dynamic higher-order aberrations, are corrected with adaptive optics.<sup>146</sup> It has been suggested that age-related changes in visual pathways account for less than half of the total decline in photopic contrast sensitivity at medium and high spatial frequencies with the remainder of the loss in sensitivity due to preretinal factors.<sup>147</sup> This point needs to be more thoroughly investigated.

Measurement of the CSF under scotopic conditions shows a pattern of loss quite different from photopic conditions, as illustrated by the squares in Fig. 9. These data were obtained following 30 min dark adaptation, using an artificial pupil and stimulus wavelength to minimize age-related variation in retinal illuminance.<sup>144</sup> The largest sensitivity loss to horizontally oriented sinusoidal gratings is at low spatial frequencies, a result that implies age-related change in neural mechanisms.

In Fig. 9, extrapolation of the CSFs to zero at high spatial frequencies provides a measure of visual resolution. At both photopic and scotopic light levels, age-related changes in the high frequency cutoff are noted even though participants were included only if they had retinæ without pathology when examined by ophthalmoscopy.

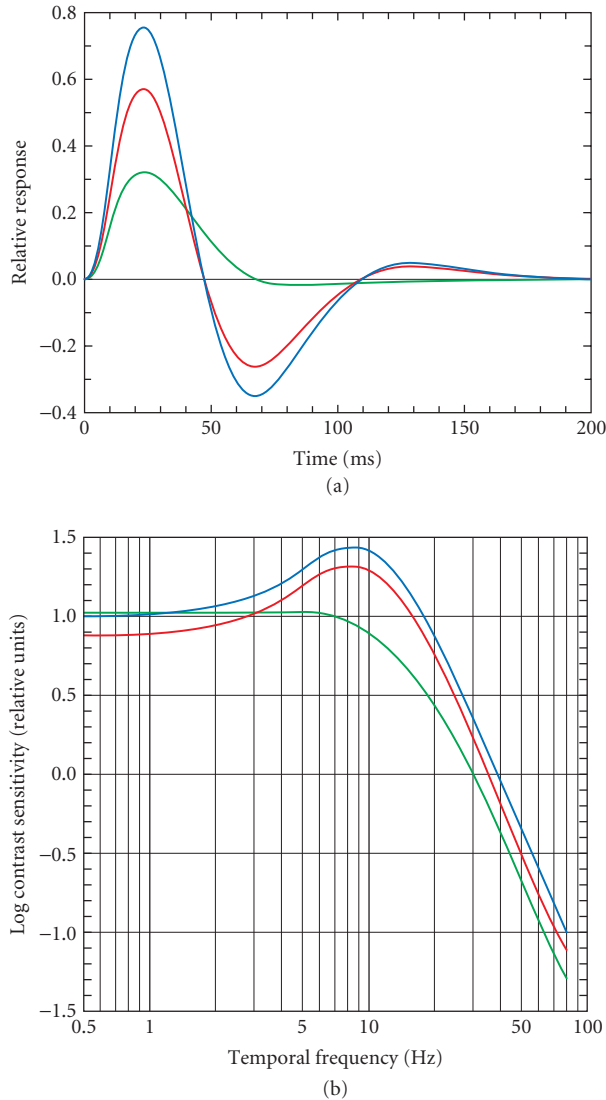
Finally, spatial visual mechanisms have been characterized in terms of their chromatic sensitivity.<sup>148</sup> To be certain that the spatial variations are strictly chromatic, it is necessary to control for chromatic aberration and individual differences in luminosity functions. With these controls in place, Hardy et al.<sup>149</sup> used stimuli modulated along separate S-cone and L-M-cone axes. They found a significant difference in chromatic contrast sensitivity between younger and older observers at all spatial frequencies and for both chromatic axes. The difference in sensitivity is greater for S-axis stimuli than for L and M varying stimuli, prior to adjustments based upon ocular media density. Contrast sensitivities measured with controls for variation in ocular media density show that differences between young and old are reduced, at least for S-cone modulation. Nevertheless, when stimuli are equated at the retina, significant differences between younger and older observers remain. The difference is similar across spatial frequency and for the two chromatic axes. From these results, it may be concluded that both optical and neural factors contribute to age-related losses in chromatic contrast sensitivity.

## Temporal Vision

The critical fusion frequency (CFF), the lowest frequency of flicker that is indiscriminable from a steady light, decreases with age.<sup>150</sup> Part of this loss in temporal resolution is due to retinal illuminance differences between young and old. Lower luminance levels are associated with a loss in temporal resolution. While the CFF measures the high temporal resolution limit, a more complete description of temporal processing is provided by the temporal contrast sensitivity function (tCSF), the amount of contrast needed for detection as a function of temporal frequency. Tyler<sup>151</sup> reported small age-related declines in temporal sensitivity for long-wavelength sinusoidal modulation, with the magnitude of the loss increasing with temporal frequency. The decrease in sensitivity began around age 20 years. Similar results were reported by Kim and Mayer,<sup>152</sup> but sensitivity losses were not observed until about the midforties, after which there was a loss in temporal sensitivity of about 0.78 decilog per decade. It was argued in both studies that these losses in sensitivity are greater than those expected from age-related changes in optical factors alone. Based on simultaneous recordings of pattern electroretinograms (PERG) and visually evoked potentials (VEPs) in young and elderly observers, Porciatti et al.<sup>153</sup> have suggested that at least some of the loss in temporal sensitivity with age is due to changes at postretinal sites.

Another approach to understanding age-related changes in temporal sensitivity is to measure the impulse response function (IRF), the theoretical response to a pulse of short duration.<sup>154</sup> This approach

is important theoretically because the IRF allows predictions of how the visual system should respond to any time-varying stimulus; it is related to the tCSF by its Fourier transform. The IRF was measured for 62 observers ranging in age from 16 to 85 years using a double-pulse method.<sup>155</sup> The IRFs with luminance modulation follow an exponentially damped sine function, with a fast excitatory response followed by an inhibitory phase.<sup>156</sup> Average IRFs and tCSFs are presented in Fig. 10 by red and blue curves. Note that while the amplitude of the IRF is reduced for older subjects, the timing is similar to



**FIGURE 10** (a) Calculated impulse response functions for theoretical 20-year-old (blue curve) and 80-year-old observer with normal (red curve) or reduced second-phase (inhibitory) amplitude (green curve). (b) Corresponding temporal contrast sensitivity functions calculated from the impulse response functions. (After Shinomori and Werner.<sup>155</sup>)

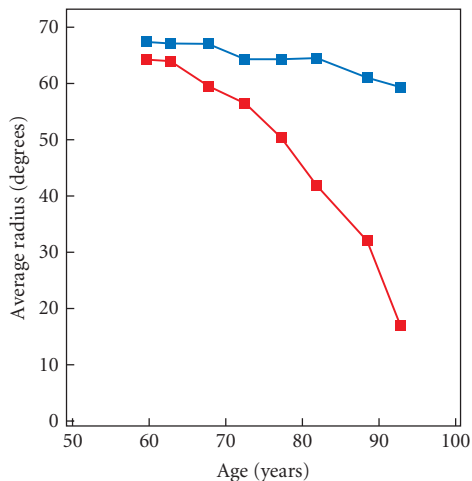
that for younger subjects. For 9 of 25 observers over 55 years of age, however, IRFs are relatively weak and show that the second (inhibitory) phase is reduced. Consequently, the IRF for these observers is quite slow and long, as shown by the IRF plotted as a green curve. The loss of inhibitory phase changes the tCSF from band-pass to low-pass as shown by the green tCSF in Fig. 10. It is not clear why older observers fall into two groups, as all observers met stringent inclusion criteria. Control conditions demonstrated that age-related changes in the IRF under these conditions cannot be ascribed to optical factors. Thus, in most cases, the human visual system maintains a stable speed of response to a flash until at least about 80 years of age, even while the response amplitude decreases with age.

## The Visual Field

The extent of the visual field narrows over adulthood, by several degrees each decade until about age 50 to 60 years and then somewhat more rapidly with increasing age.<sup>157</sup> Iwase et al.<sup>158</sup> suggested that the “volume” of the visual field decreases starting at about 30 to 40 years of age and continues for the remainder of the life span.

Senescent constriction of and reduced light sensitivity within the visual field are due in part to age-related reductions in pupil diameter and increases in light absorbance by the ocular media. Johnson et al.<sup>159</sup> performed static perimetry using experimental conditions that notably lessen the effects of lenticular senescence and reduced pupil size. They demonstrated age-related losses in sensitivity that varied from approximately 0.5 to 1 dB per decade within the central 30° visual field. Their results suggest that at least part of the visual field loss must be ascribed to neural factors.

Another approach to perimetry is to define the functional or “useful field of view,” or the area of the visual field over which information can be processed within a single fixation. The size of the useful field of view decreases when observers are engaged in additional tasks.<sup>160</sup> Ball et al.<sup>161</sup> compared sensitivity of groups of younger and older observers within the central 30°. Stimuli were presented in the presence and absence of distracting stimuli to measure the useful field of view. The older group was more impaired by the distracting stimuli, particularly at greater retinal eccentricities. This restriction of the functional visual field is apparently not due to optical or retinal factors. A more recent study of individuals ages 58 to 102 years<sup>5</sup> shows a dramatic impact of focal attention tasks on visual field changes with age, as shown in Fig. 11. For a typical light detection task, there is only a modest change even up to 90 years, but for a detection task with fixation attention demand, the visual field drops dramatically with age.



**FIGURE 11** The median average radius of the visual field is plotted as a function of age for a standard light detection task (blue symbols) and with concurrent foveal attentional demand (red symbols). (After Haegerstrom-Portnoy et al.<sup>5</sup>)



## Depth and Stereovision

Our ability to perceive the distances of objects depends upon several different kinds of information. One of the most important depth cues is stereopsis. Stereoscopic depth information is provided by retinal disparity that occurs when a single object is imaged at different retinal locations between the two eyes. There is a paucity of investigations that have examined age-related changes in stereovision. This is unfortunate because the use of retinal disparity to extract three-dimensional depth could be used to probe cortical changes that may occur with aging.<sup>162,163</sup> In addition, based on the increased number of falls and other accidents that occur among the elderly, measures of stereoscopic depth perception across adulthood might be valuable.

Jani<sup>164</sup> used a standardized test of stereovision (flashlight diastereo test) to determine the percentage of 1207 individuals who failed the test at various ages. This percentage decreased from his 0 to 9 year group and remained stable from 10 to 19 through about 40 to 49 years. With further increases in age, there was a monotonic increase in the percentage of individuals failing the test. A more sensitive indicator of stereovision requires measurement of thresholds. Bell et al.<sup>165</sup> using a Verhoeff stereopter (based on the furthest distance at which the relative distance between three rods can be detected), showed that stereo resolution decreases slowly from 30 to 40 years of age and then more rapidly after about age 40. It is not clear whether these changes in stereovision are secondary to changes in the spatial CSF. This was controlled by using random-dot stereograms which varied in their binocular correlation.<sup>166</sup> Older observers required significantly higher binocular correlation than younger observers to perceive stereodepth.

## Some Visual Functions That Change Minimally with Age

Senescent changes in visual function are widespread as might be expected given the optical and early neural changes that take place. It is perhaps surprising, therefore, that a number of visual functions change relatively little with age. The stability of color perception across the life span has already been described, but a number of other visual functions also appear resistant to aging.<sup>167</sup> For example, vernier acuity tasks have been used with subjects between 10 and 94 years of age. They were asked to align two or three lines or several points. This can be done with great precision (hence the name hyperacuity), and it does not change with age. The directional sensitivity of the retina (measured by the Stiles-Crawford effect of the first kind) also does not appear much affected by senescence. This task depends on the integrity of both physical and physiological mechanisms.

These and perhaps other (undiscovered) constancies in visual function across adulthood may provide insights into the types of neural changes that accompany age. In addition, such results might also serve as useful indices to distinguish between normal aging and changes in the same functions due to disease.

---

## 14.6 AGE-RELATED OCULAR DISEASES AFFECTING VISUAL FUNCTION

---

There are several ocular and retinal disorders that are so strongly correlated with increasing age that “age” has become a part of their name, for example, “age-related cataracts,” “age-related macular degeneration,” etc. Age, of course, represents only the passage of time, not a process, but the name does accurately imply that these problems are more likely to occur in the elderly.

## Life-Span Environmental Radiation Damage

Unlike most of the human body, the eye is unique in that it contains highly transparent tissue that allows a large proportion of incident radiation to penetrate about 2.5 cm into the body. When considering radiation damage, transparency has the advantage that the energy passing through transparent

material is not absorbed and thus can do no damage. However, all of the quanta passing through the transparent portions of the eye will ultimately be absorbed deep within the eye, and thus, unlike the rest of the body in which only the surface is vulnerable to environmental radiation hazards, tissue deep within the eye can be damaged by radiation. Therefore, environmental radiation has the potential to contribute to the aging process of ocular tissue deep within the eye. The retina is additionally vulnerable to radiation hazards because flux densities at the retina can exceed those at the cornea due to the focusing characteristics of the eye's optics. The eye's susceptibility to the accumulated effect of chronic radiation damage is further exacerbated by the need for transparency across the visible spectrum and the fact that critical cells within the eye (e.g., lens fiber cells and photoreceptors) cannot regenerate. Not surprisingly, the primary ocular targets of life-span environmental radiation damage are the lens and the retina.

With the exception of those employed in professions with high levels of UV radiation (created by some welding techniques, polymer curing, electronic circuitry imaging, fluorescent testing, etc.), the primary source of environmental UV radiation is the sun. The spectral distribution of sunlight reaching different parts of the earth's surface will differ somewhat from that shown in Fig. 2 due to such factors as latitude, altitude, cloud cover, and aerosol content as well as time of day. It is important to note that just as Rayleigh scatter makes the sky blue, the atmosphere scatters an even larger proportion of UV light, and thus the sky itself is a significant source of UV radiation.

While it is straightforward to document surface UV light, it is very difficult to calculate ocular exposure because the eye is typically shaded by the brow and upper lid and thus is rarely directly irradiated by sunlight. Instead, environmental exposure to solar radiation will come primarily from reflections off surfaces or from scatter by the atmosphere. For example, while the sunlight stays constant when moving from a snowless to a snow-covered field, ocular UV irradiance increases substantially because of the high reflectivity of the snow in the UV range.<sup>168</sup> The combined effects of high altitude (little atmosphere to absorb UV) and snow (fresh snow reflects more than 90 percent of the incident UV) can rapidly create solar keratitis in any skier who fails to wear UV-filtering eye protection.

The impact of solar radiation on the eye depends on the proportion of this radiation that is absorbed. At very short wavelengths (UVB), it is the cornea (and primarily the surface epithelium) that absorbs most of the UV, and thus is at risk for UV light damage. However, due to the rapid regeneration of corneal epithelial cells (the entire corneal epithelium is regrown every 5 to 7 days), chronic UV damage is not easily accumulated. At longer UV wavelengths, corneal transparency increases to about 50 percent, and almost all of the remaining 50 percent that passes through the cornea is then absorbed by the lens. Thus, the lens is also vulnerable to UV-damage, and particularly to chronic UV damage because cells in the lens are not replaced or recycled.

Epidemiological evidence provides support for the view that UV radiation is associated with cataract formation. For example, there is a clear correlation between incidence of cataract and other diseases (e.g., pterygium and skin cancer) and exposure to sunlight. In addition, the age of cataract decreases and its incidence increases from the poles to the equator. Young's<sup>53</sup> summary of this literature indicates prevalence rates of about 14 percent for countries above 36°, 22 percent for mid-latitude (28 to 36°) countries and 28 percent for low-latitude (10 to 26°) countries. While epidemiological studies are complex due to the many factors that vary among groups, UV radiation appears to be the only factor that changes with latitude *and* has been shown to play a significant role in aging of the lens. A similar latitude dependency appears in the age of onset of presbyopia<sup>3</sup> indicating that both of the primary age-dependent changes in lens function (presbyopia and cataract) may be regulated by levels of UV exposure.

Because of the opacity of the cornea and lens to UV radiation, relatively little UV reaches the adult retina, however, the retina is vulnerable to photochemical damage caused by UV light as well as short-wave-visible light (the "blue light hazard"). A study comprised of pseudophakic patients who had undergone bilateral cataract extraction and implantation of two IOLs differing in the amount of UV absorption suggested that increased UV exposure in the eye with the UV transparent IOL reduced visual sensitivity of short-wave-sensitive cones. It was estimated that 5 years of exposure to light through an IOL without UV absorbing chromophores produced the equivalent of more than 30 years of normal aging.<sup>7</sup> Thus, the lens pigmentation responsible for UV absorption may be an adaptation to protect the retina from photochemical effects that are damaging.

## Cataract

Cataracts are opacities of the lens that interfere with vision by reducing the amount of light transmitted while at the same time increasing its scatter within the eye. There are a number of different forms of cataract, and each form exhibits its own characteristic effects on vision and has different stages of development. For example, cataracts can form within the lens nucleus and cortex, and they may originate in either the axial or equatorial regions of the lens. Cataracts, depending on their type, can cause a host of visual problems including loss of visual acuity, changes in refractive error, anisometropia, perceived spatial distortions, disability glare, and the formation of multiple images. Often times more than one form is present in a given eye, and the lens of each eye may develop cataract at different rates.

In the United States, cataract is found in about one in seven individuals above the age of 40, and the proportion increases with age. The rate is much higher and typically occurs at an earlier age in individuals with diabetes. It is known that UV and infrared light can cause cataracts. Most individuals exhibit changes in their crystalline lenses with advancing age and senescent cataract may be an extension of this process.<sup>53</sup> In developing countries, cataracts occur much earlier in life (e.g., 14 to 18 years earlier in India than the United States), perhaps due to nutritional factors or greater ultraviolet exposure (as many developing countries have lower latitudes).

Cataract is the leading worldwide cause of blindness. Fortunately, it is treated with a high success rate by surgically removing the cataractous lens. In most cases in developed countries, an intraocular lens (IOL) is implanted. Cataract surgery in the west is one of the most successful therapies known, with visual improvement in nearly all cases. Nearly three million cataract surgeries were performed in the United States in 2007 and over 14 million worldwide.

Cataracts generate both forward and backward scatter of light, and quantification of both can explain about 50 percent of the variability in the visual acuity of older Americans.<sup>169</sup> It is the forward scatter, and not the backscatter seen by the clinician, that interferes with vision. Wide angle forward scatter would be expected to generate a uniform drop in contrast sensitivity, but in patients with cataract Hess and Woo<sup>170</sup> observed reductions in contrast sensitivity that were larger at high and middle spatial frequencies (not unlike those seen with retinal image defocus). Their results indicate that forward scatter produced by cataracts is not wide angle, but, as predicted by the larger than wavelength size of the particles formed in a cataract,<sup>171</sup> much of the forward scattered light is distributed over a small angle (Mie as opposed to Rayleigh scattering).

## Age-Related Macular Degeneration

Age-related macular degeneration (AMD) is generally a progressive, bilateral disease that compromises the integrity and functioning of the central retina. More than 13 million persons over the age of 50 in the United States have signs of macular degeneration. An individual in the early stages of AMD may experience declines in visual acuity, losses in sensitivity for central photopic<sup>172</sup> and scotopic mechanisms,<sup>173,174</sup> an inability to make fine distinctions between colors,<sup>175</sup> and metamorphopsia, a distortion in the appearance of objects in the visual scene. As the disease progresses, individuals afflicted with this condition undergo further declines in visual performance including the onset of central scotomas, defined areas of absolute or relative vision loss within the visual field, making it difficult to recognize readily faces of individuals encountered. However, because the disease does not spread past the macula or central retina, individuals can still navigate about in the environment and even “regain” some function through training to use unaffected adjacent portions of the macular region. About 1.2 million persons in the United States have reached sight-threatening late stages of AMD. This latter group is composed primarily of Caucasians aged 60 years and above, as the disease affects other ethnic groups with lower frequency. The numbers of affected individuals are increasing rapidly with the aging of the population, affecting more than 30 percent of individuals over the age of 75.

AMD is divided clinically into the exudative (wet) and nonexudative (dry) forms of the disease. In the more devastating wet form of this disease, choroidal neovascular membranes grow through

cracks in Bruch's membrane, the layer separating the choroid from the retinal pigment epithelium (RPE). Amongst other things, the RPE is responsible for regulating and providing nourishment to and removing waste products from the retina. Death of retinal cells occurs when the retina is separated for lengthy time periods from the underlying RPE as can happen when the leaky blood vessels comprising the neovascular membranes begin to exude their contents into the potential space between the RPE and photoreceptor layer of the retina. In addition, the presence of these subretinal neovascular membranes lead to scar formation beneath the macula, and this process of scarring is the primary cause of vision loss in the wet form of AMD. The wet form only comprises 10 to 15% of all cases of AMD but accounts for 90 percent of legal blindness from AMD.

In contrast to the wet form of AMD that may produce a sudden loss in vision, dry, or atrophic, AMD results in a slow progressive vision loss that usually takes place over many years. The dry form of AMD is characterized by the formation of drusen, deposits of accumulated waste products from the retina and atrophy of the RPE cells. Presently, it is unclear what causes retinal atrophy in the dry form of AMD, but it is thought to reflect ischemia from the choriocapillaris and/or a metabolic disturbance of the receptor/RPE complex leading to apoptosis.

Currently, clinical advances have been made to retard the progress of the wet form of AMD through medications and laser therapy. Fluorescein and indocyanine green angiography have been effective tools for both diagnosing wet AMD in its early stages of the disease process and defining the subretinal sites of neovascular membranes and exudative leakage from these blood vessels. By taking advantage of this valuable knowledge, laser photocoagulation is now used to destroy these vascular membranes thereby retarding the progress of the disease. For subfoveal lesions, intravitreal injection of anti-VEGF drugs can stabilize or improve vision in 95 percent of cases. Beyond these forms of treatment, visual rehabilitation has concentrated on the use of low vision aides and training AMD patients to use parafoveal and peripheral portions of their retina that are less affected by the disease.

## Diabetic Retinopathy

Diabetic retinopathy is also a common disorder associated with aging. Juvenile-onset (Type I) diabetes is generally diagnosed during the first two decades of life, whereas adult-onset (Type II) diabetes is diagnosed later in life. Type II diabetes is associated with increasing age, obesity, and a family history of the disease; Mexican Americans and African Americans are affected relatively more often than Caucasian Americans. A patient with diabetes (either Type I or Type II) develops diabetic retinal changes over time; at least 40 percent of diabetic patients will eventually develop detectable retinopathy. The incidence of diabetic retinopathy is related to both blood sugar control and the duration of the disease, with most estimates suggesting clinical evidence of retinopathy occurs about 10 years after the onset of hyperglycemia. Approximately 8000 people are blinded each year by this disease.<sup>176</sup>

There are several stages to diabetic retinopathy. Early in the disease there is patchy loss of the endothelial cells lining the retinal capillaries which leads to increased permeability of these blood vessels and the formation of microaneurysms. The appearances of small punctate hemorrhages, microaneurysms, and proteinaceous deposits within the retina are hallmarks of the early or "background" stage of diabetic retinopathy. As the disease progresses, blood vessels may become occluded resulting in retinal ischemia. In response to disrupted blood flow, the ischemic retina releases growth factors which induce the proliferation of new blood vessels or increased permeability of the damaged retinal vessels. These new blood vessels formed in the "proliferative" stage of diabetic retinopathy are fragile and can spontaneously bleed, leading to loss of vision from vitreous hemorrhage. The permeable damaged vessels can cause exudation and swelling of the macula leading to vision loss. In the latter stages of the disease there is death of ganglion cells from intraretinal hypoxia and contraction of scar tissue leading to retinal traction and subsequent retinal detachment in areas adjacent to relatively high concentrations of neovascularization. Tractional retinal detachment may cause visual distortion and blindness, and requires retinal surgery to minimize vision loss.

Initial treatment of diabetic retinopathy involves controlling blood sugar levels via dietary changes, oral agents, and/or insulin to facilitate the uptake of glucose into body tissue. Poorly controlled blood

sugar levels can lead to fluctuations in vision due to changes in the refractive power of the crystalline lens and the early formation of cataracts. The Diabetes Control and Complications Trial<sup>177</sup> demonstrated that tight control of blood glucose reduced the incidence of diabetic retinopathy. Additionally, fluctuations in blood glucose levels have been shown to influence normal functioning of S-cone pathways<sup>178</sup> perhaps accounting for decreased chromatic discrimination along a tritan axis demonstrated by diabetic individuals.<sup>179</sup>

In the latter stages of diabetic retinopathy, laser photocoagulation is used in an attempt to maintain the integrity of the central retina. Laser photocoagulation is utilized to control retinal neovascularization associated with proliferative disease and reduced vision secondary to macular edema. These treatments have been shown to decrease severe vision loss by approximately 50 percent. Because laser photocoagulation to control neovascularization is a destructive form of treatment, newer strategies are being investigated that target specific growth factors causing neovascularization.

## Glaucomas

The glaucomas are a group of chronic diseases that result in the progressive degeneration of retinal ganglion cells whose axons comprise the optic nerve. In most but not all cases of glaucoma, elevated intraocular pressure (IOP) is believed to be one of the primary risk factors for ganglion cell death. Ganglion cell death occurs through the process of apoptosis, brought on, it is believed, by the loss of trophic factors normally transported to the ganglion cell body by retrograde axoplasmic flow. In glaucoma, cell death is thought to occur following a disruption in axoplasmic flow of the retinal ganglion cell axons comprising the optic nerve; the site of this disruption is believed to be at the optic nerve head, where the insult may be mechanical<sup>180</sup> due to elevated IOP and/or a disruption in the vascular supply to the nerve head.<sup>181</sup>

Aqueous humor serves as the optically clear “nutritional supply” to the metabolically active but avascular structures of the lens and cornea. IOP is determined by a balance between production of aqueous humor (by the ciliary body, a secretory tissue that lies directly behind the iris) and drainage through the trabecular meshwork, a circumferential drainage system located just anterior to the iris at an angle formed by the cornea and iris root. Abnormalities in the trabecular meshwork are thought to underlie the elevated IOP in many forms of “open-angle” glaucoma.

Of the approximately 67 million persons affected by glaucoma worldwide, about equal numbers suffer from open-angle and angle-closure types, with the former being more prevalent in the United States and Europe and the latter being more prevalent in Asia and the Pacific Rim. Glaucoma accounts for about 10 percent of blindness in the United States. The frequencies of glaucomatous disorders increase with increasing age, particularly after the age of 40. By age 65 it affects about 10 percent of African Americans and about 3 percent of Caucasian Americans.

Angle-closure glaucoma typically occurs when a forward protruding iris covers the trabecular meshwork thus impeding the flow of aqueous humor from the eye. In this form of glaucoma, there can be sudden dramatic increases in IOP resulting in an abnormally hydrated cornea that leads to diminished visual acuity and accompanying ocular pain. In contrast, open-angle glaucoma is a slow progressive condition in which the IOP is too high for a particular eye resulting in ganglion cell death. Visual loss due to glaucoma occurs generally in the peripheral visual field and affects central vision only late in the disease process. Because the loss is gradual and usually occurs in areas overlapping “normal” vision in the fellow eye, patients are generally unaware of their visual loss until a great deal of damage has occurred. In addition to IOP, other risk factors associated with glaucoma include age, ethnicity, family history, corneal thickness, and myopia. Because increased IOP is not invariably associated with glaucoma, a complete eye examination is needed to make an accurate diagnosis. Often times these examinations will reveal pathognomonic changes such as an abnormal appearance of the optic nerve head, and on visual field testing, specific patterns of visual loss.

Because damage to optic nerve fibers is irreversible and the progress of glaucoma can, in most instances, be successfully controlled, early detection is of great importance. Decline in visual performance at earlier stages in the disease process can be detected through several forms of visual field testing targeting vulnerable ganglion cells.<sup>182,183</sup>

Current treatment regimens for glaucoma involve the use of pharmacological agents and surgical procedures aimed at controlling IOP. Drug therapy includes the use of topical and systemic agents used singly or in combination to inhibit the formation of aqueous or promote increased drainage of aqueous from the eye. Procedures involving lasers or incisional surgery typically aid in aqueous out-flow and are used when drug therapy, by itself, fails to adequately retard the progress of the disease. In addition, ongoing research is currently investigating possible therapies involving neuroprotective agents designed to reduce ganglion cell death after mechanical or chemical damage and to promote new growth.<sup>184</sup>

## 14.7 THE AGING WORLD FROM THE OPTICAL POINT OF VIEW: PRESBYOPIC CORRECTIONS

Presbyopia refers to a loss of accommodation resulting in the near point, the closest distance on which the eye can focus, receding from the eye. This is a continuous process (see Fig. 4) and generally requires some optical aid once accommodative amplitude declines to less than 3 D. Every human suffers from this senescent loss of accommodation, and thus, optical substitutes for accommodation are widely available. Because there is no cure for this condition, and no treatment is perfect, many different treatment options have been developed. Historically, spectacle lenses have been the only option for treatment of presbyopia, but recent attempts have been made to provide presbyopes with contact lenses, IOLs, and other refractive surgery corrections to expand the depth of field of the aging eye. Most of these strategies make no claim to reinstate true accommodation, and thus, can be considered examples of “pseudoaccommodation,” in that the eye’s depth of field increases without any dynamic change in optical power. Reduced pupil size, increased aberrations, bifocal optics, and residual myopia can all lead to improved near vision without any accommodation.

During the last couple of years, several IOLs and a unique surgical procedure have all made claims that they actually restore accommodation; however, almost all of the evidence reported in support of these claims (e.g., improved near vision) could be generated by pseudoaccommodation. Thus, at this time, there is no effective treatment that can restore accommodation in presbyopic humans, but considerable research effort continues with this goal in mind.

### Spectacles

To obtain a well-focused retinal image when viewing a near object, presbyopes that are emmetropic or corrected ametropes require a “near” optical correction. Since the invention of the bifocal lens by Benjamin Franklin, several ingenious approaches to this problem have been developed. With traditional bifocal lenses, the individual looks through different regions of a lens, each region having a different optical power, in order to see clearly objects at near or far. Extensions of this concept include trifocal lenses having three powers confined to three distinct zones, and aspheric lens designs in which the optical power continually changes in a prescribed manner from the top to the inferior portion of the lens. Peripheral zones of the latter designs often exhibit large amounts of astigmatism but are considered cosmetically superior. In each of these designs, the patient is able to select the preferred optical power by adjusting head and eye movements such that different regions of the lens are used to create the retinal image.

### Contact Lenses

The principles of bifocal and multifocal corrections have been incorporated into contact lenses and intraocular lenses (see next section). Like their spectacle counterparts, bifocal and aspheric contact lenses have been designed to provide the presbyope with an increased depth of field. However, unlike the case of spectacle lenses, the patient cannot select a different region of the lens with which



to view different target distances. As the eye and head move, so does the contact lens. Thus, multifocal contact and intraocular lenses include two or more optical powers within the same region of the lens, and the retinal image is generated by two or more powers simultaneously. The obvious consequence is that, in a simple bifocal lens, if 50 percent of the light is well focused on the retina, then 50 percent will be defocused. The challenge for the patient is one of seeing the focused portion of the image with the simultaneously present defocused image. Such lenses are often called “simultaneous vision” lenses.

Some designs of bifocal contact lenses consist of concentric annuli that alternate in power between near and far optical corrections while other bifocal designs include only two regions, one with a distance correction and one with a near correction. A third type of simultaneous vision lens attempts to provide increased depth of field by employing a highly aberrated “multifocal” design. A fourth type of bifocal lens employs a diffractive optical element (DOE) and is often referred to as a “diffractive” lens.<sup>185</sup> Although most designs provide improved near vision compared to a standard monofocal lens,<sup>186,187</sup> this improvement is generally accompanied by a reduction in best-corrected distance vision.

Instead of including two different corrections in the same eye, an alternate approach called “monovision” consists of different monofocal optical corrections in each eye, one for near and the other for far. Thus, unlike simultaneous bifocal vision which adds the focused and defocused images linearly (light adds) at the retina, monovision patients employ nonlinear binocular neural summation<sup>189</sup> in the cortex which biases vision to the higher contrast and better focused image. Thus, the defocused image present in one eye when viewing a distant or near target generally goes unnoticed. This failure to observe the defocused image has led to the notion that the blurred image is somehow suppressed, by a process called “blur suppression.”<sup>190</sup>

It should be noted that the retinal blur produced by monovision and simultaneous bifocal contact lens corrections leads to losses in contrast sensitivity, especially at high spatial frequencies. The magnitude of vision loss can be quite small at distance and near but large at intermediate distances where neither image is well focused.<sup>187,191</sup> Stereopsis can also be affected adversely by the unilateral blur and aniseikonia that accompany monovision corrections. In addition, patients with simultaneous vision corrections that may provide reasonable distance and near vision may experience “ghosting” caused by the defocused portion of the light.<sup>192</sup>

## Intraocular Lenses

Following extraction of a cataractous lens, a refractive error of about 17 D remains to be corrected. A person without their natural lens is called an aphake. In most cases, an intraocular lens (IOL) is inserted, and the wearer is said to be pseudophakic. IOLs have traditionally been constructed from polymethylmethacrylate, but a variety of other materials have been developed (e.g., silicone, hydrogels) so that the lens can be folded and inserted through smaller incisions than are required to insert a hard plastic lens. Many IOLs in the past 15 years have incorporated UV-absorbing chromophores, but older IOLs were virtually transparent to UV radiation from 320 to 400 nm.

Bifocal and multifocal corrections have been incorporated into IOLs. While the optical principles are mostly the same as presbyopic contact lens corrections, some different problems may emerge because of the placement of the optic and maintenance of its position. For example, IOL tilt is associated with astigmatism. However, decentering a bifocal IOL will generate a lateral shift of the defocused images. These off-set defocused images are described as ghosts which are also observed by presbyopes with bifocal contact lens corrections.

A simple segmented bifocal design might have a distance correction in the center and a near correction in the lens surround. With the correct pupil size, light might be equally distributed between the distance and near optics, but as pupil size changes, this balance will be lost. For example, if the pupil is reduced in size such that it matched the size of the central distance portion, the bifocal lens becomes a monofocal correction. To avoid this pupil size dependence, both contact lenses and IOLs have been designed with multiple zones. One type of multizone lens employs standard refractive optics but has alternating near and distance rings across the lens ensuring that both near and distant

optics will contribute to the retinal image irrespective of the pupil size, and thus, the increased depth of field will be present at all times. A second type of multizone bifocal lens, a diffractive optical element, necessarily has multiple rings as part of its design,<sup>193,194</sup> and thus is inherently resistant to pupil size changes. These diffractive lenses control their refractive power by the size of the rings and control the balance between near and distance by the size of the step between successive zones. In addition to enabling bifocal optics, these diffractive lenses have the added potential advantage of correcting some of the eye's chromatic aberration because, unlike refractive lenses (and human eyes) which will always have more power at short wavelengths, diffractive lenses have more power at longer wavelengths.

As with bifocal contact lenses, reduced image contrast in any eye using a bifocal or multifocal IOL can be expected due to the superposition of this second out-of-focus image.<sup>195</sup> The consequence of these problems may be exacerbated in older patients due to their lower contrast sensitivity. Problems with glare and halos around light sources are greater in patients with bilateral multifocal compared to bilateral monofocal IOLs, but the former patients sometimes achieve less dependence on additional spectacle corrections.<sup>196</sup>

## Noncataract Related Refractive Surgeries

There are three issues of ocular senescence related to refractive surgeries (see Chap. 16). First, as postrefractive emmetropes, many young ex-myopes will enjoy a spectacle or contact lens free life. However, as presbyopia progresses, these corrected myopes will now need reading aids for near work. Of course, if left uncorrected, many of these myopes would avoid the need for reading glasses or bifocals by simply removing their distance corrections. Second, some presbyopic patients opt for monovision refractive surgeries, essentially exchanging bilateral myopia for anisometropia. Third, some young adults opting for clear lens extraction or phakic IOLs to treat their high myopia will all experience a form of early and total presbyopia, and thus they will need a near reading aid. There is some concern that these early IOL surgeries will generate accelerated loss of endothelial cells which may lead to corneal failure in the distant future.

## Accommodation Restoration

All of the standard treatments for presbyopia attempt to increase the depth of field without any attempt to restore the lens' ability to adjust its power. That is, pseudoaccommodation is used to replace true accommodation. In the last few years, however, attempts have been made to restore accommodation in presbyopic eyes and to enable true accommodation in pseudophakic eyes (see Ref. 197 for recent comprehensive review). These approaches are based upon the assumption that some useful ciliary muscle strength persists into senescence. There have been two general approaches tried so far. In one, the sclerotic senescent lens is replaced with a lens made from a pliable elastic material (e.g., silicone) that will change shape to achieve a change in power. However, physiological incompatibilities leading to capsule opacification, and problems in the balance between power and change in power, remain.<sup>198</sup> The second approach employs fixed power optical elements that generate a change in whole eye power by movement. Interestingly, although natural accommodation generates increased positive power by a combination of changing lens shape and position, the position changes required by this type of accommodating IOL are much larger than those exhibited by the natural lens. One other technique involves surgical implantation of scleral "expansion rings" to expand the equatorial diameter of the presbyopic eye with the goal of regaining accommodation. The theory of accommodation and loss of accommodation with age behind this surgical procedure is not generally accepted to be correct,<sup>197,199</sup> and careful study of one satisfied patient found no evidence of accommodative amplitude higher than those of untreated age-matched controls.<sup>200</sup>

IOL designs that employ structural changes in the eye created by contraction of the ciliary muscle to change the position or shape of the lens are called "accommodating" IOLs. Designs that employ lens movement to change power have already made it to the market place, whereas designs that employ pliable polymers continue to be tested in the lab.



Two general strategies have been employed to generate accommodating IOLs that work via lens movements. One type uses a single plus lens supported by a hinged haptic, and when the equatorial diameter of the eye decreases as the ciliary muscle contracts, lateral forces on the outer edge of the haptic, combined with potential forward movement of the vitreous, will push the lens forward. When the positive powered IOL and positive powered cornea are closer together, the net result is more plus power. A second design strategy employs a dual optic lens containing a high powered plus lens and a compensatory negative lens. Again, as the ciliary muscle contracts, the forces produced are intended to increase the separation of the two optical components of the dual optic lens, and in so doing, increase its power.<sup>197,201</sup> Optical theory predicts that the latter design has more potential for changing power than that produced by movement of the simple lens.<sup>202</sup> Experimental and clinical studies struggle to show evidence of accommodation.<sup>197,203</sup> Some evidence for lens movement has been demonstrated in response to “pharmacologically induced accommodation” in which lens position or eye refraction is compared in eyes in which ciliary muscle contraction is either blocked (e.g., with Cyclopentolate) or activated (e.g., with pilocarpine).<sup>197,204</sup> Disappointingly, stimulus driven accommodation restoration has yet to be demonstrated,<sup>202,205</sup> and future studies need to ensure that they carefully discriminate between genuine accommodation and pseudoaccommodation.

This is a rapidly evolving field in which new designs are expected to emerge in the next few years. Lens designs that employ elastic IOLs that can change shape in response to ciliary muscle contraction offer the greatest potential,<sup>206</sup> but these face very significant challenges. The translating designs that have made it to market seem to generate little if any true accommodation.<sup>203</sup> It is important to note that most people in their 40s with perhaps only 2 or 2.5 D of accommodation can function without a reading add, therefore accommodating IOLs only need to generate about this level of accommodation for most people to function.

---

## 14.8 CONCLUSIONS

The proportion of the population above age 65 years is increasing, and there is a parallel rise in the proportion of the population that is blind or in need of refractive corrections. Senescent deterioration in the eye’s optics can lead to a wide range of changes in visual function due to reductions in retinal illuminance and degradations in image quality. Age-related changes in vision also result from neural losses. A few visual functions are spared, but substantial changes in the spatial, spectral, and temporal analysis of the retinal image have been well documented. The demographic shift in the proportion of the population that is elderly has not yet reached an equilibrium, so the social and economic consequences can be expected to be even more profound in the future. At this point in history, in spite of technical, optical, and surgical advances, the optical and visual quality of all eyes will deteriorate with advancing age, and for some, age-related pathology will precipitate serious disability due to dramatic loss of visual capability.

---

## 14.9 ACKNOWLEDGMENTS

This chapter was supported by NIA grant AG04058.

---

## 14.10 REFERENCES

1. U.S. Senate Committee on Aging, “Aging America,” (U.S. Department of Health and Human Services, Washington, D.C., 1985–86).
2. United Nations Population Division, “World Population Projections to 2150,” (New York, NY, 1998).
3. R. A. Weale, *A Biography of the Eye* (Lewis, London, 1982).
4. H. M. Leibowitz, D. E. Krueger, L. R. Maunder, R. C. Milton, M. M. Kini, H. A. Kahn, R. J. Nickerson, et al., “The Framingham Eye Study Monograph,” *Surv. Ophthalmol. (Supplement)* **24**:335–610 (1980).

5. G. Haegerstrom-Portnoy, M. E. Schneck, and J. A. Brabyn, "Seeing into Old Age: Vision Function beyond Acuity," *Optom. Vis. Sci.* **76**:141–158 (1999).
6. B. R. Hammond, E. J. Johnson, R. M. Russell, N. I. Krinsey, K. J. Yeum, R. B. Edwards, and D. M. Shodderly, "Dietary Modification of Human Macular Pigment Density," *Invest. Ophthalmol. Visual Sci.* **38**:1795–1801 (1997).
7. J. S. Werner, V. G. Steele, and D. S. Pfoff, "Loss of Human Photoreceptor Sensitivity Associated with Chronic Exposure to Ultraviolet Radiation," *Ophthalmology* **96**:1552–1558 (1989).
8. G. Haegerstrom-Portnoy, "Short-Wavelength-Sensitive-Cone Sensitivity Loss with Aging: A Protective Role for Macular Pigment?," *J. Opt. Soc. Am.* **A5**:2140–2144 (1988).
9. L. M. Verbrugge and D. L. Patrick, "Seven Chronic Conditions: Their Impact on U.S. Adults' Activity Levels and Use of Medical Services," *Am. J. Public Health* **85**:173–182(1995).
10. S. Book, "Vision Loss," (Alliance for Aging Research, 2007).
11. R. S. Ramrattan, R. C. W. Wolfs, S. Panda-Jonas, J. B. Jonas, D. Bakker, H. A. Pols, A. Hofman, and P. T. V. M. de Jong, "Prevalence and Causes of Visual Field Loss in the Elderly and Associations with Impairment in Daily Functioning," *Arch. Ophthalmol.* **119**(12):1788–1794(2001).
12. S. Brink, "Greying of Our Communities Worldwide," *Ageing Internat.* **23**:13–31 (1997).
13. H. Bouma and K. Sagawa, "Ageing Effects on Vision," (Commission Internationale de L'Eclairage, Publication No. 133, Warsaw, Poland, 1999), pp. 368–370.
14. P. E. King-Smith, B. A. Fink, R. M. Hill, K. W. Koelling, and J. M. Tiffany, "The Thickness of the Tear Film," *Curr. Eye Res.* **29**(4–5):357–368 (2004).
15. N. L. Himebaugh, A. R. Wright, A. Bradley, C. G. Begley, and L. N. Thibos, "Use of Retroillumination to Visualize Optical Aberrations Caused by Tear Film Break-Up," *Optom. Vis. Sci.* **80**(1):69–78 (2003).
16. R. Tutt, A. Bradley, C. Begley, and L. N. Thibos, "Optical and Visual Impact of Tear Break-Up in Human Eyes," *Invest. Ophthalmol. Visual Sci.* **41**(13):4117–4123 (2000).
17. D. A. Schaumberg, D. A. Sullivan, J. E. Buring, and M. R. Dana, "Prevalence of Dry Eye Syndrome among U.S. women," *Am. J. Ophthalmol.* **136**(2):318–326 (2003).
18. S. Patel and I. Wallace, "Tear Meniscus Height, Lower Punctum Lacrimale, and the Tear Lipid Layer in Normal Aging," *Optom. Vis. Sci.* **83**(10):731–739 (2006).
19. M. C. Acosta, M. L. Alfaro, F. Borrás, C. Belmonte, and J. Gallar, "Influence of Age, Gender and Iris Color on Mechanical and Chemical Sensitivity of the Cornea and Conjunctiva," *Exp. Eye Res.* **83**(4):932–938 (2006).
20. W. D. Mathers, J. A. Lane, and M. B. Zimmerman, "Tear Film Changes Associated with Normal Aging," *Cornea* **15**(3):229–234 (1996).
21. D. A. Dartt, "Dysfunctional Neural Regulation of Lacrimal Gland Secretion and its Role in the Pathogenesis of Dry Eye Syndromes," *Ocul. Surf.* **2**:76–91 (2004).
22. R. W. Young, "The Bowman Lecture, 1982. Biological Renewal. Applications to the eye," *Trans. Ophthalmol. Soc. U.K.* **102** (Pt 1):42–75 (1982).
23. J. Richard, L. Hoffart, F. Chavane, B. Ridings, and J. Conrath, "Corneal Endothelial Cell Loss after Cataract Extraction by Using Ultrasound Phacoemulsification versus a Fluid-Based System," *Cornea* **27**(1):17–21 (2008).
24. M. Millodot and I. A. Newton, "A Possible Change of Refractive Index with Age and Its Relevance to Chromatic Aberration," *Graefe's Arch. Clin. Exp. Ophthalmol.* **201**:159–167(1976).
25. M. W. Morgan, "Changes in Visual Function in the Aging Eye," in *Vision and Aging: General and Clinical Perspectives*, J. A. A. Rosenbloom, and M. W. Morgan (eds.) (Fairchild Publications, New York, 1986), pp. 121–134.
26. A. Guirao, M. Redondo, and P. Artal, "Optical Aberrations of the Human Cornea and Function of Age," *J. Opt. Soc. Am.* **A17**(10):1697–1702 (2000).
27. P. Artal, E. Berrio, A. Guirao, and P. Piers, "Contribution of the Cornea and Internal Surfaces to the Change of Ocular Aberrations with Age," *J. Opt. Soc. Am.* **A19**(1):137–143 (2002).
28. S. Lerman, "Biophysical Aspects of Corneal and Lenticular Transparency," *Curr. Eye Res.* **3**:3–14 (1984).
29. J. Lu, Z. Wang, P. Lu, X. Chen, W. Zhang, K. Shi, Y. Kang, L. Ke, and R. Chen, "Pterygium in an Aged Mongolian Population: A Population-Based Study in China," *Eye* (2007).
30. E. A. Boettner and J. R. Wolter, "Transmission of the Ocular Media," *Invest. Ophthalmol.* **1**:776–783 (1962).

31. H. Brandhorst, J. Hickey, H. Curtis, and E. Ralph, "Interim Solar Cell Testing Procedures for Terrestrial Applications," (Lewis Research Center, Cleveland, OH, 1975).
32. J. S. Werner, "The Damaging Effects of Light on the Eye and Implications for Understanding Changes in Vision Across the Life Span," in *The Changing Visual System: Maturation and Aging in the Central Nervous System*, P. Bagnoli and W. Hodos (eds.) (Plenum Press, New York, 1991), pp. 295–309.
33. B. T. Gabelt and P. L. Kaufman, "Changes in Aqueous Humor Dynamics with Age and Glaucoma," *Prog. Retin. Eye Res.* **24**:612–637 (2005).
34. A. Caldicott and W. N. Charman, "Diffraction Haloes Resulting from Corneal Oedema and Epithelial Cell Size," *Ophthal. Physiol. Opt.* **22**(3):209–213 (2002).
35. M. P. Hatton, V. L. Perez, and C. H. Dohlman, "Corneal Oedema in Ocular Hypotony," *Exp. Eye Res.* **78**(3):549–552 (2004).
36. F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *J. Physiol. (London)* **186**:558–578 (1966).
37. M. S. Banks, W. S. Geisler, and P. J. Bennett, "The Physical Limits of Grating Visibility," *Vis. Res.* **27**(11):1915–1924 (1987).
38. S. B. Laughlin, "Retinal Information Capacity and the Function of the Pupil," *Ophthal. Physiol. Opt.* **12**(2):161–164 (1992).
39. A. L. Kornzweig, "Physiological Effects of Age on the Visual Process," *Sight-Saving Rev.* **24**:130–138 (1954).
40. I. E. Loewenfeld, "Pupillary Changes Related to Age," in *Topics in Neuro-ophthalmology* H. S. Thompson, R. Daroff, L. Frisén, J. S. Glaser, and M. D. Sanders (eds.) (Williams and Wilkins, Baltimore, MD, 1979), pp. 124–150.
41. M. E. Sloane, C. Owsley, and S. L. Alvarez, "Aging, Senile Miosis and Spatial Contrast Sensitivity at Low Luminance," *Vis. Res.* **28**:1235–1246 (1988).
42. R. A. Applegate, I. Donnelly, W. J., J. D. Marsack, D. E. Koenig, and K. Pesudovs, "Three-Dimensional Relationship between High-Order Root-Mean-Square Wavefront Error, Pupil Diameter, and Aging," *J. Opt. Soc. Am. A* **24**:578–587 (2007).
43. A. Guirao, C. González, M. Redondo, E. Geraghty, S. Norrby and P. Artal, "Average Optical Performance of the Human Eye as a Function of Age in a Normal Population," *Invest. Ophthalmol. Visual Sci.* **40**:203–213 (1999).
44. R. A. Weale, *The Senescence of Human Vision* (Oxford University Press, Oxford, U.K., 1992).
45. G. Smith, "Angular Diameter of Defocus Blur Discs," *Am. J. Optom. Physiol. Opt.* **59**(11):885–889 (1982).
46. R. C. Augusteyn, "Growth of the Lens: In Vitro Observations," *Clin. Exp. Optom.* **91**(3):226–239 (2008).
47. D. A. Atchison, E. L. Markwell, S. Kasthurirangan, J. M. Pope, G. Smith, and P. G. Swann, "Age-Related Changes in Optical and Biometric Characteristics of Emmetropic Eyes," *J. Vis.* **8**:1–20 (2008).
48. L. M. Bova, M. H. Sweeney, J. F. Jamie, and R. J. Truscott, "Major Changes in Human Ocular UV Protection with Age," *Invest. Ophthalmol. Visual Sci.* **42**(1):200–205 (2001).
49. J. S. Werner, "Development of Scotopic Sensitivity and the Absorption Spectrum of the Human Ocular Media," *J. Opt. Soc. Am.* **72**:247–258 (1982).
50. R. A. Weale, "Age and the Transmittance of the Human Crystalline Lens," *J. Physiol. (London)* **395**:577–587 (1988).
51. J. Pokorny, V. C. Smith, and M. Lutze, "Aging of the Human Lens," *Appl. Opt.* **26**:1437–1440 (1987).
52. P. A. Sample, F. D. Esterson, R. N. Weinreb, and R. M. Boynton, "The Aging Lens: In Vivo Assessment of Light Absorption in 84 Eyes," *Invest. Ophthalmol. Visual Sci.* **29**:1306–1311 (1988).
53. R. Young, *Age-Related Cataract* (Oxford University Press, New York, 1991).
54. J. J. Vos and D. van Norren, "Thermal Cataract, from Furnaces to Lasers," *Clin. Exp. Optom.* **87**(6):372–376 (2004).
55. R. A. Weale, "Senile Cataract. The Case Against Light," *Ophthalmology* **90**:420–423 (1983).
56. W. N. Charman, "The Eye in Focus: Accommodation and Presbyopia," *Clin. Exp. Optom.* **91**(3):207–225 (2008).
57. B. K. Pierscionek and R. A. Weale, "Presbyopia—A Maverick of Human Aging," *Arch. Gerontol. Geriatr.* **20**(3):229–240 (1995).
58. R. Brückner, E. Batschelet, and F. Hugenschmidt, "The Basel Longitudinal Study on Aging (1955–1978)," *Doc. Ophthalmol.* **64**:235–310 (1987).

59. F. C. Donders, *On the Anomalies of Accommodation and Refraction of the Eye* (New Sydenham, London, 1864).
60. C. Oyster, *The Human Eye, Structure and Function* (Sinauer Associates Inc., Sunderland, 1999).
61. K. A. Walton, C. H. Meyer, C. J. Harkrider, T. A. Cox, and C. A. Toth, "Age-Related Changes in Vitreous Mobility as Measured by Video B Scan Ultrasound," *Exp. Eye Res.* **74**(2):173–180(2002).
62. Z. Wang, J. Dillon, and E. R. Gaillard, "Antioxidant Properties of Melanin in Retinal Pigment Epithelial Cells," *Photochem. Photobiol.* **82**(2):474–479 (2006).
63. D. M. Snodderly, J. D. Auran, and F. C. Delori, "The Macular Pigment. II. Spatial Distribution in Primate Retinas," *Invest. Ophthalmol. Visual Sci.* **25**:674–685 (1984).
64. J. J. Vos, "Tabulated Characteristics of a Proposed 2° Fundamental Observer," (Institute for Perception—TNO, Soesterberg, The Netherlands, 1978).
65. D. Moreland and P. Bhatt, "Retinal Distribution of Macular Pigment," in *Colour Vision Deficiencies VII*, G. Verriest (ed.) (Dr. W. Junk, The Hague, 1984), pp. 127–132.
66. J. S. Werner, S. K. Donnelly, and R. Kliegl, "Aging and Human Macular Pigment Density; Appended with Translations from the Work of Max Scultze and Ewald Hering," *Vis. Res.* **27**:257–268 (1987).
67. R. A. Bone, J. T. Landrum, L. Fernandez, and S. L. Tarsis, "Analysis of the Macular Pigment by HPLC: Retinal Distribution and Age Study," *Invest. Ophthalmol. Visual Sci.* **29**:843–849 (1988).
68. V. M. Reading and R. A. Weale, "Macular Pigment and Chromatic Aberration," *J. Opt. Soc. Am.* **A64**:231–234 (1974).
69. K. Kirschfield, "Carotenoid Pigments: Their Possible Role in Protecting Against Photooxidation in Eyes and Photoreceptor Cells," *Proc. R. Soc. London* **B216**:71–85 (1982).
70. X. Zhang, M. Ye, A. Bradley, and L. Thibos, "Apodization by the Stiles-Crawford Effect Moderates the Visual Impact of Retinal Image Defocus," *J. Opt. Soc. Am.* **A16**(4):812–820 (1999).
71. K. Franze, J. Grosche, S. N. Skatchkov, S. Schinking, C. Foja, D. Schild, O. Uckermann, K. Travis, A. Reichenbach, and J. Guck, "Muller Cells are Living Optical Fibers in the Vertebrate Retina," *Proc. Natl. Acad. Sci. USA* **104**(20):8287–8292 (2007).
72. M. C. Rynders, T. Grosvenor, and J. M. Enoch, "Stability of the Stiles-Crawford Function in a Unilateral Amblyopic Subject over a 38-Year Period: A Case Study," *Optom. Vis. Sci.* **72**(3):177–185 (1995).
73. P. J. DeLint, J. J. Vos, T. T. Berendschot, and D. van Norren, "On the Stiles-Crawford Effect with Age," *Invest. Ophthalmol. Visual Sci.* **38**(6):1271–1274 (1997).
74. J. M. Gorrand and F. C. Delori, "Reflectance and Curvature of the Inner Limiting Membrane at the Foveola," *J. Opt. Soc. Am.* **A16**(6):1229–1237 (1999).
75. L. Feeney-Burns, E. S. Hilderbrand, and S. Eldridge, "Aging Human RPE: Morphometric Analysis of Macular, Equatorial, and Peripheral Cells," *Invest. Ophthalmol. Visual Sci.* **25**(2):195–200(1984).
76. T. Sarna, J. M. Burke, W. Korytowski, M. Rozanowska, C. M. Skumatz, A. Zareba, and M. Zareba, "Loss of Melanin from Human RPE with Aging: Possible Role of Melanin Photooxidation," *Exp. Eye Res.* **76**(1):89–98 (2003).
77. F. C. Delori, C. K. Dorey, G. Staurenghi, O. Arend, D. G. Goger, and J. J. Weiter, "In Vivo Fluorescence of the Ocular Fundus Exhibits Retinal Pigment Epithelium Lipofuscin Characteristics," *Invest. Ophthalmol. Visual Sci.* **36**:718–729 (1995).
78. J. R. Sparrow and M. Boulton, "RPE Lipofuscin and its Role in Retinal Pathobiology," *Exp. Eye Res.* **80**:595–606 (2005).
79. J. J. Vos, "On the Cause of Disability Glare and Its Dependence on Glare Angle, Age and Ocular Pigmentation," *Clin. Exp. Optom.* **86**(6):363–370 (2003).
80. W. S. Goh and C. S. Lam, "Changes in Refractive Trends and Optical Components of Hong Kong Chinese Aged 19–39 Years," *Ophthalm. Physiol. Opt.* **14**(4):378–382 (1994).
81. L. L. Lin, Y. F. Shih, C. K. Hsiao, and C. J. Chen, "Prevalence of Myopia in Taiwanese Schoolchildren: 1983 to 2000," *Ann. Acad. Med. Singapore* **33**(1):27–33 (2004).
82. R. A. Weale, "On the Age-Related Prevalence of Anisometropia," *Ophthalm. Res.* **34**(6):389–392 (2002).
83. J. R. Lavery, J. M. Gibson, D. E. Shaw, and A. R. Rosenthal, "Refraction and Refractive Errors in an Elderly Population," *Ophthalm. Physiol. Opt.* **8**(4):394–396 (1988).

84. G. Westheimer and J. Liang, "Influence of Ocular Light Scatter on the Eye's Optical Performance," *J. Opt. Soc. Am.* **A12**:1417–1424 (1995).
85. E. Wolf, "Glare and Age," *Arch. Ophthalmol.* **64**:502–514 (1960).
86. J. k. Ijspeert, P. W. T. de Waard, T. J. T. P. van den Berg, and P. T. V. M. de Jong, "The Intraocular Straylight Function in 129 Healthy Volunteers; Dependence on Angle, Age and Pigmentation," *Vis. Res.* **36**:699–707 (1990).
87. M. L. Hennelly, J. L. Barbur, D. F. Edgar, and E. G. Woodward, "The Effect of Age on the Light Scattering Characteristics of the Eye," *Ophthalm. Physiol. Opt.* **18**:197–203 (1998).
88. B. R. Wooten and G. A. Geri, "Psychophysical Determination of Intraocular Light Scatter as a Function of Wavelength," *Vis. Res.* **27**:1291–1298 (1987).
89. D. Whitaker, R. Steen, and D. B. Elliott, "Light Scatter in the Normal Young, Elderly, and Cataractous Eye Demonstrates Little Wavelength Dependency," *Optom. Vis. Sci.* **70**(11):963–968 (1993).
90. T. J. T. P. van den Berg, J. K. Ijspeert, and P. W. T. de Waard, "Dependence of Intraocular Straylight on Pigmentation and Light Transmission Through the Ocular Wall," *Vis. Res.* **31**:1361–1367 (1991).
91. J. E. Coppens, L. Franssen, and T. J. van den Berg, "Wavelength Dependence of Intraocular Straylight," *Exp. Eye Res.* **82**(4):688–692 (2006).
92. S. Lerman and R. F. Borkman, "Spectroscopic Evaluation and Classification of the Normal, Aging and Cataractous Lens," *Ophthalm. Res.* **8**:335–353 (1976).
93. D. B. Elliott, K. C. H. Yang, K. Dumbleton, and A. P. Cullen, "Ultraviolet-Induced Lenticular Fluorescence: Intraocular Straylight Affecting Visual Function," *Vis. Res.* **33**:1827–1833 (1993).
94. J. L. Barbur, A. J. Harlow, and C. Williams, "Light Scattered in the Eye and Its Effect on the Measurement of the Colour Constancy Index," in *Colour Vision Deficiencies XIII*, C. R. Cavonius, (ed.), (Kluwer Academic Publishers, Dordrecht, 1997), pp. 439–438.
95. J. J. Vos, "Age Dependence of Glare Effects and Their Significance in Terms of Visual Ergonomics," (Lighting Research Institute, New York, 1995).
96. J. Porter, A. Guirao, I. G. Cox, and D. R. Williams, "Monochromatic Aberrations of the Human Eye in a Large Population," *J. Opt. Soc. Am.* **A18**(8):1793–1803 (2001).
97. L. N. Thibos, X. Hong, A. Bradley, and X. Cheng, "Statistical Variation of Aberration Structure and Image Quality in a Normal Population of Healthy Eyes," *J. Opt. Soc. Am.* **A19**(12):2329–2348 (2002).
98. J. L. Alio, P. Schimchak, H. P. Negri, and R. Montes-Mico, "Crystalline Lens Optical Dysfunction Through Aging," *Ophthalmology* **112**(11):2022–2029 (2005).
99. I. Brunette, J. M. Bueno, M. Parent, H. Hamam, and P. Simonet, "Monochromatic Aberrations as a Function of Age, from Childhood to Advanced Age," *Invest. Ophthalmol. Visual Sci.* **44**(12):5438–5446(2003).
100. P. Artal, M. Ferro, I. Miranda, and R. Navarro, "Effects of Aging in Retinal Image Quality," *J. Opt. Soc. Am.* **A10**:1656–1662 (1993).
101. R. I. Calver, M. J. Cox, and D. B. Elliott, "Effect of Aging on the Monochromatic Aberrations of the Human Eye," *J. Opt. Soc. Am.* **A16**:2069–2078 (1999).
102. L. Thibos, M. Ye, X. Zhang, and A. Bradley, "Chromatic Eye: A New Reduced-Eye Model of Ocular Chromatic Aberration in Humans," *Appl. Opt.* **31**:3594–3600 (1992).
103. P. P. Howarth, X. X. Zhang, A. Bradley, D. L. Still, and L. N. Thibos, "Does the Chromatic Aberration of the Eye Vary with Age?," *J. Opt. Soc. Am.* **A5**:2087–2092 (1988).
104. H. D. Whitefoot and W. N. Charman, "Hyperchromatic Lenses as Potential Aids for the Presbyope," *Ophthalm. Physiol. Opt.* **15**(1):13–22 (1995).
105. D. A. Atchison, M. Ye, A. Bradley, M. J. Collins, X. Zhang, H. A. Rahman, and L. N. Thibos, "Chromatic Aberration and Optical Power of a Diffractive Bifocal Contact Lens," *Optom. Vis. Sci.* **69**(10):797–804 (1992).
106. N. Lopez-Gil and R. Montes-Mico, "New Intraocular Lens for Achromatizing the Human Eye," *J. Cataract Refract. Surg.* **33**(7):1296–1302 (2007).
107. J. van de Kraats and D. van Norren, "Optical Density of the Aging Human Ocular Media in the Visible and the UV," *J. Opt. Soc. Am.* **A24**(7):1842–1857 (2007).
108. R. G. Dorney, R. A. McFarland, and E. Chadwick, "Dark Adaptation as a Function of Age and Time: II. A Derivation," *J. Gerontol.* **15**:267–279 (1960).

109. C. D. Coile and H. D. Baker, "Foveal Dark Adaptation, Photopigment Regeneration, and Aging," *Vis. Neurosci.* **8**:27–29 (1992).
110. G. R. Jackson, C. Owsley, and G. McGwin, "Aging and Dark Adaptation," *Vis. Res.* **39**:3975–3982 (1999).
111. E. Pulos, "Changes in Rod Sensitivity through Adulthood," *Invest. Ophthalmol. Visual Sci.* **30**:1738–1742 (1989).
112. J. F. Sturr, L. Zhang, H. A. Taub, D. J. Hannon, and M. M. Jackowski, "Psychophysical Evidence for Losses in Rod Sensitivity in the Aging Visual System," *Vis. Res.* **37**:475–481 (1997).
113. B. E. Scheffrin, M. L. Bieber, R. McLean, and J. S. Werner, "The Area of Complete Scotopic Spatial Summation Enlarges with Age," *J. Opt. Soc. Am.* **A15**:340–348 (1998).
114. J. E. E. Keunen, D. V. Norren, and G. J. V. Meel, "Density of Foveal Cone Pigments at Older Age," *Invest. Ophthalmol. Visual Sci.* **28**:985–991 (1987).
115. C. A. Curcio, C. L. Millican, K. A. Allen, and R. E. Kalina, "Aging of the Human Photoreceptor Mosaic: Evidence for Selective Vulnerability of Rods in Central Retina," *Invest. Ophthalmol. Visual Sci.* **34**(12):3278–3296(1993).
116. J. J. Plantner, H. L. Barbour, and E. L. Kean, "The Rhodopsin Content of the Human Eye," *Curr. Eye Res.* **7**:1125–1129 (1988).
117. J. Marshall, J. Grindle, P. L. Ansell, and B. Borwein, "Convolution in Human Rods: An Ageing Process," *Br. J. Ophthalmol.* **63**:181–187 (1979).
118. A. Eisner, S. A. Fleming, M. L. Klein, and W. M. Mauldin, "Sensitivities in Older Eyes with Good Acuity: Cross-Sectional Norms," *Invest. Ophthalmol. Visual Sci.* **28**:1824–1831 (1987).
119. C. A. Johnson, A. J. Adams, J. D. Twelker, and J. M. Quigg, "Age-Related Changes of the Central Visual Field for Short-Wavelength-Sensitive Pathways," *J. Opt. Soc. Am.* **A5**:2131–2139(1988).
120. J. S. Werner and V. G. Steele, "Sensitivity of Human Foveal Color Mechanisms Throughout the Life Span," *J. Opt. Soc. Am.* **A5**:2122–2130 (1988).
121. B. E. Scheffrin, J. S. Werner, M. Plach, N. Utlaut, and E. Switkes, "Sites of Age-Related Sensitivity Loss in a Short-Wave Cone Pathway," *J. Opt. Soc. Am.* **A9**:355–363 (1992).
122. J. S. Werner, B. E. Scheffrin, and M. L. Bieber, "Senescence of Foveal and Parafoveal Cone Sensitivities and Their Relations to Macular Pigment Density," *J. Opt. Soc. Am.* **A17**:1918–1932 (2000).
123. R. Lakowski, "Is the Deterioration of Colour Discrimination with Age due to Lens or Retinal Changes?" *Die Farbe* **11**:69–86 (1962).
124. Y. Ohta and H. Kato, "Colour Perception Changes with Age," *Mod. Probl. Ophthalmol.* **17**:345–352 (1976).
125. G. Verriest, "Further Studies on Acquired Deficiency of Color Discrimination," *J. Opt. Soc. Am.* **53**:185–195 (1963).
126. V. C. Smith, J. Pokorny, and A. S. Pass, "Color-Axis Determination on the Farnsworth-Munsell 100-Hue Test," *Am. J. Ophthalmol.* **100**:176–182(1985).
127. K. Knoblauch, F. Saunders, M. Kusuda, R. Hynes, M. Podor, K. E. Higgins, and F. M. de Monasterio, "Age and Illuminance Effects in the Farnsworth-Munsell 100-Hue Test," *Appl. Opt.* **26**:1441–1448 (1987).
128. J. Birch, *Diagnosis of Defective Colour Vision* (Oxford University Press, New York, 1993).
129. B. E. Scheffrin, K. Shinomori, and J. S. Werner, "Contributions of Neural Pathways to Age-Related Losses in Chromatic Discrimination," *J. Opt. Soc. Am.* **A12**:1233–1241 (1995).
130. K. Shinomori, B. E. Scheffrin, and J. S. Werner, "Age-Related Changes in Wavelength Discrimination," *J. Opt. Soc. Am.* **A18**:310–318 (2001).
131. J. M. Kraft and J. S. Werner, "Aging and the Saturation of Colors: 1. Colorimetric Purity Discrimination," *J. Opt. Soc. Am.* **A16**:223–230 (1999).
132. J. S. Werner, "Visual Problems of the Retina during Ageing: Compensation Mechanisms and Colour Constancy across the Life Span," *Prog. Retin. Eye Res.* **15**:621–645 (1996).
133. J. L. Hardy, C. M. Frederick, P. Kay, and J. S. Werner, "Color Naming, Lens Aging, and Grue: What the Optics of the Aging Eye Can Teach Us about Color Language," *Psychol. Sci.* **44**:321–327 (2005).
134. J. S. Werner and B. E. Scheffrin, "Loci of Achromatic Points Throughout the Life Span," *J. Opt. Soc. Am.* **A10**:1509–1516 (1993).
135. P. B. Delahunt, M. A. Webster, L. Ma, and J. S. Werner, "Long-Term Renormalization of Chromatic Mechanisms Following Cataract Surgery," *Vis. Neurosci.* **21**:301–307 (2004).



136. D. G. Pitts, "The Effects of Aging on Selected Visual Functions: Dark Adaptation, Visual Acuity, Stereopsis, and Brightness Contrast," in *Aging and Human Visual Functions*, R. Sekuler, D. Kline, and K. Dismukes, (eds.), (A.R. Liss, New York, 1982), pp. 131–159.
137. C. Owsley, R. Sekuler, and D. Siemsen, "Contrast Sensitivity Throughout Adulthood," *Vision Res.* **23**:689–699 (1983).
138. L. Frisén and M. Frisén, "How Good is Normal Acuity? A Study of Letter Acuity Thresholds as a Function of Age," *Graefe's Arch. Clin. Exp. Ophthalmol.* **215**:149–157(1981).
139. D. B. Elliott, K. C. H. Yang, and D. Whitaker, "Visual Acuity Changes Throughout Adulthood in Normal, Healthy Eyes: Seeing beyond 6/6," *Optom. Vis. Sci.* **72**:186–191 (1995).
140. G. Derefeldt, G. Lennerstrand, and B. Lundh, "Age Variations in Normal Human Contrast Sensitivity," *Acta Ophthalmol.* **57**:679–690 (1979).
141. K. E. Higgins, M. J. Jafe, R. C. Caruso, and F. M. deMonasterio, "Spatial Contrast Sensitivity: Effects of Age, Test Retest, and Psychophysical Method," *J. Opt. Soc. Am.* **A5**:2173–2180(1988).
142. U. Tulunay-Keeseey, J. N. VerHoeve, and C. Terkla-McGrane, "Threshold and Suprathreshold Spatiotemporal Response Throughout Adulthood," *J. Opt. Soc. Am.* **A5**:2191–2200(1988).
143. D. Elliott, D. Whitaker, and D. MacVeigh, "Neural Contribution to Spatiotemporal Contrast Sensitivity Decline in Healthy Ageing Eyes," *Vis. Res.* **30**:541–547 (1990).
144. B. E. Scheffrin, S. J. Tregear, L. O. Harvey Jr., and J. S. Werner, "Senescent Changes in Scotopic Contrast Sensitivity," *Vis. Res.* **39**:3728–3736 (1999).
145. J. D. Morrison and C. McGrath, "Assessment of the Optical Contributions to the Age-Related Deterioration in Vision," *Q. J. Exp. Psychol.* **70**:249–269 (1985).
146. S. L. Elliott, S. S. Choi, N. Doble, J. L. Hardy, J. W. Evans, and J. S. Werner, "Role of High-Order Aberrations in Senescent Changes in Spatial Vision." *J. Vis.* **9**:1–16 (2009).
147. K. B. Burton, C. Owsley, and M. E. Sloane, "Aging and Neural Spatial Contrast Sensitivity: Photopic Vision," *Vis. Res.* **33**:939–946 (1993).
148. A. Fiorentini, V. Porciatti, M. C. Morrone, and D. C. Burr, "Visual Ageing: Unspecific Decline of the Responses to Luminance and Colour," *Vis. Res.* **36**(21):3557–3566 (1996).
149. J. L. Hardy, P. B. Delahunt, K. Okajima, and J. S. Werner, "Senescence of Spatial Chromatic Contrast Sensitivity. 1. Detection under Conditions Controlling for Optical Factors," *J. Opt. Soc. Am.* **A22**:49–59 (2005).
150. N. W. Coppinger, "The Relationship between Critical Flicker Frequency and Chronological Age for Varying Levels of Stimulus Brightness," *J. Gerontol.* **10**:48–52 (1955).
151. C. W. Tyler, "Two Processes Control Variations in Flicker Sensitivity over the Life Span," *J. Opt. Soc. Am.* **A6**:481–490 (1989).
152. C. B. Y. Kim and M. J. Mayer, "Flicker Sensitivity in Healthy Aging Eyes. II. Cross-Sectional Aging Trends from 18 through 77 Years of Age," *J. Opt. Soc. Am.* **A11**:1958–1969 (1994).
153. V. Porciatti, D. C. Burr, C. Morrone, and A. Fiorentini, "The Effects of Ageing on the Pattern Electroretinogram and Visual Evoked Potential in Humans," *Vis. Res.* **32**:1199–1209(1992).
154. M. Ikeda, "Temporal Summation of Positive and Negative Flashes in the Visual System," *J. Opt. Soc. Am.* **A55**:1527–1534 (1965).
155. K. Shinomori and J. S. Werner, "Senescence of the Temporal Impulse Response to a Luminous Pulse," *Vis. Res.* **43**:617–627 (2003).
156. D. C. Burr and M. C. Morrone, "Impulse-Response Functions for Chromatic and Achromatic Stimuli," *J. Opt. Soc. Am.* **A10**:1706–1713 (1993).
157. E. Wolf, "Studies on the Shrinkage of the Visual Field with Age," *Highway Res. Rec.* **167**:1–7 (1967).
158. A. Iwase, Y. Kitazawa, and Y. Ohno, "On Age-Related Norms of the Visual Field," *Jpn. J. Ophthal.* **32**:429–437 (1988).
159. C. A. Johnson, A. J. Adams, and R. A. Lewis, "Evidence for a Neural Basis of Age-Related Visual Field Loss in Normal Observers," *Invest. Ophthalmol. Visual Sci.* **30**:2056–2064 (1989).
160. M. Ikeda and T. Takeuchi, "Influence of Foveal Load on the Functional Visual Field," *Percept. Psychophys.* **18**:255–260 (1965).
161. K. K. Ball, B. L. Beard, D. L. Roenker, R. L. Miller, and D. S. Griggs, "Age and Visual Search: Expanding the Useful Field of View," *J. Opt. Soc. Am.* **A5**:2210–2221 (1988).

162. K. O. Devaney and H. A. Johnson, "Neuron Loss in the Aging Visual Cortex of Man," *J. Gerontol.* **35**:836–841 (1980).
163. P. D. Spear, "Neural Bases of Visual Deficits during Aging," *Vis. Res.* **33**:2589–2609 (1993).
164. S. N. Jani, "The Age Factor in Stereopsis Screening," *Am. J. Optom.* **43**:653–657 (1966).
165. M. D. Bell, E. Wolf, and C. D. Bernholz, "Depth Perception as a Function of Age," *Aging Human Develop.* **3**:77–81 (1972).
166. S. Laframboise, D. DeGuise, and J. Faubert, "Effect of Aging on Stereoscopic Interocular Correlation," *Optom. Vis. Sci.* **83**:589–593 (2006).
167. J. M. Enoch, J. S. Werner, G. Haegerstrom-Portnoy, V. Lakshminarayanan, and M. Rynders, "Forever Young: Visual Functions not Affected or Minimally Affected by Aging," *J. Gerontol.: Bio. Sci.* **A54**:B336–B351 (1999).
168. D. Sliney and M. Wolbarsht, *Safety with Lasers and Other Optical Sources* (Plenum Press, New York, 1980).
169. W. J. Donnelly, 3rd, K. Pesudovs, J. D. Marsack, E. J. Sarver, and R. A. Applegate, "Quantifying Scatter in Shack-Hartmann Images to Evaluate Nuclear Cataract," *J. Refract. Surg.* **20**(5):S515–S522 (2004).
170. R. Hess and G. Woo, "Vision Through Cataracts," *Invest. Ophthalmol. Visual Sci.* **17**(5):428–435 (1978).
171. M. J. Costello, S. Johnsen, K. O. Gilliland, C. D. Freel, and W. C. Fowler, "Predicted Light Scattering from Particles Observed in Human Age-Related Nuclear Cataracts Using Mie Scattering Theory," *Invest. Ophthalmol. Visual Sci.* **48**(1):303–312(2007).
172. A. E. Elsner, S. A. Burns, and J. J. Weiter, "Cone Photopigment in Older Subjects: Decreased Optical Density in Early Age-Related Macular Degeneration," *J. Opt. Soc. Am.* **A19**(1):215–222 (2002).
173. G. Haegerstrom-Portnoy and B. Brown, "Two-Color Increment Thresholds in Early Age Related Maculopathy," *Clin. Vis. Sci.* **4**:165–172 (1989).
174. J. S. Sunness, R. W. Massof, M. A. Johnson, N. M. Bressler, S. B. Bressler, and S. L. Fine, "Diminished Foveal Sensitivity may Predict the Development of Advanced Age-Related Macular Degeneration," *Ophthalmology* **96**:375–381 (1989).
175. R. A. Applegate, A. J. Adams, J. C. Cavender, and F. Zisman, "Early Color Vision Changes in Age-Related Maculopathy," *Appl. Opt.* **26**:1458–1462 (1987).
176. National Society to Prevent Blindness, "Prevent Blindness America" (Schaumburg, IL, 1994).
177. Diabetes Control and Complications Trial Research Group, "The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus," *New Eng. J. Med.* **329**:977–986 (1993).
178. V. J. Volbrecht, M. E. Schneck, A. J. Adams, J. A. Linfoot, and E. Ai, "Diabetic Short-Wavelength Sensitivity: Variations with Induced Changes in Blood Glucose Level," *Invest. Ophthalmol. Visual Sci.* **35**:1243–1246 (1994).
179. R. Lakowski, P. A. Aspinall, and P. R. Kinnear, "Association between Colour Vision Losses and Diabetes Mellitus," *Ophthalm. Physiol. Opt.* **4**:145–159 (1972/73).
180. Y. Yang, J. C. Downs, C. Girken, L. Sakata, A. Bellezza, H. Thompson, and C. F. Burgoyne, "3-D Histomorphometry of the Normal and Early Glaucomatous Monkey Optic Nerve Head: Lamina Cribosa and Peripapillary Scleral Position and Thickness," *Invest. Ophthalmol. Visual Sci.* **48**:4597–4607 (2007).
181. J. Flammer and S. Orgül, "Optic Nerve Blood-Flow Abnormalities in Glaucoma," *Prog. Retin. Eye Res.* **17**:267–289 (1998).
182. C. A. Johnson and S. J. Samuels, "Screening for Glaucomatous Visual Field Loss with Frequency-Doubling Perimetry," *Invest. Ophthalmol. Visual Sci.* **38**:413–425 (1997).
183. H. Sun, M. W. Dul, and W. H. Swanson, "Linearity Can Account for the Similarity among Conventional, Frequency-Doubling, and Gabor-Based Perimetric Tests in the Glaucomatous Macula," *Optom. Vis. Sci.* **83**(7):455–465 (2006).
184. S. J. Chew and R. Ritch, "Neuroprotection: The Next Breakthrough in Glaucoma? Proceedings of the 3rd Annual Optic Nerve Rescue and Restoration Think Tank," *J. Glaucoma* **6**:263–266 (1997).
185. A. L. Cohen, "Diffractive Bifocal Lens Designs," *Optom. Vis. Sci.* **70**(6):461–468 (1993).
186. M. A. Bullimore and R. J. Jacobs, "Subjective and Objective Assessment of Soft Bifocal Contact Lens Performance," *Optom. Vis. Sci.* **70**(6):469–475 (1993).
187. A. Bradley, H. Abdul Rahman, P. S. Soni, and X. Zhang, "Effects of Target Distance and Pupil Size on Letter Contrast Sensitivity with Simultaneous Vision Bifocal Contact Lenses," *Optom. Vis. Sci.* **70**(6):476–481 (1993).



188. S. Sanislo, D. Wicker, and D. G. Green, "Contrast Sensitivity Measurements with the Echelon Diffractive Bifocal Contact Lens as Compared to Bifocal Spectacles," *Claeo J.* **18**(3):161–164 (1992).
189. G. E. Legge and G. S. Rubin, "Binocular Interactions in Suprathreshold Contrast Perception," *Percept. Psychophys.* **30**(1):49–61 (1981).
190. C. Schor, L. Landsman, and P. Erickson, "Ocular Dominance and the Interocular Suppression of Blur in Monovision," *Am. J. Optom. Physiol. Opt.* **64**(10):723–730 (1987).
191. R. Legras, V. Hornain, A. Monot, and N. Chateau, "Effect of Induced Anisometropia on Binocular Through-focus Contrast Sensitivity," *Optom. Vis. Sci.* **78**(7):503–509 (2001).
192. A. Back, T. Grant, and N. Hine, "Comparative Visual Performance of Three Presbyopic Contact Lens Corrections," *Optom. Vis. Sci.* **69**(6):474–480 (1992).
193. J. T. Holladay, H. Van Dijk, A. Lang, V. Portney, T. R. Willis, R. Sun, and H. C. Oksman, "Optical Performance of Multifocal Intraocular Lenses," *J. Cataract Refract. Surg.* **16**(4):413–422 (1990).
194. M. Simpson, "Diffractive Multifocal Intraocular Lens Image Quality," *Appl. Opt.* **31**(19):3621–3626 (1992).
195. E. Peli and A. Lang, "Appearance of Images Through a Multifocal Intraocular Lens," *J. Opt. Soc. Am.* **A18**(2):302–309 (2001).
196. J. C. Javitt, F. Wang, D. J. Trentacost, M. Rowe, and N. Tarantino, "Outcomes of Cataract Extraction with Multifocal Intraocular Lens Implantation," *Ophthalmology* **104**:589–599 (1997).
197. A. Glasser, "Restoration of Accommodation: Surgical Options for Correction of Presbyopia," *Clin. Exp. Optom.* **91**(3):279–295 (2008).
198. O. Nishi and K. Nishi, "Accommodation Amplitude after Lens Refilling with Injectable Silicone by Sealing the Capsule with a Plug in Primates," *Arch. Ophthalmol.* **116**(10):1358–1361 (1998).
199. P. L. Kaufman, "Scleral Expansion Surgery for Presbyopia," *Ophthalmology* **108**(12):2161–2162 (2001).
200. L. A. Ostrin, S. Kasthurirangan, and A. Glasser, "Evaluation of a Satisfied Bilateral Scleral Expansion Band Patient," *J. Cataract Refract. Surg.* **30**(7):1445–1453 (2004).
201. S. D. McLeod, V. Portney, and A. Ting, "A Dual Optic Accommodating Foldable Intraocular Lens," *Br. J. Ophthalmol.* **87**(9):1083–1085 (2003).
202. S. D. McLeod, "Optical Principles, Biomechanics, and Initial Clinical Performance of a Dual-Optic Accommodating Intraocular Lens (an American Ophthalmological Society thesis)," *Trans. Am. Ophthalmol. Soc.* **104**:437–452 (2006).
203. O. Findl and C. Leydolt, "Meta-Analysis of Accommodating Intraocular Lenses," *J. Cataract Refract. Surg.* **33**(3):522–527 (2007).
204. A. Langenbacher, B. Seitz, S. Huber, N. X. Nguyen, and M. Kuchle, "Theoretical and Measured Pseudophakic Accommodation after Implantation of a New Accommodative Posterior Chamber Intraocular Lens," *Arch. Ophthalmol.* **121**(12):1722–1727 (2003).
205. J. S. Cumming, S. G. Slade, and A. Chayet, "Clinical Evaluation of the Model AT-45 Silicone Accommodating Intraocular Lens: Results of Feasibility and the Initial Phase of a Food and Drug Administration Clinical Trial," *Ophthalmology* **108**(11):2005–2009; discussion 2010 (2001).
206. G. Marchini, P. Mora, E. Pedrotti, F. Manzotti, R. Aldigeri, and S. A. Gandolfi, "Functional Assessment of two Different Accommodative Intraocular Lenses Compared with a Monofocal Intraocular Lens," *Ophthalmology* **114**(11):2038–2043 (2007).

---

# ADAPTIVE OPTICS IN RETINAL MICROSCOPY AND VISION

---

Donald T. Miller

*School of Optometry  
Indiana University  
Bloomington, Indiana*

Austin Roorda

*School of Optometry  
University of California  
Berkeley, California*

---

## 15.1 GLOSSARY

---

**Adaptive optics.** Opto-electro-mechanical method for dynamic closed-loop compensation of wavefront aberrations that reside between the object and image.

**Bimorph mirror.** Class of wavefront corrector distinguished by its piezoelectric material that is sandwiched between a continuous top electrode (reflective) and a bottom, patterned electrode array.

**Contrast sensitivity.** Reciprocal of the minimum perceptible contrast.

**Deformable mirror.** Mirror whose surface profile is controlled by applying voltages to a set of adjacent electrodes or actuators.

**Discrete-actuator deformable mirror.** Class of wavefront corrector distinguished by its continuous reflective surface and array of underlying actuators that alter the surface shape.

**Higher-order aberrations.** Aberrations higher than second order.

**Influence function.** Deflection of the corrector surface when a unit voltage is applied to a single actuator.

**Low-order aberrations.** Aberrations of first and second order.

**Membrane mirror.** Class of wavefront corrector distinguished by its edge-clamped, flexible, reflective membrane that is sandwiched between a transparent top electrode and an underlying array of electrodes.

**Ophthalmoscope.** Instrument for viewing and recording images of the back of the eye.

**Optical coherence tomography.** Interferometric-based imaging modality that optically sections tissue with micron-scale axial resolution. Axial resolution is inversely related to the spectrum of the imaging light source.

**Photoreceptor.** Cell that initiates the visual process in the retina by transducing the captured photons into an encoded neural signal.

- Point spread function.** Response (image) of an optical system to a point source of light (object).
- Psychophysics.** A branch of science that deals with the relationship between physical stimuli and sensory response.
- Retinal microscopy.** Observation of retinal features at the cellular level in the living eye.
- Scanning laser ophthalmoscope.** Imaging modality that raster scans a point source across the object and filters the reflected light with a confocal pinhole for enhanced contrast and axial sectioning.
- Segmented corrector.** Class of wavefront corrector distinguished by its array of small closely spaced reflective or transmissive elements that induce local piston and tilt on the incident wavefront.
- Shack-Hartmann wavefront sensor.** An objective wavefront sensor that measures the wave aberrations of the eye with principle components being a light source beacon, lenslet array, and areal CCD.
- Spectacles.** Ophthalmic appliance that corrects prism, defocus (sphere), and astigmatism (cylinder).
- Strehl ratio.** Ratio of the maximum light intensity in the aberrated point spread function to that in the aberration-free point spread function with the same pupil diameter. Values range from 0 to 1.
- Visual acuity.** Capacity for seeing distinctly the details of an object. In units of the Snellen fraction, average clinical visual acuity varies between 20/15 and 20/20.
- Wave aberration.** Optical deviations of a wavefront that degrade image quality and that can vary temporally and spatially across the pupil of the optical system.
- Zernike polynomials.** Specific set of polynomials frequently used to mathematically represent the wave aberration of the eye.

## 15.2 INTRODUCTION

---

Vision is a most acute human sense. Yet one need only to examine the first step in the visual process—the formation of an image on the retina—to realize that it is often far from perfect. Image quality at the retina is limited fundamentally by the wave aberrations intrinsic to the cornea and crystalline lens and diffraction due to the finite size of the eye's pupil. Together, aberrations and diffraction limit not only what the eye sees looking out, but also determine the smallest internal structures that can be observed when looking into the eye with a microscope. Conventional corrective methods such as spectacles, contact lenses, and refractive surgery provide a static amelioration of low-order sphere (defocus) and cylinder (astigmatism). However, ocular image quality can be significantly improved by dilating the pupil to minimize diffraction and correcting the aberrations across the larger pupil. The presence of aberrations in the human eye<sup>1</sup> has been recognized for some time. Indeed, Hermann von Helmholtz, commenting on the eye 150 years ago, put it as follows: “Now, it is not too much to say that if an optician wanted to sell me an instrument which had all these defects, I should think myself quite justified in blaming his carelessness in the strongest terms, and giving him back his instrument.”<sup>2</sup>

Spectacles have been used to correct defocus since at least as early as the thirteenth century. They began to be used to correct astigmatism shortly after 1801, when Thomas Young discovered that condition in the human eye. However, it has only been in the last decade and a half—two centuries later—that additional, higher-order aberrations have been corrected. There are three reasons for this long period of seemingly little progress.

First, the most significant and troublesome aberrations in the eye are defocus and astigmatism. Correcting these two usually improves vision to an acceptable level. Second, for most of this period defocus and astigmatism were the only aberrations of the eye that could be easily measured. Third, even when higher-order aberrations could be measured by the cumbersome techniques available at that time, there was no simple or economical way to correct them. In 1961, for instance, M. S. Smirnov<sup>3</sup> measured many of the higher-order aberrations in the human eye for the first time. But his psychophysical method required 1 to 2 h of measurements per eye, followed by an additional 10 to 12 h of calculations. On the strength of his results, Smirnov surmised that it would be possible to manufacture custom lenses to compensate for the higher-order aberrations of individual eyes.

Recent technological advances, notably in adaptive optics (AO),<sup>4</sup> have provided solutions to the problems Smirnov encountered. The concept of AO was proposed in 1953 by astronomer Horace Babcock<sup>5</sup> as a means of compensating for the wavefront distortion induced by atmospheric turbulence.<sup>6</sup> Transforming AO into a noninvasive vision tool fundamentally consists of measuring the aberrations of the eye and then compensating for them. But the technique must be able to deal with the problem of variation of ocular and retinal tissue for a given person, as well as from person to person, and it must also take into account human safety.

Prior to the development of AO for the eye, several techniques showed promise for enhancing vision and improving the resolution of retinal imaging. These included use of interference fringes to probe the visual system without optical blurring;<sup>7</sup> a postdetection imaging technique coined speckle interferometry to recover high spatial frequencies of the cone mosaic;<sup>8</sup> and precise correction of defocus and astigmatism in conjunction with quick flashes of quasi-monochromatic light to image individual cone photoreceptors.<sup>9</sup> Success with the latter was confined to eyes of relatively low higher-order aberrations. In 1989, the first wavefront corrector was applied to the eye, a 13-actuator membrane mirror that corrected the astigmatism in one subject's eye using his conventional prescription for spectacles.<sup>10</sup> Correction of additional aberrations was not possible as there was no effective means at that time for measuring the aberrations.

These early techniques shared the common problem that the aberrations beyond defocus and astigmatism were largely unknown. To address this limitation, Liang et al.<sup>11</sup> developed a wavefront sensor based on the principles of Shack-Hartmann wavefront sensing (SHWS) and further refined it for measuring aberrations up through the first 10 radial orders (corresponding to the first 65 Zernike modes).<sup>12</sup> The SHWS was a watershed. It provided—for the first time—a rapid, automated, noninvasive, and objective measure of the wave aberrations of the eye. The method required no feedback from the patient; its measurements could be collected in a fraction of a second; and its near-infrared illumination was less intrusive to the subject than earlier techniques had been. That same year, the first AO retina camera was developed based on a SHWS and a 37-actuator deformable mirror.<sup>13</sup> The system included an additional channel for collecting images of the living retina and measuring the visual performance of subjects. With a pupil 6 mm in diameter, the system substantially corrected aberrations up through the fourth-Zernike order, where most of the significant aberrations in the eye reside. Thus, Liang et al. demonstrated for the first time that AO allows coma, spherical aberration, and other irregular aberrations to be corrected in the eye. The closed-loop speed of their system was on the order of tens of seconds and was incapable of tracking temporal fluctuations in the eye's aberrations. Their system may be more accurately characterized as an active rather than adaptive optics system.

In little more than a decade since AO for the eye has experienced exponential growth.<sup>14</sup> Essentially all facets of AO have undergone substantive development for use with the eye, including its wavefront sensor, wavefront corrector, and control algorithm. In addition, AO has been integrated into a variety of retina camera architectures to increase their resolution and sensitivity including the three principle types: flood-illuminated ophthalmoscope,<sup>13,15–18</sup> scanning laser ophthalmoscope (SLO),<sup>19–24</sup> and optical coherence tomography (OCT).<sup>25–35</sup> AO for the eye has also entered the commercial arena. The SHWS for measuring ocular aberrations has been commercially available for several years; commercial AO systems for vision testing for the last couple of years; and commercial AO retina cameras now entering the market.

While early AO retina cameras focused on imaging cone photoreceptors owing to their high reflectance and contrast, many other cellular structures have since been imaged in the living human retina including nerve fiber bundles, individual blood cells flowing through the smallest retinal capillaries, fibers of Henle, retinal pigment epithelium cells, and astrocytes. Ganglion cells and their axons have also been visualized with AO in primates when used in conjunction with a contrast-enhancing fluorescent dye. There is also an ever expanding list of fundamental and clinical research applications that are underway and which have been made possible by the increased resolution and sensitivity afforded by AO.

Several reviews of AO for vision science already exist,<sup>36–40</sup> including a textbook devoted to the topic.<sup>41</sup> Nevertheless, the field continues to grow rapidly and much has changed since the first *OSA Handbook* chapter on this topic was written in 2000.<sup>42</sup> Our aim here is to summarize the current understanding of the ocular aberration properties (pertinent for designing AO systems for the eye), detail the status of AO technologies for the eye, and review some of the latest scientific and clinical uses of this powerful technology.

## 15.3 PROPERTIES OF OCULAR ABERRATIONS

Aberrations of the eye vary spatially across the eye's pupil, vary over time, and vary with field location at the retina (isoplanatism). These fundamental properties (spatial distribution, temporal distribution, and field dependence) dictate the performance requirements of AO for effective correction of ocular aberrations. Here we summarize the most pertinent aspects of each for AO.

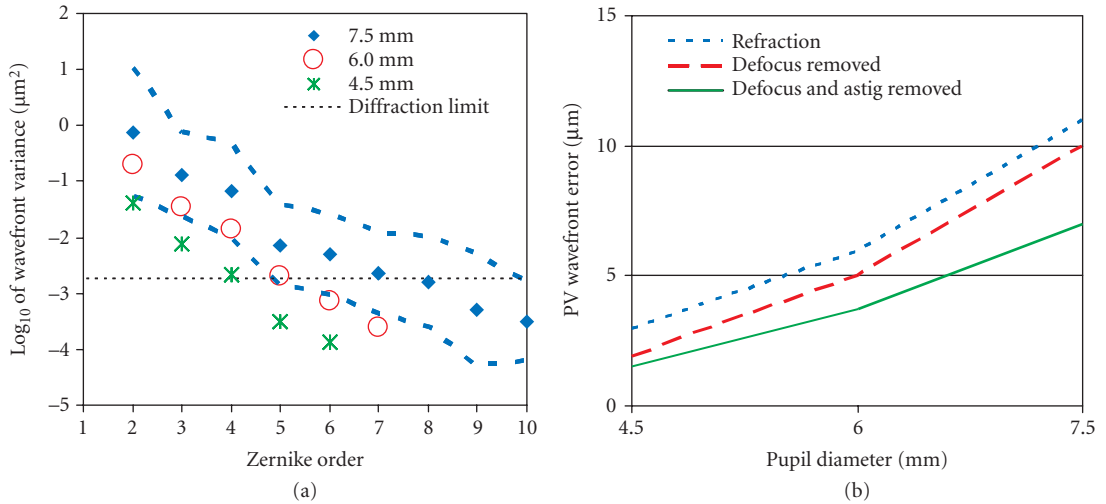
The spatial distribution of ocular aberrations varies considerably among subjects, and often does not correspond to the classical aberrations of manufactured optical systems (e.g., Seidel aberrations). Instead, Zernike polynomials are routinely used for representing ocular aberrations. The first-order Zernike polynomials represent tilt, the second order correspond to defocus and astigmatism, the third order are coma and trefoil, the fourth order are spherical aberration, secondary astigmatism, and quadrafoil, and so on. Third order and higher are collectively termed higher-order aberrations or irregular aberrations. Zernikes are not the most efficient polynomial representation of the eye's aberrations, but are only slightly less compact than the principal modes that derive from a principal components analysis.<sup>43,44</sup> Zernikes are also mathematically simpler and have well-established properties. Certainly the long history of Zernike polynomials in astronomy for representing atmospheric turbulence<sup>45</sup> influenced their early use by the vision community. The now ubiquitous use of Zernike polynomials is perhaps best reflected by the community's establishment and universal acceptance of a single naming convention for the Zernike coefficients and polynomials as described by the OSA/VSIA Standards Taskforce.<sup>46</sup>

Today detailed measurements of the spatial distribution of wave aberrations have been made in large populations of normal, healthy adult eyes with large pupils,<sup>43,44,47</sup> including the pooling of data collected at 10 laboratories (2560 eyes).<sup>48</sup> More recent studies have expanded this to include the effect of age, accommodation, refractive state, refractive surgery, and disease.<sup>49–57</sup>

The spatial characteristics of the ocular aberrations most important for AO are spatial fidelity (frequency composition of the aberrations) and magnitude [peak-to-valley (PV) wavefront error]. Both are captured by the Zernike representation—spatial fidelity by the Zernike order and magnitude by the Zernike coefficient, and both are strongly influenced by pupil size. As an example, Fig. 1a shows the wavefront variance decomposed by Zernike order for a population of about 100 normal subjects, most in their early twenties.<sup>44</sup> Data points are shown for three pupil sizes: 4.5, 6.0, and 7.5 mm, which span the range over which AO is typically applied to the human eye. For comparison, the maximum physiological pupil size for young subjects is nominally 8 mm. In Fig. 1a, the power decreases monotonically (approximately linear on a logarithmic scale) with second-order aberrations dominating the total wavefront error. Power also decreases monotonically with decreasing pupil size. The 7.5-mm pupil represents the most demanding condition for AO. For this pupil size, correction of Zernike polynomials up through at least 10th order is necessary to reach diffraction-limited imaging [ $\lambda/14$  root-mean-square (RMS) error] in approximately 95 percent of the population, that is, two times the standard deviation of the  $\log_{10}$  (wavefront variance).

Figure 1b shows the corresponding PV wavefront error that encompasses 95 percent of the population as a function of pupil size (4.5 to 7.5 mm). Three curves are shown that correspond to three different second-order states. As expected, the PV error increases monotonically with pupil size and depends strongly on the second-order state. The PV error for the 7.5-mm pupil ranges from 7 to 11  $\mu\text{m}$  depending on the second-order state. The largest error of 11  $\mu\text{m}$  represents the most demanding condition for AO correction. Note that for this study, each subject was meticulously refracted with trial lenses. Typically, AO systems operate under less ideal conditions in which the second-order aberrations are only coarsely corrected or not at all. Under these more realistic conditions, second-order aberrations can readily surpass the PV error shown in the figure by several times. As such, careful design of an AO system must take into account the anticipated refractive state of the population and the expected manner in which trial lenses or translating lenses will be applied.

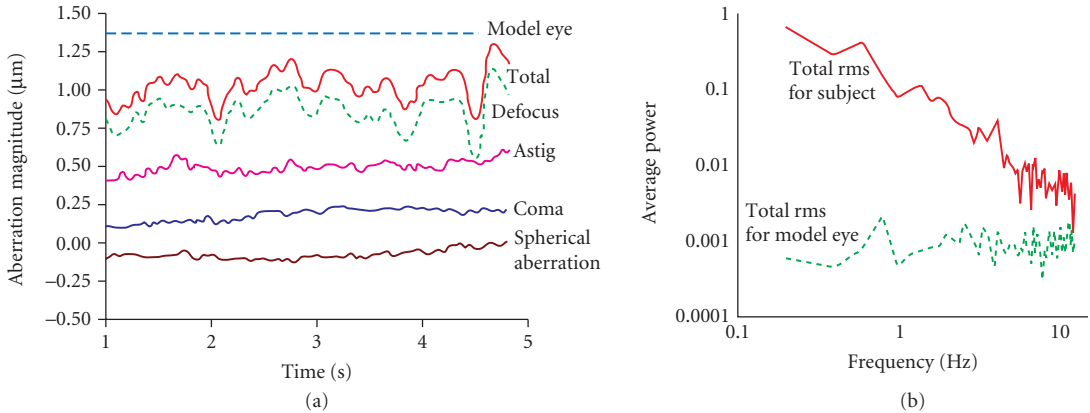
The population measurements in Fig. 1 represent essentially static wave aberrations and therefore do not capture the temporal behavior of the ocular media. Two early studies investigated the temporal dynamics with attention to their impact on AO performance.<sup>58,59</sup> Both studies employed a SHWS to track the dynamics. Hofer et al. measured aberration changes up to approximately 5 to 6 Hz, and



**FIGURE 1** Spatial properties of ocular aberrations in a large population of 100 normal eyes.<sup>44</sup> (a)  $\log_{10}$  of the wavefront variance after a conventional refraction using trial lenses is plotted as a function of Zernike order and pupil size (4.5, 6.0, and 7.5 mm). Diamond and corresponding dashed curves represent the mean and mean  $\pm$  two times the standard deviation of the  $\log_{10}$  (wavefront variance), respectively, for a 7.5-mm pupil. Star and open circle correspond to 4.5- and 6.0-mm pupils. Thin, horizontal dashed line corresponds to  $\lambda/14$  RMS for  $\lambda = 0.6 \mu\text{m}$ . (b) PV wavefront error that encompasses 95 percent of the population is plotted as a function of pupil diameter. Three second-order states are shown: (i) residual aberrations after a conventional refraction using trial lenses (short blue dashed line) (ii) all aberrations present with zeroed Zernike defocus (long red dashed line), and (iii) all aberrations present with zeroed defocus and astigmatism (solid green line).

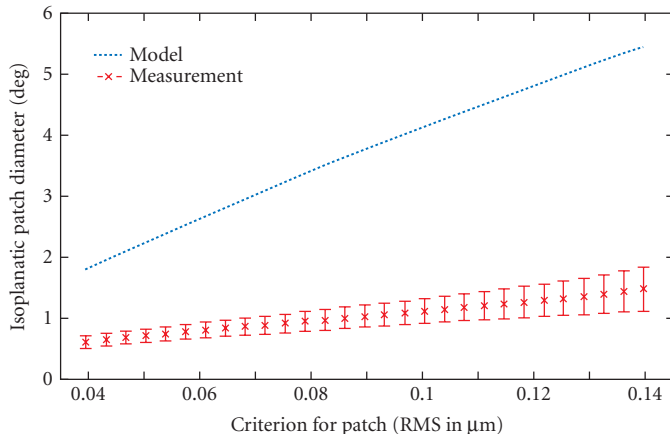
Lois-Diaz et al. up to approximately 30 Hz. Figure 2a shows representative temporal traces of the total RMS wavefront error and several Zernike terms for one subject and a model eye over a 5 s interval. Figure 2b shows corresponding average temporal power spectra which reveal the frequency content of the traces. As depicted in the figure, temporal fluctuations are found in all of the eye's aberrations, not just defocus, even when the eye's accommodation is paralyzed. In both studies, the reported temporal fluctuations decreased at  $f^{-4/3}$  or equivalently four dB/octave, where  $f$  is the temporal frequency. This decrease is evident in the temporal power spectrum shown in Fig. 2b. The vast majority of the aberration power lies below 1 to 2 Hz, suggesting effective AO correction needs only a temporal bandwidth of a couple of hertz. For comparison, the 1 to 2 Hz is approximately two orders smaller than that of atmospheric turbulence for ground-based telescopes.

Finally, aberrations of the eye also vary with field location at the retina. That is, image quality (or more precisely the point spread function or optical transfer function) at one point on the retina is not the same as at another point when the two points are separated sufficiently. This difference in image quality stems from the fact that the ocular aberrations originate at the cornea and crystalline lens rather than at the pupil of the eye. Rays originating from different field locations on the retina will therefore take slightly different paths through the ocular media, accumulating different phase delays or equivalently different aberrations. In general there is a local region about any point on the retina in which the path differences are sufficiently small that the ocular aberrations are effectively constant. This region is termed the isoplanatic patch.<sup>60</sup> The diameter of the patch depends on the properties of the eye, but also on the definition of image quality, as for example, use of the stringent Maréchal criterion,  $\lambda/14\text{-}\mu\text{m}$  RMS (Strehl ratio = 0.8) yields a narrower isoplanatic diameter than the more relaxed criterion of 1 rad<sup>2</sup> (Strehl ratio = 0.37). Which criterion to choose depends on the application, though it remains to be seen what criterion is optimal for common AO retinal imaging uses as well as AO psychophysical uses.



**FIGURE 2** Temporal properties of ocular aberrations.<sup>58</sup> (a) Temporal traces of the total RMS wavefront error and Zernike terms: defocus, astigmatism, coma, and spherical aberration for one subject. A trace of the total RMS wavefront error for an artificial eye is also shown and reflects the sensitivity of the instrument. (b) Average temporal power spectra are shown for the fluctuations in the total RMS wavefront error for a model eye and one subject whose accommodation was paralyzed. Aberrations were computed for a 4.7-mm pupil size. (Reproduced with permission from Ref. 58.)

Anecdotal evidence in the AO literature indicates the isoplanatic patch diameter for the human eye is a couple of degrees. Experiments to directly measure the patch size have only recently been conducted. A detailed survey of efforts in this area coupled with a more thorough analysis were recently published by Bedggood et al.<sup>61</sup> Figure 3 shows a representative result from this study, plotting isoplanatic patch diameter as a function of the residual wavefront RMS at the patch edge. As expected, the patch size increases as the criterion relaxes, that is, larger residual RMS. As an example, AO systems



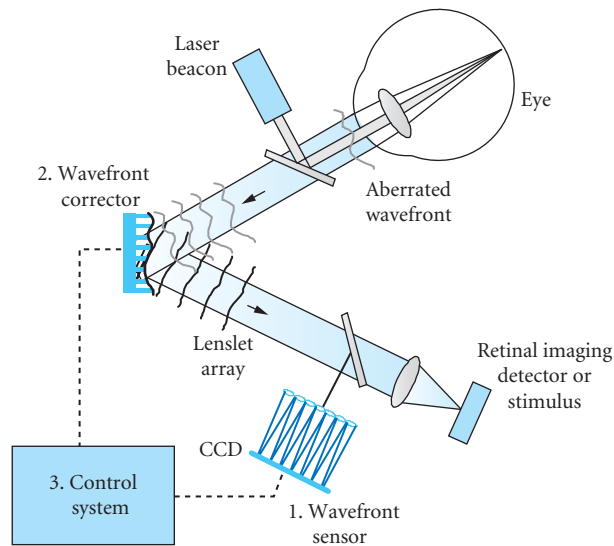
**FIGURE 3** Isoplanatic properties of ocular aberrations. Isoplanatic patch diameter is plotted versus residual wavefront RMS at the patch edge after perfect correction of ocular aberrations at the patch center. Data points are average measurements across seven healthy subjects, aged 22 to 39, with error bars at  $\pm 1$  standard deviation. Predicted patch diameter is also shown using the Liou Brennan schematic eye.<sup>62</sup> Pupil size for both is 6 mm. (Reproduced with permission from Ref. 61.)



typically correct the eye (with a large pupil) to within a wavefront RMS error of approximately  $0.1\ \mu\text{m}$ . Using this value as a realistic criterion for the residual RMS at the patch edge in Fig. 3, the corresponding isoplanatic diameter is approximately  $1.1^\circ$ , consistent with anecdotal evidence and other measurements reported in the literature. For comparison, this patch diameter is three orders of magnitude larger than that encountered with ground-based telescopes. Interestingly, optical models of the eye (e.g., the Liou Brennan schematic eye in Fig. 3) are not sufficiently developed yet to predict experimental findings, both in terms of the patch size and criterion dependence. As evident in Fig. 3, the Liou Brennan model overestimates the patch diameter by several times.

## 15.4 IMPLEMENTATION OF AO

The principle components of an AO system are the wavefront sensor, wavefront corrector, and control system. Figure 4 shows a simplified schematic of AO applied to the eye and highlights the principle components. The effectiveness of AO depends fundamentally on its ability to measure, correct,



**FIGURE 4** Concept schematic of AO applied to the eye. The three principle components of AO are (1) the wavefront sensor, (2) the wavefront corrector, and (3) the control system. The wavefront sensor (shown here as a Shack-Hartmann implementation) works by focusing a laser beam onto the subject's retina. This is analogous to the artificial-guide-star approach now being used to measure the wave aberrations of atmospheric turbulence.<sup>63</sup> The reflection from the retinal spot is distorted as it passes back through the refracting media of the eye. A two-dimensional lenslet array, placed conjugate with the eye's pupil, samples the exiting wavefront forming an array of images of the retinal spot. A CCD sensor records the displacement of the spots, from which first local wavefront slopes and then global wavefront shape are determined. Wavefront compensation is realized with a deformable mirror. The mirror lies in a plane conjugate with the subject's pupil and the lenslet array of the SHWS. For retinal imaging, a light source (not shown) illuminates the retina, some of which is reflected out of the eye, reflects from the wavefront corrector, and forms an aerial image at the retinal imaging detector. For visual testing, the retinal imaging detector is replaced with a visual stimulus that is viewed by the subject through the AO system.



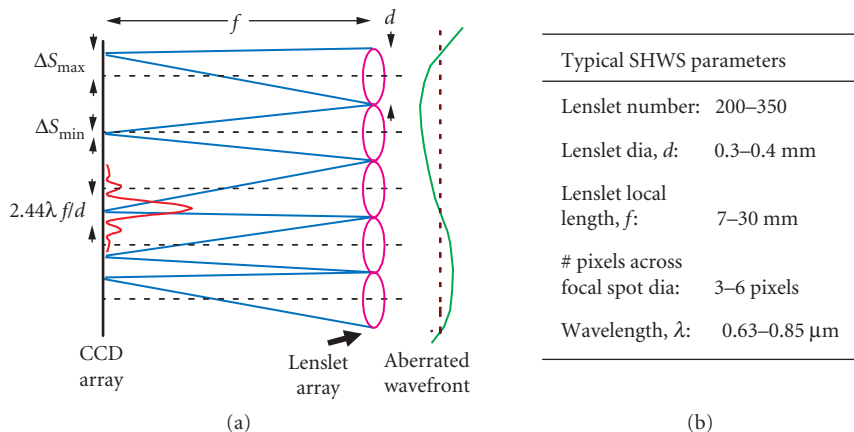
and track the ocular aberrations, the extent to which depends heavily on the spatial and temporal properties of the aberrations as described in the previous section. Measurement is accomplished by the wavefront sensor, correction by the wavefront corrector, and tracking by the control system in conjunction with the sensor and corrector. The status of each of these components and their requirements for effective use with the eye are described in order below.

## Wavefront sensor

There is an abundance of optical sensors for wavefront measuring with many representing highly mature devices. These include the Shack-Hartman sensor, shearing interferometer, pyramid sensor, curvature sensor, phase diversity, laser ray tracing, phase-shifting interferometer, and common path interferometer. While several of these have been applied to the eye and others continue to attract interest, to date all operational AO systems for the eye have been constructed around the SHWS. As such, we confine our attention to this sensor type, though much of which will be applicable to the other sensors. The discussion that follows will reference Fig. 5, which shows an enlarged schematic of the SHWS (lenslet and CCD array) along with typical values of several key parameters.

A first-order design of the SHWS entails evaluation of several key properties: (1) spatial sampling, (2) dynamic range, (3) accuracy, and (4) signal-to-noise. These are primarily set by the characteristics of the lenslet and CCD arrays of the sensor, and therefore selection of these two components is critical. Figure 5*b* shows typical values in use. Spatial sampling of the sensor is specified by the number of lenslets that fills the pupil of the eye. The necessary number of lenslets for reliable representation of the ocular aberrations depends on the composition of the ocular aberrations, pupil size of the eye, and to a lesser extent the SH wavelength. It should be noted that while we assume here the lenslets to be uniformly distributed (as is the case of SH sensors in current AO systems), several nonuniform distributions have also been evaluated.<sup>64</sup>

Theoretical modeling in conjunction with measured aberrations of the eye indicates a one-to-one relationship between the number of Zernike modes and number of lenslets necessary to reliably reconstruct these modes.<sup>65</sup> Therefore, the number of required lenslets that fill the pupil must be at



**FIGURE 5** (a) A two-dimensional lenslet array, placed conjugate with the eye's pupil, samples the exiting wavefront forming an array of images of the retinal spot. A CCD sensor records the displacement of the spots. Maximum and minimum detectable displacements of the spots are indicated by  $\Delta S_{\max}$  and  $\Delta S_{\min}$ . (b) Key SHWS parameters and typical values for use with the eye. Lenslet number refers to the number that fills the pupil of the eye. Lenslet diameter is referenced to the eye pupil. Pixels refer to the detector unit of the CCD array.  $\lambda$  is the wavelength of the SHWS laser beacon.

least equal to the number of Zernike modes, which is related to the maximum Zernike order by  $(N+1)(N+2)/2 - 3$ .  $N$  is the maximum Zernike order and piston, tip, and tilt are not included. Based on this, reliable representation up through 10th Zernike order requires at least 63 lenslets, that is, equals  $(10+1)(10+2)/2 - 3$ , that uniformly cover the pupil.

Figure 1a provides additional information—at least for normal healthy eyes—as to the number of Zernike orders necessary for reliably representing ocular aberrations. As an example, for the stringent condition of diffraction-limited imaging through a large 7.5-mm pupil in 95 percent of the population, Fig. 1a indicates Zernike representation up through at least 10th order is necessary. For this scenario, at least 63 lenslets are therefore required to reliably sample the aberrations. Fewer lenslets are necessary for a smaller pupil or a less stringent image quality metric. In practice, SHWSs routinely sample the pupil much higher. There is little cost to do so, since there is an over abundance of light for the sensor. Oversampling makes the measurements more robust (to pupil edge effects, eye motion, drying of the precorneal tear film, and system noise) and applicable to eyes with a wider range of aberrations.

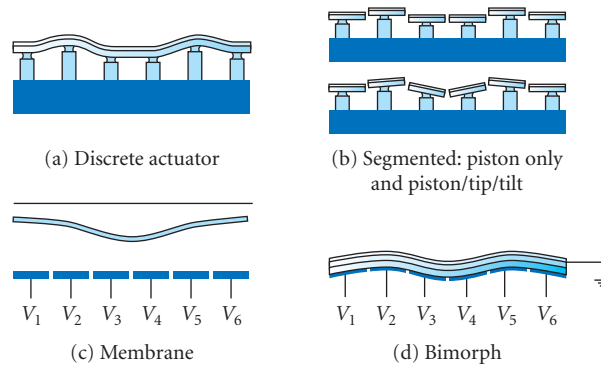
Dynamic range refers to the maximum wavefront slope  $\Delta\theta_{\max}$  that can be measured by the SHWS (depicted in Fig. 5a).  $\Delta\theta_{\max}$  is typically expressed as  $\Delta S_{\max}/f$ , where  $\Delta S_{\max}$  is the maximum displacement of the lenslet spot (= lenslet radius) and  $f$  the lenslet focal length. Since spherical and cylindrical refractive errors typically consume much of the dynamic range of the sensor, it is often more useful to convert  $\Delta S_{\max}$  to a maximum measurable defocus (diopters), expressed by  $D = \Delta\theta_{\max} /$  (pupil radius). For example, a lenslet diameter of 0.4 mm and a lenslet focal length of 24 mm, gives a  $\Delta\theta_{\max}$  and  $D$  of 8.3 mrad and 2.45 diopters, respectively, for a 6.8-mm eye pupil. Note that  $D$  represents the dioptic range at the wavefront sensor. The corresponding range at the eye is often different depending on the eye-to-sensor magnification. In general, the dynamic range can be increased by increasing the lenslet diameter, decreasing the lenslet focal length, or decreasing the pupil diameter. The pupil diameter, however, is usually not adjustable, being defined early in the design process of the retina camera.

Accuracy refers to the minimum wavefront slope,  $\Delta\theta_{\min}$ , that can be measured by the SHWS.  $\Delta\theta_{\min}$  can be expressed as  $\Delta S_{\min}/f$ , where  $\Delta S_{\min}$  is the minimum detectable displacement of the lenslet spot (see Fig. 5a).  $\Delta S_{\min}$  depends on the pixel size of the SHWS detector (typically a CCD array), diameter of lenslet spot at the detector (which is a function of the lenslet diameter and focal length), accuracy of the centroiding and thresholding algorithms, and signal-to-noise of the captured SHWS spot image. In general, subpixel accuracy is routinely achieved.<sup>66</sup> The above example (lenslet diameter = 0.4 mm and focal length = 24 mm) with a CCD pixel size of 26  $\mu\text{m}$  yields approximately 4.4 pixels across the focal spot and  $\Delta\theta_{\min} \ll 1.1$  mrad. While the SHWS parameters can be manipulated to increase accuracy, dynamic range of the sensor depends inversely on the lenslet diameter and focal length, thereby creating a clear trade-off between dynamic range and accuracy. Additional discussion on this topic can be found in Ref. 66.

Signal-to-noise of the captured SHWS spot images is limited by the amount of light that can be safely directed into the eye, number of pixels per lenslet focal spot, quantum efficiency of the SHWS CCD, and throughput efficiency the SHWS. Unlike astronomical AO systems, however, that operate under light-starved conditions and covet each photon to achieve the approximately 100 photons/lenslet required to close the AO loop, vision AO systems operate under comparatively relaxed light requirements. A typical vision AO system has a minimum of approximately 500,000 detected photons/lenslet, this with hundreds of lenslets that sample the pupil. Such high light levels permit use of relatively low quantum efficient CCDs and tens of pixels that sample the core of the lenslet focal spot.

## Wavefront Corrector

The effectiveness of AO to correct ocular aberrations is typically limited by the performance of the wavefront corrector, which is also by far the most expensive component. This device dynamically imparts an ideally conjugate aberration profile onto the passing wavefront, thus canceling the original aberrations. Numerous types of wavefront correctors have been employed in AO systems for the eye. Historically, correctors have been selected somewhat arbitrarily for the eye, that is, driven



**FIGURE 6** Four main classes of wavefront correctors. (a) Discrete actuator deformable mirrors consist of a continuous, reflective surface, and an array of actuators, each capable of producing a local deformation in the surface. (b) Piston-only segmented correctors consist of an array of small planar mirrors whose axial motion (piston) is independently controlled. LC-SLMs modulate the wavefront in a similar fashion, but rely on changes in refractive index rather than the physical displacement of a mirror surface. Piston/tip/tilt segmented correctors add independent tip and tilt motion to the piston-only correctors. (c) Membrane mirrors consist of a grounded, flexible, reflective membrane sandwiched between a transparent top electrode, and an underlying array of patterned electrodes, each of which is capable of producing a global deformation in the surface. (d) Bimorph mirrors consist of a layer of piezoelectric material sandwiched between a continuous top electrode and a bottom, patterned electrode array. A top mirrored layer is added to the top continuous electrode. An applied voltage causes a deformation of the top mirrored surface. (Reproduced with permission from Ref. 41, chap. 4, Fig. 4.2.)

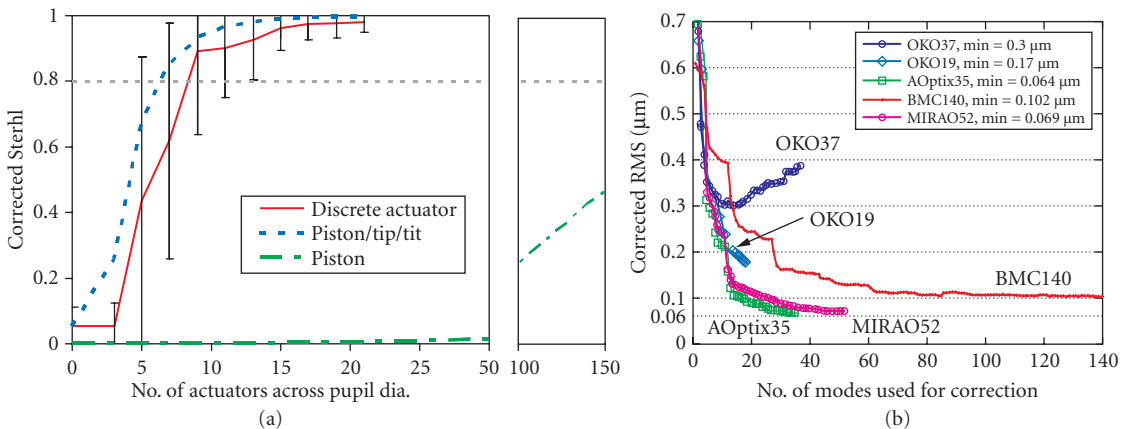
primarily by availability and cost rather than performance, with the expectation that image quality will improve, the extent of which was empirically determined. More recently, several extensive theoretical studies have evaluated the performance of general wavefront corrector classes<sup>67,68</sup> as well as specific commercial devices.<sup>69,70</sup> The four main classes of wavefront correctors are depicted in Fig. 6. These include discrete actuator, membrane, bimorph, and segmented piston and piston/tip/tilt wavefront correctors. Regardless of class, performance is governed primarily by the corrector's actuator number, stroke, and influence function. Stroke refers to the dynamic range of the corrector and limits the largest wavefront error that can be corrected. Larger stroke provides better correction of large-magnitude aberrations. Actuator number across the eye's pupil and actuator influence function (localized deflection of the mirror surface that results when a single actuator is pushed or pulled) determine the fidelity of the correction. More actuators and a more localized influence provide better correction of high-spatial-frequency aberrations (i.e., aberration, that are of higher order). Other important corrector parameters include temporal response, surface reflectivity, diameter of mirror surface, and cost.

Wavefront correctors first applied to the eye were developed primarily for compensation of atmospheric turbulence. A common example being macroscopic discrete actuator deformable mirrors (DMs), such as those manufactured by Xinetics, Inc.<sup>71</sup> Specifically, their actuator number, stroke, influence function, and speed were tailored to the spatial and temporal properties of the atmosphere<sup>4</sup> rather than that of the eye.<sup>43,44,58,59</sup> The dynamic range of such devices often limited the compensation of ocular aberrations, especially second-order contributions, and their kilohertz response was overkill for the 1 to 2 Hz fluctuations of the ocular aberrations. While wavefront correctors represent a small fraction of the total cost of ground-based telescopes in which they are employed, they represent a significant fraction of the total cost of most retina cameras. Atmospheric wavefront correctors are also generally bulky, with large mirror surfaces (~several centimeters or more) that require long focal length relay optics to magnify the pupil of the eye. A smaller corrector comparable to the dilated pupil of the eye (4 to 8 mm) can substantially reduce the instrument size, which is attractive for research retina cameras and mandatory for commercial ones.

Alternative wavefront corrector technologies, which are more cost effective and smaller, have become commercially available and span the four corrector classes depicted in Fig. 6. Large stroke (up to  $16\ \mu\text{m}$ ) bimorph mirrors having 13 to 35 actuators have been investigated by several groups.<sup>15–17,28,30,72</sup> An even larger stroke ( $50\ \mu\text{m}$ ) corrector, a magnetic membrane mirror with 52 actuators was recently evaluated by Fernandez et al.<sup>73</sup> Microelectromechanical systems (MEMS) promises batch fabrication of low cost, compact wavefront correctors. Bulk micromachined membrane MEMS mirrors, employing 37 electrodes and  $3.5\text{-}\mu\text{m}$  stroke have been successfully applied to the eye.<sup>74,75</sup> Although bimorph and membrane mirrors have a large dynamic range for correction of low-order aberrations, the effective range drops rapidly with increasing order of the aberration mode.

Surface micromachined devices are another class of MEMS mirror whose mode of operation is comparable to discrete actuator DMs. Early performance results with a surface micromachined MEMS DM was reported by Doble et al.<sup>76</sup> Today, this mirror type is employed in numerous AO systems for the eye.<sup>20,21,24,28,30,32</sup> Liquid spatial light modulators (LC-SLMs) are an alternative wavefront corrector technology. Transmissive, pixelated designs with 69 and 127 pixels were examined by Thibos et al.<sup>77</sup> and Vargas et al.,<sup>78</sup> respectively. Prieto et al.<sup>79</sup> and Fernandez et al.<sup>29</sup> investigated an optically addressed LC-SLM. This reflective device has high spatial resolution ( $480 \times 480$  piston-only pixels) and operates with low control voltages ( $\sim 5\ \text{V}$ ). LC-SLMs in general, however, are limited to  $2\pi$  phase modulation of polarized light. Extended ranges can be achieved with phase wrapping.

Theoretical studies to predict the performance of wavefront correctors and guide their use with the eye have been undertaken with representative results from two such studies given in Fig. 7. Figure 7a shows the predicted corrected Strehl for three general types of wavefront correctors (discrete actuator, piston/tip/tilt, and piston) in conjunction with the measured ocular aberrations of Fig. 1. In this study, the effect of actuator stroke was separated from that of the actuator number and influence function by assuming sufficient stroke in the Fig. 7a plot. For the discrete actuator corrector in the figure, at least a  $13 \times 13$  array of actuators is required to achieve diffraction-limited imaging (Strehl = 0.8) in 95 percent of the population, that is,  $\pm 2$  standard deviations. Piston/tilt correctors require somewhat less and piston correctors require substantively more. As shown in Fig. 7a, 100 to 150 piston



**FIGURE 7** (a) Predicted corrected Strehl ratio for three general types of wavefront correctors (discrete actuator, piston/tip/tilt, and piston) as a function of actuator number across the pupil diameter.<sup>68</sup> Pupil diameter at the eye is 6 mm, wavelength is  $0.6\ \mu\text{m}$ , and the ocular aberrations are given in Fig. 1. Sufficient corrector stroke is assumed. The error bars for the single representative curve correspond to  $\pm 2$  standard deviations. (b) Predicted corrected RMS for several commercial correctors is shown as a function of number of corrector modes used for correction.<sup>70</sup> Model includes thresholding to account for the finite mirror stroke. Devices are from OKO Technologies (OKO19 and OKO37), Boston Micromachines Corp. (BMC140), AOptix Technologies (AOptix35), and Imagine Eyes (MIRA052). Pupil diameter at the eye is 6 mm and the randomly generated ocular aberrations are based on the population statistics of Fig. 1.

segments (maximum sampling density of the study) across the pupil diameter are required to reach a corrected Strehl of only 0.4. As for the required stroke of the corrector, it should be at least one-half the PV error of the aberrations to be corrected, for example, the PV error in Fig. 1*b*. For example, in Fig. 1, the PV error for the 7.5 mm pupil ranges from 7 to 11  $\mu\text{m}$  depending on the second-order state. Thus the required stroke must be at least one half of this (3.5 to 5.5  $\mu\text{m}$ ). Note that the required stroke can increase significantly if a larger portion of the second-order aberrations is corrected with AO rather than for example trial lenses. As an example, a 1-D spherical error across a 7.5-mm pupil increases the required stroke by an additional 3.5  $\mu\text{m}$  based on the equation: stroke = 1/4 (refractive error in diopters) (pupil radius in millimeters)<sup>2</sup>.

Figure 7*b* shows results from another study in which performance was predicted for several commercially available correctors. The modeling took into account the actual influence functions, finite stroke of each corrector (thresholding), and number of corrector modes included in the correction. The commercial devices evaluated included three membrane mirrors (OKO19, OKO37, and MIRA052), one bimorph mirror (AOptix35), and a surface micromachined MEMS deformable mirror (BMC140). Of these only two (AOptix35 and MIRA052) approach diffract limited performance.

In practice, AO retina cameras with a single wavefront corrector have not achieved sufficient correction to yield diffraction-limited imaging across large pupils ( $\geq 6$  mm), more so in the presence of typical levels of second-order aberrations. Current correctors—while having improved noticeably in recent years—struggle to provide the necessary high spatial fidelity and high dynamic range to fully compensate both low- and higher-order aberrations of the eye. To overcome this barrier, other approaches have been explored, including a multipass configuration that traverses the same corrector more than once<sup>80</sup> and a cascade of two wavefront correctors, one of high spatial fidelity and the other of large stroke.<sup>34,81,82</sup>

## Control System

The control system performs the critical step of rapidly and repeatedly converting the raw output of the wavefront sensor into voltage commands that shape the corrector surface, that is, the deformation created by the sum response of the corrector actuators. The control loop takes into account the number, arrangement, and interaction of lenslets and actuators; the influence function and stroke of actuators; temporal response of the AO components; and a desired performance criteria, for example, minimize the RMS wavefront error. The control loop can be characterized by distinct spatial and temporal properties, which are each discussed in order further.

Spatial control starts with the layout of lenslets and actuators. For AO astronomical telescopes, the lenslet-actuator arrangement typically follows a Southwell or Fried configuration in which each lenslet is centered on one actuator or centered between four adjacent actuators. Both geometries yield one lenslet per actuator. This ratio maximizes the number of photons per lenslet, critical for the photon starved conditions under which astronomical AO systems operate, while preserving sensitivity to actuator motion. The relative coarse sampling of the lenslets, however, makes the AO sensitive to misalignment of the lenslet and actuator arrays, and the wavefront sensor cannot detect certain corrector modes, as for example waffle mode. As discussed in section “Wavefront Sensor,” the eye provides substantially more light than astronomical applications, allowing many more lenslets per actuator. Typical AO systems for the eye sample each actuator with two to ten lenslets. This oversampling greatly increases the tolerance to alignment errors and permits detection of all corrector modes.

Spatial (static) control of the wavefront corrector is normally realized through a multiplication of two matrices, one representing the wavefront sensor output (slope measurements) and the other a reconstructor matrix representing the interaction of each actuator with each lenslet. At its most basic level, the wavefront sensor output is an areal image of focus spots arranged in a quasi-regular pattern and resting on a noise floor, one spot formed behind each lenslet. The critical information in the image is the displacement of the focus spots relative to predetermined reference positions. Determining these displacements is nontrivial owing to noise in the image, and the finite width

and often irregular shape of the focus spots. Collectively, these reduce the accuracy to locate the spot centers. Centroiding based on center of mass is a common solution to this problem and is effective as it takes into account the full energy distribution of the spot. Following the mathematical notation by L. Chen,<sup>83</sup> the horizontal and vertical centroid positions for the  $k$ th lenslet in the array and a unit voltage applied to the  $m$ th actuator in the corrector can be expressed as

$$\Delta x_{k,m} = \frac{\sum_{i,j \in k} x_{i,j} I_{i,j,m}}{\sum_{i,j \in k} I_{i,j,m}} - x_{k,ref} \quad \text{and} \quad \Delta y_{k,m} = \frac{\sum_{i,j \in k} y_{i,j} I_{i,j,m}}{\sum_{i,j \in k} I_{i,j,m}} - y_{k,ref} \quad (1)$$

$i$  and  $j$  represent coordinates for a pixelated detector, such as a CCD array.  $I_{i,j,m}$  is the intensity distribution of the focus spots;  $x_{k,ref}$  and  $y_{k,ref}$  are reference coordinates, for example, those that minimize the wavefront error. The corresponding horizontal and vertical wavefront slopes are determined from the centroids by

$$s_{km_x} = \frac{\Delta x_{k,m}}{f} \quad \text{and} \quad s_{km_y} = \frac{\Delta y_{k,m}}{f} \quad (2)$$

where  $f$  is the focal length of the lenslets (see Fig. 5). In matrix notation, the wavefront slope output of the sensor can be expressed as

$$S_m = (s_{1mx}, s_{1my}, s_{2mx}, s_{2my}, \dots, s_{kmx}, s_{kmy}, \dots, s_{Kmx}, s_{Kmy})^T \quad (3)$$

where  $K$  is the total number of lenslets. Systematically applying a unit voltage to each of the  $M$  actuators in the corrector and recording the resulting displacement of the focus spots  $S_m$ , a  $2K \times M$  matrix is constructed called the influence function matrix:

$$A = (S_1, \dots, S_m, \dots, S_M) \quad (4)$$

Because  $A$  is generated by systematically applying unit voltages to the actuators,  $A$  is mathematically related to the applied voltages,  $V = (v_1, v_2, \dots, v_M)^T$ , and the resulting wavefront slopes,  $S = (s_{1x}, s_{1y}, \dots, s_{Kx}, s_{Ky}, \dots, s_{Kx}, s_{Ky})^T$ , by

$$S = AV \quad (5)$$

Equation (5) can be solved for  $V$  by first finding an inverse for  $A$ . Since  $A$  is typically singular (e.g., a nonsquare matrix), no inverse exists and mathematical techniques such as singular value decomposition (SVD) are commonly used to find an appropriate pseudoinverse, denoted  $A^+$ :

$$V = A^+ S \quad (6)$$

$A^+$  is often referred to as a reconstructor matrix as a single matrix multiplication with  $S$  reconstructs the actuator voltages necessary to correct the measured aberrations. The single matrix multiplication is fast and efficient, and for SVD represents a least squares solution.

The control model described above assumes the AO operates in a linear fashion. While this is overly simplistic for actual AO systems, the AO control loop is much more robust than suggested, converging even when nonlinearities are present owing to its closed-loop operation. Equation (6) refers to what is called the direct slope reconstruction method. Another approach that is sometimes employed is the modal reconstruction method in which wavefront slope measurements are first converted to Zernike aberration modes that in turn are converted to applied voltages. While this approach is slower and less efficient, it has the advantage of allowing explicit control of individual modes, including which ones are sent to the corrector.

Equation (6) does not capture the closed-loop performance of AO as it neglects the temporal dynamics of the ocular aberrations and the time delays associated with measuring the wavefront slopes, and computing and applying the control voltages. The resulting error can be mathematically

expressed as  $\phi_{\text{resid}}(x, t) = \phi_{\text{aberr}}(x, t) - \phi_{\text{correct}}(x, t)$ , where  $\phi_{\text{aberr}}(x, t)$  is the ocular aberrations at time  $t$ ,  $\phi_{\text{correct}}(x, t)$  the applied correction based on slope measurements acquired at an earlier time  $t - \tau$ , and  $\phi_{\text{resid}}(x, t)$  the uncorrected portion at time  $t$ .  $\tau$  is the closed-loop period of the AO. Note that  $\phi_{\text{correct}}(x, t)$  is directly related to the voltages,  $V$ , that are applied iteratively to the corrector and expressed as  $V(t) = V(t - \tau) - G V_{\text{estimate}}(t)$ .  $V(t - \tau)$  are voltages already on the corrector as determined from the previous loop iteration,  $V_{\text{estimate}}(t)$  are adjustments to the voltages based on Eq. (6), and  $V(t)$  represent the updated voltages.  $G$  is the variable loop gain (adjusted between 0 and 1) that controls the extent to which  $V_{\text{estimate}}(t - \tau)$  is applied and represents the trade-off between stability and sensitivity of the AO system, for example, higher gain increases sensitivity to temporal fluctuations in the aberrations, but at the expense of decreased system stability. Typical values of  $G$  for AO vision science instruments are between 0.1 to 0.4.

Predicting the temporal performance of an AO system is critical for optimizing system design for the eye and is accomplished by modeling the system as a cascade of transfer functions, each representing an independent component of the system. Figure 8 shows a block diagram representation of an AO control loop with input/output parameters  $M(s)$ ,  $X(s)$ , and  $R(s)$  that depict the time-varying aberrations of the eye, residual aberrations after correction, and wavefront corrector voltages.  $s$  is a complex frequency and defined as  $i2\pi f$ , where  $f$  is the temporal frequency and  $i^2 = -1$ . Noise in the system is assumed negligible. For analysis, the AO system can be decomposed as a linear cascade of independent transfer functions, in this case the four AO stages that most often limit temporal performance in AO vision science systems: SHWS exposure, sensor readout and computation of  $V$ , integral compensator, and finally holding mirror position for one AO loop period. Common expressions for the four transfer functions are given by

$$H_{\text{exp}}(s) = \frac{1 - e^{-sT_1}}{sT_1} \quad (7)$$

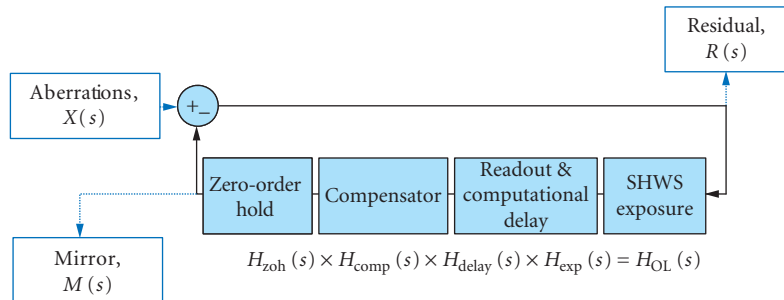
$$H_{\text{delay}}(s) = e^{-st_1} \quad (8)$$

$$H_{\text{comp}}(s) = \frac{G}{1 - e^{-sT_2}} \quad (9)$$

$$H_{\text{zoh}}(s) = \frac{1 - e^{-sT_2}}{sT_2} \quad (10)$$

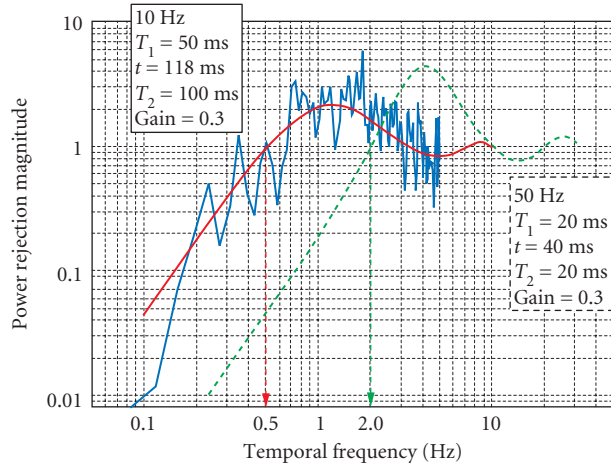
where  $T_1$  is the exposure duration of the SHWS detector,  $T_2$  is the sampling period of the AO loop, and  $t_1$  is the total delay to readout the sensor and compute  $V$ .

Using the notation in Fig. 8, the open-loop transfer function (single iteration),  $H_{\text{OL}}(s)$ , is equal to  $M(s)/R(s)$  and corresponds to the multiplication of the four transfer functions, that is,  $H_{\text{exp}}(s) H_{\text{delay}}(s) H_{\text{comp}}(s) H_{\text{zoh}}(s)$ . While this quantifies the extent to which fluctuations in the aberrations are



**FIGURE 8** Block diagram of the AO system for modeling temporal performance.





**FIGURE 9** Experimental and theoretical curves for the power rejection magnitude of a representative AO system for vision science. Experimental curve (jagged blue line) was obtained on one eye using a gain of 0.3 for a 6.8-mm pupil. The corresponding theoretical curve (solid, red) was based on the actual system parameters as highlighted in the leftmost box. The second theoretical curve (dashed, green) predicts the performance of an AO system that is 5 times faster. System parameters are highlighted in the rightmost box. Cutoff frequencies for the two configurations are 0.5 and 2.0 Hz.

reliably transferred to the corrector, a more useful transfer function is the error transfer function,  $H_{\text{error}}(s)$ , defined as  $R(s)/X(s)$ .  $H_{\text{error}}(s)$  quantifies the extent to which fluctuations in the aberrations are reduced by the AO.  $H_{\text{error}}(s)$  is related to  $H_{\text{OL}}(s)$  by the expression  $1/[1 + H_{\text{OL}}(s)]$ . As an example, Fig. 9 shows the power rejection magnitude (defined as  $|H_{\text{error}}(s)|^2$ ) for two representative AO configurations, one being 5 times faster than the other. Both configurations employ double buffering, a common technique realized with a frame-transfer CCD camera, in which one SHWS frame is processed while the next frame is being acquired. The power rejection magnitude permits determining the cutoff frequency of the AO system, being defined as the frequency at which the power rejection magnitude first reaches 1. For the two configurations in Fig. 9, cutoff frequencies occur at 0.5 and 2.0 Hz, with the former consistent with the experimental measurements shown. Note that a cutoff of 0.5 Hz indicates that temporal frequencies below 0.5 Hz are (at least partially) corrected by the system, while those above 0.5 Hz are (at least partially) amplified. Because the vast majority of the aberration power in the eye lies below 1 to 2 Hz (see Fig. 2 and related discussion), the 2.0-Hz bandwidth AO configuration should be adequately fast and should provide somewhat improved performance over that of the 0.5-Hz system.

## 15.5 APPLICATION OF AO TO THE EYE

The sole function of AO in ophthalmic applications is to compensate or control the aberrations in the optics that lie between the eye's fundus and the rest of the world. As such, AO can be implemented into any vision or imaging application that employs these optics. This section will be split into applications where the goal is to get a better view of the retina and applications where the goal is to reduce or manipulate the blur of light delivered to the retina.

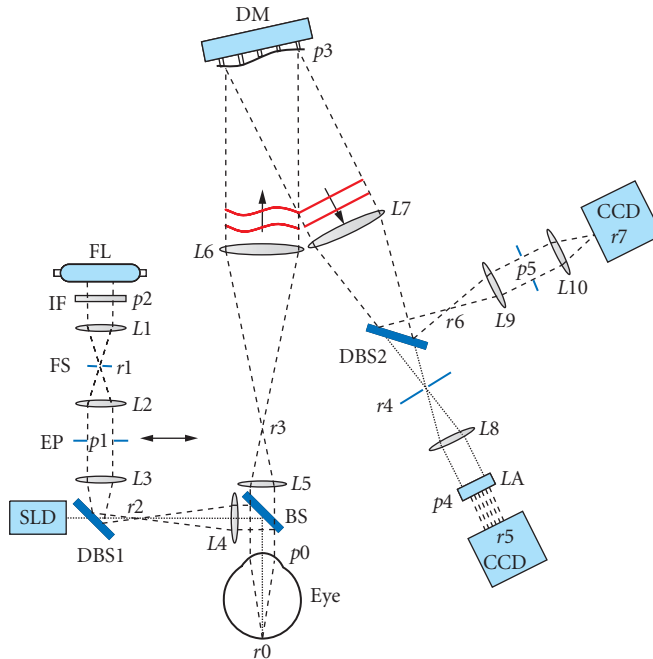


## Applications Involving Imaging of the Retina

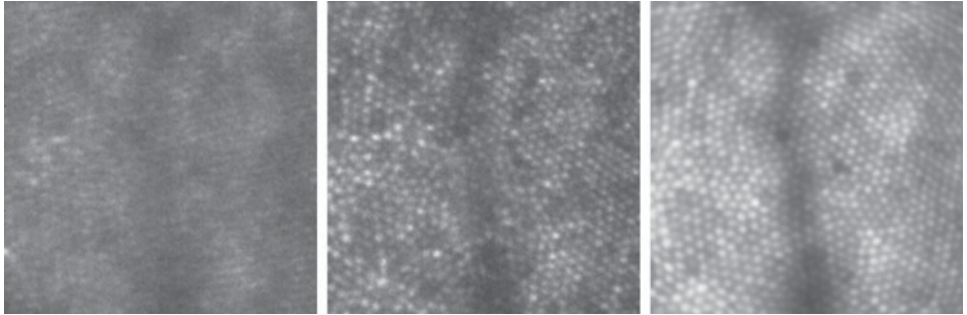
AO's role in a retina camera is to achieve diffraction-limited imaging through the highest possible numerical aperture in the eye. Most imaging applications that use AO correct over pupil sizes of 5 mm or greater, thereby taking advantage of the reduced diffraction. For example, a corrected pupil of 6 mm has a Rayleigh resolution limit about  $2\ \mu\text{m}$ , which is on the scale of the smallest cells in the retina (rods and foveal cones, for example).

**Flood-Illuminated AO Ophthalmoscope** The first ophthalmic imaging modality that demonstrated a high-order aberration correction with AO was a flood-illuminated ophthalmoscope.<sup>13</sup> In this system (Fig. 10), the camera looks at the retina through an optical system where the deformable mirror is placed in the path in a plane that is conjugate to the pupil of the eye. A low coherent collimated light source (SLD, superluminescent diode) is generally used to project a beacon onto the retina. The scattered light passes through the system, off the DM and to the wavefront sensor (LA & CCD). The AO system operates in a closed loop to correct the eye's aberration. At the same time, a separate light source is used to illuminate a patch of retina limited by the field stop (FS) at the location of the beacon. The scattered light emerges, its wavefront is compensated by the DM, and the light is directed to the science camera after reflection from a dichroic beamsplitter (DBS2) to form a sharp retinal image. The camera can be moved to focus on different layers of the retina. Keeping track of the retinal and pupil conjugate positions is critical and they are labeled throughout the schematic.

The flood-illuminated AO ophthalmoscope has been used effectively to measure basic optical properties of photoreceptors. A combination of AO imaging with retinal densitometry was used to provide the first maps of the arrangement of the three cone classes in the human retina (see Fig. 11).<sup>84,85</sup> It was



**FIGURE 10** Schematic of a flood-illuminated AO retina camera.

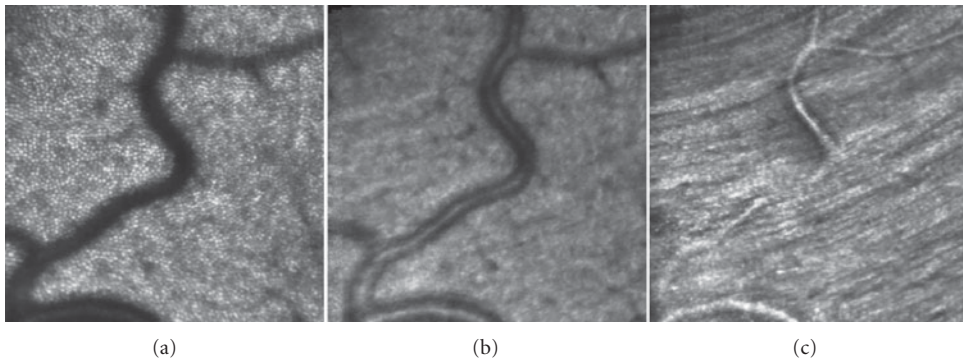


**FIGURE 11** The three images are of the same location in a human retina about  $1^\circ$  ( $300\ \mu\text{m}$ ) from the foveal center. The left image is taken after careful correction of defocus and astigmatism. The center image is a single frame taken with higher-order aberrations correction. The right image is a registered sum of about 20 frames. The small spots are single cone photoreceptors and in the registered frame virtually all cone photoreceptors are resolved.

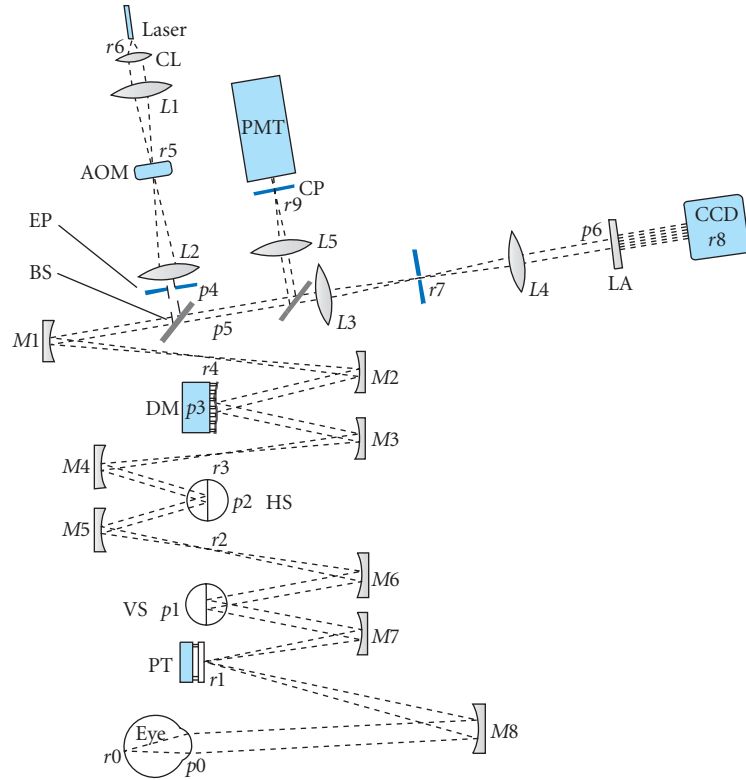
also used to measure the directional properties of individual cones,<sup>86</sup> measure the locus of fixation relative to the maximum density of foveal cones,<sup>87</sup> measure changes in reflectance of cones over a diurnal period,<sup>88</sup> reveal new information about photoreceptor packing and structure in eyes with retinal degenerations,<sup>89–91</sup> and detect fast scattering changes of cones in response to visual stimulation.<sup>92,93</sup>

**Adaptive Optics Scanning Laser Ophthalmoscope** The scanning laser ophthalmoscope (SLO)<sup>94</sup> is a special application of the scanning laser microscope, invented in 1955 by Marvin Minsky.<sup>95</sup> The main difference with an SLO is that the eye is used as the objective lens and the retina is always the sample being imaged. As such, the lateral and axial resolution of the SLO imaging is limited by the quality of the objective lens (cornea and lens), which is why AO can provide a significant benefit.

The main motivation to employ AO in an SLO is because of its confocal property. Unlike flood-illuminated cameras, confocal systems block scattered light from features that are outside the focal plane, thereby improving the contrast of features in the focal plane of interest. Moreover, by moving the focal plane, one can optically “section” the retinal tissue (see Fig. 12 for example). These features offer significant benefits for ophthalmoscopy because the retina is a thick, multilayered structure, which scatters throughout its depth.



**FIGURE 12** Confocal optical sectioning with an AO SLO. The sequence of AO-SLO images shows a variation in retinal structure as the focal plane of the instrument is systematically shifted from the (a) photoreceptor layer, (b) the blood vessels, and (c) the nerve fiber layer.



**FIGURE 13** Schematic of an adaptive optics scanning laser ophthalmoscope. Key: CL—collimating lens; AOM—acousto-optic modulator; EP—entrance pupil; BS1—beamsplitter 1; DM—deformable mirror; HS—horizontal scanning mirror; VS—vertical scanning mirror; PT—pupil tracking mirror; LA—lenslet array; CP—confocal pinhole; PMT—photomultiplier tube. Pupil and retinal conjugates are labeled  $p$  and  $r$ , respectively. Mirrors and lenses are labeled  $M\#$  and  $L\#$  along the optical path. Telescope lens/mirror-pairs for relaying the pupil through the path are  $L1-L2$ ,  $L3-L4$ ,  $M1-M2$ ,  $M3-M4$ ,  $M5-M6$ , and  $M7-M8$ .

An SLO image is acquired over time as the scattered light is recorded from a focused spot on the retina as it scans in a raster pattern across the retina. The image is formed digitally in a frame grabber, which combines horizontal and vertical position information from the scanning mirrors with the digitized values of the analog intensity stream from the detector, to form an extended field image.

The concept of applying AO to an SLO was first proposed by Dreher et al.<sup>10</sup> but was not fully implemented with a wavefront sensor and a mirror capable of correcting high-order aberrations until Roorda et al.<sup>19</sup> The layout of a representative AO-SLO instrument is shown in Fig. 13.

In a confocal scanning laser imaging system, the effective PSF is computed as

$$\text{PSF}_{\text{SLO}} = \text{PSF}_{\text{in}} \times (\text{PSF}_{\text{out}} \otimes D) \quad (11)$$

where  $D$  represents the confocal aperture. When the confocal aperture approaches a size equal to the radius of the airy disk of the collection path of the system, the effective PSF is simply the product of the ingoing and outgoing PSFs. Under these optimal conditions, the lateral resolution exceeds conventional imaging by about 40 percent, as assessed by the full width at half maximum (FWHM)

of the PSF. If the confocal aperture is large, then image quality is governed only by  $\text{PSF}_{\text{in}}$ , effective optical sectioning disappears, and lateral resolution is the same as a flood-illuminated imaging system. Axial resolution in confocal SLOs can be defined and computed in several ways. The standard way to determine axial resolution is to measure the detected intensity from a flat, diffusely scattering surface as it moves relative to the focal plane of the SLO.<sup>10</sup> The FWHM of the resulting intensity distribution is a measure of the axial resolution of the SLO.<sup>96</sup>

The AO SLO employs a series of telescopes to relay the light to the elements that act on the beam. For example, wavefront sensing ( $p6$  in Fig. 13), wavefront correcting ( $p3$ ), and tip and tilt adjustments ( $p1$  and  $p2$ ) for raster scanning the beam all need to be done in pupil-conjugate planes. At present, each of these actions is done with a separate component and so a telescope is required to relay conjugate images of the pupil to the various components. Maintaining conjugacy of the pupil planes is critical in an AO SLO. If, for example, the eye's pupil was not conjugate to the scanning mirror, the beam would scan across the pupil, rather than pivoting about the pupil, and neither the wavefront sensor nor deformable mirror would see a stable aberration pattern. To avoid back reflections, the telescopes employ focusing mirrors rather than lenses. A consequence of using mirrors in the optical path is that they have to be used off-axis, which generates unwanted aberrations in the optical path. However, with the help of optical design software, each mirror in the system can be adjusted to compensate for the coma generated by the previous reflection, and the remaining astigmatism can be corrected by a cylindrical lens placed somewhere in the path.<sup>97</sup>

Wavefront sensing and compensation is done in a unique way in an AO SLO. First, AO is effective in both directions; to focus the light to a sharp point on the retina and to take the scattered light from the eye and focus it to a sharp point at the confocal pinhole. In the AO SLO, the wave aberration is measured with the same light that is used to form the image. This is possible because, although the light is being scanned in a raster pattern on the retina, the light is descanned on the return path, rendering it stationary again. Thus, the wavefront sensor sees the light from the retina as though it were coming from a single spot, which makes an aberration measurement possible. This method has several advantages. First, it automatically implements the method of Hofer et al. in which scanning on the retina is employed to remove speckle from the short-exposure retinal images.<sup>58</sup> A second advantage is that the average aberration is measured over the entire field of view of the system, thereby ensuring more uniformity in the correction over the field that is being imaged. The final advantage is that the wavefront sensing and the imaging portions of the system use the same light path and light source, which reduces noncommon path errors and eliminates noncommon aberrations between the wavefront sensor and imaging camera due to chromatic effects.

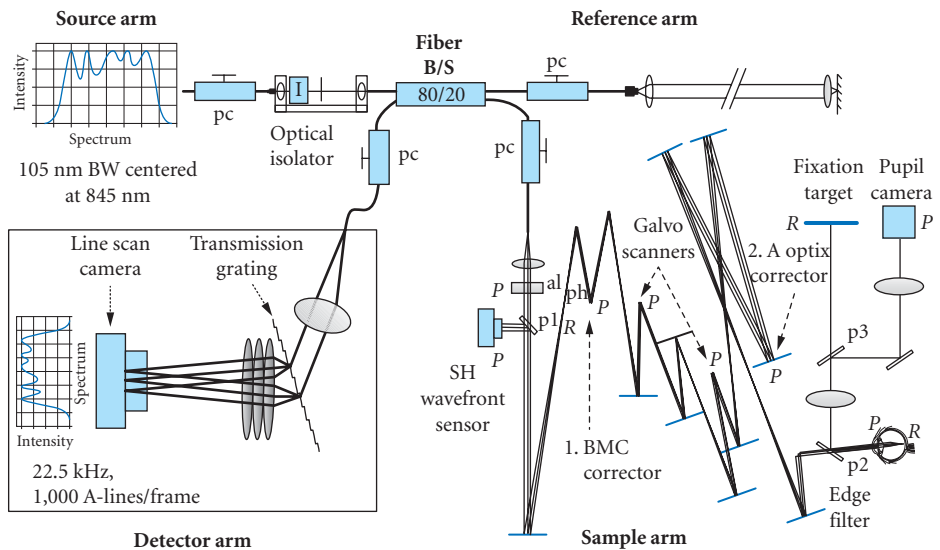
**Adaptive Optics Optical Coherence Tomography** OCT is a noninvasive, interferometric imaging modality that provides significantly higher-axial resolution and sensitivity than the flood-illuminated ophthalmoscope and SLO modalities discussed previously. The term OCT was coined in 1991 in reference to the first optical B-scan ( $xz$  plane) images collected of the in vitro human retina using interferometry.<sup>98</sup> Note that this was preceded by substantial work in the 1980s on a technique called low-coherence interferometry that provided depth-resolved A-scan ( $z$ ) images of the retina using essentially the same interferometric principles.<sup>99,100</sup> Since the early 1990s, OCT technology and knowledge have grown rapidly. This has led to an increasingly large and diverse array of OCT designs that fall into two broad categories: time domain and spectral (or Fourier) domain. These domains refer to the temporal and spectral detection of the OCT signal, respectively. Details of time domain and spectral domain OCT, associated underlying theory, and other applications can be found in Chap. 18.

In an AO-OCT combination, OCT provides the high-axial resolution, and AO the complementary high-transverse resolution. Together, the AO-OCT combination provides a powerful imaging tool whose 3D resolution and sensitivity in the eye substantially surpass that of any current retinal imaging modality. Current state-of-the-art AO-OCT instruments have an isotropic 3D resolution approaching  $3 \times 3 \times 3 \mu\text{m}^3$  in retinal tissue. Most OCT designs have already been combined with AO and have demonstrated both an increase in transverse resolution and sensitivity beyond that realized with OCT alone. AO-OCT combinations include time-domain en face ( $xy$ ) flood-illumination OCT using an areal CCD,<sup>25</sup> time-domain tomographic scanning ( $xz$ ) ultrahigh-resolution OCT,<sup>26</sup> time-domain en face scanning OCT,<sup>31,33</sup> high-resolution spectral-domain OCT,<sup>27,28,30–32</sup> ultrahigh-resolution spectral-domain OCT,<sup>29,34,35,82</sup> and swept

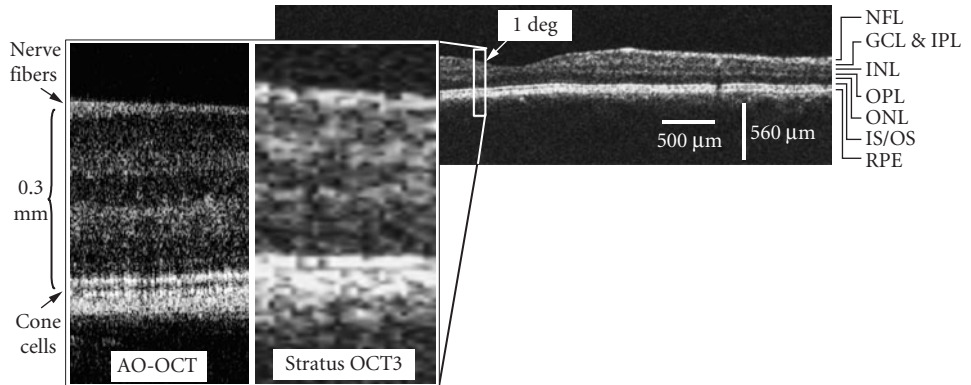
source OCT.<sup>101</sup> AO-OCT instruments have been evaluated with a variety of wavefront correctors that cover the four types depicted in Fig. 6: discrete actuator, membrane, bimorph, and LC-SLM wavefront correctors, all in combination with a SHWS.

As a representative example, Fig. 14 shows a schematic of the Indiana AO-OCT instrument based on ultrahigh-resolution spectral-domain OCT. Use of a single mode fiber beam splitter provides true confocality [see Eq. (11)]. The sample arm is characteristic of an AO-SLO (see Fig. 13 and corresponding text), raster scanning a focused spot across the retina. The slower acquisition of OCT, however, precludes use of a resonant scanner, which is used with the SLO. For the AO OCT shown in the figure, improved aberration correction is realized by cascading two wavefront correctors, a bimorph that corrects low-order, large-magnitude ocular aberrations and a discrete actuator deformable mirror for higher-order, small-magnitude aberrations. As an added benefit, the large stroke of the bimorph reduces the need for trial lens correction of individual refractive error, which is important for OCT because trial lenses produce unequal dispersion in the sample and reference channels. A third advantage of two wavefront correctors is that the focus of the AO correction can be varied within the retina for optimal lateral resolution at a particular retinal layer of interest.

Figure 15 illustrates the optical benefit of AO OCT compared to the widely used OCT clinical instrument, the Stratus OCT3. The inset shows essentially the same cross-sectional patch of retina imaged with both instruments. While the cellular structures are clearly too small for the clinical instrument to resolve, the same structures (in particular the lateral and axial extent of individual cone photoreceptor outer segments) are observed with AO OCT with focus at the photoreceptor layer. An additional benefit of AO—readily apparent in the inset—is the much smaller speckle size due to the larger pupil through which the image is captured, in this case more than four times larger in diameter. The larger pupil also increases substantially the light collection capacity of the instrument.



**FIGURE 14** Schematic of an AO-OCT instrument based on ultrahigh-resolution spectral-domain OCT.<sup>82</sup> The AO system is located in the sample arm and consists of a Shack-Hartmann wavefront sensor, a 140-element Boston Micromachines MEMS deformable mirror, and a 37-element AOptix bimorph mirror. The MEMS corrects high-order, small-magnitude ocular aberrations and complements the low-order, large-magnitude correction of the AOptix DM. The AOptix DM is positioned close to the eye to prevent refractive errors of the eye from distorting the circular beam at the SHWS.<sup>30</sup> Key: al—customized achromatizing lens; B/S—beamsplitter; p1–3—pellicles; pc—polarization controller; ph—pinhole; P/R—pupil and retina conjugate planes. Footprint of the instrument is approximately 2.5 ft × 4 ft.



**FIGURE 15** Cross-sectional images of the retina in the same subject obtained with (a) AO high-resolution spectral-domain OCT and (b) clinical OCT. Images were acquired at  $1^\circ$  retinal eccentricity. Key: NFL—nerve fiber layer; GCL—ganglion cell layer; IPL and OPL—inner and outer plexiform layers; INL and ONL—inner and outer nuclear layers; IS/OS—inner segment/outer segment junction; RPE—retinal pigment epithelium.

For example, a 4 times increase in pupil diameter (e.g., 6 mm compared to 1.5 mm) leads to a nominal 12-dB upper limit increase in sensitivity. The actual increase is lower depending on the scattering properties of the retina and the throughput efficiency of the sample channel, which contains many more elements than conventional OCT.

Current AO-OCT systems have been used successfully to create 3D retinal images of structures previously only visible with histology, including the foveal microvasculature, bundles within the retinal nerve fiber layer, the fibers of Henle, the 3D photoreceptor mosaic, drusen in age-related macular degeneration, and the tiny pores of the lamina cribrosa of the optical nerve in normal and glaucoma patients. AO OCT is beginning to be employed for studying a number of other clinical conditions, including age-related macular degeneration, hereditary retinal dystrophies, and optic neuropathies.

**Contrast and Resolution** Image contrast is a key requirement for visualizing features in retinal images. There are three elements to achieving an image with sufficient contrast. The first is to transfer as much of the available contrast from the object to the image. This is achieved by making the modulation transfer function (MTF) as high as possible and where the role of AO is crucial. Additional increases in the MTF are obtained by judicious selection of the imaging modality. AO SLO, for example, uses confocal imaging to increase the lateral and axial MTF.<sup>96,102</sup> OCT uses coherence-gated detection to increase the MTF in the axial direction. The second key is to make the most of the signal that reaches the image plane. Proper selection of detectors and detection modalities can be used to achieve high signal-to-noise ratios with low levels of light. Flood-illuminated cameras use cooled scientific grade CCD cameras with low readout noise, AO SLOs use sensitive photomultiplier tubes and avalanche photodiodes with carefully designed amplification electronics,<sup>103</sup> and OCT uses heterodyning techniques to amplify weak signals. The third element is to have contrast in the object itself. Not only do retinal features need to scatter light, but they need to scatter light in such a way as to reveal their intrinsic structure. Cone photoreceptors, for example, have very high intrinsic contrast and lend themselves very well to imaging. The fiber optic property of cones<sup>86</sup> makes them appear as an array of regularly spaced point sources of nearly 100 percent contrast. But most other features in the retina scatter very little light and have less intrinsic contrast. The index of refraction of a ganglion cell body, for example, is well matched to its surrounding media and therefore scatters very little light. Under these situations, manipulation of the imaging wavelength and/or polarization might be used to improve contrast. Blood vessels, for example, are best viewed with shorter wavelengths<sup>104</sup> and the contrast of the photoreceptors and other retinal features can be enhanced by careful control of the incident and detected polarization.<sup>105</sup>



A common method to achieve contrast in many imaging systems is to use fluorescence, both intrinsic and extrinsic. Recently, an AO SLO at the University of Rochester has been equipped with a fluorescence channel where they imaged the autofluorescent lipofuscin to reveal the complete and contiguous array of retinal pigment epithelium cells in human and monkeys. By using extrinsic contrast agents in a monkey, they could image retinal capillaries (fluorescein angiography), and ganglion cell bodies and their dendrites.<sup>21</sup>

## Applications Involving AO-controlled Light Delivery to the Retina

The pioneers of AO for the human eye appreciated that the benefits of AO work in both directions.<sup>13</sup> Human vision has always been limited by diffraction for small pupils and aberrations for large pupils. Consequently, the optimal pupil size for sharpest vision in a typical human eye often lies somewhere between 2 and 4 mm,<sup>12,106–110</sup> depending on image quality criteria and the specific individual. Even for the optimal pupil size, the image quality remains very poor by comparison with its diffraction limited condition and is far from what could be achieved if the aberrations of the largest anatomical pupil size were corrected.

Before AO, the only method to generate high frequency, high-contrast images on the retina was using interferometry. Two mutually coherent point sources, laterally displaced in the pupil would produce sinusoidal interference fringes of 100 percent contrast at spatial frequencies governed by the separation of the two points.<sup>111–113</sup> These techniques have been used effectively to understand some aspects of the limits of human vision and can produce higher-contrast images than any AO system. But, aside from a few modifications,<sup>114</sup> the stimuli were limited largely to sinusoidal gratings. The contrast of images projected through the AO system is still limited by the MTF of large pupils, but can generate complex images with more contrast and resolution than ever before. Moreover, an AO system need not be used only to correct the aberration, but can be used to manipulate them as well.

**Challenges for AO Systems That Project Light into the Eye** All AO systems face the challenge of how to ensure that the AO is working properly. In astronomy, the performance of the AO system is gauged by the quality of the images of distant stars, which are effective point sources. In the retinal image, where no point sources exist, two metrics are typically used to gauge how well an AO system is working, the residual wavefront error, and the relative improvements of the retinal image. Neither metric provides absolute, objective measures of system performance. In postprocessing, multiframe blind deconvolution can produce reasonable estimates of the residual point spread function,<sup>115</sup> but these cannot be used as feedback in closed-loop operation. When projecting light into the eye, it is even more difficult because feedback on image quality is not available in most cases, and we can only rely on residual wavefront error and the subject's perception or visual performance.

The challenge is made worse because the stimulus is viewed along a different optical path, it extends over a finite field angle, and often employs wavelengths or a span of wavelength that are different than the wavefront sensing beacon. The main challenges that one faces in developing an AO system for vision science are:

**Longitudinal chromatic aberration (LCA)** The eye has a lot of chromatic aberration. Refractive power from 400 to 900 nm will change by about 2.6 D<sup>116–118</sup> and the foci span an axial focus range in the eye approaching 1 mm. By comparison, the thickness of the neural retina is about 250  $\mu\text{m}$ . This means that if one measures and corrects aberrations with IR light and tests vision with visible light, then they must adjust for the focus of the visual target. Fortunately, the LCA of the eye is fairly consistent between individuals so one can estimate where the correct focus must be.<sup>116,119</sup> But there is no element of the AO closed loop that will be able to provide feedback on the fine tuning of the visible wavelength. This forces the experimenter to use the subject's perceptual responses as feedback for alignment of the system.

**Transverse chromatic aberration (TCA)** Images of different wavelengths projected onto the retina will be displaced laterally because of TCA. Consider, for example, a collimated white light

beam entering the eye. Relative image shifts between both ends of the white light spectrum can be significant, depending on where the beam enters the eye. TCA arises because the chief ray of the beam does not correspond with the achromatic axis of the eye. Image displacements of  $\pm 3$  arcmin can occur with pupil shifts of  $\pm 1$  mm.<sup>120–121</sup> So, whenever a broadband visual stimulus is used off the achromatic axis of the eye, the spread can significantly degrade image quality and defeat the benefits of AO.

*Alignment* A final challenge in achieving the best AO performance for stimuli delivered to the retina is the alignment of the stimulus itself. The stimulus channel may be subject to noncommon path errors (discussed earlier) and therefore will present new aberrations that the AO system cannot detect or correct. Since there is no feedback from this optical path, the experimenter must be obsessive about its optical design and alignment. Perceived sharpness of the image is not sufficient to judge image quality as the human visual system is limited by its retina to detection of spatial frequencies at 60 cycles/degree at best, whereas the optics can transmit frequencies over 200 cycles/degree.<sup>113</sup> Moreover, perceived sharpness depends on context and on whether you are preadapted to a blurred or sharp stimulus.<sup>122</sup>

*Conventional AO Vision Systems* The most basic implementation of an AO system for vision testing is one in which a subject is given a view of a target through a deformable mirror that is conjugate to the optics of the eye. The optics are no different than for a AO flood-illuminated ophthalmoscope (Fig. 10), except that there is no illumination source and the retina camera is replaced by the stimulus. The first experiments were designed simply to test the extent to which the correction of aberration improved vision.<sup>13,123</sup> After correction of aberration and removal of chromatic aberration (by using monochromatic light), improvements of 20 times were reported for contrast sensitivity of a 32 cycles/degree gratings through a 6-mm pupil.<sup>123</sup> Of course, the largest improvements were realized for the largest pupils, for which diffraction is low and aberrations are high. As for visual acuity, improvements were significant, but absolute performance was limited because of the low luminances available in the display. The latter result begs the question about the practical visual benefit of correcting all aberrations. Large pupils, for which the visual benefits are greatest, are only reached under dim light conditions—the same conditions for which the retina starts to limit vision.<sup>124–126</sup> In a practical task (detection of the contrast threshold for a 20/60 letter) the visual benefit for all luminances after taking typical pupil sizes into account was improved with AO, but the improvements were relatively modest.<sup>127</sup>

*AO-Controlled Stimulus Delivery in an AO SLO* An alternate way to test vision with AO-corrected stimuli is to project the image directly onto the retina. By modulating the laser power during the raster scan, complex AO-controlled stimuli can be projected directly on the retina and be seen by the subject. Moreover, the same modulation patterns will appear in the retinal image.<sup>128</sup> This method for delivering stimuli to the retina offers some important advantages. First, the scanning beam and the stimulus can employ the same light source and are therefore not affected by TCA or LCA. Second, the recorded image can be used as a metric to determine the quality of the projected image as well as where the stimulus was presented, both axially and laterally. In some cases, two lasers might be used, one primarily for imaging (e.g., an infrared channel which is used to guide the location of the stimulus) and a second for delivery of the stimulus (e.g., small spot stimulus of a visible wavelength). In this situation, LCA and TCA do need to be taken into account so the operator knows exactly where the visible stimulus is landing.<sup>23</sup> Applications that employ stimulus delivery in an SLO have been used on numerous occasions,<sup>94,129,130</sup> but the application of the technique with AO SLO is relatively new and has been implemented on only a few systems.

### *What Will Vision AO Systems Be Used for?*

*The role of aberrations and visual performance* How much can vision improve? Experiments with AO SLO have reported visual acuities better than 20/8 (letter size for 80% correct judgments of the orientation of a tumbling E).<sup>128,131</sup> Ultimately, acuity is limited by other factors in the visual system,



such as cone photoreceptor spacing, eye movements, and cortical processing limits. By bypassing the optics and simultaneously providing unprecedented views of retina structure as well as eye movements, AO SLO and other modalities provide the best tools to help us understand these limits. AO SLO studies of the role of eye movements on visual acuity are currently underway.

*Measuring activity of individual cones* The eye was never designed to allow stimulation of single cones. Nevertheless, recording sensations from single cones can help reveal how cones are connected to neurons further downstream in the human visual system. There are several applications where stimulation on the scale of single cones is useful. Hofer et al, for example, used an AO system to deliver brief flashes of light to individual cones, and recorded subject's perceived color sensation on each trial.<sup>132</sup> A simple hypothesis, termed the "elemental sensation hypothesis" was that color percepts would fall into three categories, red, green, and blue, corresponding to the three cone types in the eye. Instead, viewers required up to eight different color names, requiring a rethinking of how color sensations are developed.<sup>133</sup> From a clinical perspective, recording the sensitivity of individual cones can provide important relationships between retinal structure and function. In an elegant set of experiments, Makous et al. recorded light sensations in brief flashes to confirm that individuals with a rare genetic mutation leading to apparent loss in a subset of photoreceptors had small, cone-sized gaps across their visual field.<sup>134</sup> The most recent application involves the direct and targeting stimulation of individual cones in a monkey while simultaneously recording electrical activity in downstream neurons.<sup>135</sup> Although neuroscientists have been recording activity of single cells in the visual system for decades, AO allows, for the first time, stimulation of single cones as the input and promises to be a useful tool for mapping connections between the retina and brain.

*Using AO to generate aberrations* There are many cases where the presence of aberrations may be beneficial to human vision. Considering that most people are not really handicapped by their ocular aberrations, it is sensible to consider how one could work with tolerable levels of aberration to increase visual performance in some other way. The most promising application for controlling aberrations is for the relief of presbyopia, which refers to the age-related loss in ability to deform ones natural lens to focus on near objects. Although aberrations cannot restore accommodation, they can increase the eyes depth of focus and relieve these effects. AO vision systems are currently being used to test the benefits of specific aberration profiles, which can later be implemented into contact lenses or intraocular lenses.<sup>136–138</sup>

## 15.6 ACKNOWLEDGMENTS

The authors thank Nathan Doble of IrisAO, Inc. for assistance with the Fig. 1 data. Financial support was provided by the National Eye Institute grants 1R01 EY018339 and 5R01 EY014743 to Donald T. Miller and EY014375 to Austin Roorda. The authors are also supported in part by the National Science Foundation Science and Technology Center for Adaptive Optics, managed by the University of California at Santa Cruz under cooperative agreement No. AST-9876783.

## 15.7 REFERENCES

1. For reviews, see W. N. Charman, "Optics of the Eye," *Handbook of Optics*, (eds.), McGraw-Hill, New York, 2008. See also P. Artal, J. M. Bueno, A. Guirao, and P. M. Prieto, "Aberration Structure of the Eye," *Adaptive Optics for Vision Science: Principles, Practices, Design and Applications*, J. Porter, H. Queener, J. Lin, K. Thorn, and A. Awwal (eds.), Wiley, New York, 2006. See also R. R. Krueger, R. A. Applegate, and S.M. MacRae, *Wavefront Customized Visual Correction: The Quest for Super Vision II*, Slack, Inc., Thorofare, New Jersey, 2004.
2. H. von Helmholtz, *Popular Scientific Lectures*, M. Kline (ed.), Dover Publications, New York, 1962.
3. M. S. Smirnov, "Measurement of the Wave Aberration of the Human Eye," *Biophysics* 7:766–795 (1962).
4. R. K. Tyson, *Principles of Adaptive Optics*, 2nd ed., Academic, 1998.

5. H. W. Babcock, "The Possibility of Compensating Astronomical Seeing," *Pub. of the Astronomical Soc. of the Pacific* **65**:229–236 (1953).
6. L. A. Thompson, "Adaptive Optics in Astronomy," *Phys. Today* 24–31 (December, 1994).
7. F. W. Campbell and D. G. Green, "Optical and Retinal Factors Affecting Visual Resolution," *J. Physiol. (London)* **181**:576–593 (1965).
8. P. Artal and R. Navarro, "High-Resolution Imaging of the Living Human Fovea: Measurement of the Intercenter Cone Distance by Speckle Interferometry," *Opt. Lett.* **14**:1098–1100 (1989).
9. D. T. Miller, D. R. Williams, G. M. Morris, and J. Liang, "Images of the Cone Mosaic in the Living Human Eye," *Vis. Res.* **36**:1067–1079 (1996).
10. A. W. Dreher, J. F. Bille, and R. N. Weinreb, "Active Optical Depth Resolution Improvement of the Laser Tomographic Scanner," *Appl. Opt.* **24**:804–808 (1989).
11. J. Liang, B. Grimm, S. Goelz, and J. F. Bille, "Objective Measurement of the Wave Aberrations of the Human Eye Using a Shack-Hartmann Wavefront Sensor," *J. Opt. Soc. Am. A* **11**:1949–1957 (1994).
12. J. Liang and D. R. Williams, "Aberrations and Retinal Image Quality of the Normal Human Eye," *J. Opt. Soc. Am. A* **14**, 2873–2883 (1997).
13. J. Liang, D. R. Williams, and D. T. Miller, "Supernormal Vision and High Resolution Retinal Imaging through Adaptive Optics," *J. Opt. Soc. Am. A* **14**:2884–2892 (1997).
14. D. R. Williams and C. Max, Figure F.1 in "Foreword," *Adaptive Optics for Vision Science: Principles, Practices, Design and Applications*, J. Porter, H. Queener, J. Lin, K. Thorn, and A. Awwal (eds.), Wiley, New York, xviii (2006).
15. V. Larichev, P. V. Ivanov, N. G. Iroshnikov, V. I. Shmalhauzen, and L. J. Otten, "Adaptive System for Eye-Fundus Imaging," *Quantum Electron.* **32**:902–908 (2002).
16. N. Ling, Y. Zhang, X. Rao, X. Li, C. Wang, Y. Hu, and W. Jiang, "Small Table-Top Adaptive Optical Systems for Human Retinal Imaging," in *High-Resolution Wavefront Control: Methods, Devices, and Applications IV*, J. D. Gonglewski, M. A. Vorontsov, M. T. Gruneisen, S. R. Restaino, and R. K. Tyson (eds.), *Proc. SPIE* **4825**: 99–108 (2002).
17. M. Glanc, E. Gendron, F. Lacombe, D. Lafaille, J. F. Le Gargasson, and P. Lena, "Towards Wide-Field Imaging with Adaptive Optics," *Opt. Commun.* **230**:225–238 (2004).
18. J. Rha, R. S. Jonnal, K. E. Thorn, J. Qu, Y. Zhang, and D. T. Miller, "Adaptive Optics Flood-Illumination Camera for High Speed Retinal Imaging," *Opt. Express* **14**:4552–4569 (2006).
19. A. Roorda, F. Romero-Borja, W. J. Donnelly, H. Queener, T. J. Hebert, and M. C. W. Campbell, "Adaptive Optics Scanning Laser Ophthalmoscopy," *Opt. Express* **10**:405–412 (2002).
20. Y. Zhang, S. Poonja, and A. Roorda, "MEMS-Based Adaptive Optics Scanning Laser Ophthalmoscopy," *Opt. Lett.* **31**:1268–1270 (2006).
21. D. C. Gray, W. Merigan, J. I. Wolfing, B. P. Gee, J. Porter, A. Dubra, T. H. Twietmeyer et al., "In Vivo Fluorescence Imaging of Primate Retinal Ganglion Cells and Retinal Pigment Epithelial Cells," *Opt. Express* **14**:7144–7158 (2006).
22. D. X. Hammer, R. D. Ferguson, C. E. Bigelow, N. V. Iftimia, T. E. Ustun, and S. A. Burns, "Adaptive Optics Scanning Laser Ophthalmoscope for Stabilized Retinal Imaging," *Opt. Express* **14**:3354–3367 (2006).
23. K. Grieve, P. Tiruveedhula, Y. Zhang, and A. Roorda, "Multi-Wavelength Imaging with the Adaptive Optics Scanning Laser Ophthalmoscope," *Opt. Express* **14**:12230–12242 (2006).
24. S. A. Burns, R. Tumber, A. E. Elsner, R. D. Ferguson, and D. X. Hammer, "Large-Field-of-View, Modular, Stabilized, Adaptive-Optics-Based Scanning Laser Ophthalmoscope," *J. Opt. Soc. Am. A* **24**:1313–1326 (2007).
25. D. T. Miller, J. Qu, R. S. Jonnal, and K. Thorn, "Coherence Gating and Adaptive Optics in the Eye," in *Coherence Domain Optical Methods and Optical Coherence Tomography in Biomedicine VII*, V. Valery, V. Tuchin, J. A. Izatt, J. G. Fujimoto (eds.), *Proc. SPIE* **4956**:65–72 (2003).
26. B. Hermann, E. J. Fernández, A. Unterhuber, H. Sattmann, A. F. Fercher, W. Drexler, P. M. Prieto, and P. Artal, "Adaptive-Optics Ultrahigh-Resolution Optical Coherence Tomography," *Opt. Lett.* **29**:2142–2144 (2004).
27. Y. Zhang, J. Rha, R. S. Jonnal, and D. T. Miller, "Adaptive Optics Parallel Spectral Domain Optical Coherence Tomography for Imaging the Living Retina," *Opt. Express* **13**:4792–4811 (2005).
28. R. J. Zawadzki, S. Jones, S. S. Olivier, M. Zhao, B. A. Bower, J. A. Izatt, S. S. Choi, S. Laut, and J. S. Werner, "Adaptive-Optics Optical Coherence Tomography for High-Resolution and High-Speed 3D Retinal In Vivo Imaging," *Opt. Express* **13**:8532–8546 (2005).

29. E. J. Fernández, B. Považay, B. Hermann, A. Unterhuber, H. Sattmann, P. M. Prieto, R. Leitgeb, P. Ahnelt, P. Artal, and W. Drexler, "Three-Dimensional Adaptive Optics Ultrahigh-Resolution Optical Coherence Tomography Using a Liquid Crystal Spatial Light Modulator," *Vis. Res.* **45**:3432–3444 (2005).
30. Y. Zhang, B. Cense, J. Rha, R. S. Jonnal, W. Gao, R. J. Zawadzki, J. S. Werner, S. Jones, S. Olivier, and D. T. Miller, "High-Speed Volumetric Imaging of Cone Photoreceptors with Adaptive Optics Spectral Domain Optical Coherence Tomography," *Opt. Exp.* **14**:4380–4394 (2006).
31. D. Merino, C. Dainty, A. Bradu, and A. G. Podoleanu, "Adaptive Optics Enhanced Simultaneous En-Face Optical Coherence Tomography and Scanning Laser Ophthalmoscopy," *Opt. Exp.* **14**:3345–3353 (2006).
32. C. E. Bigelow, N. V. Iftimia, R. D. Ferguson, T. E. Ustun, B. Bloom, and D. X. Hammer, "Compact Multimodal Adaptive-Optics Spectral-Domain Optical Coherence Tomography Instrument for Retinal Imaging," *J. Opt. Soc. Am. A* **24**:1327–1336 (2007).
33. M. Pircher, R. J. Zawadzki, J. W. Evans, J. S. Werner, and C. K. Hitzenberger, "Simultaneous Imaging of Human Cone Mosaic with Adaptive Optics Enhanced Scanning Laser Ophthalmoscopy and High-Speed Transversal Scanning Optical Coherence Tomography," *Opt. Lett.* **33**:22–24 (2008).
34. R. J. Zawadzki, B. Cense, Y. Zhang, S. S. Choi, D. T. Miller, and J. S. Werner, "Ultrahigh-Resolution Optical Coherence Tomography with Monochromatic and Chromatic Aberration Correction," *Opt. Express* **16**:8126–8143 (2008).
35. E. J. Fernández, B. Hermann, B. Považay, A. Unterhuber, H. Sattmann, B. Hofer, P. K. Ahnelt, and W. Drexler, "Ultrahigh Resolution Optical Coherence Tomography and Pancorrection for Cellular Imaging of the Living Human Retina," *Opt. Express* **16**:11083–11094 (2008).
36. D. R. Williams, J. Liang, D. T. Miller, and A. Roorda, "Wavefront Sensing and Compensation for the Human Eye," *Adaptive Optics Engineering Handbook*, R. K. Tyson (ed.), Marcel Dekker, New York (2000).
37. A. Roorda and D. R. Williams, "Retinal Imaging Using Adaptive Optics," in *Wavefront Customized Visual Correction: The Quest for Supervision II*, R. R. Krueger, R. A. Applegate, and S. M. MacRae (eds.), Slack, Incorporated, Thorofare, 2004, pp. 43–51.
38. J. Carroll, D. C. Gray, A. Roorda, and D. R. Williams, "Recent Advances in Retinal Imaging with Adaptive Optics," *Optics & Photonics News* 36–42 (January 2005).
39. N. Doble, "High-Resolution, In Vivo Retinal Imaging Using Adaptive Optics and its Future Role in Ophthalmology," *Expert Rev. Med. Devices*, **2**:205–216 (2005).
40. A. Roorda, K. Venkateswaran, F. Romero-Borja, D. R. Williams, J. Carroll, and H. Hofer, "Adaptive Optics Ophthalmoscopy," in *Atlas of Posterior Segment Imaging*, D. Huang, P. K. Kaiser, C. Y. Lowder, and E. Traboulsi (eds.), Elsevier Science, 2005.
41. J. Porter, H. Queener, J. Lin, K. Thorn, and A. Awwal (eds.), *Adaptive Optics for Vision Science: Principles, Practices, Design and Applications*, Wiley, New York (2006).
42. D. T. Miller, "Adaptive Optics in Retinal Microscopy and Vision," *OSA Handbook of Optics Vol. III.*, M. Bass, J. M. Enoch, E. W. Van Stryland, and W. L. Wolfe (eds.), McGraw-Hill, New York, (2000).
43. J. Porter, A. Guirao, I. G. Cox, and D. R. Williams, "Monochromatic Aberrations of the Human Eye in a Large Population," *J. Opt. Soc. Am. A* **18**:1793–1803 (2001).
44. L. N. Thibos, X. Hong, A. Bradley, and X. Cheng, "Statistical Variation of Aberration Structure and Image Quality in a Normal Population of Healthy Eyes," *J. Opt. Soc. Am. A* **19**:2329–2348 (2002).
45. M. C. Roggemann and B. Welch, *Imaging through Turbulence*, M. J. Weber (ed.), CRC Press, Boca Raton, 1996.
46. L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, R. Webb, and VISA Standards Taskforce Members, "Standards for Reporting the Optical Aberrations of Eyes," *J. Refract. Surg.* **18**:S652–S660 (2002).
47. L. Wang and D. D. Koch, "Ocular Higher-Order Aberrations in Individuals Screened for Refractive Surgery," *J. Cataract Refract. Surg.* **29**:1896–1903 (2003).
48. T. O. Salmon and C. van de Pol, "Normal-Eye Zernike Coefficients and Root-Mean-Square Wavefront Errors," *J. Cataract Refract. Surg.* **32**:2064–2074 (2006).
49. M. J. Collins, C. F. Wildsoet, and D. A. Atchison, "Monochromatic Aberrations and Myopia," *Vis. Res.* **35**:1157–1163 (1995).
50. X. Hong and L. Thibos, "Longitudinal Evaluation of Optical Aberrations Following Laser in Situ Keratomileusis Surgery," *J. Refract. Surg.* **16**:S647–S650 (2001).

51. X. Cheng, A. Bradley, X. Hong, and L. Thibos, "Relationship between Refractive Error and Monochromatic Aberrations of the Eye," *Optom. Vis. Sci.* **80**:43–49 (2003).
52. S. Amano, Y. Amano, S. Yamagami, T. Miyai, K. Miyata, T. Samejima, and T. Oshika, "Age-Related Changes in Corneal and Ocular Higher-Order Wavefront Aberrations," *Am. J. Ophthalmol.* **137**:988–992 (2004).
53. J. McLellan, S. Marcos, and S. A. Burns, "Age-Related Changes in Monochromatic Wave Aberrations of the Human Eye," *IOVS* **42**:1390–1395 (2001).
54. H. Cheng, J. K. Barnett, A. S. Vilupuru, J. D. Marsack, S. Kasthurirangan, R. A. Applegate, and A. Roorda, "A Population Study on Changes in Wave Aberrations with Accommodation," *J. Vis.* **4**:272–280 (2004).
55. J. Porter, G. Yoon, S. MacRae, I. Cox, and D.R. Williams, "Aberrations Induced in Customized Laser Refractive Surgery due to Shifts between Natural and Dilated Pupil Center Locations," *J. Cataract Refract. Surg.* **32**:21–32 (2006).
56. H. Radhakrishnan and W. N. Charman, "Age-Related Changes in Ocular Aberrations with Accommodation," *J. Vis.* **7**:11–21 (2007).
57. S. Pantanelli, S. MacRae, T. M. Jeong, and G. Yoon, "Characterizing the Wave Aberration in Eyes with Keratoconus or Penetrating Keratoplasty Using a High Dynamic Range Wavefront Sensor," *Ophthalmology* **114**:2013–2021 (2007).
58. H. Hofer, P. Artal, B. Singer, J. L. Aragón, and D. R. Williams, "Dynamics of the Eye's Wave Aberration," *J. Opt. Soc. Am. A* **18**:497–506 (2001).
59. L. Diaz-Santana, C. Torti, I. Munro, P. Gasson, and C. Dainty, "Benefit of Higher Closed-Loop Bandwidths in Ocular Adaptive Optics," *Opt. Express* **11**:2597–2605 (2003).
60. D. L. Fried, "Anisoplanatism in Adaptive Optics," *J. Opt. Soc. Am.* **72**:52–61 (1982).
61. P. Bedggood, M. Daaboul, R. Ashman, G. Smith, and A. Metha, "Characteristics of the Human Isoplanatic Patch and Implications for Adaptive Optics Retinal Imaging," *J. Biomed. Opt.* **13**:024008 (2008).
62. H.-L. Liou and N. A. Brennan, "Anatomically Accurate, Finite Model Eye for Optical Modeling," *J. Opt. Soc. Am. A* **14**:1684–1694 (1997).
63. R. Q. Fugate, D. L. Fried, G. A. Ameer, B. R. Boeke, S. L. Browne, P. H. Roberts, R. E. Ruane, G. A. Tyler, and L. M. Wopat, "Measurement of Atmospheric Wavefront Distortion Using Scattered Light from a Laser Guide-Star," *Nature* **353**:144–146 (1991).
64. L. Llorente, S. Marcos, C. Dorronsoro, and S. A. Burns, "Effect of Sampling on Real Ocular Aberration Measurements," *J. Opt. Soc. Am. A* **24**:2783–2796 (2007).
65. G. Yoon, "Wavefront Sensing and Diagnostic Uses," in *Adaptive Optics for Vision Science: Principles, Practices, Design and Applications*, J. Porter, H. Queener, J. Lin, K. Thorn, and A. Awwal (eds.), Wiley, New York, 2006.
66. D. R. Neal, J. Copland, and D. Neal, "Shack-Hartmann Wavefront Sensor Precision and Accuracy," in *Advanced Characterization Techniques for Optical, Semiconductor, and Data Storage Components*, A. Duparré and B. Singh (eds.), Proc. SPIE **4779**:148–160 (2002).
67. D. T. Miller, L. N. Thibos, and X. Hong, "Requirements for Segmented Correctors for Diffraction-Limited Performance in the Human Eye," *Opt. Express* **13**:275–289 (2005).
68. N. Doble, D. T. Miller, G. Yoon, and D. R. Williams, "Requirements for Discrete Actuator and Segmented Wavefront Correctors for Aberration Compensation in Two Large Populations of Human Eyes," *Appl. Opt.* **46**:4501–4514 (2007).
69. G. T. Kennedy and C. Paterson, "Correcting the Ocular Aberrations of a Healthy Adult Population Using Microelectromechanical (MEMS) Deformable Mirrors," *Opt. Commun.* **271**:278–284 (2007).
70. T. Farrell, E. Daly, E. Dalimier, and C. Dainty, "Task-Based Assessment of Deformable Mirrors," *Proc. SPIE* **6467**:64670F (2007).
71. B. R. Oppenheimer, D. L. Palmer, R. G. Dekany, A. Sivaramakrishnan, M. A. Ealey, and T. R. Price, "Investigating a Xinetics Inc. Deformable Mirror," *Proc. SPIE* **3126**:569–579 (1997).
72. D. A. Horsley, H. Park, S. P. Laut, J. S. Werner, "Characterization of a Bimorph Deformable Mirror Using Stroboscopic Phase-Shifting Interferometry," *Sens. Actuators A* **134**:221–230 (2006).
73. E. J. Fernández, L. Vabre, B. Hermann, A. Unterhuber, B. Povazay, and W. Drexler, "Adaptive Optics with a Magnetic Deformable Mirror: Applications in the Human Eye," *Opt. Express* **14**:8900–8917 (2006).
74. E. J. Fernández and P. Artal, "Membrane Deformable Mirror for Adaptive Optics: Performance Limits in Visual Optics," *Opt. Exp.* **11**:1056–1069 (2003).

75. E. J. Fernández, I. Iglesias, and P. Artal, "Closed-Loop Adaptive Optics in the Human Eye," *Opt. Lett.* **26**:746–748 (2001).
76. N. Doble, G. Yoon, L. Chen, P. Bierden, B. Singer, S. Olivier, and D. R. Williams, "The Use of a Microelectromechanical Mirror for Adaptive Optics in the Human Eye," *Opt. Lett.* **27**:1579–1581 (2002).
77. L. N. Thibos and A. Bradley, "Use of Liquid-Crystal Adaptive Optics to Alter the Refractive State of the Eye," *Optom. Vis. Sci.* **74**:581–587 (1997).
78. F. Vargas-Martín, P. M. Prieto, and P. Artal, "Correction of the Aberrations in the Human Eye with a Liquid-Crystal Spatial Light Modulator: Limits to Performance," *J. Opt. Soc. Am. A* **15**:2552–2562 (1998).
79. P. M. Prieto, E. J. Fernández, S. Manzanera, and P. Artal, "Adaptive Optics with a Programmable Phase Modulator: Applications in the Human Eye," *Opt. Exp.* **12**:4059–4071 (2004).
80. R. H. Webb, M. J. Albanese, Y. Zhou, T. Bifano, and S. A. Burns, "Stroke Amplifier for Deformable Mirrors," *Appl. Opt.* **43**:5330–5333 (2004).
81. D. C. Chen, S. M. Jones, D. A. Silva, and S. S. Olivier, "High-Resolution Adaptive Optics Scanning Laser Ophthalmoscope with Dual Deformable Mirrors," *J. Opt. Soc. Am. A* **24**:1305–1312 (2007).
82. B. Cense, E. Koperda, J. M. Brown, O. P. Kocaoglu, W. Gao, R. S. Jonnal, and D. T. Miller, "Volumetric Retinal Imaging with Ultrahigh-Resolution Spectral-Domain Optical Coherence Tomography and Adaptive Optics: Comparison of Two Broadband Light Sources" *Opt. Exp.* **17**:4095–4111 (2009).
83. L. Chen, "Control Algorithms," in *Adaptive Optics for Vision Science: Principles, Practices, Design and Applications*, J. Porter, H. Queener, J. Lin, K. Thorn, and A. Awwal (eds.), Wiley, New York, 2006.
84. A. Roorda and D. R. Williams, "The Arrangement of the Three Cone Classes in the Living Human Eye," *Nature* **397**:520–522 (1999).
85. H. Hofer, J. Carroll, J. Neitz, M. Neitz, and D. R. Williams, "Organization of the Human Trichromatic Cone Mosaic," *J. Neurosci.* **25**:9669–9679 (2005).
86. A. Roorda and D. R. Williams, "Optical Fiber Properties of Individual Human Cones," *J. Vis.* **2**:404–412 (2002).
87. N. M. Putnam, H. Hofer, N. Doble, L. Chen, J. Carroll, and D. R. Williams, "The Locus of Fixation and the Foveal Cone Mosaic," *J. Vis.* **5**:632–639 (2005).
88. A. Pallikaris, D. R. Williams, and H. Hofer, "The Reflectance of Single Cones in the Living Human Eye," *Invest Ophthalmol. Vis. Sci.* **44**:4580–4592 (2003).
89. J. I. Wolfing, M. Chung, J. Carroll, A. Roorda, and D. R. Williams, "High-Resolution Retinal Imaging of Cone-Rod Dystrophy," *Ophthalmology* **113**:1014–1019 (2006).
90. J. Carroll, M. Neitz, H. Hofer, J. Neitz, and D. R. Williams, "Functional Photoreceptor Loss Revealed with Adaptive Optics: An Alternate Cause of Color Blindness," *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8461–8466 (2004).
91. S. S. Choi, N. Doble, J. L. Hardy, S. M. Jones, J. L. Keltner, S. S. Olivier, and J. S. Werner, "In Vivo Imaging of the Photoreceptor Mosaic in Retinal Dystrophies and Correlations with Visual Function," *Invest Ophthalmol. Vis. Sci.* **47**:2080–2092 (2006).
92. K. Grieve and A. Roorda, "Intrinsic Signals from Human Cone Photoreceptors," *Invest Ophthalmol. Vis. Sci.* **49**:713–719 (2008).
93. R. S. Jonnal, J. Rha, Y. Zhang, B. Cense, H. Gao, and D. T. Miller, "In Vivo Functional Imaging of Human Cone Photoreceptors," *Opt. Exp.* **15**:16141–16160 (2007).
94. R. H. Webb, G. W. Hughes, and O. Pomerantz, "Flying Spot TV Ophthalmoscope," *Appl. Opt.* **19**:2991–2997 (1980).
95. M. Minsky, "Memoir on Inventing the Confocal Scanning Laser Microscope," *Scanning* **10**:128–138 (1988).
96. F. Romero-Borja, K. Venkateswaran, A. Roorda, and T. J. Hebert, "Optical Slicing of Human Retinal Tissue In Vivo with the Adaptive Optics Scanning Laser Ophthalmoscope," *Appl. Opt.* **44**:4032–4040 (2005).
97. W. J. Donnelly, "Improving Imaging in the Confocal Scanning Laser Ophthalmoscope," M.S. dissertation, (University of Houston, Houston, TX, 2001).
98. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, et al., "Optical Coherence Tomography," *Science* **254**:1178–1181 (1991).
99. A. F. Fercher and E. Roth, "Ophthalmic Laser Interferometry," in *Optical Instrumentation for Biomedical Laser Applications*, G. J. Mueller (ed.), *Proc. SPIE* **658**:48–51 (1986).
100. A. F. Fercher, K. Mengedocht, and W. Werner, "Eye Length Measurement by Interferometry with Partially Coherent Light," *Opt. Lett.* **13**:186–188 (1988).

101. D. T. Miller, B. Cense, Y. Zhang, W. Gao, J. Jiang, and A. Cable, "Retinal Imaging at 850 nm with Swept Source Optical Coherence Tomography and Adaptive Optics," *Invest. Ophthalmol. Vis. Sci.* **48**:E-Abstract 2769 (2007).
102. C. J. R. Sheppard and M. Gu, "The Significance of 3-D Transfer Functions in Confocal Scanning Microscopy," *J. Microsc.* **165**:377–390 (1991).
103. Y. Zhang and A. Roorda, "Photon Signal Detection and Evaluation in the Adaptive Optics Scanning Laser Ophthalmoscope," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **24**:1276–1283 (2007).
104. J. A. Martin and A. Roorda, "Pulsatility of Parafoveal Capillary Leukocytes," *Exp. Eye Res.* **8**(3):356–360 (2008).
105. H. Song, Y. Zhao, X. Qi, Y. T. Chui, and S. A. Burns, "Stokes Vector Analysis of Adaptive Optics Images of the Retina," *Opt. Lett.* **33**:137–139 (2008).
106. F. W. Campbell and D. G. Green, "Optical and Retinal Factors Affecting Visual Resolution," *J. Physiol.* **181**:576–593 (1965).
107. F. W. Campbell and R. W. Gubisch, "Optical Quality of the Human Eye," *J. Physiol.* **186**:558–578 (1966).
108. H. C. Howland and B. Howland, "A Subjective Method for the Measurement of Monochromatic Aberrations of the Eye," *J. Opt. Soc. Am.* **67**:1508–1518 (1977).
109. G. Walsh, W. N. Charman, and H. C. Howland, "Objective Technique for the Determination of Monochromatic Aberrations of the Human Eye," *J. Opt. Soc. Am. A* **1**:987–992 (1984).
110. W. J. Donnelly III and A. Roorda, "Optimal Pupil Size in the Human Eye for Axial Resolution," *J. Opt. Soc. Am. A* **20**:2010–2015 (2003).
111. M. A. Arnulf and M. O. Dupuy, "La Transmission des Contrastes par le Système Optique de L'oeil et les Seuils de Contrastes Retinines," *C. R. Acad. Sci. (Paris)* **250**:2757–2759 (1960).
112. G. Westheimer, "Modulation Thresholds for Sinusoidal Light Distributions on the Retina," *J. Physiol.* **152**:67–74 (1960).
113. D. R. Williams, "Aliasing in Human Foveal Vision," *Vis. Res.* **25**:195–205 (1985).
114. J. Liang and G. Westheimer, "Method for Measuring Visual Resolution at the Retinal Level," *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* **10**:1691–1696 (1993).
115. J. C. Christou, A. Roorda, and D. R. Williams, "Deconvolution of Adaptive Optics Retinal Images," *J. Opt. Soc. Am. A* **21**:1393–1401 (2003).
116. A. G. Bennett and R. B. Rabbetts, *Clinical Visual Optics*, 2nd ed. (Butterworths, London, 1989).
117. L. N. Thibos, M. Ye, X. Zhang, and A. Bradley, "The Chromatic Eye: A New Reduced-Eye Model of Ocular Chromatic Aberration in Humans," *Appl. Opt.* **31**:3594–3600 (1992).
118. E. J. Fernández, A. Unterhuber, B. Povazay, B. Hermann, P. Artal, and W. Drexler, "Chromatic Aberration Correction of the Human Eye for Retinal Imaging in the Near Infrared," *Opt. Exp.* **14**:6213–6225 (2006).
119. L. N. Thibos, A. Bradley, D. L. Still, X. X. Zhang, and P. A. Howarth, "Theory and Measurement of Ocular Chromatic Aberration," *Vis. Res.* **30**:33–49 (1990).
120. P. Simonet and M. C. W. Campbell, "The Optical Transverse Chromatic Aberration on the Fovea of the Human Eye," *Vis. Res.* **30**(2):187–206 (1990).
121. L. N. Thibos, "Calculation of the Influence of Lateral Chromatic Aberration on Image Quality Across the Visual Field," *J. Opt. Soc. Am. A* **4**(8):1673–1680 (1987).
122. M. A. Webster, J. S. Werner, and D. Field, "Adaptation and the Phenomenology of Perception," in *Fitting the Mind to the World: Adaptation and Aftereffects in High Level Vision*, C. Clifford and G. Rhodes (eds.), Oxford University Press, 2005, pp. 241–277.
123. G. Y. Yoon and D. R. Williams, "Visual Performance after Correcting the Monochromatic and Chromatic Aberrations of the Eye," *J. Opt. Soc. Am. A* **19**:266–275 (2002).
124. D. H. Kelly, "Spatial Frequency Selectivity in the Retina," *Vis. Res.* **15**:665–672 (1975).
125. A. M. Rohaly and G. Buchsbaum, "Global Spatiochromatic Mechanism Accounting for Luminance Variations in Contrast Sensitivity Functions," *J. Opt. Soc. Am. A* **6**:312–317 (1989).
126. A. Stockman and L. T. Sharpe, "Into the Twilight Zone: The Complexities of Mesopic Vision and Luminous Efficiency," *Ophthalmic Physiol. Opt.* **26**:225–239 (2006).
127. E. Dalimier, C. Dainty, and J. L. Barbur, "Effects of Higher-Order Aberrations on Contrast Acuity as a Function of Light Level," *J. Mod. Opt.* **55**:791–803 (2008).



128. S. Poonja, S. Patel, L. Henry, and A. Roorda, "Dynamic Visual Stimulus Presentation in an Adaptive Optics Scanning Laser Ophthalmoscope," *J. Refract. Surg.* **21**:S575–S580 (2005).
129. G. T. Timberlake, M. A. Mainster, R. H. Webb, G. W. Hughes, and C. L. Trempe, "Retinal Localization of Scotomata by Scanning Laser Ophthalmoscopy," *Invest. Ophthalmol. Vis. Sci.* **22**:91–97 (1982).
130. M. A. Mainster, G. T. Timberlake, R. H. Webb, and G. W. Hughes, "Scanning Laser Ophthalmoscopy. Clinical Applications," *Ophthalmology* **89**:852–857 (1982).
131. E. A. Rossi, P. Weiser, J. Tarrant, and A. Roorda, "Does the Correction of Higher-Order Aberrations Improve Visual Performance in Myopia?," *J. Vis.* **7**(8):14; 1–14 (2007).
132. H. Hofer, B. Singer, and D. R. Williams, "Different Sensations from Cones with the Same Pigment," *J. Vision* **5**:444–454 (2005).
133. D. H. Brainard, D. R. Williams, and H. Hofer, "Trichromatic Reconstruction from the Interleaved Cone Mosaic: Bayesian Model and the Color Appearance of Small Spots," *J. Vis.* **8**(5):15; 1–23 (2008).
134. W. Makous, J. Carroll, J. I. Wolfing, J. Lin, N. Christie, and D. R. Williams, "Retinal Microscotomas Revealed with Adaptive-Optics Microflashes," *Invest. Ophthalmol. Vis. Sci.* **47**:4160–4167 (2006).
135. L. C. Sincich, Y. Zhang, P. Tiruveedhula, J. C. Horton, and A. Roorda, "Resolving Single Cone Inputs to Visual Receptive Fields," *Nature Neurosci.* in press (2009).
136. P. A. Piers, E. J. Fernández, S. Manzanera, S. Norrby, and P. Artal, "Adaptive Optics Simulation of Intraocular Lenses with Modified Spherical Aberration," *Invest. Ophthalmol. Vis. Sci.* **45**:4601–4610 (2004).
137. P. A. Piers, S. Manzanera, P. M. Prieto, N. Gorceix, and P. Artal, "Use of Adaptive Optics to Determine the Optimal Ocular Spherical Aberration," *J. Cataract Refract. Surg.* **33**:1721–1726 (2007).
138. H. Guo, D. A. Atchison, and B. J. Birt, "Changes in Through-Focus Spatial Visual Performance with Adaptive Optics Correction of Monochromatic Aberrations," *Vis. Res.* **48**:1804–1811 (2008).

---

# REFRACTIVE SURGERY, CORRECTION OF VISION, PRK AND LASIK

---

## L. Diaz-Santana

*Department of Optometry and Visual Science  
City University  
London, United Kingdom*

## Harilaos Ginis

*Institute of Vision and Optics  
University of Crete, Greece*

---

## 16.1 GLOSSARY

---

### Definitions

**Acellular.** Containing no cells.

**Accommodation.** The ability of the crystalline lens to change its optical power allowing the eye to bring into focus the images of objects at different distances.

**Collagen.** The main structural protein found in animal connective tissues, yielding gelatin when boiled.

**Endothelium.** The tissue that forms a single layer of cells lining various organs and cavities of the body, especially the blood vessels, heart, and lymphatic vessels. It is formed from the embryonic mesoderm.

**Epithelium.** The thin tissue forming the outer layer of a body's surface.

**Fibroblast.** Type of cell that synthesizes and maintains the extracellular matrix of many animal tissues. Fibroblasts provide a structural framework (stroma) for many tissues, and play a critical role in wound healing. They are the most common cells of connective tissue in animals.

**Hydrophilic.** Having a tendency to mix with, dissolve in, or be wetted by water.

**Keratocyte.** A flattened connective tissue cell lying between the fibrous tissue lamellae which constitute the cornea. The keratocyte has branching processes that communicate with those of other keratocytes.

**Lamella.** A thin layer, membrane, scale, or platelike tissue.

**Nomogram.** A graphical calculating device, a two-dimensional diagram, designed to allow the approximate graphical computation of a function.

**Refraction.** Measurement of the focusing characteristics of an eye or eyes.



## Equations

**Equation (1).** This is the field of a monochromatic scalar propagating wave as

$U(r)$  modulus of the wave  
 $r$  position  
 $\phi(r)$  wavefront  
 $k$  wave number  
 $\pi$  3.14159  
 $\nu$  optical frequency  
 $t$  time

**Equation (2).** This is the definition of the wave number.

$\pi$  3.14159  
 $\nu$  optical frequency  
 $c$  speed of light  
 $\lambda$  optical wavelength

**Equation (3).** This is the wavefront aberration as a polynomial expansion.

$\omega$  wavefront aberration  
 $x$  horizontal coordinate  
 $y$  vertical coordinate  
 $a_n$   $n$ th weighting value or coefficient  
 $P_n$   $n$ th polynomial from an orthogonal set

**Equation (4).** This relates the back focal length of a correcting lens with the vertex distance and the eye's far point.

$k$  distance from the eye to its far point  
 $f'_c$  back focal length of the correcting lens  
 $d$  correcting lens' vertex distance

**Equation (5).** This is the power of a correcting lens placed at a finite distance from the eye.

$F'_c$  optical power of the correcting lens  
 $K$   $1/k$   
 $k$  distance from the eye to its far point  
 $d$  correcting lens' vertex distance

**Equation (6).** This is an empirical equation to describe the dependence of the ablation rate on the fluence.

$A$  ablation rate  
 $F$  fluence  
 $F_{\text{thr}}$  ablation threshold  
 $m$   $0.3 \mu\text{m}/\text{pulse}$

## 16.2 INTRODUCTION

### Optics of the Human Eye

The human eye can be thought of as a photographic camera; an optical element(s) forms a real image of an object onto a photosensitive surface, the retina. The eye, like any other optical system, is formed from several different materials (tissues), each with a different refractive index and transmittance. The total optical power of the eye is about +60 diopters (D) of which +48 D are provided by its frontal part, the cornea. This tissue has refractive index of  $n = 1.376$ . The cornea is in contact with the air, giving it strong optical power.<sup>1-3</sup>

The eye is self-contained in an almost spherical “bag” of opaque white tissue, except on the cornea, called the sclera. The cornea is approximately oval in shape with average dimensions of 12.6-mm horizontal diameter and 11.7-mm vertical diameter. It absorbs most of the UV radiation with a peak at 270 nm, but transmits radiation with wavelengths from 310 to 2500 nm approximately.<sup>1-3</sup>

Immediately behind the cornea is a chamber known as the anterior chamber filled with a watery fluid: the aqueous humour. This fluid replenishes nutrients and takes away metabolic wastes from the posterior cornea, crystalline lens, and anterior vitreous. It also helps to keep intraocular pressure stable. Its refractive index is  $n = 1.336$ .

The iris marks the end of the anterior chamber. The volume behind it is known as the posterior chamber. The iris is the aperture stop of the eye. In a normal young eye the iris can expand or contract the pupil diameter from 2 or 3 mm to 7 or 8 mm. This reduces or increases the amount of light passing through the system and increases or decreases the depth of focus of the eye. The iris can be easily identified as it gives the eye its characteristic colour: blue, green, brown, and so on. In later life the ability of the iris to expand is severely reduced.<sup>2-3</sup>

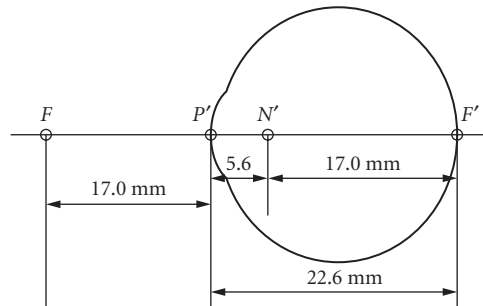
Next to the iris is the crystalline lens of the eye. It has a complex layered structure and is remarkably transparent. It completely lacks a blood supply and innervations. It continues to grow in size and mass throughout life. It has a gradient refractive index which varies from about 1.406 in the inner core to approximately 1.386 at the outer core. It is a flexible structure and can change shape in order to provide a fine focusing mechanism.<sup>2</sup> By combining the cornea and the lens into a single optical system it is possible to treat the system as a thin lens in a medium. The nodal point of such a system lies at 17 mm from the retina and its posterior principal point lies on the cornea, 5.6 mm from the nodal point. A schematic eye representing this system is shown in Fig. 1.

The posterior chamber is filled with the vitreous. With a volume of almost 4 mL, it represents about the 80 percent of the total volume of the eye. It is a transparent collagen-based substance with a refractive index of 1.336. The vitreous plays an important role in several metabolic eye processes like oxygenation, depository for metabolic wastes, such as lactic acid, and as a medium for active transport of different substances throughout the eye. Immediately behind the vitreous lies the retina, the surface where the image is formed. The retina is a very complex multilayered structure, that not only captures the light, but some of the initial visual processes occur there. The central part of the retina, the fovea, contains the highest density of photoreceptors and is the region where the image formed by the optics of the eye must be of highest quality.<sup>2,3</sup>

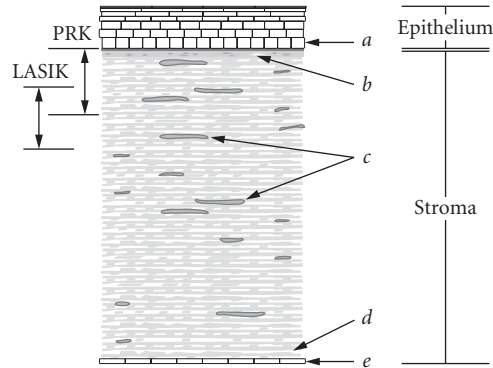
The optics of the eye are covered in greater detail in Chap. 1.

## The Cornea

The first optical surface of the eye is the cornea, and hence the interface between air and tissue produces the highest refractive power, and makes the cornea the most important optical surface in the eye. The most common refractive surgery procedures achieve a change in the overall optical power of the eye by changing the shape of the cornea. For this reason we will give more attention to its cellular structure.



**FIGURE 1** Schematic eye treating the whole optical system as a single thin lens. (Adapted from Ref. 1.)



**FIGURE 2** Schematic representation of the corneal layers.

The cornea is a remarkably transparent and mechanically stable tissue. The cornea features five layers with distinct roles as seen in Fig. 2. The anterior surface of the cornea is covered by the corneal epithelium (*a* in Fig. 2). The corneal epithelium is the outermost surface of the cornea and it is the surface that is wetted by the tear film.

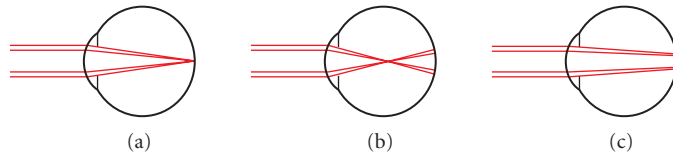
The corneal epithelium (stratified squamous epithelium), has a thickness of approximately  $50\ \mu\text{m}$ . It is a multicellular layer, comprising fast regenerating cells that provide a barrier to water, ions, and bacteria. The epithelium consists of progressively flatter cells, so that its outer surface is capable of smoothing (to some extent) minor underlying irregularities or roughness. The epithelium is not a permanent tissue, in the sense that it is continuously regenerated. Consequently, in order to produce permanent changes to the corneal geometry, any modifications must be applied to the underlying layers.

Underneath the epithelium there is a layer of dense collagen (*b* in Fig. 2) having a thickness of about  $12\ \mu\text{m}$ , called Bowman's layer. This layer has no cells and it has been hypothesized that this structure provides mechanical strength to the cornea. The main part of the cornea representing approximately 90 percent of its thickness is the corneal stroma. The corneal stroma consists mainly of elongated bands of Type I and V collagen arranged in a lamellar configuration. Consecutive lamellae are organized in random orientations perpendicular to the optical axis to avoid anisotropy in the transverse orientations. The space around the collagen fibers is filled with a hydrophilic glycosaminoglycan and water. In the corneal stroma there are relatively sparse keratocytes (*c* in Fig. 2) that are the fibroblastic cells of the cornea. Keratocyte activity sustains proper collagen spacing and collagen fibril diameter in the cornea. In the case of injury, keratocytes produce collagen at accelerated rates in an attempt to restore stromal integrity. Often this new collagen is substantially less organized than normal collagen and is characterized by decreased transparency.

Descemet's membrane (also known as the posterior limiting lamina) (*d* in Fig. 2) of the cornea is an acellular membrane made mostly of collagen. The corneal endothelium is a monolayer of specialized cells that lines the posterior surface of the cornea. The corneal endothelium acts like a fluid pump across the posterior surface of the cornea that actively maintains the cornea in the relatively dehydrated state that maintains dense collagen organization and therefore transparency of the cornea.

## Refractive Errors

The function of the optics of the eye is to form an image of the outside world onto the retina. In an *emmetropic* eye and in the absence of accommodation, the image of an object at infinity is sharply focused onto the retina. When the unaccommodated eye fails to produce a well-focused image of a distant object on the retina, it is said to be *ametropic* or to have a refractive error. The degradation in



**FIGURE 3** Schematic representation of refractive errors. (a) Normal emmetropic eye, (b) myopic eye, and (c) hyperopic eye.

the image seen by an ametropic eye is produced purely by the optics of the eye, and hence, in principle, it is possible to correct it by optical means alone.

The definition of refractive errors or ametropias characterize the excess or lack of optical power in the eye. In these definitions it is understood that the crystalline lens is meant to be relaxed, or *unaccommodated*. The case of an eye free of refractive errors or emmetropic is shown in Fig. 3a; the image of an object at infinity is formed on the retina. In ametropia this is no longer the case, an object at infinity appears blurred on the retina. Ametropias are divided in two main types: spherical ametropia and astigmatism.

The reader should be aware that *presbyopia* is not considered to be an ametropia. Presbyopia is the decrease in our ability to accommodate as we grow old. This process starts from youth and continues till the age of about 60 when our ability to accommodate disappears completely. In presbyopia, the crystalline lens does not change shape any longer and we are left with a single focal plane. In addition to presbyopia, an eye can have any of the refractive errors or ametropias that will be defined in the following sections.

**Spherical Ametropias** In spherical ametropia, the optical power of the eye is symmetric around its optical axis, but there is a mismatch between the optical power of the eye and its axial length. It is divided in two cases: *myopia* and *hypermetropia*. Myopia occurs when the optical power of the eye is too large for its length and the image of an object at infinity is formed in front of the retina. This is shown in Fig. 3b. Hypermetropia, sometimes referred to as hyperopia, occurs when the optical power of the eye is too small for its length so that the image plane lies behind the retina. This situation is shown in Fig. 3c.

The point conjugated with the retina in the absence of accommodation is called the *far point*. In an unaccommodated emmetropic eye the far point lies at infinity. Hence, accommodation is only required when seeing objects closer than infinity. In myopia, the far point is placed at a finite distance from the eye. The larger the myopia the closer the far point will be. Accommodation can only bring into focus objects closer to the eye than the far point, as accommodation can only increase the optical power of the eye. In a myopic eye any object beyond the far point will always be out of focus.

In contrast, the hypermetropic eye must always accommodate. The far point lies behind the retina, and hence power must always be added in order to bring an image into focus. The young hypermetrope can accommodate, and this is usually done unconsciously. Hence hypermetropia can pass unnoticed for many years.

**Astigmatism** Myopia and hyperopia affect the optical power of the eye equally in all meridians. In *astigmatism* this is not the case. An astigmatic optical surface has two perpendicular principal meridians; the curvature varying from a maximum in one meridian to a minimum in the other. The astigmatism of such a surface is defined as the difference between the optical powers at each one of these meridians. Hence astigmatism as a refractive error has the same meaning as that of on-axis astigmatism normally found in geometrical optics. Oblique astigmatism, arises when the angle of incidence of a pencil of rays is not parallel to its optical axis. This can also arise in the eye, but it is not a refractive error and hence it will not be dealt in this chapter.

The simplest possible case of an astigmatic system is that of a cylindrical lens in which one meridian has zero curvature. A lens with a toroidal surface is another example, and is commonly used to make spectacle lenses. It is generated by the revolution of an arch  $AB$  around an axis  $CD$  not passing through the arch's center of curvature  $C_c$ .

In general astigmatism can be regarded as a combination of spherical and cylindrical elements. Optically this is equivalent to an sphero-cylindrical lens with cylindrical power equal to the difference

between the two principal powers. Additionally, in the case of toroidal surfaces the spherical power is that of the weakest principal meridian.

The angle of astigmatism is measured on a plane perpendicular to the eye's optical axis. It is defined as the angle at which the meridian with the maximum curvature lies. In optometric practice this angle is measured counterclockwise starting from the horizontal, when looking into the eye. The same system is used in both eyes. For horizontal astigmatism the angle used is 180, not 0. The  $^\circ$  symbol is not used to avoid confusion with the number zero, so that, for example,  $13^\circ$  is not mistaken with 130.

Astigmatism is present in practically all human eyes. There are several reasons for this. Firstly, the cornea is seldom perfectly spherical. It usually presents certain degree of astigmatism. Likewise, the crystalline lens may present astigmatic surfaces. Moreover, the mutual alignment between crystalline lens and cornea is not perfect. There is always certain amount of tilt or decentration between these elements.

**Ocular Wavefronts** The eye is a biological system, and as such it is not perfectly symmetric. The refractive errors discussed so far are only a first approximation. The eye presents other more irregular errors in its refractive power. Older texts refer to these errors as *irregular astigmatism*. The more modern literature and current research papers refer to these errors as *higher-order aberrations*. Strictly speaking refractive error is a measure of the error in optical power in the eye and has units of diopters. Instead, higher-order aberrations are a measure of phase error and can have units of microns or radians. The term higher-order aberrations originates from the description of an optical wavefront.

The field of a monochromatic scalar propagating wave may be written explicitly as

$$\mathbf{u}(r, t) = \text{Re}[U(r)\exp(-ik\phi(r))\exp(-i2\pi\nu t)] \quad (1)$$

where  $U(r)$  and  $k\phi(r)$  are, respectively, the modulus and phase of the wave at position  $r$ ,  $k$  is the wave number

$$k = 2\pi \frac{\nu}{c} = \frac{2\pi}{\lambda} \quad (2)$$

and  $\text{Re}[\cdot]$  is short notation meaning the "real part of." In the definition of the wave number  $\nu$  is the optical frequency,  $\lambda$  is the optical wavelength, and  $c$  is the speed of light in free space.

The time dependence of the optical wave of fixed wave number in Eq. (1) is known a priori. Consequently the complex function  $U(r) = U(r)\exp[-ik\phi(r)]$  is enough to describe the optical disturbance. The function  $\phi(r)$  is typically referred to as the wavefront of the optical perturbation and the optical perturbation  $U(r)$  is referred to as the optical field, the field, or the propagating wave.

A perfect optical system, in the presence of a point source produces at the exit pupil plane a perfectly spherical wavefront converging to an ideal focusing point. Any departures from this perfectly spherical shape are known as *aberrations* and the function describing them is usually represented by  $\omega(x, y)$ . If the reference spherical wavefront is known, then the wavefront aberration  $\omega(x, y)$  not only contains all the information about the aberrations in the system, but it can be used to describe the phase of the propagating optical wave, keeping in mind that the reference sphere has been subtracted from this function.

The quantity  $\omega(x, y)$  is a real function. In Cartesian coordinates, it can be expressed as a polynomial expansion of the form

$$\omega(x, y) = \sum_{n=0}^{\infty} a_n P_n(x, y) \quad (3)$$

where  $P_n$  are the expansion polynomials and  $a_n$  are the weighting values for each polynomial. In principle  $\{P_n\}$  can be any polynomial base, however, in vision sciences the Zernike polynomials have gained popularity and are widely used. The American National Standards Institute (ANSI) has produced a standard for reporting optical aberrations of eyes which uses Zernike polynomials.<sup>4</sup>

When using Zernike polynomials, aberrations are split into higher- and lower-order aberrations. Table 1 shows the first 14 Zernike polynomials according to the ANSI standard (having ignoring the piston or constant term, as it is not possible to measure it in the eye). The first five terms are the lower-order

**TABLE 1** The First 14 Zernike Polynomials According to the ANSI Standard;<sup>4</sup>  $\rho$  and  $\theta$  Are Radial Coordinates on the Pupil Plane

No.	Polynomial	Description
1	$2\rho\sin\theta$	$x$ tilt or horizontal prism
2	$2\rho\cos\theta$	$y$ tilt or vertical prism
3	$\sqrt{6}\rho^2\sin 2\theta$	45° astigmatism
4	$\sqrt{3}(2\rho^2 - 1)$	Focus or sphere
5	$\sqrt{6}\rho^2\cos 2\theta$	0° astigmatism
6	$\sqrt{8}\rho^3\sin 3\theta$	Trefoil
7	$\sqrt{8}(3\rho^3 - 2\rho)\sin\theta$	$x$ Coma
8	$\sqrt{8}(3\rho^3 - 2\rho)\cos\theta$	$y$ Coma
9	$\sqrt{8}\rho^3\cos 3\theta$	Trefoil
10	$\sqrt{10}\rho^4\sin 4\theta$	-
11	$\sqrt{10}(4\rho^4 - 3\rho^2)\sin(2\theta)$	-
12	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$	Spherical aberration
13	$\sqrt{10}(4\rho^4 - 3\rho^2)\cos(2\theta)$	-
14	$\sqrt{10}\rho^4\cos 4\theta$	-

aberrations and correspond to prism, sphere or defocus, and astigmatism. Aberrations described by any polynomials of order higher than two are referred to as higher-order aberrations.

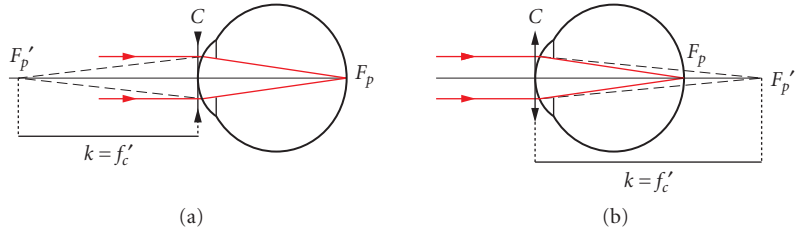
Study of the wavefront error and higher-order aberrations on the eye has been possible thanks to the development of objective techniques to measure the ocular wavefront aberrations. The most popular of these techniques is the use of a Shack-Hartmann sensor, based the work by Liang et al.<sup>5</sup> Another popular technique is *laser ray tracing* which was developed simultaneously by Moreno-Barriuso et al.<sup>6</sup> and by Molebny.<sup>7</sup> Several other techniques exist; see for instance Refs. 8 to 10.

Currently, there is no consensus about how to estimate refractive error from wavefront data. A number of different metrics have been proposed.<sup>11</sup> Also a methodology to represent wavefront maps as refractive power maps has recently been proposed by Iskander et al.<sup>12</sup> The details of these works fall outside the scope of this chapter and will not be addressed here.

## Correcting Refractive Errors

Refractive correction must be done for the unaccommodated eye and turn the compound system formed by the eye and its correction into an emmetropic “eye”. That is, a well-corrected eye must be able to form a sharp image of an object at infinity on the retina in the absence of accommodation. We have already defined the far point of the eye as the optical conjugate of the retina in the absence of accommodation. Hence, refractive correction must shift that point to infinity. Alternatively, applying the reversibility of the optical path, a correcting lens must create an image of an object at infinity on the far point of the uncorrected eye.

Figure 4a shows the optical principle of the correction of a myopic eye. The dashed line shows the optical rays converging to the far point ( $F'_p$ ) in the absence of correction. The red lines show the



**FIGURE 4** Optical principle for the correction of ametropias. (a) Myopia and (b) hypermetropia. (Adapted from Ref. 13)

light path after refractive correction. A lens of negative optical power is placed in front of the eye. Its back focal length ( $f'_p$ ) is equal to the distance between the eye and its far point ( $k$ ). Under these circumstances a pencil of rays coming from an object at infinity will diverge after passing through the lens, and a virtual image will form on the eye's far point  $F'_p$ . The optical system of the eye will now be able to bring this image into focus on the retina  $F'_p$ .

The principle of correction for a hypermetropic eye is exactly the same as above, but in this case a correcting lens with positive power is required. This situation is shown Fig. 4b. The correcting positive lens creates the image of an object at infinity behind the eye, on its far point. As in the previous case, the unaccommodated eye can now form a sharp image of the object on the retina.

The inverse of the distance to the far point  $1/k \equiv K$  is then equal to the optical power  $F'_p$  of the lens required to correct the refractive error and we have

$$k = f'_c \Rightarrow K = F'_c$$

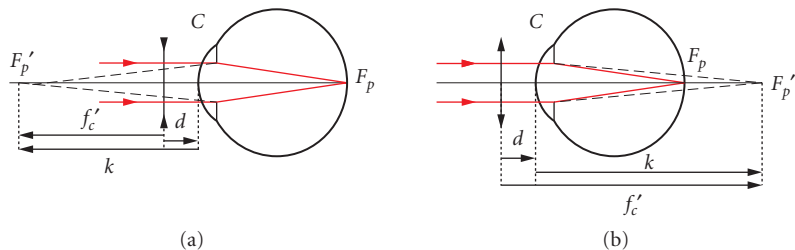
The diagrams in Fig. 4 are both for the case when the correction lens is in contact with the eye, like the case of a contact lens. When correction is achieved with spectacle correction one must take into account the distance between the lens and the eye. This distance is always measured from the back vertex of the lens (the one closest to the eye), and is called the vertex distance and denoted by  $d$ . Diagrams illustrating this situation are shown in Fig. 5. It is clear that in this case

$$k = f'_c - d \quad (4)$$

and the power of the correcting lens must be adjusted as follows:

$$F'_c = \frac{K}{1 + dK} \quad (5)$$

It should be clear from Fig. 5 that if a refractive correction is measured with a lens at a distance  $d$ , changing this distance to  $d'$  will have an effect on the power of the correcting lens. The larger the power of



**FIGURE 5** Optical principle for the correction of ametropias with a spectacle lens. (a) Myopia and (b) hypermetropia. (Adapted from Ref. 13.)

the lens the larger this impact. However, since the far point of the eye must coincide with the focal point of the correcting lens, it is straightforward to adjust the prescription. If the lens is to be moved by  $d - d'$  mm closer to the eye's far point, then the focal length  $f'_p$  of the correcting lens must be reduced by  $d - d'$  mm.

This is particularly important for refractive surgery, because refractive errors are typically prescribed using trial lenses placed 10 to 14 mm away from the cornea. The most common type of refractive surgery is aimed at changing the optical power of the cornea itself, so this distance must be taken into consideration.

An astigmatic eye can be treated as if it had two separate far points, each one corresponding to one of the principal meridians. The correction of astigmatism is achieved treating each principal meridian independently using the same principles described above. Namely, using an astigmatic lens, the focal point of each meridian of the lens must coincide with the far point of the corresponding meridian in the eye.

Correction of higher-order aberrations has been proposed by several authors.<sup>14–18</sup> Partial correction of aberrations has also been proposed as a way to improve visual performance.<sup>19</sup> Although these studies have shown an improvement in contrast sensitivity after correction of higher-order aberrations, this has only been possible in laboratory conditions. Moreover, Dalimier et al. found that aberration correction is most effective when the target was viewed in bright light. For lower light levels, the improvement in optical quality was masked out by a decrease in neural sensitivity. That is, for lower light levels, resolution is limited by neural mechanisms and not by optical quality.

These results mean that correction of higher-order aberrations, as an everyday solution, may not be practical for the majority of people. During the day, the pupil constricts reducing the benefit that correction could bring. At night the pupil dilates, but neural sensitivity is reduced. Higher-order aberration correction may only benefit patients with pathologically abnormal corneas where they play a dominant role in image degradation. Nevertheless, understanding of important visual mechanisms like the triggering of accommodation may largely benefit from carefully designed studies where aberration-free images are projected on the retina.

Refractive corrections are covered in greater detail in Chap. 12.

## 16.3 REFRACTIVE SURGERY MODALITIES

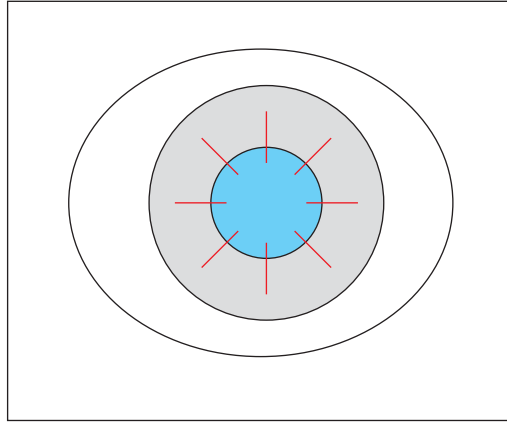
As mentioned earlier, refractive surgical procedures are attempting to alter the optical power of the eye with permanent means in order to compensate for the eye's refractive error. There are three major categories of refractive procedures: *Intraocular lens (IOL) implantation*, *corneal incisions/implants*, and *laser corneal ablation*.

### Intraocular Lenses, Corneal Incisions, and Implants

**Intraocular Lenses** Intraocular lenses (IOLs) were first developed for the substitution of the crystalline lens after cataract extraction. The first lenses were made of appropriate glass and were substantially heavy.<sup>20,21</sup> Modern lenses are made of appropriate biocompatible polymers (such as silicone or acrylic). Some types are designed to be implanted in the eye while the natural lens is present and therefore to compensate for possible refractive errors. These IOLs (phakic IOLs<sup>22–24</sup>) are either implanted in the posterior chamber (between the iris and the crystalline lens) or in the anterior chamber supported by the iris. The latter is the most common type currently used. The main advantage of IOL implantation is that the correction is reversible in the sense that the implant can be removed. Generally high corrections can be achieved.

**Corneal Incisions/Implants** A different approach for correcting ametropias and, historically, one of the first to be widely applied corneal incisions procedures. Corneal incisions are performed (typically using a diamond knife) either in a radial or in a circumferential manner. Depth of the incisions is typically adjusted to be 90 percent of the local thickness of the cornea. The rationale of corneal incisions is that local relaxation of mechanical stress results to a change in corneal geometry. Radial keratotomy is one of the most common incisional techniques (Fig. 6).



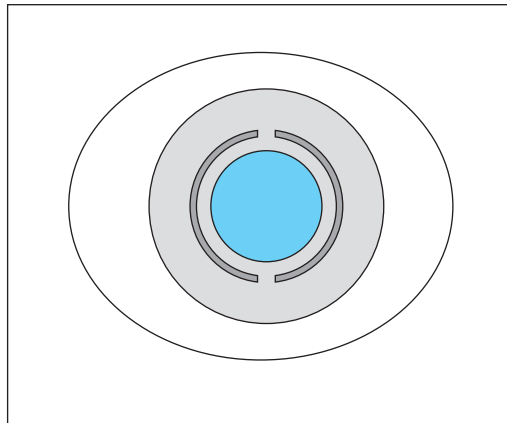


**FIGURE 6** Schematic representation of radial incisions for the correction of myopia (radial keratotomy—RK).

The main drawbacks of radial keratotomy are the low predictability,<sup>25</sup> the large amount of induced aberrations,<sup>26</sup> and the compromised mechanical strength of the eye globe.<sup>27</sup> The low predictability of incisional surgery is associated with the fact that the long-term redistribution of stress in the corneal surface is not stable over time but is varying during the healing process. The final outcome must be examined several months after the incisions when healing has reached a stable condition.

The variability of the healing process may be attributed to individual factors such as the particular anatomy of the given cornea, age, and other biochemical and biological factors. In order to reduce this variability the desired refractive change is achieved by using empirical nomograms with different number, length, and positioning of the incisions that additionally take into consideration the factors mentioned above. Circumferential (arcuate) incisions have been employed for the correction of astigmatism usually in pairs across the steep meridian of the corneal astigmatism.

Another modality that has been employed to correct low refractive errors is the implantation of intracorneal ring segments.<sup>28</sup> These semicircular arcs are implanted in circumferential tunnels (created either mechanically or by means of femtosecond laser photodisruption) in the corneal stroma (Fig. 7).



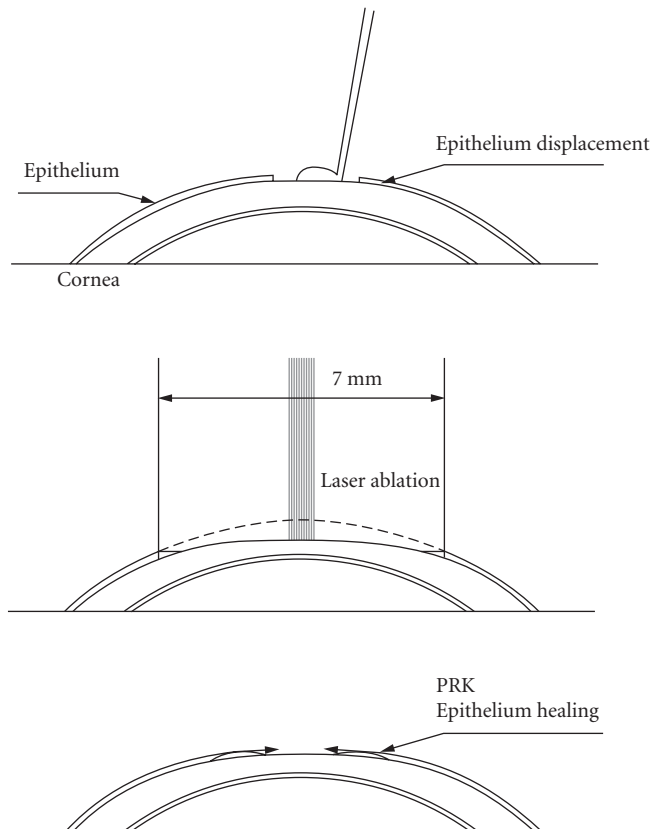
**FIGURE 7** Schematic representation of intracorneal ring segments (ICRS) implanted around the pupil.

These segments are made of stiff material like polymethyl methacrylate (PMMA) and their rigid shape provides a new set of boundary conditions for corneal geometry. Their refractive effect is achieved by elevating the circumferential portion of the cornea over their implantation tunnel. The fact that they are rigid has been utilized in order to provide mechanical support to corneas with compromised mechanical stability such as in keratoconus<sup>29</sup> or keratectasia.<sup>30</sup>

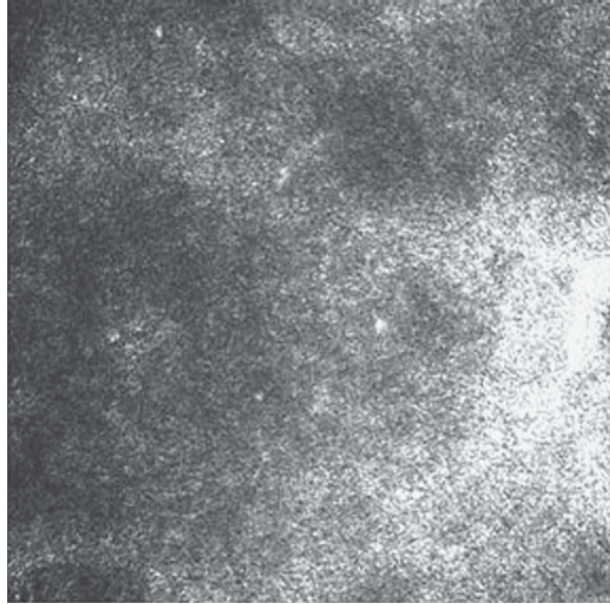
## Laser Corneal Refractive Procedures

The majority of the refractive procedures performed today are based on controlled ablation of corneal tissue in order to modify its central curvature. Among all the different approaches, the most critical differentiation is the depth at which the ablation is performed. In surface ablations (PRK, Epi-LASIK, LASEK) tissue removal involves the Bowman's layer and superficial stroma after appropriate removal of the corneal epithelium. In LASIK, tissue is removed from deeper stroma after creation of a hinged flap. Surface ablations are considered less invasive than LASIK in the sense that for any given correction a thinner part of the stroma is affected.

**Photorefractive Keratectomy (PRK)** PRK was the first laser procedure to be widely available.<sup>31,32</sup> It involves removal of corneal epithelium with a blunt instrument such as a spatula or a brush. After the Bowman's layer is exposed, laser ablation is performed using an appropriate superposition of a number of laser pulses. A schematic representation of this procedure is shown in Fig. 8. After



**FIGURE 8** Schematic representation of photorefractive keratectomy.



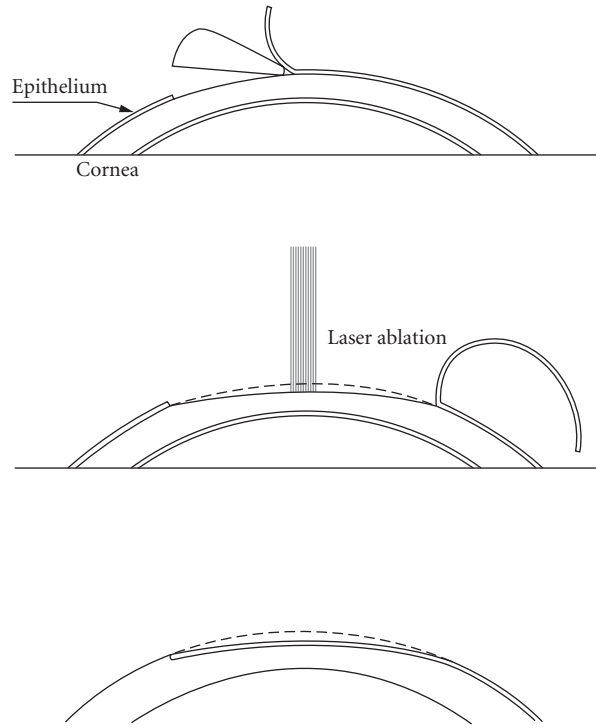
**FIGURE 9** Confocal microscope image of the subepithelial scar tissue after PRK. Side is 400  $\mu\text{m}$ , depth is at the epithelial—stromal interface (50  $\mu\text{m}$ ).

ablation, the corneal epithelium regenerates slowly and within 3 to 5 days completely covers the reshaped corneal surface. During this period of epithelial healing (that is characterized by pain or discomfort to the patient) the cornea is covered by a contact lens to reduce discomfort. Antibiotics are also administered as drops during this period of time to prevent infection of the exposed cornea. Additionally to antibiotics, anti-inflammatory treatment often in the form of steroids is prescribed. Time duration of the steroid treatment is not standard. Several months may be necessary depending on the physician's preferred scheme.

The technique is surgically simple and straightforward and is considered safe and effective for the treatment of low to moderate refractive errors.<sup>33</sup> Since treatment of the cornea with the excimer laser results to trauma and a transient inflammation, there is a healing response that not only modulates the refractive outcome, but additionally influences the transparency of the cornea due to the low organization of newly formed collagen. Figure 9 shows a confocal microscope image of corneal scar. This postoperative deterioration of corneal transparency (known as corneal haze) is not expressed with significant severity in the majority of the patients.<sup>34</sup>

In order to avoid scar tissue formation several different approaches have been developed, collectively called advanced surface ablations. These techniques are variants of the PRK where either special pharmaceutical treatment is applied in order to control the production of new collagen or the epithelial layer is preserved in order to cover the treated area immediately after laser ablation. The most common pharmaceutical is mitomycin C.<sup>35</sup> Mitomycin C is an anticancer drug known also to effectively inhibit keratocyte proliferation and therefore haze formation. However, there is skepticism about its use by many practitioners because its long-term effects and safety have not yet been documented. In order to reduce postoperative exposure to inflammation factors the LASEK procedure was developed.

**LASEK/Epi-LASIK** In these procedures, the epithelium is preserved as a whole sheet and is repositioned in order to cover the treated portion of the cornea. Presumably, this would reduce



**FIGURE 10** Schematic representation of the Epi-LASIK technique.

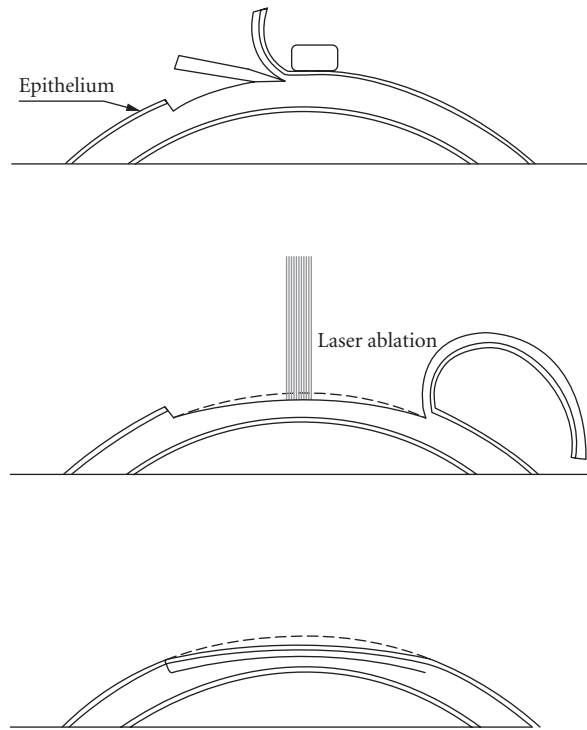
exposure of keratocytes to biochemical factors in the tears that trigger and enhance the healing response. In LASEK (laser subepithelial keratomileusis),<sup>36</sup> the epithelium is loosened by the application of an alcohol solution for a short-time interval (typically 20% alcohol in water for 30 s). After the treatment with alcohol, the epithelium can be delaminated using a blunt spatula in a manner that preserves its continuity. After laser ablation, the epithelium is repositioned onto the cornea. This epithelium is replaced in 3 to 5 days by new epithelial cell migrating from the periphery.

Epi-LASIK (epithelial laser in situ keratomileusis), is a similar technique (Fig. 10) where the epithelium is mechanically separated from the underlying Bowman's layer by means of a special device called epikeratome.<sup>37,38</sup>

Although promising, these techniques have not been proved to be more effective than PRK in terms of postoperative risk for haze development. Both LASEK and Epi-LASIK being essentially surface ablations are mostly preferred for low to moderate refractive errors.

**Laser In Situ Keratomileusis (LASIK)** The most common laser technique for the correction of moderate and high refractive errors is laser in situ keratomileusis (LASIK). LASIK was invented in the late 1980s by Pallikaris et al.<sup>39,40</sup> and has gained popularity due to its faster rehabilitation and less postoperative pain. In LASIK (Fig. 11) a thin (typically 120  $\mu\text{m}$ ) flap is created exposing deeper stromal layers while preserving epithelium and the Bowman's layer.

The flap is created using a special instrument called "microkeratome." The microkeratomomes feature an oscillating blade and an applanator that flattens a portion of the cornea where the cut is performed. By this instrument, a flap of constant thickness can be produced where this thickness is controlled by the gap between the applanator and the knife's edge.



**FIGURE 11** Schematic representation of laser in situ keratomileusis (LASIK).

After laser ablation, the flap is repositioned on the reshaped corneal stromal bed. This, effectively isolates the treated area from inflammation factors of the environment and therefore corneal transparency is less affected by the healing response. The creation of the flap along with the ablation may influence the mechanical stability of the cornea. There are different safety criteria related to how much ablation can be performed where the most common is that the remaining corneal stromal thickness under the ablation should be no thinner than 300  $\mu\text{m}$ .

**Ablation Profiles** After the corneal stroma is exposed with either of the above-mentioned techniques, a sequence of overlapping laser pulses is delivered in order to excise corneal tissue in a controlled manner and modify the curvature of the corneal surface. The desired ablation profile is achieved by the appropriate superposition of—typically—a few tens of thousands of pulses, depending on the diameter of the treated zone, the desired refractive correction, and the type of ametropia.

Typically, only the portion of the cornea in front of the pupil is treated in order to minimize the tissue that needs to be excised. For this purpose, the diameter of the pupil in scotopic conditions is measured and the ablation profile is calculated for a zone preferably slightly larger than the scotopic pupil diameter.

For the correction of myopia, the refractive power of the cornea has to be reduced and therefore the ablation profile is designed to excise more tissue centrally than peripherally and therefore flatten the central cornea. Similarly, the correction of hyperopia requires steepening of the central cornea by the removal of peripheral tissue. The removed tissue has the shape of a positive or negative meniscus

lens respectively. In the case of astigmatism, the ablation profile has a toric shape introducing a relative change in dioptric power across two perpendicular meridians.

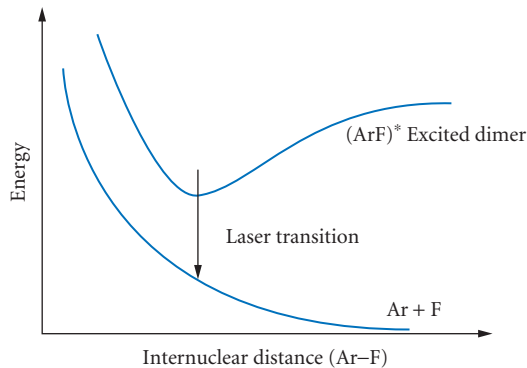
The ablation profiles typically attempt the modification of corneal curvature in a circular zone centred on the line-of-sight that has diameter of the order of 6 to 8 mm. This zone is known as the optical zone of the profile. The edges of the optical zone are treated with additional pulses in order to smooth out the border between corrected and noncorrected portion of the cornea creating a transition zone. The width of the transition zone is of the order of 1 mm and depends on the magnitude of the discontinuity that needs to be smoothed. It is clear that the wider the optical zone the more the tissue needs to be excised for any given dioptric correction. The exact shape that needs to be removed in any case is calculated by geometrical considerations known as the Munnerlyn's equations.<sup>41,42</sup>

Laser systems usually incorporate further empirical adjustments to compensate for biological factors such as the biomechanical response of the cornea and/or the effect of healing (especially at the border of the optical zone). Ablation profiles can be either calculated directly from the spherocylindrical refractive error or can be generated based on measurements of corneal topography or wavefront aberration of the eye. The advent of topography and wavefront-guided ablation profiles generated an interest in creating supernormal (aberration-free) optical systems for the human eye.<sup>43</sup> However, it is now understood that the outcome variability that originates from individual factors such as the healing response does not allow the precise sculpturing of corneal surface in order to eliminate high-order aberrations in a predictable manner.<sup>35</sup>

In the next section we turn our attention to the characteristics of the lasers used to perform the refractive surgery procedures and how they interact with the living tissue.

## 16.4 LASER ABLATION

Excimer lasers are pulsed ultraviolet (UV) lasers where the amplification medium is a biatomic system existing only at an electronically excited state.<sup>44</sup> Excited dimers (e.g., ArF) can be formed during electric discharges through gas mixtures of the corresponding elements and they do not exist in their ground electronic state (Fig. 12).<sup>44,45</sup> Excited dimers have a typical lifetime of a few nanoseconds. Their spontaneous dissociation is accompanied by the emission of a UV photon. In excimer lasers, typically excitation is provided by a transverse electric discharge (perpendicular to the optical axis) through a mixture of gases of pressure similar to that of atmospheric pressure. The dissociation of



**FIGURE 12** Potential energy for ArF in electronically excited and ground state.

the excited dimers renders the medium transparent to its emission wavelength. Typically, excimer lasers are characterized by high gain and low feedback therefore the beam has low coherence and generally has considerable inherent divergence.

The most common excimer lasers and their respective emission wavelengths are XeF (351 nm), XeCl (308 nm), KrF (248 nm), and ArF (193 nm). Typically, for surgical systems the pulse duration is of the order of 20-ns FWHM and the total energy per pulse is of the order of tens or hundreds of millijoule. The electric efficiency for a typical ArF laser system is of the order of 2 percent.<sup>45</sup> Applications of excimer lasers include material's processing, micromachining, and the excision of tissue.

## Corneal Photoablation

The first experiments for the evaluation of excimer lasers as means to excise corneal tissue, were performed in the late 1980s.<sup>46–48</sup> These studies concluded that the most suitable wavelength was 193 nm (ArF excimer laser) due to its relatively limited thermal damage to the underlying tissue, the accuracy of the excision, and the smoothness of the resulting surface.<sup>48</sup> The interaction of high-power pulsed UV radiation with corneal tissue results from the process of ablative photodecomposition (or simply photoablation) described just below.

The UV pulse is absorbed by a thin surface layer. Precise measurement of the absorption coefficient in real conditions has not been accomplished; however, the penetration depth is estimated to be a few microns.<sup>49</sup> As energetic photons (6.4 eV) are absorbed, the molecules that form the extracellular matrix are decomposed (Fig. 13a). This process is primarily photochemical.<sup>48</sup>

According to a model proposed by Dougherty et al. to describe the role of corneal hydration in the photoablation process,<sup>50</sup> radiation with a wavelength of 193 nm is primarily absorbed by collagen, while water plays a more passive role absorbing a small fraction of the radiation and—mostly—affecting the thermal and mechanical properties of the tissue. As radiation is absorbed and molecular fragments are formed, more photons are absorbed by the fragments and this results to a portion of tissue from the surface to be effectively vaporized. For typical conditions (pulse energy/duration) the pressure near the surface may reach levels of the order of 100 bar.<sup>51</sup> The ablation products expand rapidly and are ejected perpendicularly to the surface with supersonic velocities.<sup>52</sup>

Based on the description above, it is reasonable to expect that the deeper a collagen fiber the larger the fragments and the less the kinetic energy of each fragment. Therefore, there is a critical depth (depending on the pulse energy density) for which fragments are energetic enough to be expelled from the material. Under this depth, the collagen has undergone direct photochemical damage but not to the extent required for ablation. For energy densities below a certain threshold photoablation does not occur. This threshold has been experimentally determined to be approximately 40 mJ/cm<sup>2</sup> for normal cornea and typical pulse durations.<sup>53</sup>

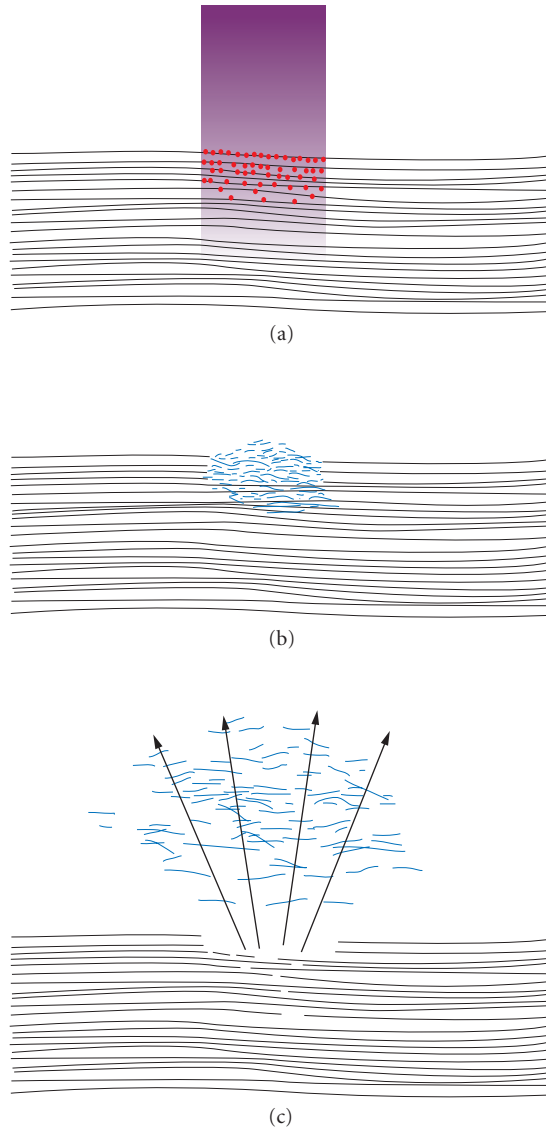
## Ablation Rate

At any given conditions (pulse duration, pulse repetition rate) the ablation rate is defined as the depth of tissue excised by each pulse and is generally expressed in  $\mu\text{m}/\text{pulse}$ . As mentioned above, ablation rate depends on the incident energy density. Given that for a particular laser system the pulse duration is constant, the quantity that is empirically used to quantify the incident laser energy density is the surface density (fluence) typically expressed in mJ/cm<sup>2</sup>.

The empirical equation to describe the dependence of the ablation rate on fluence is logarithmic<sup>53</sup>

$$A = m \cdot \ln \frac{F}{F_{\text{thr}}} \quad (6)$$

where  $A$  is the ablation rate in  $\mu\text{m}/\text{pulse}$ ,  $F$  is the fluence,  $F_{\text{thr}}$  the ablation threshold, (in this equation it is equal to 50 mJ/cm<sup>2</sup>) and  $m$  is an empirical constant equal to 0.3  $\mu\text{m}/\text{pulse}$

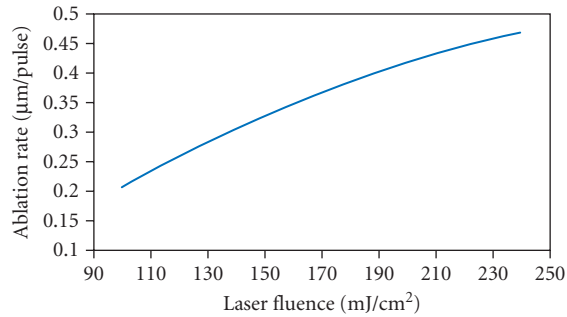


**FIGURE 13** Schematic representation of the corneal photoablation process.

(Fig. 14). Typically, fluences employed in refractive surgery applications range<sup>54</sup> from 120 to 225 mJ/cm<sup>2</sup>.

In practice, parameters such as pulse duration, beam profile, the presence of a gas stream for ablation debris aspiration, and the repetition rate may influence the ablation rate. These parameters as well as the temperature and humidity in the operating room are factors that are standardized in refractive surgery in order to minimize the variability of the delivered ablation. As mentioned earlier, the appropriate ablation profiles (that are designed to modify corneal geometry in a predetermined manner) are created with an appropriate superposition of pulses. Typical repetition rates are of the





**FIGURE 14** Corneal ablation rate as a function of fluence.

order of 400 Hz. For the removal of tissue necessary for the correction of moderate myopia, the total number of pulses depend on all of the parameters mentioned above and are of the order of a few tens of thousands of pulses. Besides standardization of laser and environmental parameters during ablation, the accuracy of tissue excision is enhanced by the utilization of empirical nomograms<sup>55</sup> where other parameters are taken into account such as attempted refractive correction, diameter of the treatment zone, patient age, corneal radius of curvature, corneal thickness, and others.

The main variable that may influence the effective ablation rate is corneal hydration.<sup>50</sup> Standardization of the procedures minimizes the variability associated to corneal hydration changes during surgery. It has been suggested that hydration could be measured spectroscopically and ablation parameters could be adjusted dynamically to compensate for possible changes of hydration.<sup>56</sup>

### Thermal, Photochemical, and Photoacoustic Effects

As mentioned above, photoablation can be primarily seen as a fast photochemical decomposition of the structural elements of the cornea (the extracellular matrix) during which high-pressure gradients are developed removing fragments from the surface of the irradiated tissue.<sup>46–48</sup> In parallel to photoablation there are phenomena that their potential implications in the acute postoperative inflammation and long-term healing response of the cornea is not fully understood.

A significant part of the laser energy is used for the photochemical decomposition of the extracellular matrix. However, there is a part of the laser energy directly dissipated as heat (e.g., for depths greater than the ablation rate where energy density is subthreshold). Moreover, the highly energetic fragments deposit part of their energy to the surface as they are expanding. As a result of these two mechanisms each laser pulse results to a certain amount of heat dissipated on the free surface of the ablated cornea.

The temperature rise associated with this heat dissipation is highly variable and depends on the particular conditions (repetition rate, fluence, beam diameter pulse superposition pattern, etc.). In practical conditions the temperature rise has been reported to range from 0 to 8°C.<sup>57,58</sup> One of the parameters taken into account in pulse superposition pattern design is the effective reduction of repetition rate in the sense that irradiation of any point with consecutive pulses is avoided. In this manner, minimal changes in surface temperature of the cornea can be achieved.<sup>59</sup>

In principle, ArF excimer laser radiation is not capable of producing damage to the DNA of keratocytes as the penetration depth is smaller than the dimensions of one cell.<sup>60</sup> In that sense if the nucleus of a cell is exposed to this UV radiation a significant portion of the cell is already ablated and therefore the cell is not viable. It is the secondary radiation (fluorescence) that bears a potential risk for mutagenesis.<sup>60,61</sup> Fluorescence is primarily emitted at wavelengths in the region between 260 and 320 nm.<sup>61</sup> Photoreactivation experiments with yeast cell cultures showed a significant amount of DNA repair activity after 193-nm excimer laser irradiation. Although the role of

secondary radiation in the healing response is not yet fully understood<sup>62</sup> there seems to be minimal concern among the physicians as clinical experience of more than 20 years has not shown that the potential risk is significant.

Additionally, it has been hypothesised<sup>51</sup> that the mechanical stress of the cornea due to the pressure gradients associated to photoablation may result to structural changes in collagen and to the formation of scar tissue and loss of corneal transparency.

## 16.5 ACKNOWLEDGMENTS

The authors would like to thank Loukia Leonidou for assistance in preparing the figures in this chapter.

## 16.6 REFERENCES

1. D. W. Meltzer, ed., *Optics, Refraction and Contact Lenses*, Basic and Clinical Science Course, Sec. 3, American Academy of Ophthalmology, 1994.
2. W. M. Hart, ed., *Adler's Physiology of the Eye, Clinical Application*, 9th ed., Mosby Year Book, St. Louis, 1993.
3. R. M. Boynton, *Human Color Vision*, Optical Society of America, 1992.
4. *American National Standard for Ophthalmics, ANSI-Z80.28-2004, Methods for Reporting Optical Aberrations of Eyes*, American National Standards Institute, Inc., 2004.
5. J. Liang, B. Grimm, S. Goelz, and J. F. Bille, "Objective Measurement of Wave Aberration of the Human Eye with the Use of a Hartmann-Shack Wavefront Sensor," *J. Opt. Soc. Am. A.* **11**:1949–1957 (1994).
6. E. Moreno-Barriuso, S. Marcos, R. Navarro, and S. A. Burns, "Comparing Laser Ray Tracing, Spatially Resolved Refractometer and Hartmann-Shack Sensor to Measure the Ocular Wave Aberration," *Optom. Vis. Sci.* **78**:152–156 (2001).
7. V. Molebny, "Principles of Ray-tracing Aberrometry," in *1st International Congress of Wavefront Sensing and Aberration-Free Refractive Correction*, vol. 16, pp. 572–575 (2000).
8. F. Diaz-Doutón, J. Pujol, M. Arjona, and S. O. Luque, "Curvature Sensor for Ocular Wavefront Measurement," *Opt. Lett.* **31**:2245–2247 (2006).
9. C. Torti, S. Gruppetta, and L. Diaz-Santan., "Wavefront Curvature Sensing for the Human Eye," *J. Mod. Opt.* **55**:691–702 (2008).
10. I. Iglesias, R. Ragazzoni, Y. Julien, and P. Artal, "Extended Source Pyramid Wavefront Sensor for the Human Eye," *Opt. Express* **10**:419–428 (2002).
11. Jason D. Marsack, Larry N. Thibos, and Raymond A. Applegate, "Metrics of Optical Quality Derived from Wave Aberrations Predict Visual Performance," *J. Vis.* **4**(4):322–328 (2004).
12. D. R. Iskander, B. A. Davis, M. J. Collins, and R. Franklin, "Objective Refraction from Monochromatic Wavefront Aberrations via Zernike Power Polynomials," *Ophthalmic. Physiol. Opt.* **27**(3):245–255 (2007).
13. A. G. Bennett and R. B. Rabbetts, *Clinical Visual Optics*, Butterworths, London, 1984.
14. J. Liang and D. R. Williams, "Supernormal Vision and High Resolution Retinal Imaging through Adaptive Optics," *J. Opt. Soc. Am. A.* **14**:2884–2892 (1997).
15. P. Artal, F. Vargas, and I. Iglesias, "Detection of the Wavefront Aberration in the Human Eye and Its Partial Correction with a Liquid Crystal Light Modulator," in *International Workshop: Adaptive Optics for Industry and Medicine*, Shatura, Rusia, 1997.
16. M. B. McDonald, "Summit—Autonomous CustomCornea Laser in situ Keratomileusis Outcomes," in *1st International Congress of Wavefront Sensing and Aberration-Free Refractive Correction*, vol. 16, pp. 617–618, 2000.
17. H. Hofer, L. Chen, G. Y. Yoon, B. Singer, Y. Yamauchi, and D. R. Williams, "Improvement in Retinal Image Quality with Dynamic Correction of the Eye's Aberrations," *Opt. Exp.* **8**:631–643 (2001).
18. E. Dalimier, C. Dainty, and J. L. Barbur, "Effects of Higher-Order Aberrations on Contrast Acuity as a Function of Light Level," *J. Mod. Opt.* **55**(4):791–803 (2008).

19. Patricia A. Piers, S. Manzanera, P. M. Prieto, N. Gorceix, and P. Artal, "Use of Adaptive Optics to Determine the Optimal Ocular Spherical Aberration," *J. Cataract Refract. Surg.* **33**:1721–1726 (2007).
20. H. Ridley, "Intraocular Acrylic Lenses after Cataract Extraction," *Lancet* **I**(6699):118–121 (1952).
21. H. Ridley, "The Treatment of Cataract," *Practitioner* **178**(1067):525–533 (1957).
22. C. F. Lovisolo and D. Z. Reinstein, "Phakic Intraocular Lenses," *Surv. Ophthalmol.* **50**(6):549–587 (2005).
23. L. Espandar, J. J. Meyer, and M. Moshirfar, "Phakic Intraocular Lenses," *Curr. Opin. Ophthalmol.* **19**(4):349–356 (2008).
24. D. H. Chang and E. A. Davis, "Phakic Intraocular Lenses," *Curr. Opin. Ophthalmol.* **17**(1):99–104 (2006).
25. D. R. Sanders, M. R. Deitz, and D. Gallagher, "Factors Affecting Predictability of Radial Keratotomy," *Ophthalmology* **92**(9):1237–1243 (1985).
26. R. A. Applegate, H. C. Howland, R. P. Sharp, A. J. Cottingham, and R. W. Yee, "Corneal Aberrations and Visual Performance after Radial Keratotomy," *J. Refract. Surg.* **14**(4):397–407 (1998).
27. P. F. Vinger, W. F. Mieler, J. H. Oestreicher, and M. Easterbrook, "Ruptured Globes Following Radial and Hexagonal Keratotomy Surgery," *Arch. Ophthalmol.* **114**(2):129–134 (1996).
28. T. E. Burris, "Intrastromal Corneal Ring Technology: Results and Indications," *Curr. Opin. Ophthalmol.* **9**(4):9–14 (1998).
29. G. D. Kymionis, C. S. Siganos, N. S. Tsiklis, A. Anastasakis, S. H. Yoo, A. I. Pallikaris, N. Astyrakakis, and I. G. Pallikaris, "Long-term Follow-up of Intacs in Keratoconus," *Am. J. Ophthalmol.* **143**(2):236–244 (2007).
30. G. D. Kymionis, N. S. Tsiklis, A. I. Pallikaris, G. Kounis, V. F. Diakonou, N. Astyrakakis, and C. S. Siganos, "Long-term Follow-up of Intacs for Post-Lasik Corneal Ectasia," *Ophthalmology* **113**(11):1909–1917 (2006).
31. C. R. Munnerlyn, S. J. Koons, and J. Marshall, "Photorefractive Keratectomy: A Technique for Laser Refractive Surgery," *J. Cataract Refract. Surg.* **14**(1):46–52 (1988).
32. M. B. McDonald, J. M. Frantz, S. D. Klyce, R. W. Beuerman, R. Varnell, C. R. Munnerlyn, T. N. Clapham, B. Salmeron, and H. E. Kaufman, "Central Photorefractive Keratectomy for Myopia. The Blind Eye Study," *Arch. Ophthalmol.* **108**(6):799–808 (1990).
33. C. W. Jr. Flowers, P. J. McDonnell, and S. D. McLeod, "Excimer Laser Photorefractive Keratectomy," *Ophthalmol. Clin. North. Am.* **14**(2):274–283 (2001).
34. P. Fagerholm, "Phototherapeutic Keratectomy: 12 Years of Experience," *Acta Ophthalmol Scand.* **81**(1):19–32 (2003).
35. W. B. Trattler and S. D. Barnes, "Current Trends in Advanced Surface Ablation," *Curr. Opin. Ophthalmol.* **19**(4):330–334 (2008).
36. M. Camellin, "Laser Epithelial Keratomileusis for Myopia," *J. Refract. Surg.* **19**(6):666–670 (2003).
37. I. G. Pallikaris, V. J. Katsanevaki, M. I. Kalyvianaki, and I. I. Naoumidi, "Advances in Subepithelial Excimer Refractive Surgery Techniques: Epi-Lasik," *Curr. Opin. Ophthalmol.* **14**(4):207–212 (2003).
38. I. G. Pallikaris, M. I. Kalyvianaki, V. J. Katsanevaki, and H. S. Ginis, "Epi-Lasik: Preliminary Clinical Results of an Alternative Surface Ablation Procedure," *J. Cataract Refract. Surg.* **31**(5):879–885 (2005).
39. I. G. Pallikaris, M. E. Papatzanaki, E. Z. Stathi, O. Frenschcock, and A. Georgiadis, "Laser in situ Keratomileusis," *Lasers Surg. Med.* **10**(5):463–468 (1990).
40. I. G. Pallikaris, M. E. Papatzanaki, D. S. Siganos, and M. K. Tsilimbaris, "A Corneal Flap Technique for Laser in situ Keratomileusis. Human Studies," *Arch. Ophthalmol.* **109**(12):1699–1702 (1991).
41. C. R. Munnerlyn, S. J. Koons, and J. Marshall, "Photorefractive Keratectomy: A Technique for Laser Refractive Surgery," *J. Cataract Refract Surg.* **Jan**;14(1):46–52 (1988).
42. P. J. McDonnell, H. Moreira, J. Garbus, T. N. Clapham, J. D'Arcy, and C. R. Munnerlyn, "Photorefractive Keratectomy to Create Toric Ablations for Correction of Astigmatism," *Arch. Ophthalmol.* **May**;109(5):710–3 (1991).
43. T. Seiler, M. H. Dastjerdi, "Customized Corneal Ablation," *Curr. Opin. Ophthalmol.* **Aug**;13(4):256–260 (2002).
44. D. Basting, ed., *Excimer Laser Technology: Laser Sources, Optics, Systems and Applications*, Lambda Physik, Gottingen, 2001.
45. V. M. Borisov, I. E. Bragin, A. Yu. Vinokhodov, and V. A. Vodchits, "Pumping Rate of Electric-Discharge Excimer Lasers," *Quantum Electronics* **25**(6):507–510 (1995).

46. R. Srinivasan and E. Sutcliffe, "Dynamics of the Ultraviolet Laser Ablation of Corneal Tissue," *Am. J. Ophthalmol.* **103**(3 pt 2):470–471 (1987).
47. R. Srinivasan, "Ablation of Polymers and Biological Tissue by Ultraviolet Lasers," *Science* **234**(4776):559–65 (1986).
48. S. L. Trokel, R. Srinivasan, and B. Braren, "Excimer Laser Surgery of the Cornea," *Am. J. Ophthalmol.* **96**(6):710–715 (1983).
49. G. B. Petit and M. N. Ediger, "Corneal-tissue Absorption Coefficients for 193- and 213-nm Ultraviolet Radiation," *App. Opt.* **35**:3386–3391 (1995).
50. P. J. Dougherty, K. L. Wellish, and R. K. Maloney, "Excimer Laser Ablation Rate and Corneal Hydration," *Am. J. Ophthalmol.* **118**:169–176 (1994).
51. O. Kermani and H. Lubatschowski, "Structure and Dynamics of Photo-Acoustic Shock-waves in 193 nm Excimer Laser Photo-ablation of the Cornea," *Fortschr Ophthalmol.* **88**(6):748–753 (1991).
52. D. W. Hahn, M. N. Ediger, and G. H. Pettit, "Dynamics of Ablation Plume Particles Generated during Excimer Laser Corneal Ablation," *Lasers Surg. Med.* **16**(4):384–389 (1995).
53. T. Seiler and P. J. McDonnel, "Excimer Laser Photorefractive Keratectomy," *Surv. Ophthalmol.* **40**:89–118 (1995).
54. Surgical Lasers Operation Manuals and Data sheets. (Various.)
55. M. H. Feltham and R. J. Wolfe, "Some Variables to Consider to Avoid the Need for Lasik Surgical Enhancements," *Clin. Exp. Optom.* **83**(2):76–81 (2000).
56. I. G. Pallikaris, H. S. Giniş, G. A. Gounis, D. Anglos, T. G. Papazoglou, and L. P. Naoumidis, "Corneal Hydration Monitored by Laser-Induced Breakdown Spectroscopy," *J. Refract. Surg.* **14**(6):655–660 (1998).
57. C. Maldonado-Codina, P. B. Morgan, and N. Efron, "Thermal Consequences of Photorefractive Keratectomy," *Cornea* **20**(5):509–515 (2001).
58. T. Bende, T. Seiler, and J. Wollensak, "Side Effects in Excimer Corneal Surgery: Corneal Thermal Gradients," *Graefes Arch. Clin. Exp. Ophthalmol.* **226**:277–280 (1988).
59. M. Vetrugno, A. Maino, E. Valenzano, and E. Cardia, "Corneal Temperature Changes during Photorefractive Keratectomy using the Laserscan 2000 Flying Spot Laser," *J. Refract. Surg.* **17**(4):454–459 (2001).
60. T. Seiler, T. Bende, K. Winckler, and J. Wollensak, "Side Effects in Excimer Corneal Surgery. DNA Damage as a result of 193 nm Excimer Laser Radiation," *Graefes Arch. Clin. Exp. Ophthalmol.* **226**(3):273–276 (1988).
61. N. Muller-Stolzenburg, S. Schrunder, J. Helfmann, H. J. Buchwald, and G. I. Muller, "Fluorescence Behavior of the Cornea with 193 nm Excimer Laser Irradiation," *Fortschr Ophthalmol.* **87**(6):653–658 (1990).
62. Z. Z. Nagy, J. Toth, A. Nagymihaly, and I. Suveges, "The Role of Ultraviolet-b in Corneal Healing Following Excimer Laser in situ Keratomileusis," *Pathol. Oncol. Res.* **8**(1):41–46 (2002).

*This page intentionally left blank*

---

# THREE-DIMENSIONAL CONFOCAL MICROSCOPY OF THE LIVING HUMAN CORNEA

---

Barry R. Masters

*Department of Biological Engineering  
Massachusetts Institute of Technology  
Cambridge, Massachusetts*

---

## 17.1 GLOSSARY

---

**Adaptive optics.** An optical system that consists of a wavefront sensor that can measure the wavefront of the light, that is, a Shack-Hartmann sensor, a deformable mirror that can alter the wavefront of the light, and a closed loop-feedback control system to minimize the wavefront error. An example is the use of adaptive optics in a scanning laser ophthalmoscope, in which the wave front distortions due to the cornea and the ocular lens are corrected by the adaptive optics.

**Confocal microscope.** A microscope that is based on spatial filtering with two pinholes or slit apertures located in conjugate planes. The word confocal means that both the point source of light and the point detector are cofocused on the same focal volume within the specimen. A point source of light (source aperture) illuminates a diffraction limited focal volume within the specimen. The scattered or emitted light from this focal volume is focused on a detection pinhole located in a plane that is conjugate to the source aperture. A large area detector is placed behind the photodetector pinhole.

**Diffraction limited imaging.** The theoretical resolution of an optical microscope is limited by the diffraction of light. Abbe defined this diffraction limit for microscopic resolution. The diffraction limited resolution is proportional to numerical aperture of the microscope objective, that is, its ability to collect light, and inversely proportional to the wavelength of the light from the specimen. A microscope that can image at the theoretical limit is a diffraction limited imaging system. Today, there are optical microscopes that exceed the Abbe diffraction limit.

**Laser scanning confocal microscope.** A confocal microscope in which the laser beam is scanned over the back focal plane of the microscope objective with the simultaneous scanning of the diffraction limited beam over the specimen. The scattered or emitted light from the focal volume is collected by an epi-illumination system, descanned, and passes through an aperture placed in front of a large area photodetector.

**Linear optics.** The response of the illuminated material to the electromagnetic field is linear with the amplitude of the field. Examples are one-photon absorption, scattering, and fluorescence.

**Micro lens Nipkow disk confocal microscope.** This is a major improvement over the Nipkow disk tandem scanning confocal microscope that has an illumination efficiency of only 1 to 2 percent. In this new device a second Nipkow disk which contains about 20,000 microlenses is added to the

**Nipkow disk microscope.** The lower Nipkow disk contains another 20,000 pinholes that are arranged in the same pattern as the microlenses on the upper disk. The lower pinhole disk is located in the focal plane of the upper microlens disk. This arrangement increases the illumination efficiency to 50 percent and results in a high sensitivity even for weak reflecting specimens. The same pinholes are used for illumination (laser excitation light) and spatial filtering of the fluorescence light, and a dichroic mirror placed between the two disks separates the excitation light from the fluorescence that is focused on the detector. Both disks rotate together on the same axis to form a real-time scanned image of the specimen.

**Nipkow disk tandem-scanning confocal microscope.** A real-time, direct-view, tandem-scanning microscope based on a rotating Nipkow disk, contains sets of conjugate pinholes arranged in a set of spirals about the axis of rotation. On a given diameter of the disk there is one pinhole for the illumination and one pinhole for detection of the scattered or emitted light from the specimen. As the disk rotates, many conjugate sets of illumination and detection pinholes (arranged in conjugate planes on opposite sides of the Nipkow disk) operate in parallel to simultaneously scan the specimen with a diffraction limited illumination spot, and to detect the descanned scattered or emitted light.

**Nonlinear optics.** The response of the illuminated material to the electromagnetic field is nonlinear with the amplitude of the field. At high light intensities, the incident field modifies the optical properties of the material such that the light waves can interact and exchange energy and momentum. Examples of nonlinear optics are two-photon excitation fluorescence, second-harmonic generation, and third-harmonic generation.

**Numerical aperture.** The numerical aperture (NA) of a microscope objective is a measure of the range of angles that the lens can accept or emit light. The NA of a microscope objective is defined as:  $NA = n \sin \theta$ , where  $n$  is the index of refraction of the medium between the objective and the specimen, and  $\theta$  is the half-angle of the maximum cone of light that can enter the objective.

**Optical aberrations.** They represent the failure of an optical system to produce a perfect image. Modern microscope objectives are made to minimize five categories of optical aberrations: spherical aberrations (results in a lack of sharp focus), coma (results when light is focused at points off the optical axis), astigmatism, field curvature, and distortion. In addition to these monochromatic Seidel aberrations, there are also axial chromatic aberrations and lateral chromatic aberrations in which different colors of light focus at different positions.

**Scanning slit confocal microscope.** A real-time, confocal microscope in which a slit of incident light is scanned over the back focal plane of the microscope objective. The reflected and the emitted light from the focal volume within the specimen is collected by the microscope objective, descanned by the scanning system, and detected by a wide-area detector.

**Second-harmonic generation (SHG).** A nonlinear process in which an incident wave of frequency  $\omega$  interacts with a material that does not contain an inversion symmetry, and generates a new signal at the frequency  $2\omega$ .

**Slit-lamp microscope.** A microscope that is based on the principle of focal illumination. A slit of light is used to illuminate a volume within the eye. A biomicroscope collects the scattered and reflected light from a volume that intersects the illumination volume. The intersection of the illumination volume and the collection volume is in focus.

**Specular microscope.** A modification of the slit lamp in which one half of the microscope objective is used for illumination of the specimen and one half of the microscope objective is used to collect the scattered and the reflected light. For a plane reflecting surface when the angle between the normal to the surface and the illumination beam of light is equal to the angle between the normal to the surface and the reflected beam of light, the condition of specular reflection exists. The microscope that images this specular reflection is called a specular microscope. It is used to image the endothelial cells due to the strong specular reflection at the endothelial cell and the aqueous humor interface.

**Third-harmonic generation (THG).** A nonlinear process in which an incident wave of frequency  $\omega$  interacts with any material, and generates a new signal at the frequency  $3\omega$ .

## 17.2 INTRODUCTION

This chapter presents the optical principles of the confocal microscope and its applications to the design of clinical confocal microscopes for imaging the living human cornea. A series of technical advances resulted in new optical designs from the early slit lamp with its focal plane illumination to the modern scanning slit and laser scanning clinical confocal microscopes. A carefully edited selection of papers, books, and commercial Web sites is presented as a further resource of clinical images, instrument designs, and clinical protocols. The emphasis of this chapter is to critically examine the limitations of the various designs of clinical instruments, and to present the author's perspective on the state of the art instruments and the proposed future instrument development that will incorporate the suggested improvements as well as those based on the emerging field of nonlinear optical microscopy,<sup>1,2</sup> specifically second- and third-generation harmonic microscopy. It is further suggested that future clinical biomicroscopes will incorporate correlative microscopy, that is, both linear reflected confocal microscopy together with nonlinear higher-harmonic microscopy.

## 17.3 THEORY OF CONFOCAL MICROSCOPY

The principles of the confocal microscope in both the transmission mode and in the epi-illumination mode were clearly stated and illustrated in the 1961 patent granted to Marvin Minsky.<sup>3</sup> To facilitate easy access to the key papers and the important patents in the field of confocal microscopy and its many biological applications SPIE Press published an edited collection of these documents.<sup>3</sup> Further theoretical and practical details on instrument design, theories of transverse, and axial optical resolution are contained in a modern book.<sup>4</sup>

A widefield optical microscope has no capability for axial optical sectioning; therefore, when the specimen is a thick, highly scattering tissue the images at the focal plane are degraded by out-of-focus contributions to the image. The confocal microscope solves this problem with the method of spatial filtering.<sup>3</sup> A point source of light from an incoherent or a coherent source illuminates a point in the specimen via the microscope objective. The back-scattered or the fluorescence light from that point is collected by the microscope objective and passes an aperture that is in a conjugate plane to the point source of illumination. Only light from the focal plane of the illumination spot passes the conjugate aperture and is detected by the detector located directly behind the detection aperture. Light from the out-of-focus plane cannot pass the detection aperture and therefore it is not detected. This confocal principle is shown in Fig. 1. The use of spatial filtering provides an enhanced axial resolution ( $z$  axis of the microscope) and is the source of the axial optical sectioning in the confocal microscope.

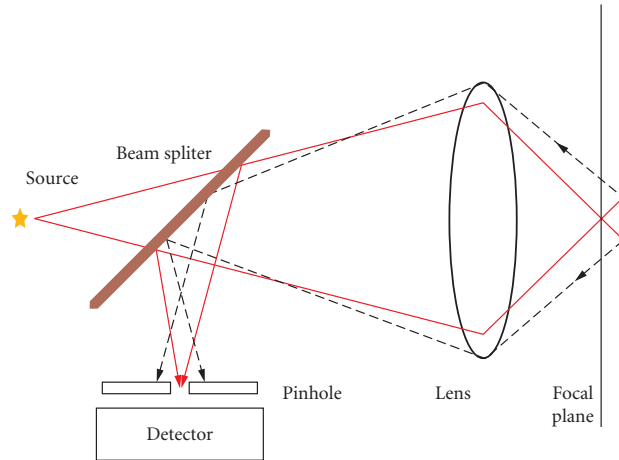
In order to acquire two-dimensional images of the focal plane it is necessary to scan the light beam over the back focal plane of the microscope objective and to descann the scattered or emitted light that is collected from the microscope objective. Typically two sets of oscillating galvanometer driven mirrors, one for the horizontal scan and one for the vertical scan, are used for laser scanning and descanning.

## 17.4 THE DEVELOPMENT OF CONFOCAL INSTRUMENTS

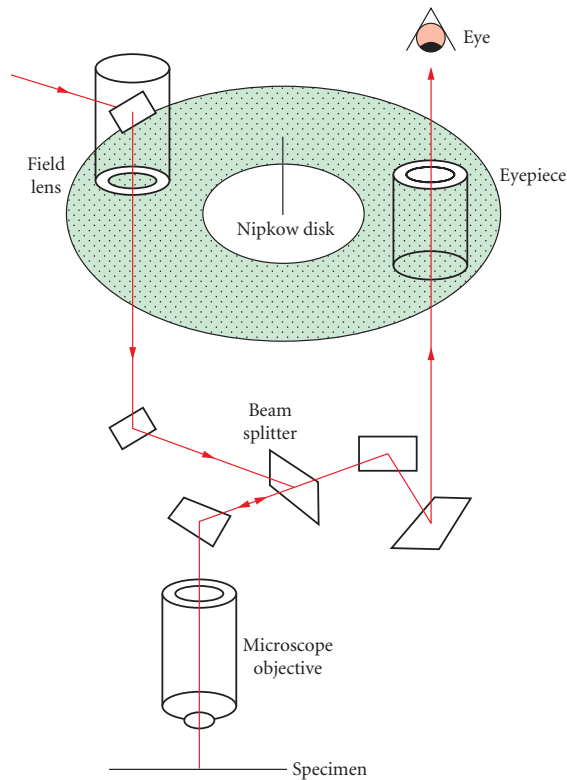
The late David Maurice is honored by the vision research community for his outstanding contributions to the development of optical ophthalmic instrumentation.<sup>5</sup> The development of the scanning slit confocal microscope is based on a series of linked technical advances that include parallel invention as well as unique creative invention, and the story of the development of live cell imaging in ophthalmology is described in a recent book.<sup>6</sup>

Two groups of confocal microscopes were developed for ophthalmology: the first was a laser scanning ophthalmoscope for imaging the retina was developed by Webb and his collaborators<sup>7</sup> and the





**FIGURE 1** The principle of the confocal microscope. This schematic diagram shows the depth discrimination capability of a confocal microscope and how it discriminates against reflected light from out of the focal plane. The dashed vertical line to the right of the focal plane represents an out-of-focus plane. Only the light rays (solid lines) from the focal plane pass the pinhole and are detected. (Reproduced with permission from Ref. 6.)

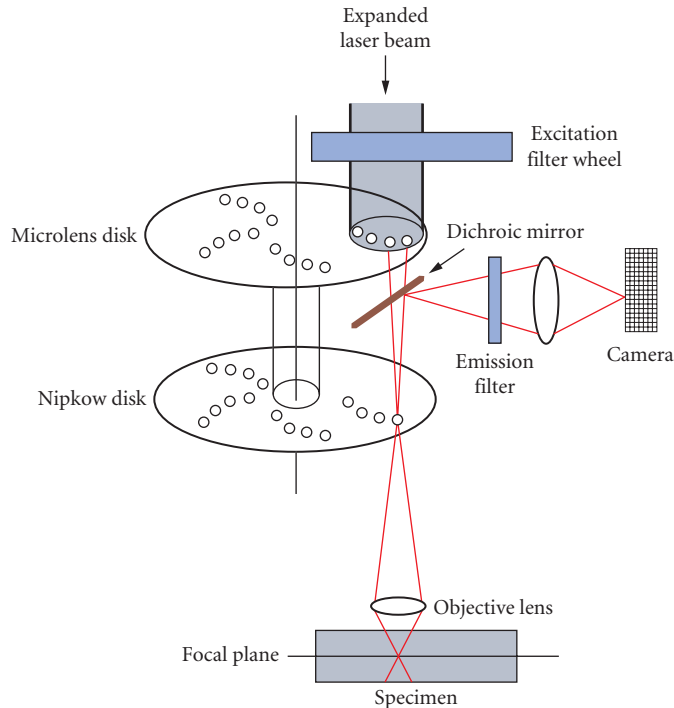


**FIGURE 2** The principle of Nipkow disk confocal microscope. This schematic shows the real-time, direct-view, tandem-scanning Nipkow disk confocal microscope. The light source from the upper left is a mercury arc lamp or a tungsten filament lamp. (Reproduced with permission from Ref. 6.)

second group of confocal microscopes that typically used an incoherent light source was developed for imaging the cornea.<sup>8</sup>

Several early types of *in vivo* confocal microscopes were based on the spinning Nipkow disk that contained spiral arrays of multiple sets of conjugate pinholes, one pinhole was used for illumination, and a conjugate pinhole was used for spatial filtering in front of the detector.<sup>3,6</sup> The design of the Nipkow disk based confocal microscope is shown in Fig. 2.

The Nipkow disk-based confocal microscope has a very poor illumination efficiency (1 to 2 percent) and therefore weak reflecting corneal cell layers such as the basal epithelial cell layer cannot be imaged. The solution to this problem was made by Akira Ichihara and his coworkers at the Yokogawa Institute Corporation in Japan with their invention of a microlens Nipkow disk confocal microscope.<sup>6</sup> The addition of an upper disk that contained 20,000 microlenses which focused the light on the 20,000 pinholes on the lower disk. The key point is that the lower pinhole disk is aligned and located in the focal plane of the upper disk. The resulting illumination efficiency is about 50 percent. Figure 3 shows the principle of the microlens Nipkow disk confocal microscope.



**FIGURE 3** The principle of microlens Nipkow disk confocal microscope. Schematic drawing of the Yokogawa microlens Nipkow disk confocal microscope. The two disks rotate on a common axis. The upper microlens disk contains 20,000 microlenses arranged in a series of spirals. The lower Nipkow disk contains 20,000 pinholes arranged in a series of spirals. The lower disk containing the pinholes is in the focal plane of the microlenses in the upper disk. A dichroic mirror placed between the two disks passes the incident laser beam, but reflects the beam of emitted light to the emission filter and then it is focused on the camera. (Reproduced with permission from Ref. 6.)

## 17.5 THE SCANNING SLIT AND LASER SCANNING CLINICAL CONFOCAL MICROSCOPES

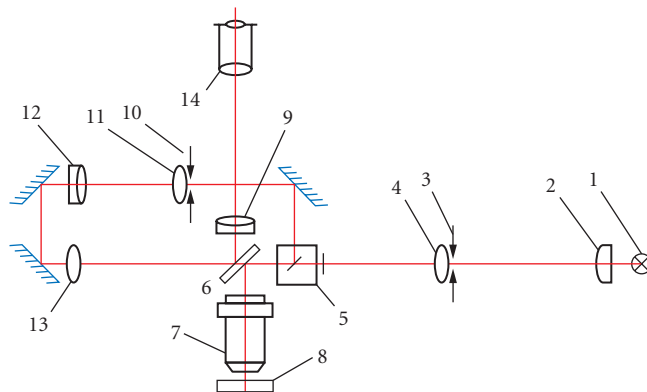
Based on the Svishchev confocal microscope design of a two-sided oscillating mirror for scanning a slit of light across the back focal plane of a high NA, water immersion microscope objective, a clinical confocal microscope with an incoherent halogen lamp light source was designed and constructed.<sup>9</sup> Figure 4 shows the principle of the two-sided, oscillating mirror in the scanning slit confocal microscope.

The advantage of this scanning slit confocal microscope is that the slit widths of the two conjugate slits, one for the illumination side and one for the detection side, are completely adjustable. They could be opened to increase the reflected signal from weak scattering corneas, or could be reduced to increase the resolution of the microscope.<sup>10</sup> Later commercial designs used fixed slits.

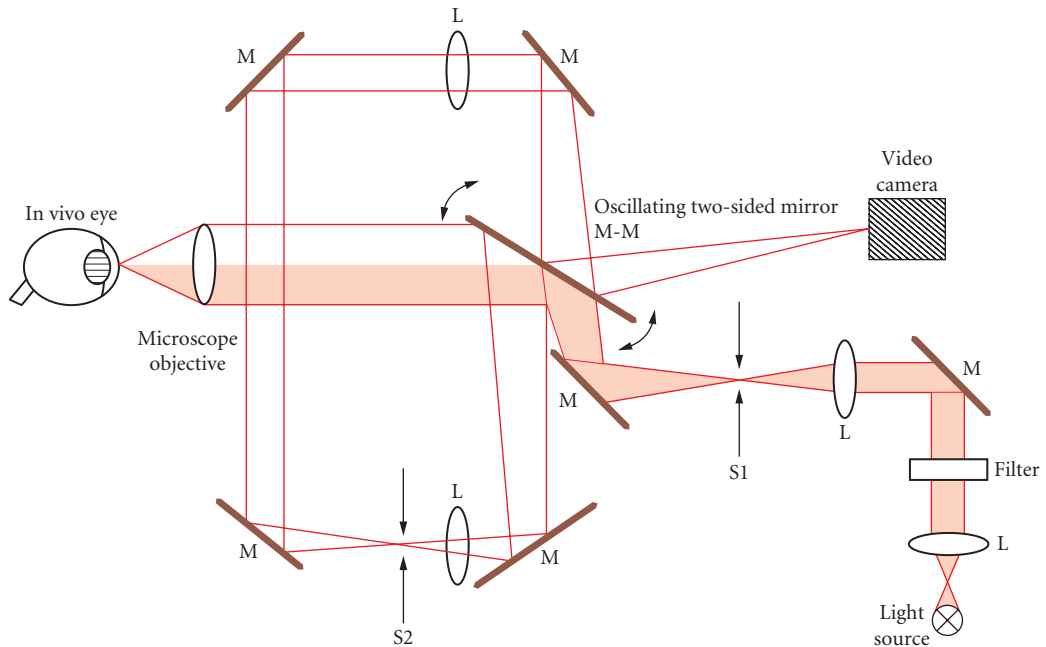
This scanning slit confocal microscope was placed in the clinic and was used to study a variety of normal and clinical conditions of the *in vivo* human cornea. Figure 5 shows the principle of the clinical scanning slit confocal microscope based on the Svishchev oscillating two-sided mirror design.

For example, this scanning slit confocal microscope was successfully used to study the basal epithelial cell layer, and its adjacent wing cell layer. In the clinic the confocal microscope was used to relocate the same cells within the cornea over different periods of time, this provided a method to track cell migration over time. Since the microscope objective contacted the anterior surface of the cornea there is the need for both an anesthetic and the application of a drop of an index matching gel to the tip of the microscope objective.

In recent years the scanning slit confocal microscope has been made into a commercial product that is now in the fourth generation.<sup>11</sup> One useful feature of the commercial version, the Confoscan 4, is the 20x microscope objective that works under noncontact conditions; therefore, there is no need



**FIGURE 4** The principle of the Svishchev two-sided, oscillating mirror, scanning slit confocal microscope. The light source, 1, is projected by a condenser lens, 2, onto the first slit, 3; the light passes through a prism cube, 5; an image of the first slit is scanned over the back focal plane of the microscope objective, 7, by the two-sided oscillating mirror, 6, which descans the reflected light from the focal plane of the specimen. The second slit, 10, is conjugate with the first slit (confocal) and excludes the light that is not in the specimen's focal plane, 8. The two-sided oscillating mirror, 6, performs three functions: it scans the image of the slit 3 over the back focal plane of the microscope objective, descans the beam from the object, and rescans the beam for observation in the ocular, 14. The parts labeled 2, 4, 9, 11, 12, and 13, are lenses. (Reproduced with permission from Ref. 6.)



**FIGURE 5** The optical principles of the clinical, video-rate, scanning-slit, in vivo, confocal microscope with two adjustable slits: S1 for illumination and S2 for imaging. An oscillating two-sided mirror M-M scans S1 over the back focal plane of the objective, descans the collected light and directs it to S2, then rescans the beam, which is imaged on the photocathode of a video camera. The dark ray path indicated the illumination light rays. The light ray path indicates the reflected light that is collected by the microscope objective. One half of the NA of the microscope objective is used for illumination, and one half of the NA is used for collection of the reflected light form the specimen. (Reproduced with permission from Ref. 6.)

for anesthetic nor for an index matching gel on the tip of the objective. Numerous publications based on the slit scanning confocal microscope are replete with confocal images of normal corneas and the diseased and altered corneas.<sup>12–16</sup> In addition, the Nidek Web site features many examples of confocal microscopic images of the cornea.<sup>11</sup>

Modern clinical confocal microscopes can acquire stacks of confocal images through the full thickness of the human cornea and produce three-dimensional visualizations of the a volume from the anterior surface to the posterior endothelial cell layer. This approach of three-dimensional visualization of the living cornea was first demonstrated on rabbit corneas by Masters and coworkers.<sup>17–19</sup>

An alternative to the commercial scanning slit confocal microscope is a laser scanning confocal microscope designed for retinal imaging by Heidelberg Engineering that can be modified with a Rostock Cornea Module (RCM) microscope that permits confocal microscopy of the cornea.<sup>20</sup> The Heidelberg Retinal Tomograph (HRT) is a confocal laser scanning microscope with a diode laser light source at 670 nm. The optical principle of this confocal microscope is shown in Fig. 1.

The RCM unit has the ability to change the focal plane within the cornea. A single-use planar cap is fitted over the tip of the module and it is in contact with the anterior surface of the cornea. The plastic cap stabilizes the distance between the module and the cornea. An index matching gel is placed between the instrument tip and cornea. Between the movable plastic cap is an index matching gel that serves to increase the numerical aperture of the objective which results in increased resolution and contrast. When it is desirable to use the confocal microscope in the noncontact mode, the water immersion microscope objective is removed together with its contact cap and is replaced with a long

focal length dry objective. Typically a noncontact Nikon 50X, NA 0.45 microscope objective is used to image the cornea in the noncontact mode. The specifications from the manufacturer include a focus range of 4 mm, the lateral image size of 380 by 380  $\mu\text{m}$ , the lateral optical resolution of about 1  $\mu\text{m}$ , and the depth of focus of about 1 to 2  $\mu\text{m}$ .

## 17.6 CLINICAL APPLICATIONS OF CONFOCAL MICROSCOPY

The clinical utility of a clinical confocal microscope for the examination of the *in vivo* human cornea depends on the optical and mechanical design parameters of the confocal microscope, the optical condition of the subject's cornea, and finally, on the skill and the experience of the clinician who is making the examination.

A general introduction to the principles and the practice of clinical confocal microscopy is a rich resource of clinical protocols. A review on the confocal microscopy of the cornea by Böhnke and Masters provides the accumulated clinical experience of many years of patient examination.<sup>21</sup> This practical discussion of clinical techniques includes the following topics: technical setup of the confocal microscope, the technique of biomicroscopy with the confocal microscope, the suggested clinical examination procedure, patient data archiving and retrieval, and the dissemination, documentation, and publication of microscopic findings. This paper also contains an atlas of confocal images from the normal cornea, the effects of aging, the effects of contact lens wear, the cornea with known pathologies, and the postsurgical cornea.

Creative scientists, engineers, and clinicians have worked together in the last decades to develop new clinical confocal microscopes that have successfully been incorporated into clinical practice. Optics and optical engineering are in the forefront of the design and development of medical devices for both treatment and diagnostics providing rich career opportunities for highly trained people.

Two books are available with many clinical examples of the application of clinical confocal microscopy to the cornea.<sup>22,23</sup> These books are replete with confocal images from a wide variety of normal and pathological corneas and cover the effects of photorefractive surgery on the structure of the cornea. Another source of clinical corneal images is the World Wide Web. I include two commercial sites of companies that manufacture clinical confocal instruments.<sup>11,20</sup> Their Web sites include many examples of confocal images of the cornea in a variety of normal and pathological conditions.

As an exemplar of the use of the clinical confocal microscopy to investigate new clinical findings the following study is presented.<sup>24</sup> This paper illustrates the confounding effect of acquiring clinical confocal images of the human cornea with inappropriate optical resolution. The authors posed the following question: what are the long-term effects of contact lens wear on the human cornea? They answered this question by first acquiring confocal images through the full thickness of the *in vivo* human cornea, and then by a frame by frame analysis to determine morphological changes within the cornea that are correlated with long-term contact lens wear and do not occur in subjects that do not wear contact lenses.

The scanning slit confocal microscope with a 50X/1.0 NA water immersion objective was used to investigate the corneal morphology in long-term contact lens wearers. The authors investigated 13 patients with a history of up to 26 years of soft contact lens wear, 11 patients with a history of up to 25 years of rigid gas permeable contact lens wear, and a control group of 29 normal subjects without a history of contact lens wear. For contact lens wearers epithelial microcystic changes and alterations of endothelial cell morphology were found as described previously. The significant new finding was there were highly reflective panstromal microdot (submicron) deposits in the entire thickness of the stroma for the contact lens wearers. It was concluded that this newly observed stromal microdot degeneration scales with the years of contact lens wear and may be the early stage of a significant corneal disease. Further correlative microscopic studies involved electron microscopy of *ex vivo* human corneas, and spectroscopic studies of the microdots. Initially, other groups were not able to observe the stromal microdot deposits because they used a Nipkow disk confocal microscope with a low power and a low NA microscope objective. The low power, low NA microscope objective did not

provide the necessary resolution to image the stromal microdot deposits. This study illustrates the requirement of the appropriate optical resolution in studies with the clinical confocal microscope, as well as correlative electron microscopy to validate the clinical conclusions.

## 17.7 PERSPECTIVES

In the linear domain confocal microscopes can be developed that use different colored light-emitting diodes as the light source. The degree of penetration of the incident light is a function of the wavelength and the amount of absorbing and scattering material in a particular cornea. The selection of various colors of illumination light would provide the clinician with improved image contrast since the illumination could be tuned to optimize either the penetration depth or the contrast.

The use of noncontact confocal microscopes such as the Heidelberg Retina Tomograph II with the Rostock Corneal Module is a major improvement over the previous designs with an applanating microscope objective that flattens the corneal surface. The noncontact confocal microscope is less likely to transmit microbes from patient to patient and causes less trauma to the cornea. The Nidek ConfoScan4 clinical confocal microscope also operates in a noncontact mode.

Another problem that still requires further development to reach a solution is a method to determine the position of the focal plane within the cornea. The current use of movable lenses together with accurate position sensors do not provide the true position of the focal plane within the cornea with respect to the corneal surface.

Typically, the technical specifications of commercial clinical confocal microscopes do not include information on the point-spread function, or other metrics of the transverse and the optical resolution of the microscope with a stated frequency of illumination, and a stated numerical aperture of the microscope objective. Measurements of the image, a series of optical sections, through a subdiffraction of gold particle or a 0.1- $\mu\text{m}$  fluorescent bead are readily made and would provide useful information.

Microscope images obtained through thick, scattering, and absorbing tissues such as a cornea containing scar tissues result in further degradation of the image quality. Even when a gel is used to approximately match the refractive index of the cornea and the microscope objective there are induced optical aberrations that degrade the image quality. For example, the refractive index varies within the cornea, both in depth and in the imaging plane of the microscope. One option to mitigate this problem is to incorporate adaptive optics (a closed loop-feedback system that corrects the wavefront distortions due to the cornea). Adaptive optics have successfully been incorporated into scanning-laser ophthalmoscopes for retinal imaging.<sup>25</sup>

All of the clinical confocal microscopes in use today are based on linear optics and generate contrast from the absorption and scattering of the incident light. There are many published studies of nonlinear higher-harmonic generation microscopy that were performed on *ex vivo* corneas, but there is a paucity of *in vivo* animal studies and no *in vivo* human studies have been reported.<sup>26</sup> Second- and third-harmonic generation can be explained by the theory of the nonlinear susceptibility in which for the time domain the polarization is expanded in a power series of the sum of products of the linear susceptibility and the electric field, the second-order susceptibility and the square of the electric field, and the third-order susceptibility and the cube of the electric field and higher-order terms. Second-harmonic generation (SHG) is described by the second-order susceptibility and third-harmonic generation (THG) is described by the third-order susceptibility. THG has a broad potential for tissue imaging since it occurs in all materials, including dielectric materials with inversion symmetry. In general THG is a weak process; however, it is dipole allowed. THG microscopy has an emerging potential for cell and tissue imaging. As new techniques are being developed for the surface-enhanced THG at interfaces, the technique may become useful for imaging the cornea. Another emerging development is the use of enhanced THG microscopy with nanogold particles to utilize the surface-plasmon-resonance effect.

Both SHG and THG imaging do not exhibit the saturation effects or the photobleaching effects that are associated with multiphoton excitation fluorescence microscopy. These microscopic techniques can

penetrate into millimeters of tissue and provide submicron three-dimensional optical sectioning. Collagen has a noncentrosymmetric structure and can convert the incident ultrashort laser pulse to its second harmonic, thus providing a noninvasive imaging modality for the cornea. The backward-scattered SHG signal may provide a sensitive clinical method to study corneal haze following photorefractive surgery.

Structures within the cornea, for example, collagen fibrils that are a few nanometers in diameter, can only be imaged with an electron microscope. When imaged with a confocal microscope or with a SHG or THG microscope the collagen fibrils, and all subdiffraction limit structures appear at the diffraction limit of the microscope. This optical phenomena is rarely explained in many papers on nonlinear microscopy of the cornea.

Finally, the image quality of a clinical corneal microscope is dependent on the microscope objective. Olympus Corporation has developed a new 25X, NA 1.05 water immersion microscope objective that is optimized for nonlinear microscopy based on femtosecond pulses of laser light. This lens that produced tightly focused light in the focal volume, a high collection efficiency due to the high NA, and a long free-working distance may enhance the performance of corneal microscope that operates in the nonlinear optical domain.

---

## **17.8 SUMMARY**

The goals of biomicroscopy of the living human cornea are to improve our understanding of the cellular structure and function in the normal cornea and in diseased, aging, and surgically altered corneas. Biomicroscopy of the anterior segment of the eye has provided the clinician with a series of improved instruments from the early slit lamps to a variety of confocal microscopes; with each new design there were improvements in both the image resolution and the contrast. Modern clinical confocal microscopes provide transverse resolution and axial resolution of several microns. Both the scanning slit confocal microscope and the laser scanning confocal microscope provide high-resolution images with high contrast and are ideal to investigate nerve regeneration in the postsurgical cornea, the formation of scar tissue, and other morphological anomalies due to disease, contact lens wear, overnight lid closure, or surgical procedures such as photorefractive surgery. Routine confocal microscopy of the cornea reveals the cellular and subcellular structure of all the cells in the full thickness of the cornea. Nonlinear microscopy, especially higher-harmonic generation microscopy has the potential for molecular imaging with the benefit that it is based on a scattering process and, therefore, there is no net energy deposition within the ocular tissue. Future clinical biomicroscopes can incorporate correlative microscopy, that is, both linear reflected confocal microscopy and nonlinear higher-harmonic microscopy.

---

## **17.9 ACKNOWLEDGMENTS**

The author and Professor M. Böhnke shared the 1999 Alfred Vogt-Prize for Ophthalmology (the highest award in Switzerland for scientific research in ophthalmology) from the Alfred Vogt-Stiftung zur Förderung der Augenheilkunde Zürich, for their work: "Confocal Microscopy of the Cornea." The author thanks Professor M. Böhnke, formerly of the Department of Ophthalmology, University of Bern, for many years of collaboration, and Dr. Andreas A. Thaeer for his collaboration in the development of the clinical confocal microscope.

---

## **17.10 REFERENCES**

1. B. R. Masters and P. T. C. So, *Handbook of Biomedical Nonlinear Optical Microscopy*, Oxford University Press, New York, 2008.

2. R. W. Boyd, *Nonlinear Optics*, 3rd ed., Academic Press, San Diego, CA, 2008.
3. B. R. Masters, *Selected Papers on Confocal Microscopy*, Milestone Series MS 131, SPIE Optical Engineering Press, Bellingham, WA, 1996.
4. T. R. Corle and G. S. Kino, *Confocal Scanning Optical Microscope and Related Imaging Systems*, Academic Press, San Diego, CA, 1996.
5. B. R. Masters, "David Maurice's Contributions to Optical Ophthalmic Instrumentation: Roots of the Scanning Slit Confocal Microscope," *Experimental Eye Research* **78**:315–326 (2004).
6. B. R. Masters, *Confocal Microscopy and Multiphoton Excitation Microscopy: The Genesis of Live Cell Imaging*, SPIE Optical Engineering Press, Bellingham, WA, 2006.
7. R. H. Webb, "Scanning Laser Ophthalmoscope," In B. R. Masters, (ed.), *Noninvasive Diagnostic Techniques in Ophthalmology*, Springer Verlag, New York, pp. 438–540, 1990.
8. B. R. Masters, *Noninvasive Diagnostic Techniques in Ophthalmology*, Springer Verlag, New York, 1990.
9. B. R. Masters and A. A. Thaeer, "Real-Time Scanning Slit Confocal Microscopy of the In Vivo Human Cornea," *Applied Optics*, **33**(4):695–701 (1994).
10. B. R. Masters, "Scanning Slit Confocal Microscopy of the In Vivo Cornea," *Optical Engineering* **34**(3):684–692 (1995).
11. <http://usa.nidek.com>; accessed May 26, 2009.
12. B. R. Masters and A. A. Thaeer, "In Vivo Human Corneal Confocal Microscopy of Identical Fields of Subepithelial Nerve Plexus, Basal Epithelial and Wing Cells at Different Times," *Microscopy Research and Technique* **29**:350–356 (1994).
13. B. R. Masters and A. A. Thaeer, "In Vivo, Real-Time Confocal Microscopy of the Continuous Wing Cell Layer Adjacent to the Basal Epithelium in the Human Cornea: a New Benchmark for In Vivo Corneal Microscopy," *Bioimages* **3**(1):7–11 (1995).
14. B. R. Masters and A. A. Thaeer, "Real-Time Confocal Microscopy of In Vivo Human Corneal Nerves," *Bioimages*, **4**(3):129–134 (1997).
15. B. R. Masters and M. Böhnke, "Three-Dimensional Confocal Microscopy of the Human Cornea in Vivo," *Ophthalmic Research* **33**(3):125–135 (2001).
16. B. R. Masters and M. Böhnke, "Three-Dimensional Confocal Microscopy of the Living Eye," *Annual Review of Biomedical Engineering*, Annual Reviews, Palo Alto, CA, **4**:69–91 (2002).
17. B. R. Masters and S. W. Paddock, "Three-Dimensional Reconstruction of the Rabbit Cornea by Confocal Scanning Optical Microscopy and Volume Rendering," *Applied Optics* **29**:3816–3822 (1990).
18. B. R. Masters and M. A. Farmer, "Three-Dimensional Confocal Microscopy and Visualization of the In Situ Cornea," *Computerized Medical Imaging and Graphics*, **17**(3):211–219 (1993).
19. B. R. Masters, "Three-Dimensional Confocal Microscopy of the Living In Situ Rabbit Cornea," *Optics Express* **3**(9):351–355 (1998).
20. <http://www.heidelbergengineering.com>; accessed May 26, 2009.
21. M. Böhnke and B. R. Masters, "Confocal Microscopy of the Cornea," *Progress in Retina & Eye Research* **18**(5): 553–628 (1999).
22. L. Mastropasqua and M. Nubile, *Confocal Microscopy of the Cornea*, SLACK Incorporated, Thorofare, NJ, 2002.
23. R. F. Guthoff, C. Baudouin, and J. Stave, *Atlas of Confocal Laser Scanning In-Vivo Microscopy in Ophthalmology*, Springer Verlag, Berlin, 2006.
24. M. Böhnke and B. R. Masters, "Long Term Contact Lens Wear Induces a Corneal Degeneration with Micro-dot Deposits in the Corneal Stroma," *Ophthalmology* **104**:1887–1896 (1997).
25. J. Porter, H. Queener, J. Lin, K. Thorn, and A. Awwal, (eds.), *Adaptive Optics for Vision Science*, Wiley-Interscience, Hoboken, NJ, 2006.
26. B. R. Masters, "Correlation of Histology and Linear and Nonlinear Microscopy of the Living Human Cornea," *Journal of Biophotonics* **2**:127–139 (2009).



*This page intentionally left blank*

---

# DIAGNOSTIC USE OF OPTICAL COHERENCE TOMOGRAPHY IN THE EYE

---

Johannes F. de Boer

*Department of Physics  
VU University, Amsterdam  
Rotterdam Ophthalmic Institute  
Rotterdam, The Netherlands*

---

## 18.1 GLOSSARY

---

**A-line.** Depth scan along the beam direction in an OCT measurement.

**Axial resolution.** Resolution along the beam direction in an OCT measurement.

**Birefringence.** Optical property of a material where the refractive index of the material depends on the polarization state of light.

**Bruchs membrane.** The innermost layer of the choroid, connected to the retinal pigmented epithelium.

**Capillary.** The smallest blood vessels, approximately 5 to 10  $\mu\text{m}$  in diameter.

**Coherence length.** The length along the beam path over which light is correlated.

**Diattenuation or dichroism.** Optical property of a material where the light absorption of the material depends on the polarization state of light.

**Fourier transform.** Operation that transforms one function of a variable into another, where the new function describes which frequencies are present in the original function.

**Fovea.** The area in the back of the inside of the eye located in the center of the macula, responsible for sharp central vision.

**Choroid.** The vascular layer of the eye lying between the retina and the sclera.

**Jones matrix formalism.** In the Jones matrix formalism the electromagnetic wave before and after interaction is described by Jones vectors with two complex-valued entries, and their transformation is described by the complex-valued  $2 \times 2$  Jones matrix.

**Michelson interferometer.** Most common configuration for optical interferometry where the source light is split by a beam splitter into a sample and a reference arm. Reflected light is combined in the detection arm.

**Mueller matrix formalism.** In the Mueller matrix formalism the polarization properties of light are described by Stokes vectors with four real-valued entries, and their transformation is described by the real-valued  $4 \times 4$  Mueller matrix.

**Optical circulator.** An optical circulator is a three-port device that allows light to travel in only one direction—from port 1 to port 2, then from port 2 to port 3.

**Optical coherence tomography (OCT).** Optical interferometric technique to create cross-sectional images of scattering media.

**Optical frequency domain imaging (OFDI) or swept source OCT (SS-OCT).** OCT configuration in which the reference arm length is stationary and the interference in the detection arm is spectrally resolved by scanning a narrow band source over a wavelength range.

**Optic nerve head.** The circular area in the back of the inside of the eye where the optic nerve connects to the retina.

**Phase retardation.** The retardation in phase that originates from the difference in refractive index for polarized light in a birefringent material.

**Poincaré sphere.** Graphical representation of the Stokes vector on a sphere.

**Relative intensity noise.** Noise associated with the instability in the power level of a laser or light source.

**Retinal pigmented epithelium.** The pigmented cell layer between the choroid and the photoreceptors.

**Semiconductor optical amplifier.** Semiconductor device that amplifies light by several orders of magnitude.

**Shot noise.** Fundamental noise generated by statistical fluctuations in a stream of particles (photons, electrons).

**Signal to noise ratio.** Ratio of a signal power to the noise power corrupting the signal.

**Spectral or Fourier domain OCT.** OCT configuration in which the reference arm length is stationary and the interference in the detection arm is spectrally resolved.

**Stokes vector.** Four element vector that describes the polarization properties of light.

**Temporal coherence.** The feature that light at different time points exhibit correlation.

**Time domain OCT.** OCT configuration in which the reference arm length is mechanically scanned.

**Unitary matrix.** A unitary matrix is an  $n$  by  $n$  complex matrix  $U$  satisfying the condition  $UU^* = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$ , where  $\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}$  is the identity matrix and  $U^*$  is the conjugate transpose (also called the Hermitian adjoint) of  $U$ .

**Vitreous humor.** The clear gel that fills the space between the lens and the retina.

**White noise.** Noise that has a flat frequency distribution (all frequencies are equally present in the noise).

## 18.2 INTRODUCTION

Optical coherence tomography (OCT) is a low coherence interferometric method for imaging of biological tissue.<sup>1,2</sup> OCT creates cross-sectional images of tissue by measuring the location (lateral and depth) where light is backscattered. OCT has become a very powerful imaging tool in ophthalmology, since it provides high resolution cross-sectional images of the retina and the cornea that approach the resolution of histology. The development of objective criteria to characterize ophthalmic diseases based on OCT images is an active area of research.

OCT is analogous to ultrasound imaging, where a sound pulse is sent into tissue and the time delay is measured of reflected sound waves. The time delay is converted to a distance using the propagation speed of sound in tissue. The speed of light ( $3 \times 10^8$  m/s) is 5 orders of magnitude higher

than the speed of sound ( $1.5 \times 10^3$  m/s), and therefore it is not possible to measure directly the small time delay of a reflected optical pulse. This time delay can be measured however by interfering the light reflected from the tissue with a reference in a Michelson type interferometer. This principle will be explained in more detail further.

For more than a decade after its inception, the dominant implementation of OCT has been time domain OCT (TD-OCT), in which the length of a reference arm in a Michelson type interferometer is rapidly scanned. TD-OCT has been used extensively in ophthalmology, where cross-sectional OCT images of the retina have provided useful information regarding the presence or progression of specific ocular diseases.<sup>3</sup> The acquisition rate of clinical and preclinical TD-OCT systems is limited by sensitivity and the maximum permissible incident power on the eye<sup>4</sup> to about 400 depth profiles/s, preventing comprehensive screening of large retinal areas. Over the past 5 years, the dominant implementation of OCT in ophthalmology has become spectral or Fourier domain OCT (SD/FD-OCT). As will be discussed, SD/FD-OCT has a significant sensitivity advantage over TD-OCT. The sensitivity improvement of SD-OCT allows for dramatically increased acquisition speeds by approximately a factor of 100 without compromising image quality. Three-dimensional data sets can therefore be rapidly acquired, opening the possibility of comprehensive screening. The first SD/FD-OCT implementation was reported in 1995.<sup>5</sup> In SD-OCT the reference arm is kept stationary, and the depth information is obtained by a Fourier transform of the spectrally resolved interference fringes measured with a spectrometer in the detection arm of a Michelson interferometer.

An alternative method to SD/FD-OCT that provides the same sensitivity advantage is optical frequency domain imaging (OFDI) or swept source OCT.<sup>6–9</sup> In OFDI, the spectrally resolved interference fringes are not measured by a spectrometer in the detection arm of a Michelson interferometer, but by rapidly tuning a narrowband source over a wavelength range. A detector in the reference arm measures the spectrally resolved interference as a function of time.

OCT images provide structural information based on the scattering properties of the tissue under investigation. OCT can be enhanced by functional extensions such as Doppler sensitivity to detect flow, and polarization sensitivity to detect birefringent structures (structures for which the refractive index depends on the polarization state of light). In this chapter, the aforementioned techniques will be discussed and explained in more detail.

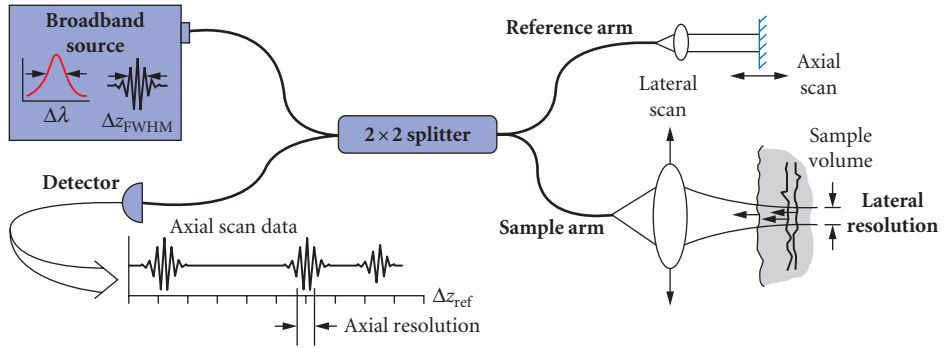
### 18.3 PRINCIPLE OF OCT: TIME DOMAIN OCT

Optical coherence tomography is a very sensitive interferometric method to measure the location of structures that reflect light with an axial resolution of 1 to 10  $\mu\text{m}$ . OCT exploits the low temporal coherence properties of a broadband light source. For a low temporal coherence light source, electromagnetic waves will not interfere when the time delay between two beams is larger than the coherence time. The product of the coherence time and the speed of light gives the coherence length. This property can be used to measure the path length difference between two arms of a Michelson interferometer. When the path length difference is larger than the coherence length, light will not interfere. The path length difference for which the two arms match in length within the coherence length of the source produces an interference pattern. Figure 1 shows the principle of an OCT system with a scanning mirror in the reference arm.

In the sample arm is a structure with three reflecting layers. As the reference arm mirror is translated, the detector in the detection arm measures an interference pattern each time the reference arm path length matches the path length to one of the reflecting layers in the sample arm. In OCT the envelope of the interference pattern is detected and a single scan of the reference arm mirror provides a single axial (along the beam direction) reflectivity profile of the sample. A single depth profile is also called an A-line. Repeating many depth scans (A-lines) at lateral locations creates a cross-sectional image of, for example, the retina, as shown in Fig. 2.

In a more mathematical description the intensity at the detector is given by

$$\langle I(\Delta z) \rangle = \langle I_r \rangle + \langle I_s \rangle + \langle E_r^* E_s(\Delta z) \rangle + \langle E_r E_s^*(\Delta z) \rangle \quad (1)$$



**FIGURE 1** Basic OCT configuration. Light from a broadband source is split into a sample and a reference arm. The reference arm mirror scans in depth, creating interference fringes when the reference arm path length matches the sample arm path length to one of the three reflecting layers in the sample.

where  $E_{r,s}$ ,  $E_{r,s}^*$  are the electric field component and its complex conjugate,  $\Delta z$  is the path length difference between the two arms of the interferometer, and subscripts  $r$  and  $s$  denote fields reflected from the reference and sample arm, respectively. The angular brackets denote time averaging. The first two terms on the right-hand side of Eq. (1) give the intensity reflected from the reference and sample arm, respectively.

The last two terms of Eq. (1) correspond to the interference between reference and sample arm light. In our analysis, the electric field amplitude is represented by a complex analytic function,<sup>10</sup>  $E(z)$ , with

$$E(z) = \int \tilde{e}(k) \exp(-ikz) dk \quad (2)$$

where  $\tilde{e}(k)$  is the field amplitude as a function of free space wave number  $k = 2\pi/\lambda$ , with

$$\tilde{e}(k) = 0 \quad \text{if} \quad k < 0 \quad (3)$$

From the Wiener-Khinchine theorem, it follows,

$$\langle \tilde{e}(k) \tilde{e}(k') \rangle = S(k) \delta(k - k') \quad (4)$$

which defines  $\tilde{e}(k)$  in terms of the source power spectral density  $S(k)$ . Using the Wiener-Khinchine theorem [Eq. (4)], the interference term in Eq. (1) is given by

$$I_{r,s}(z, \Delta z) = E_r^* E_s + E_r E_s^* \propto \sqrt{R(z)} \int \cos(2k\Delta z) S(k) dk \quad (5)$$



**FIGURE 2** OCT image of a human retina. Arrow indicates the position of a blood vessel. (Reproduced from Ref. 11 with permission from the Optical Society of America.)

with  $z$  the depth in the tissue and  $\Delta z$  the optical path length difference between sample and reference arms,  $\Delta z = z_r - z_s$ . We will assume a Gaussian power spectral density for the source

$$S(k) \propto \exp\left[-\left(\frac{k-k_0}{\kappa}\right)^2\right] \quad (6)$$

with the FWHM spectral bandwidth of the source given by  $\kappa 2\sqrt{\ln 2}$ . The integration over  $k$  in Eq. (5) can now easily be performed

$$I_{r,s}(z, \Delta z) \propto \sqrt{R(z)} \cos(2k_0 \Delta z) \exp[-(\Delta z / \Delta l)^2] \quad (7)$$

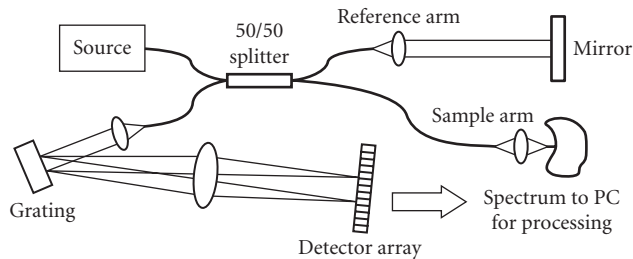
with the FWHM of the interference fringes envelope given  $\Delta l / 2\sqrt{\ln 2}$

$$\Delta l = \frac{1}{\kappa} = \frac{\lambda_0^2 \sqrt{\ln 2}}{\pi \Delta \lambda} \quad (8)$$

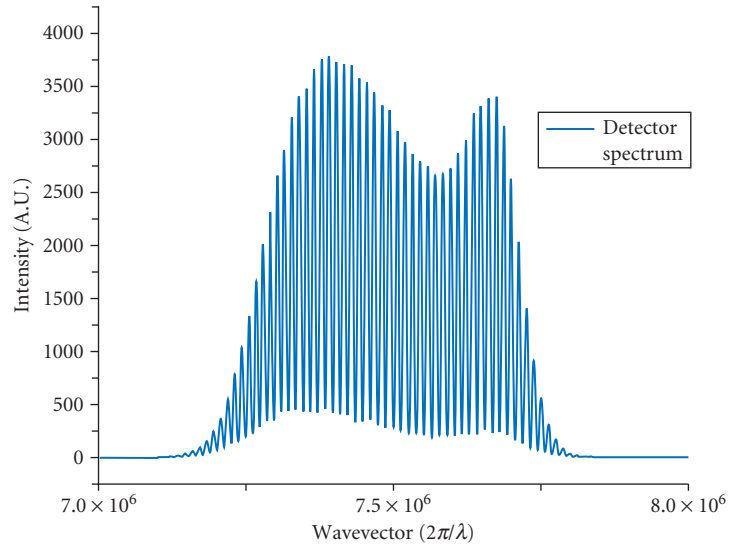
and  $\Delta \lambda$  the spectral FWHM of the source in wavelength. Equation (7) is an important result; it describes the interference fringe intensity that is observed for a reflective structure in the sample arm when the reference arm is scanned. The right-hand side of Eq. (7) consists of the reflected amplitude at depth  $z$  ( $\sqrt{R(z)}$ ), the Doppler shift or carrier frequency generated by the variation of the optical path length difference between sample and reference arm ( $\cos(2k_0 \Delta z)$ ), and the interference fringes envelope ( $\exp(-(\Delta z / \Delta l)^2)$ ), which describes the width of the interference pattern, respectively. The depth resolution is determined by the width  $\Delta l$  of the interference pattern, and Eq. (8) shows the relation between axial or depth resolution and the spectral width of the source in OCT. The broader the source (larger  $\Delta \lambda$ ), the smaller the coherence length  $\Delta l$ , that is, the better the axial resolution. The availability of sources with very broad spectral widths over the past decade has increased the axial (depth) resolution of OCT from 15  $\mu\text{m}$  to approximately 2 to 3  $\mu\text{m}$ .<sup>12</sup> The lateral resolution in OCT is determined by the lateral size of the focused spot, in analogy to confocal microscopy.

## 18.4 PRINCIPLE OF OCT: SPECTRAL DOMAIN OCT

Spectral domain OCT is an alternative implementation of the time domain OCT technology. The first SD/FD-OCT implementation was reported in 1995.<sup>5</sup> Figure 3 shows the basic configuration of an SD/FD-OCT system.



**FIGURE 3** Basic configuration of an SD/FD-OCT system. Light from a broadband light source is split into a sample and a reference arm. The mirror in the reference arm is stationary. After reflection, the sample and reference arm light is recombined and interferes. In the detection arm, the light is dispersed by a grating and each wavelength is focused on an element in a detector array. (Reproduced from Ref. 13 with permission from the Biophysical Society.)



**FIGURE 4** Spectrally resolved interference pattern as detected by the spectrometer in the detection arm of an SD-OCT system with a single reflector in the sample arm. (Reproduced from Ref. 13 with permission from the Biophysical Society.)

The source light is split into a reference and sample arm. At recombination in the detection arm, the light interferes. In the detection arm, the light is spectrally dispersed by a grating, and each wavelength is focused on an element in a linear detection array by a lens. The path length difference between the sample and reference arm is determined by the spectrally resolved interference detected by the linear detection array. In Fig. 4, an example of such a spectrally resolved interference pattern is shown for a single reflector in the sample arm. The power reflected from the sample and reference arm was approximately equal, giving nearly full constructive and destructive interference fringes.

The interference pattern shown in Fig. 4 can be understood as follows: for wavelengths for which the path length difference between sample and reference arm is exactly a multiple of the wavelength we observe constructive interference, and for wavelengths for which the path length difference between sample and reference arm is  $(N+0.5)\lambda$  we observe destructive interference. Mathematically, the interference spectrum is described by

$$I(k) = I_r(k) + I_s(k) + 2\sqrt{I_s(k)I_r(k)} \sum_n \alpha_n \cos(kz_n) \quad (9)$$

where  $I_r(k)$  and  $I_s(k)$  are the wavelength-dependent intensities reflected from reference and sample arms, respectively, and  $k$  is the wave number. The last term on the right-hand side of Eq. (9) represents the interference between light returning from the reference and sample arms.  $\alpha_n$  is the square root of the sample reflectivity at depth  $z_n$ , and  $z_n$  is the path length difference between sample and reference arm. The summation sums over all reflective structures at location  $z_n$  in the sample arm. Equation (9) shows that the path length difference is encoded on the spectrum by the modulation term  $\cos(kz_n)$ . The modulation frequency as a function of wave vector is determined by the path length difference  $z_n$ . The larger the path length difference, the higher the modulation frequency. This also illustrates that the maximum path length difference (depth range) that can be measured is determined by the ability to measure high modulation frequencies as a function of wave vector. The spectral resolution of the spectrometer plays a crucial role in measuring these high modulation frequencies. In SD-OCT, the depth range  $z$  is inversely proportional to the spectral resolution  $\delta\lambda$  of the spectrometer and given by<sup>14</sup>

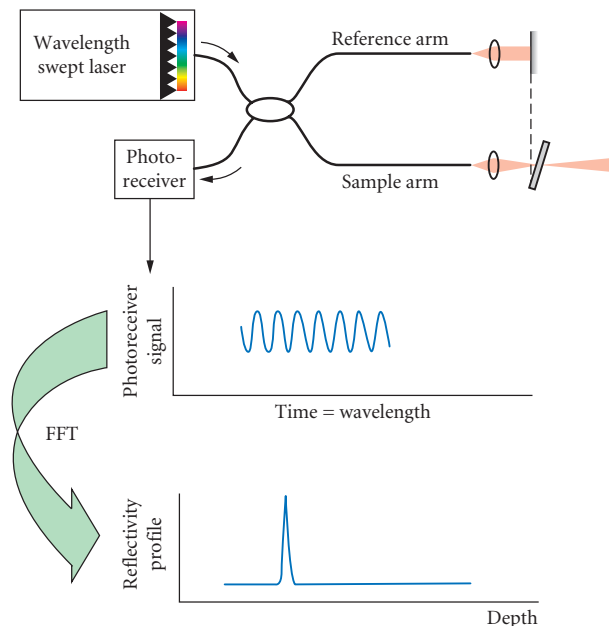
$$z = \lambda_0^2 / 4n\delta\lambda \quad (10)$$

A depth profile or depth scan is obtained from the spectrum through a Fourier transformation [see Eq. (18) later in this chapter].

SD-OCT provides a significant advantage in sensitivity or equivalently signal to noise ratio (SNR), that despite reports as early as 1997<sup>14,15</sup> has taken about half a decade to be recognized fully by the OCT community.<sup>7,16,17</sup> The first demonstration of SD-OCT for in vivo retinal imaging in 2002<sup>18</sup> was followed by a full realization of the sensitivity advantage by video rate in vivo retinal imaging,<sup>10</sup> including high speed three-dimensional volumetric imaging,<sup>19</sup> ultra high resolution video rate imaging,<sup>20,21</sup> and Doppler blood flow determination in the human retina.<sup>22,23</sup> The superior sensitivity of SD-OCT, combined with the lack of need for a fast mechanical scanning mechanism, has opened up the possibility of much faster scanning without loss of image quality and provided a paradigm shift from point sampling to volumetric mapping of biological tissue in vivo. The technology has been particularly promising in ophthalmology.<sup>24,25</sup> Later in this chapter a detailed analysis of the sensitivity advantage of SD-OCT is presented.

## 18.5 PRINCIPLE OF OCT: OPTICAL FREQUENCY DOMAIN IMAGING

In SD-OCT the spectrally resolved interference fringes are detected by a spectrometer in the detection arm. The same information can be obtained by scanning a narrowband source over a large wavelength range, and detecting the light with a single detector. In this case, the wavelength is encoded as a function of time. This technique, optical frequency domain imaging (OFDI) or swept source OCT (SS-OCT), gives the same sensitivity advantage as SD-OCT.<sup>7,8</sup> Figure 5 gives a basic

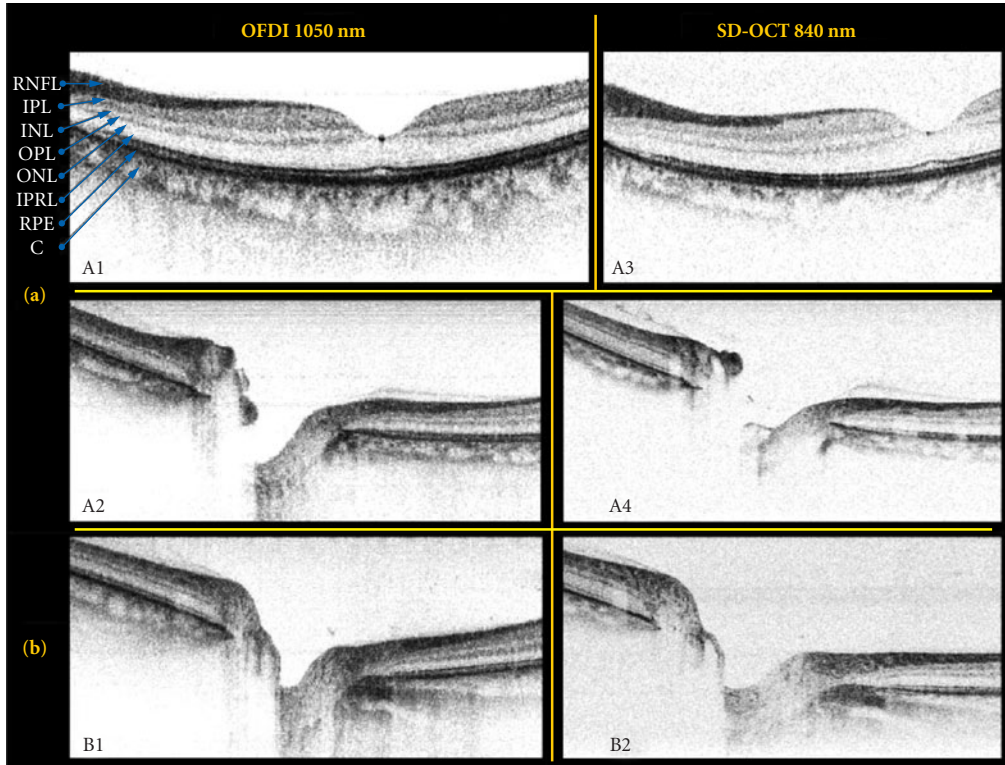


**FIGURE 5** Basic OFDI configuration. A narrowband light source rapidly tunes over a wavelength range. Sample and reference arm light interferes in the detection arm. The light is detected by the photoreceiver as a function of time, where time is equivalent to wavelength or wave vector, giving the spectrally resolved interference fringes. A Fourier transform gives a reflectivity profile of the sample.



configuration of an OFDI or SS-OCT system. A narrowband laser source rapidly scans over a large wavelength range. The photo receiver detects the interference between sample and reference arm light as a function of time, where time is equivalent to wavelength or wave vector. A single wavelength scan gives the wavelength-resolved interference, which after a Fourier transform provides a single reflectivity profile of the sample.

OFDI has been demonstrated at 1300 nm in a variety of tissues. In ophthalmology, OFDI has been demonstrated at a wavelength of 1050 nm<sup>26</sup> and 840 nm.<sup>27</sup> Particularly 1050 nm is an interesting wavelength for ophthalmology. It is well known that longer wavelengths penetrate deeper into biological tissue. This suggests that 1300 nm would provide deeper penetration in retinal tissue; however, the water absorption of 1300 nm is too large to image the retina through the vitreous humor. At 1050 nm there is a window where the water absorption is lower than at 1300 nm. The better penetration of 1050 nm over 800 nm into the retinal layers and the choroid is demonstrated in Fig. 6 and Ref. 28 in a comparison between an OFDI system at 1050 nm and an SD-OCT system at 840 nm. The better visualization below of the retinal pigmented epithelium layer and into the choroid could be particularly useful for the diagnosis of age-related macular degeneration (AMD).<sup>29</sup>



**FIGURE 6** Comparison of two imaging systems (OFDI at 1050 nm and SD-OCT at 840 nm). **A1** and **A2**: OFDI images at fovea and optic nerve head, respectively, from volunteer A, 36-year-old Asian male. **A3** and **A4**: SD-OCT images from the same volunteer at similar tissue locations. **B1** and **B2**: OFDI and SD-OCT images, respectively, obtained from volunteer B, 41-year-old Caucasian male. OFDI images exhibit considerably deeper penetration in tissue than SD-OCT images in all the data sets. The OFDI image (**A1**) shows the anatomical layered structure: RNFL, retinal nerve fiber layer; IPL, inner plexiform layer; INL, inner nuclear layer; OPL, outer plexiform layer; ONL, outer nuclear layer; IPRL, interface between the inner and outer segments of the photoreceptor layer; RPE, retinal pigmented epithelium; and C, choriocapillaris and choroid. (Reproduced from Ref. 26 with permission from the Optical Society of America.)

## 18.6 SD-OCT VERSUS OFDI

The best choice for a particular OCT technology in ophthalmology depends on a number of factors. Currently, SD-OCT is the preferred implementation at wavelengths around 800 nm, due to the availability of high performance silicon array detectors and the difficulty to create robust rapid tunable sources in this wavelength range. At wavelengths of 1050 and 1300 nm, OFDI is the preferred implementation due to the availability of a key component for a rapid tunable source at these wavelengths, a high performance semiconductor optical amplifier (SOA), and the relatively poor performance of InGaAs detector arrays. Other considerations that play a role are the relative ease with which to construct a broadband light source compared to a broadband rapid tunable source, favoring SD-OCT. On the other hand, SD-OCT suffers from sensitivity decay as a function of depth, limiting the effective depth range to about 1 mm, that will be detailed further below. Experimental OFDI systems have in general demonstrated an effective depth range that is 2 to 4 times better than SD-OCT. In cases where a large effective depth range is required, for example, imaging the cornea, or imaging patients with deep optic nerve head cupping, OFDI has an advantage.

## 18.7 SENSITIVITY ADVANTAGE OF SD-OCT OVER TD-OCT

In the standard time-domain (TD) implementation of OCT, the position of the reference mirror in the interferometer is rapidly scanned by mechanical means in order to obtain a depth profile (A-line) within a sample. In SD/FD-OCT, no mechanical scanning of the reference arm is required. Instead, the cross-spectral density at the detection arm of the interferometer is measured by means of a spectrometer.<sup>5,14,30</sup> Although this method has long been proposed and demonstrated, only recently have there been efforts to explicitly show that SD-OCT can produce a better detection sensitivity than the time domain method.<sup>7,16,17</sup> SD-OCT does not require scanning of the reference arm length and therefore has the potential for faster image acquisition rates. High speed line scan cameras in the spectrometer with an integration time as small as 34 and even 16  $\mu$ s permit the acquisition of A-lines at a rate of 30 to 60 kHz. Recent work has experimentally demonstrated a 148-fold (21.7 dB) sensitivity improvement of SD-OCT<sup>10</sup> and a near shot noise limited performance. Shot noise is a fundamental noise source that cannot be improved upon.

In essence the SNR advantage of SD-OCT over TD-OCT is based on the significant reduction of noise obtained by replacing the single-element detector with a multielement array detector. In a TD-OCT system, each wavelength is uniquely encoded as a frequency, and shot noise has a white noise characteristic. In a single detector TD-OCT system, the shot noise generated by the power density at one particular wavelength is present at all frequencies, and therefore adversely affects the SNR at all other wavelengths. By spectrally dispersing each wavelength to a separate detector, the cross shot noise term is eliminated in both hybrid and fully parallel SD-OCT systems.<sup>17</sup>

## 18.8 NOISE ANALYSIS OF SD-OCT USING CHARGE COUPLED DEVICES (CCDs)

The core of an SD-OCT system is the spectrometer in the detection arm. In general, light detection in the spectrometer is achieved by a charge coupled device (CCD) like a line array. Alternative implementations have been suggested that require individual processing of pixel charges.<sup>31</sup> Here we will treat the noise analysis based on a simple integrating CCD. To facilitate the noise analysis in the

case of detection by CCDs, the signal and noise terms will be expressed in charge squared ( $e^2$ ). In SD-OCT the signal  $S_{SD}$  given by,<sup>16,17</sup>

$$S_{SD} = \frac{\eta e^2 P_{ref} P_{sample} \tau_i^2}{E_v^2} (e^2) \quad (11)$$

with  $\eta$  the quantum efficiency of the detector,  $e$  the electron charge,  $P_{ref}$  and  $P_{sample}$ , respectively, the reference arm and sample arm power at the detection arm fiber tip,  $\tau_i$  the integration time, and  $E_v$  the photon energy.

In the noise analysis, it is generally assumed that the reference arm power is much larger than the sample arm power, and the shot noise and relative intensity noise (RIN) contributions of the sample arm power to the noise are neglected.<sup>32</sup> The read-out and dark noise, shot noise, and RIN contributions to the overall noise in electrons squared per read-out cycle are then given by, respectively,<sup>16,17</sup>

$$\sigma_{noise}^2 = \sigma_{r+d}^2 + \frac{\eta e^2 P_{ref} \tau_i}{E_v} + \left( \frac{\eta e P_{ref}}{E_v} \right)^2 \tau_i \tau_{coh} (e^2) \quad (12)$$

with  $\sigma_{r+d}^2$  the sum of read-out noise and dark noise, and  $\tau_{coh} = \sqrt{2 \ln 2} / \pi \lambda_0^2 / c \delta \lambda$  the coherence time, with  $c$  the speed of light.<sup>33</sup> A close look at Eq. (12) shows that the noise consists of a constant term ( $\sigma_{r+d}^2$ ), a term that linearly increases with the reference arm power (shot noise) and a term that quadratically increases with the reference arm power (RIN). The signal [Eq. (11)] increases linearly with the reference arm power. The signal to noise ratio (SNR) determines the performance of the system. Since the shot noise is the limiting noise term that cannot be improved upon, the optimal signal to noise performance is achieved when shot noise dominates both read-out noise and RIN.<sup>32</sup> When shot noise dominates the noise, both the signal [Eq. (11)] and the noise [Eq. (12)] increase linearly with the reference arm power and the SNR is constant. Shot noise dominates read-out noise and dark noise when  $\eta e^2 P_{ref} \tau_i / \sigma_{r+d}^2 E_v > 1$ , and shot noise dominates RIN when their ratio is larger than 1, that is,  $E_v / \eta P_{ref} \tau_{coh} > 1$ . The optimal reference arm power is found when read-out noise and dark noise are equal to the RIN.<sup>34</sup>

$$\sigma_{r+d}^2 = \left( \frac{\eta e P_{ref}}{E_v} \right)^2 \tau_i \tau_{coh} \quad (13)$$

Thus for a system to operate close to shot-noise-limited performance, shot noise should dominate thermal and RIN noise at the optimal reference arm power

$$P_{ref} = \frac{\sigma_{r+d} E_v}{\eta e \sqrt{\tau_i \tau_{coh}}} \quad (14)$$

At this optimal reference arm power, the inequalities describing shot noise dominance over read-out noise and RIN, respectively, reduce to the same equation

$$\frac{e \sqrt{\tau_i}}{\sigma_{rd} \sqrt{\tau_{coh}}} > 1 \quad (15)$$

In general, one would like to choose the integration time  $\tau_i$  as short as possible, and the coherence time  $\tau_{coh}$  as long as possible. The coherence time is inversely related to the spectral resolution of the spectrometer which in turn relates linearly to the maximum depth range of the system. In conclusion, the parameter that most determines the system performance is the read-out and dark noise of the detector  $\sigma_{rd}$ .

## 18.9 SIGNAL TO NOISE RATIO AND AUTOCORRELATION NOISE

In the shot noise limit, the following expressions are found for the SNR in the time domain<sup>32</sup> and the spectral domain,<sup>17</sup> where the expression for the spectral domain can also be derived by the ratio of Eqs. (11) and (12)

$$\text{SNR}_{\text{TD}} = \frac{\eta P_{\text{sample}}}{E_v \text{BW}} \quad \text{SNR}_{\text{SD}} = \frac{\eta P_{\text{sample}} \tau_i}{E_v} \quad (16)$$

where  $\eta$  is the spectrometer efficiency,  $P_{\text{sample}}$  is the sample arm power returning to the detection arm, BW is the electronic detection bandwidth in a time domain system,  $\tau_i$  is the detector integration time, and  $E_v$  is the photon energy. The electronic signal bandwidth BW is centered at the carrier frequency given by the Doppler shift [see Eq. (7)] and the bandwidth is proportional to the spectral width of the source and the velocity of the reference arm mirror. A detailed numerical comparison between the SNR in the time and spectral domain showed a more than 2 orders better SNR in the spectral domain.<sup>10</sup> Unlike the SNR in the time domain, Eq. (16) demonstrates that in the shot noise limit,  $\text{SNR}_{\text{SD}}$  is independent of the spectral width of the source. This implies that the axial resolution can be increased at no penalty to the SNR, provided that the full spectral width of the source can be imaged onto an array detector. However, this result should be interpreted with some care. The sample arm power returning to the detection arm is assumed to come from a single reflecting surface. In tissue, however, the reflected power comes from multiple structures along a depth profile. The SNR for a particular position along the depth profile is given on average by the total power reflected by all structures within the coherence length of the source. As the resolution increases (the coherence length decreases), the total reflected power within the coherence length decreases. As a consequence, the SNR at a particular position along the depth profile will reduce as the resolution increases by increasing the source optical bandwidth.

In SD-OCT the depth information is obtained by a Fourier transform of the spectrally resolved interference fringes. The detected interference signal at the spectrometer may be expressed as<sup>14</sup>

$$I(k) = I_r(k) + I_s(k) + 2\sqrt{I_s(k)I_r(k)} \sum_n \alpha_n \cos(kz_n) \quad (17)$$

where  $I_r(k)$  and  $I_s(k)$  are the wavelength-dependent intensities reflected from reference and sample arms, respectively, and  $k$  is the wave number. The last term on the right-hand side of Eq. (17) represents the interference between light returning from reference and sample arms.  $\alpha_n$  is the square root of the sample reflectivity at depth  $z_n$ . Depth information is retrieved by performing an inverse Fourier transform of Eq. (16), yielding the following convolution<sup>14</sup>

$$\left| \text{FT}^{-1}[I(k)] \right|^2 = \Gamma^2(z) \otimes \left\{ \delta(0) + \sum_n \alpha_n^2 \delta(z - z_n) + \sum_n \alpha_n^2 \delta(z + z_n) + O\left[ I_s^2 / I_r^2 \right] \right\} \quad (18)$$

with  $\Gamma(z)$  representing the envelope of the coherence function. The first term in the braces on the right-hand side describes the autocorrelation signal from the reference arm and has magnitude unity. The second and third terms are due to interference between light returning from reference and sample arms and form two images, where each has magnitude on the order of  $I_s/I_r$ . These two terms provide mirror images, where one is retained. The final term, with magnitude on the order of  $I_s^2/I_r^2$ , describes autocorrelation noise due to interference within the sample arm.<sup>14,18</sup>  $I_s$  and  $I_r$  represent the total intensity reflected from sample and reference arm, respectively. Sample autocorrelation noise is generated by the interference of light reflected at different depth locations in the sample. Equation (18) indicates that the relative contribution of sample autocorrelation noise can be reduced by increasing the reference arm power with respect to the signal. Decreasing the detector

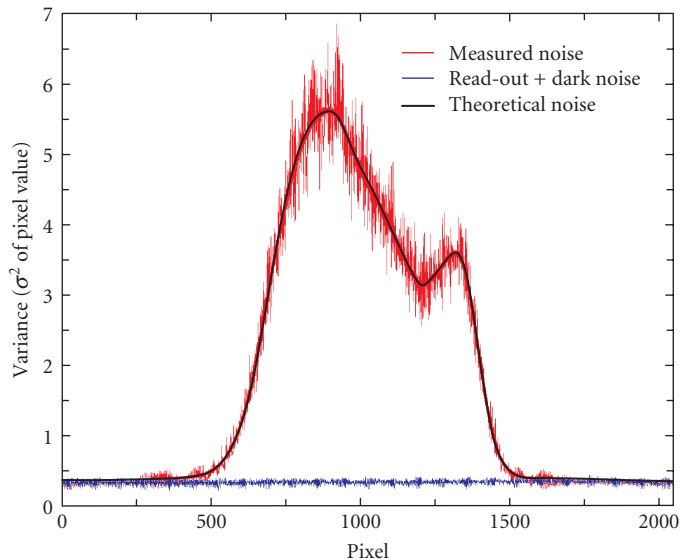
integration time permits an increase in the reference arm power without saturating the detector, decreasing the ratio  $I_s^2/I_r^2$  and consequently reducing the contribution of autocorrelation noise in ultrahigh-speed SD-OCT.

## 18.10 SHOT-NOISE-LIMITED DETECTION

In this section we take a closer look at the different noise components described in Eq. (12) for an actual system. In an SD-OCT system the sample arm was blocked, and only the reference arm light was detected by the spectrometer. One thousand spectra were recorded at an acquisition rate of 29.3 kHz. The read-out and shot noise are shown in Fig. 7.

The noise was determined by calculating the variance at each camera pixel over 1000 consecutive spectra. Dark noise measurements were taken with the source light off. Only light returning from the reference arm was used to measure the shot noise and RIN in the system. The shot noise and RIN are given by the second and third term on the right-hand side of Eq. (12), expressed in number of electrons squared. Taking into account only the shot and dark + read-out noise, the measured variance is proportional to

$$\sigma^2(\lambda) \sim \frac{P_{\text{ref}}(\lambda)}{E_v} + \sigma_{r+d}^2 \quad (19)$$

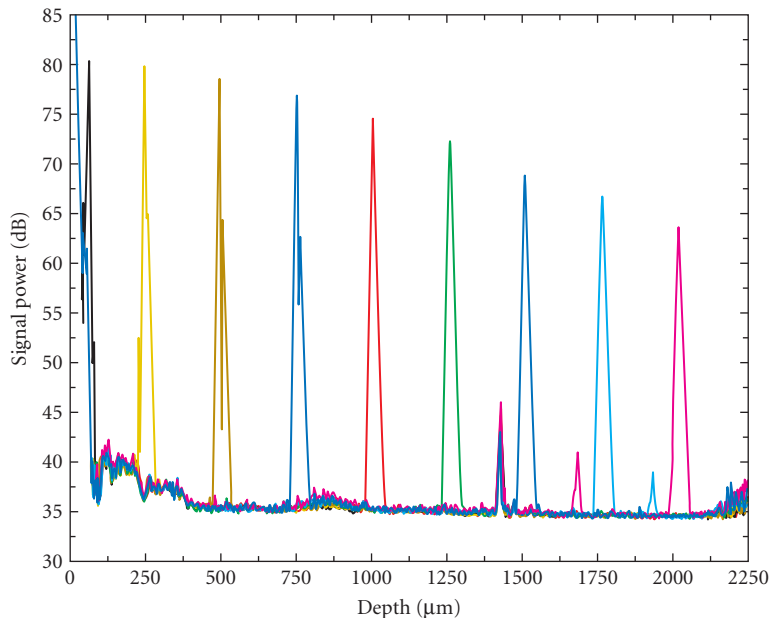


**FIGURE 7** Noise analysis of an SD-OCT system. The variance (vertical axis) was calculated as a function of wavelength (pixel number, horizontal axis) for 1000 consecutive spectra. Both the read-out + dark noise (blue curve, no illumination) and the shot noise (red curve, illumination by reference arm only) were determined. The theoretical shot noise curve was fit using Eq. (19) to the measured noise. The excellent fit of the theoretical shot noise curve to the measured noise demonstrates that RIN was significantly smaller than shot noise and did not contribute. (Reproduced from Ref. 19 with permission from the Optical Society of America.)

The first term on the right-hand side of Eq. (19) is the shot noise contribution which is linearly proportional to the reference arm power, and the second term is the dark and read-out contribution to the noise. Equation (19) was fit to the measurements, limiting the fit to the central 700 pixels. Relative intensity noise (RIN) was not dominant in this setup, as experimentally demonstrated by the excellent fit of only the shot noise and dark and read-out noise to the measured noise in Fig. 7, and theoretically since the maximum power per pixel (4.6 nW) at a 34.1  $\mu\text{s}$  integration time does not meet the criteria for RIN dominated noise.<sup>10</sup> This demonstrates shot-noise-limited performance of an SD-OCT system.

## 18.11 DEPTH DEPENDENT SENSITIVITY

In SD-OCT, signal sensitivity is strongly dependent on depth within an image. To characterize system sensitivity as a function of ranging depth, 1000 A-lines were acquired at an acquisition speed of 34.1  $\mu\text{s}/\text{A-line}$  for 9 different positions of a weak reflector in the sample arm. The reflected sample arm power was 1.18 nW for all reflector positions. The noise floor decayed by 5 dB between a depth of 500  $\mu\text{m}$  and 2 mm, and the peak signal dropped by 21.7 dB over the first 2 mm. Due to fixed-pattern noise, the true noise floor could not be determined between 0 and 500  $\mu\text{m}$ . After zero-padding before the FFT to correct wavelength-mapping errors from pixel to wave vector, a 16.7 dB loss in peak signal was noted across the first 2 mm, whereas the noise level dropped by only 0.4 dB between 500  $\mu\text{m}$  (35.1 dB) and 2 mm (34.7 dB) (Fig. 8). The zero-padding method produced a nearly constant noise level and improved the signal by more than 5 dB at the greatest depths in the scan. Although zero-padding did not change the local SNR, this method eliminated the shoulders that are present at larger scan depths.<sup>7</sup> The decay in both the signal and the noise level across the entire scan



**FIGURE 8** The depth dependent loss in signal sensitivity from a weak reflector. The signal decayed 16.7 dB between 0 and 2 mm. The peaks at 1.4, 1.6, and 1.85 mm are fixed-pattern noise. (Reproduced from Ref. 19 with permission from the Optical Society of America.)

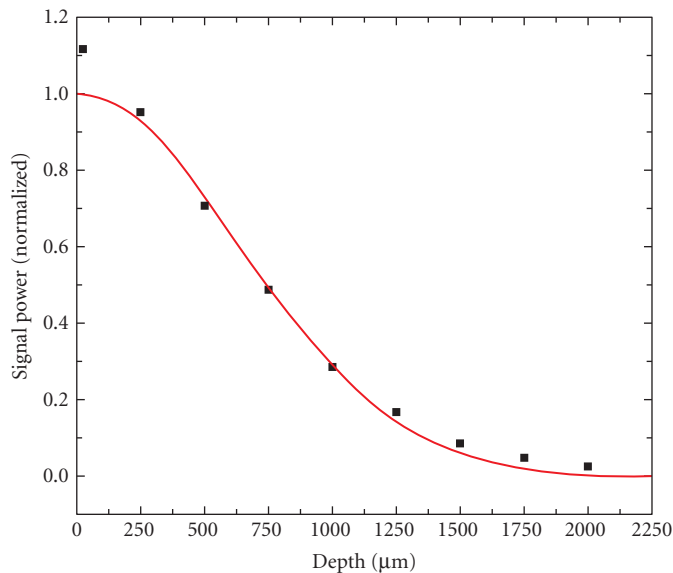
length of 2.4 mm have been theorized to amount to 4 dB as a result of the finite pixel width.<sup>16</sup> As demonstrated by the experimental data, the noise level decayed by less than 4 dB over the entire scan length, which we attribute to the statistical independence of the shot noise between neighboring pixels of the array. Thus, the finite pixel width does not introduce a decay of the noise level.

The finite spectrometer resolution introduces a sensitivity decay<sup>34</sup> similar to that introduced by the finite pixel size.<sup>16</sup> Convolution of the finite pixel size with the Gaussian spectral resolution yields the following expression for the sensitivity reduction,  $R$ , as a function of imaging depth,  $z$ <sup>34</sup>

$$R(z) = \frac{\sin^2(\pi z/2d)}{(\pi z/2d)^2} \exp\left[-\frac{\pi^2 \omega^2}{8 \ln 2} \left(\frac{z}{d}\right)^2\right] \quad (20)$$

where  $d$  is the maximum scan depth and  $\omega$  is the ratio of the spectral resolution to the sampling interval. Equation (20) was fit to the signal decay data presented in Fig. 8 with  $\omega$  as a free parameter, and the result is shown in Fig. 9. Due to its proximity to the autocorrelation peak, the first data point was not included in the fit. The value for  $\omega$  obtained from the fit was 1.85, demonstrating that the working spectral resolution was 0.139 nm.

The SNR was determined by the ratio of the peak at 250  $\mu\text{m}$  (79.8 dB) and the noise level. Due to the fixed-pattern noise at 250  $\mu\text{m}$ , the noise level was determined to be 35.2 dB by extrapolation of the linear region between 0.5 and 2 mm. The resulting SNR of 44.6 dB for 1.18 nW returning to the detection arm was 2.2 dB below the theoretical value given by Eq. (16) of 46.8 dB, for an integration time of 34.1  $\mu\text{s}$ , a central wavelength of 840 nm, and a spectrometer efficiency of 28 percent. With 600  $\mu\text{W}$  of power incident on an ideal reflector in the sample arm, the measured power returning to the detection arm was 284  $\mu\text{W}$ . The sum of the SNR at 1.18 nW (44.6 dB) and the 10 log ratio of maximum (284  $\mu\text{W}$ ) over measured (1.18 nW) power (53.8 dB) gives a sensitivity of 98.4 dB.



**FIGURE 9** Decay of sensitivity across the measurement range. Symbols: Peak intensities of data presented in Fig. 8. Solid line: Fit of Eq. (20) to the data points. (Reproduced from Ref. 19 with permission from the Optical Society of America.)

## 18.12 MOTION ARTIFACTS AND FRINGE WASHOUT

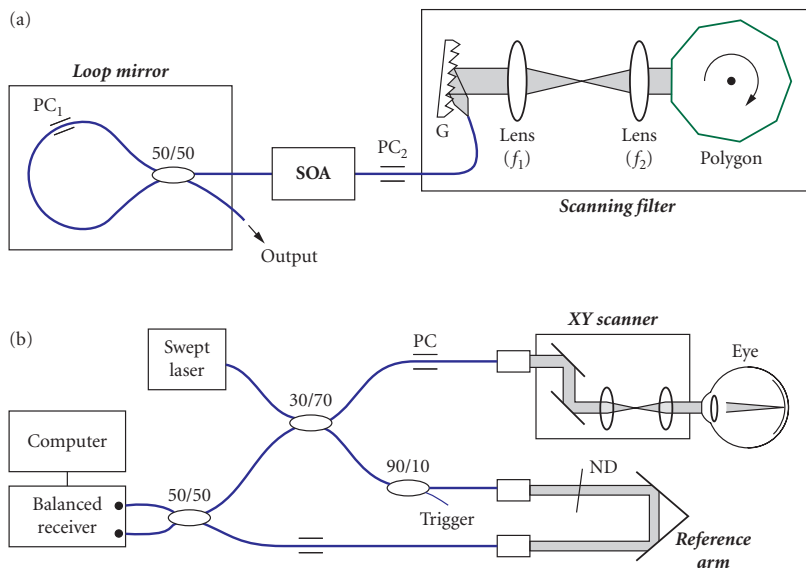
As OCT utilizes lateral point-scanning, motion of the sample or scanning beam during the measurement causes SNR reduction and image degradation in SD-OCT and OFDI.<sup>35</sup> Yun et al. theoretically investigated axial and lateral motion artifacts in continuous wave (CW) SD-OCT and swept-source OFDI, and experimentally demonstrated reduced axial and lateral motion artifacts using a pulsed source and a swept source in endoscopic imaging of biological tissue.<sup>35,36</sup> Stroboscopic illumination in full field OCT was demonstrated, resulting in reduced motion artifacts for in vivo measurement.<sup>37</sup> In ophthalmic applications of SD-OCT, SNR reduction caused by high speed lateral scanning of the beam over the retina may be dominant over axial patient motion. Using pulsed illumination reduces lateral motion artifacts and provides a better SNR for in vivo high-speed human retinal imaging.<sup>38</sup>

## 18.13 OFDI AT 1050 NM

An alternative technique to SD-OCT is optical frequency domain imaging (OFDI).<sup>8</sup> In OFDI, a rapidly tuned laser source is used and the spectrally resolved interference fringes are recorded as a function of time in the detection arm of the interferometer. Published results in healthy volunteers have shown that OFDI has better immunity to sensitivity degradation due to lateral and axial eye motion, and has an effective ranging depth that is 2 to 2.5 times better than SD-OCT (depth dependent sensitivity decay of 6 dB over 2 to 2.5 mm).<sup>8,26,27</sup>

More importantly, recent research at 1050 nm spectral range has demonstrated a better retinal penetration depth,<sup>26,28,39</sup> particularly important for detecting retinal abnormalities at or below the retinal pigment epithelium (RPE). A wavelength of 1050 nm has less attenuation from scattering in opaque media, commonly seen in cataract patients.<sup>40</sup> Although the water absorption at 1050 nm is higher than in the 850 nm region, this is partially compensated by the approximately 3 times higher maximum permissible exposure according to the ANSI standards (1.9 mW at 1050 nm).<sup>4</sup>

Figure 10a depicts a schematic of a laser source for OFDI in a linear cavity configuration.<sup>8,26</sup> The gain medium was a commercially available, bi-directional semiconductor optical amplifier



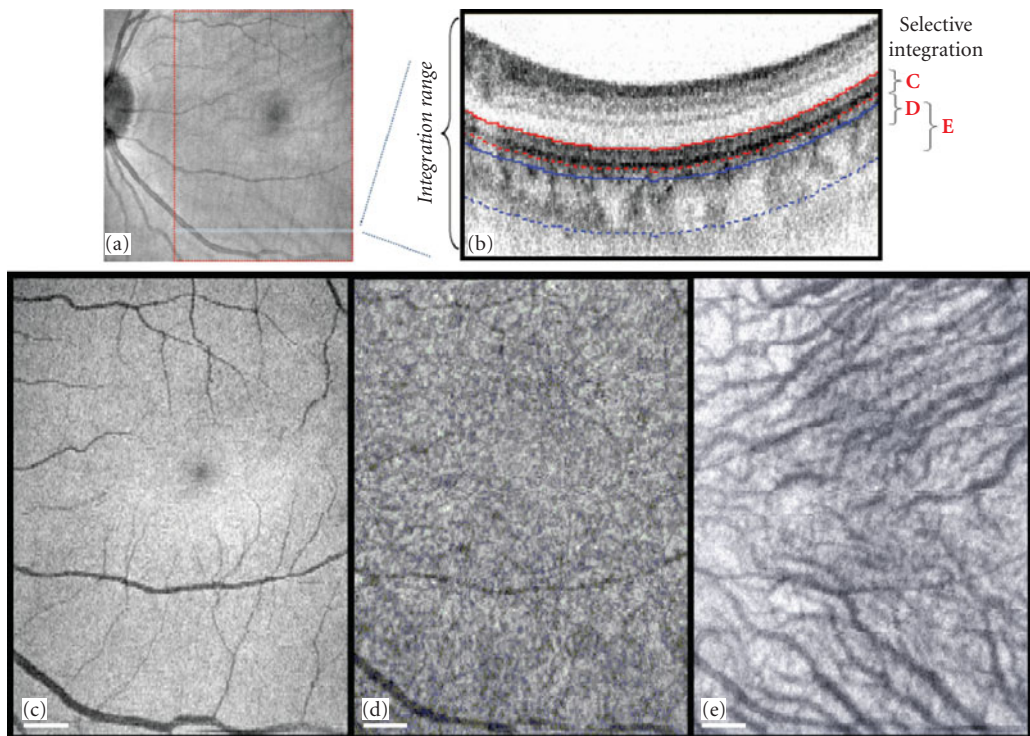
**FIGURE 10** Experimental setup: (a) wavelength-swept laser and (b) OFDI system. (Reproduced from Ref. 26 with permission of the Optical Society of America.)



(QPhotonics, Inc., QSOA-1050) driven at an injection current level of 400 mA. One port of the amplifier was coupled to a wavelength-scanning filter<sup>41</sup> that comprises a diffraction grating (1200 lines/mm), a telescope ( $f_1 = 100$  mm,  $f_2 = 50$  mm), and a polygon mirror scanner (Lincoln Lasers, Inc., 40 facets). The design bandwidth and free spectral range of the filter were approximately 0.1 and 61 nm, respectively. The amplifier's other port was spliced to a loop mirror made of a 50/50 coupler. Sweep repetition rates of up to 36 kHz were possible with 100 percent duty cycle.

Figure 10b depicts the complete ophthalmic OFDI system.<sup>26</sup> The effective ranging depth was 2.4 mm (depth dependent sensitivity decay of 6 dB over 2.4 mm) due to the finite coherence length of the laser output.

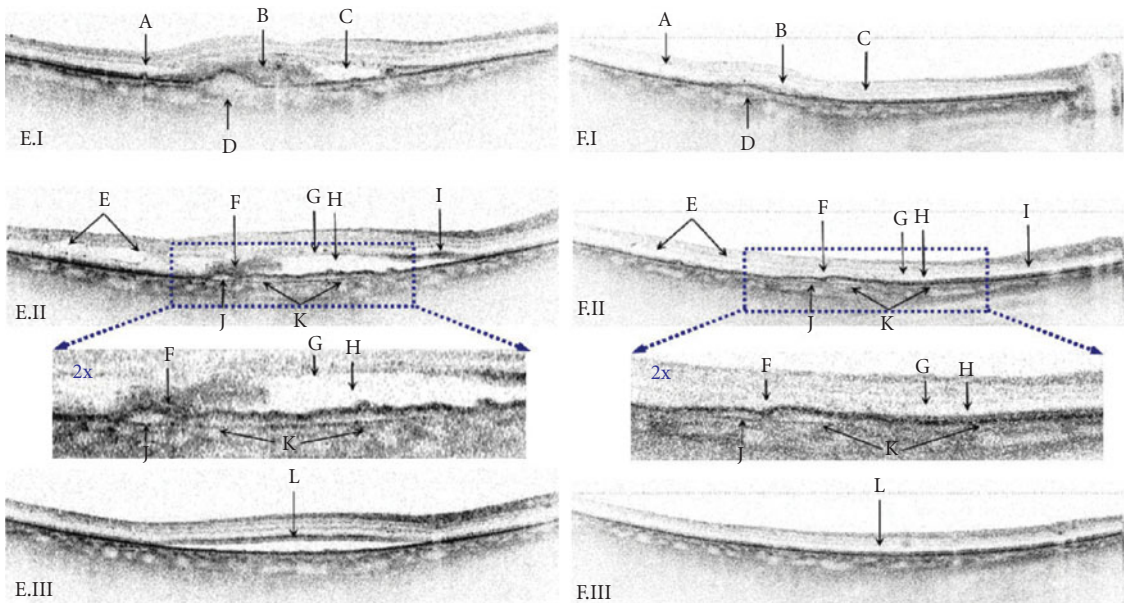
The OFDI system acquired data as the focused sample beam was scanned over an area of 6 (horizontal) by 5.2 mm (vertical) across the macular region in the retina. Each image frame in the three-dimensional volume was constructed from a thousand A-line scans. Given three-dimensional tomographic data of the eye's posterior segment, integrating the pixel values along the entire depth axis readily produces a two-dimensional fundus-type reflectivity image.<sup>42,43</sup> Figure 11a depicts an integrated reflectivity



**FIGURE 11** The retinal and choroidal vasculature extracted from a three-dimensional OFDI data set. (a) Two-dimensional reflectance image ( $5.3 \times 5.2$  mm<sup>2</sup>) obtained with the conventional full-range integration method. Higher (lower) reflectivity is represented by white (black) in the grayscale. (b) Illustration of the axial-sectioning integration method, with the different integration regions labeled C,D,E corresponding to the following fundus-type reflectivity images, respectively: (c) retinal reflectivity image showing the shadow of retinal vasculature ( $3.8 \times 5.2$  mm<sup>2</sup>), (d) reflectivity image obtained from the upper part of the choroid, and (e) reflectivity image from the center of the choroid revealing the choroidal vasculature. Shadows of retinal vasculature are also visible in (d) and (e). Scale bars: 0.5 mm. (Figure reproduced from Ref. 26 with permission from the Optical Society of America.)

image generated from the entire OFDI image sequence. The image visualizes the optic nerve head, fovea, retinal vessels, and the faint outline of the deep choroidal vasculature; however, the depth information is completely lost. To overcome this limitation of the conventional method, we integrated only selective regions based on anatomical structures. For example, to visualize the retinal vasculature with maximum contrast, we used automatic image segmentation techniques<sup>43</sup> and integrated the reflectivity in the range between IPRL and RPE (marked by red lines and labeled C in Fig. 11*b*), where the shadow or loss of signal created by the retinal vessels above appeared most distinctly.<sup>42</sup> Integrating over the entire retina, including the vessel, often results in a lower contrast in the vasculature because retinal blood vessels produce large signals by strong scattering. Figure 11*c* depicts the fundus-type reflectivity image (shadow) of the retinal vessels produced with this depth-sectioning method. The choriocapillary layer contains abundant small blood vessels and pigment cells. Using a thin integration region in the upper part of choroid (labeled D in Fig. 11*b*), we also obtained an image of the choriocapillary layer (Fig. 11*d*). To obtain an image of the complete choroid region, we used the bottom integration region (marked by blue lines and labeled E in Fig. 11*b*). The choroidal vasculature is clearly visualized in the resulting reflectivity image (Fig. 11*e*).

Figure 12 demonstrates OFDI at 1050 nm in an AMD patient. The left and right panel show images at the same location pre- and posttreatment with anti-VEGF therapy. The advantage of 1050 nm is a better penetration below the retinal pigmented epithelium (RPE), providing detailed information of what is believed to be type I vascularization between the RPE and Bruchs membrane (presumed sub-RPE CNV).



**FIGURE 12** OFDI images at 1050 nm of an AMD patient pretreatment: E.I-III, and posttreatment: F.I-III. Features—A: drusen, B: blood clot, C: subretinal fluid, D: RPE detachment, E: cystic changes, F: blood clot, G: weak scattering in photoreceptors, H: subretinal fluid, I: strong scattering in photoreceptors, J: RPE detachment, K: presumed sub-RPE CNV, and L: strong scattering from photoreceptors in the periphery of the subretinal fluid. (Reproduced from Ref. 29 with permission from the Association for Research in Vision and Ophthalmology.)

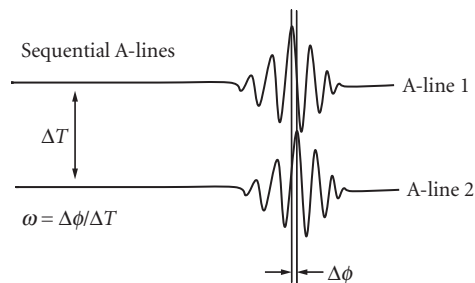
## 18.14 FUNCTIONAL EXTENSIONS: DOPPLER OCT AND POLARIZATION SENSITIVE OCT

Optical coherence tomography is an interferometric technique capable of noninvasive high-resolution cross-sectional imaging by measuring the intensity of light reflected from within tissue.<sup>2</sup> This results in a noncontact imaging modality that provides images similar in scale and geometry to histology. Just as different stains can be used to enhance the contrast in histology, various extensions of OCT allow for visualization of features not readily apparent in traditional OCT. For example, optical Doppler tomography<sup>44</sup> can enable depth-resolved imaging of flow by observing differences in phase between successive depth scans.<sup>45–47</sup> Polarization-sensitive OCT (PS-OCT) utilizes depth-dependent changes in the polarization state of detected light to determine the light polarization changing properties of a sample.<sup>48–53</sup>

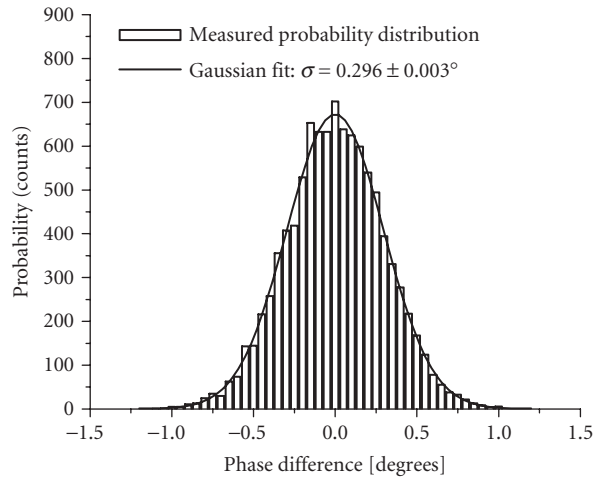
## 18.15 DOPPLER OCT AND PHASE STABILITY

In the past, phase-resolved optical Doppler tomography (ODT) based on time domain OCT (TD-OCT) has proven able to make high-resolution, high-velocity-sensitivity cross-sectional images of in vivo blood flow.<sup>45–47,54–59</sup> ODT measurements of blood flow in the human retina have been demonstrated,<sup>60,61</sup> yet the accuracy and sensitivity was compromised by a slow A-line rate and patient motion artifacts, which can introduce phase inaccuracy and obscure true retinal topography. Originally, Doppler shifts were detected by the shift in the carrier frequency in TD-OCT systems, where a trade-off between Doppler sensitivity and spatial resolution had to be made.<sup>54–56,62</sup> The detection of Doppler shifts improved significantly with the method pioneered by Zhao et al.<sup>45,46</sup> In this method two sequential A-lines are acquired at the same location. Phase-resolved detection of the interference fringes permits the determination of small phase shifts between the interferograms. The phase difference  $\Delta\phi$  divided by the time laps  $\Delta T$  between the sequential A-lines gives a Doppler shift  $\Delta\omega = \Delta\phi / \Delta T$ , associated with the motion of the scattering particle. Figure 13 gives a graphical example of this method.

Combining optical Doppler tomography with the superior sensitivity and speed of SD-OCT has allowed a significant improvement in detecting Doppler signals in vivo. In the first combination of these technologies, velocity of a moving mirror and capillary tube flow was demonstrated,<sup>63</sup> followed by in vivo demonstration of retinal blood flow.<sup>22,23</sup>



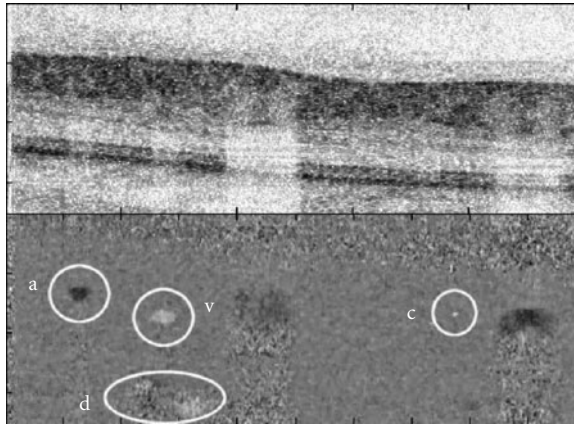
**FIGURE 13** Principle of fast Doppler OCT. OCT depth profiles are acquired sequentially at the same location. Small motion of the scattering object results in a phase shift of the interference fringes. The phase difference divided by the time lapse between the sequential A-lines gives a Doppler shift due to the motion of the scattering object.



**FIGURE 14** Probability distribution of the measured phase difference between adjacent A-lines, with a stationary reflector in the sample arm. Bars: Counted phase difference for 9990 A-lines. Bin size =  $0.05^\circ$ . Solid line: Gaussian fit to the distribution, with a measured standard deviation of  $0.296 \pm 0.003^\circ$ . (Reproduced from Ref. 23 with permission from the Optical Society of America.)

In SD-OCT, a phase-sensitive image is generated by simply determining the phase difference between points at the same depth in adjacent A-lines. The superior phase stability of SD-OCT, due to the absence of moving parts, is demonstrated in Fig. 14. The data was acquired with a stationary mirror in the sample arm, without scanning the incident beam. Ideally, interference between sample and reference arm light should have identical phase at the mirror position for all A-lines. This condition underlies the assumption that any phase difference between adjacent A-lines is solely due to motion within the sample. The actual phase varies in a Gaussian manner about this ideal, as demonstrated in Fig. 14, where we present the measured probability distribution of phase differences with a standard deviation of  $0.296 \pm 0.003^\circ$ . This value is over 25 times lower than previously quantified figures for time domain optical Doppler tomography systems,<sup>59,64</sup> and at an acquisition speed of 29 kHz corresponds to a minimum detectable Doppler shift of  $\pm 25$  Hz. With a time difference of  $34.1 \mu\text{s}$  between acquired A-lines, phase wrapping occurs at Doppler shifts greater than 15 kHz. Thus, the system dynamic range described by the ratio of maximum to minimum detectable Doppler shifts before phase wrapping occurs is a factor of 600.

In vivo images of structure and Doppler flow were acquired at 29 frames per second (1000 A-lines per frame) and subsequently processed. The images presented in Fig. 15 is 1.6 mm wide and has been cropped in depth to  $580 \mu\text{m}$ , from their original size of 1.7 mm. The layers of the retina visible in the intensity image have been identified and described previously,<sup>3</sup> with the thick, uppermost layer being the nerve fiber layer, and the thinner, strongly scattering deep layer being the retinal pigmented epithelium. One can see the pulsatility of blood flow in the artery (a), while the flow in the vein (v) is less variable (see Ref. 23). At the lower left-center of the image, it is possible to distinguish blood flow deep within the retina (d). With reference to the intensity image, one can see that this blood flow is being detected *below* the retinal pigmented epithelium, and we believe this is the first time that optical Doppler tomography imaging techniques have been able to observe and localize blood flow within the choroid. To the left of the large vessel on the right-hand side of the image, note the appearance of a very small vessel (c). The diameter of this vessel is slightly under  $10 \mu\text{m}$ .



**FIGURE 15** Movie of structure (top panel) and bi-directional flow (bottom panel) acquired in vivo in the human eye at a rate of 29 frames per second. The sequence contained 95 frames, totaling 3.28 s (see Ref. 23). Image size is 1.6 mm wide by 580  $\mu\text{m}$  deep. a: artery; v: vein; c: capillary; and d: choroidal vessel. (Reproduced from Ref. 23 with permission from the Optical Society of America.)

## 18.16 POLARIZATION SENSITIVE OCT

Polarization-sensitive OCT (PS-OCT) utilizes depth-dependent changes in the polarization state of detected light to determine the light polarization changing properties of a sample.<sup>48–53</sup> These material properties, including birefringence, dichroism, and optic axis orientation, can be determined by studying the depth evolution of the Stokes parameters,<sup>49–52,65–69</sup> or by using the changing reflected polarization states to first determine Jones or Mueller matrices.<sup>53,70–74</sup> PS-OCT provides additional contrast to identify tissue structures. Nearly all linear tissue structures, like collagen, nerve fibers, and muscle fiber exhibit birefringence. PS-OCT has been used in a wide variety of applications. In dermatology the reduced birefringence in tumor tissue was observed due to the destruction of the extracellular collagen matrix by the tumor.<sup>75</sup> The reduced birefringence due to thermally denatured collagen was observed in burns.<sup>67</sup> In ophthalmology, the birefringence of the retinal nerve fiber was measured and found to depend on the location around the optic nerve head.<sup>76,77</sup> The onset and progression of caries lesions was associated with changes in birefringence of dental tissue.<sup>78</sup>

We will first treat the Jones formalism to describe polarization properties and use the Jones formalism to determine the polarization properties of tissue. The Jones formalism provides a convenient mathematical description of polarized light and polarization effects.<sup>79</sup> The complex electric field vector  $\mathbf{E}$  can be decomposed into a pair of orthonormal basis vectors to yield

$$\begin{aligned}\mathbf{E} &= E_p \hat{\mathbf{e}}_p + E_{\perp} \hat{\mathbf{e}}_{\perp} \\ E_p &= a_p e^{-i\delta_p} \quad E_{\perp} = a_{\perp} e^{-i\delta_{\perp}}\end{aligned}\tag{21}$$

where  $\hat{\mathbf{e}}_p$  and  $\hat{\mathbf{e}}_{\perp}$  are unit vectors along the horizontal and vertical,  $a_p$  and  $a_{\perp}$  are the amplitude along the horizontal and vertical, and  $\delta_p$  and  $\delta_{\perp}$  are the phase along the horizontal and vertical, respectively. In this case, the vibrational ellipse of the electric field can be reformulated as

$$\mathbf{E}_{\text{vib}}(t) = a_p \cos(\omega t + \delta_p) \hat{\mathbf{e}}_p + a_{\perp} \cos(\omega t + \delta_{\perp}) \hat{\mathbf{e}}_{\perp}\tag{22}$$



with  $\omega$  the angular frequency and  $t$  time. The overall irradiance, or intensity, of the beam of light then can be expressed as the scalar quantity

$$I = a_p^2 + a_\perp^2 \quad (23)$$

It is worth noting that while the overall irradiance of a beam is not dependent on its polarization state, it is possible to measure irradiance along a particular orientation (e.g., the intensity of a beam in the horizontally polarized direction).

Linear polarization states occur for phase differences  $\Delta\delta = \delta_p - \delta_\perp = m\pi$ , where  $m \in \mathbb{Z}$ , as the vibrational ellipse collapses to a line described by

$$\mathbf{E}_{\text{vib}}(t) = (a_p \hat{\mathbf{e}}_p + (-1)^m a_\perp \hat{\mathbf{e}}_\perp) \cos(\omega t + \delta_p) \quad (24)$$

The orientation of the linear polarization state depends on the ratio of amplitudes  $a_p$  and  $a_\perp$ . The polarization state of light is horizontal or vertical when  $a_\perp = 0$  or  $a_p = 0$ , respectively, and oriented at  $\pm 45^\circ$  if  $|a_p| = |a_\perp|$ . An orientation angle  $\theta$  can then be defined according to the relations

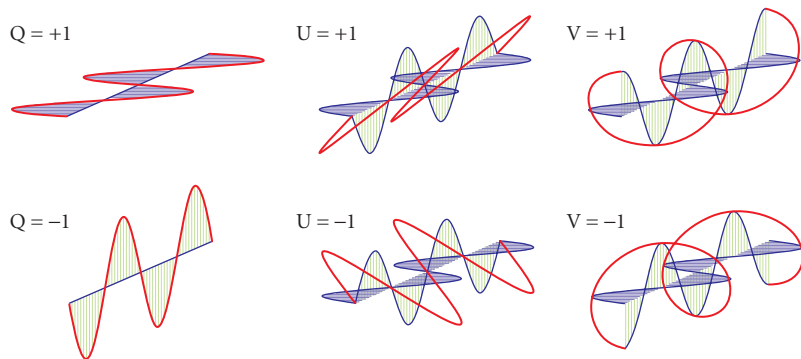
$$\begin{aligned} a_p &= a \cos\theta & a_\perp &= a \sin\theta \\ a &= \sqrt{a_p^2 + a_\perp^2} & \theta &= \tan^{-1} \frac{a_p}{a_\perp} \end{aligned} \quad (25)$$

where  $a$  can be thought of as an overall amplitude of the electric field and  $\Delta\delta = 0$ .

Circular polarization states are obtained when  $|a_p| = |a_\perp|$  and  $\Delta\delta = \frac{\pi}{2} + n\pi$ , which is evident through the form of the resultant vibrational ellipse

$$\mathbf{E}_{\text{vib}}(t) = a_p (\cos(\omega t + \delta_p) \hat{\mathbf{e}}_p \pm \sin(\omega t + \delta_p) \hat{\mathbf{e}}_\perp) \quad (26)$$

This describes a circle, the handedness (left- or right-circular) of which is determined by the sign between the orthogonal components. Circular polarization states differ only in their phase difference compared to linearly polarized light at  $45^\circ$ , and so the phase difference  $\Delta\delta$  between orthogonal electric field components reflects the ratio between circular and linear components of the polarization state. Figure 16 gives a graphical representation of the different vibrational ellipses



**FIGURE 16** Vibrational ellipses for various polarization states  $Q = 1$  and  $Q = -1$  correspond to horizontal and vertical linear polarized light.  $U = 1$  and  $U = -1$  correspond to linear polarized light at  $+45^\circ$  and  $-45^\circ$ , and  $V = 1$  and  $V = -1$  correspond to circular polarized light, respectively.

The electric field decomposition in Eq. (21) can be rewritten as a complex 2-vector such that

$$\mathbf{E} = \begin{bmatrix} E_p \\ E_{\perp} \end{bmatrix} = \begin{bmatrix} a_p e^{-i\delta_p} \\ a_{\perp} e^{-i\delta_{\perp}} \end{bmatrix} = a e^{-i\delta_p} \begin{bmatrix} \cos\theta \\ e^{i\Delta\delta} \sin\theta \end{bmatrix} \quad (27)$$

while the time-invariant electric field vector  $\mathbf{E}$ , also known as a Jones vector, does depend on the amplitude and exact phase of the electric field components, it should be noted that the polarization state itself is completely determined by the orientation angle  $\theta$  and the phase difference  $\Delta\delta$ .

Just as two vectors of length  $n$  can be related using a matrix of dimension  $n \times n$ , two polarization states can be related using a complex  $2 \times 2$  matrix known as a Jones matrix. The polarization properties of any nondepolarizing optical system can be described using a Jones matrix. The transmitted polarization state  $\mathbf{E}'$  as a result of an optical system represented by a Jones matrix  $\mathbf{J}$  acting on an incident polarization state  $\mathbf{E}$  can be determined by

$$\mathbf{E}' = \begin{bmatrix} E'_p \\ E'_{\perp} \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} E_p \\ E_{\perp} \end{bmatrix} = \mathbf{J}\mathbf{E} \quad (28)$$

Subsequent transmission of  $\mathbf{E}'$  through an optical system  $\mathbf{J}'$  results in a polarization state  $\mathbf{E}'' = \mathbf{J}'\mathbf{E}' = \mathbf{J}'(\mathbf{J}\mathbf{E}) = \mathbf{J}'\mathbf{J}\mathbf{E}$ . As a result, the combined polarization effect of a cascade of optical elements,  $\mathbf{J}_1, \mathbf{J}_2, \dots, \mathbf{J}_n$ , can be described by the product  $\mathbf{J} = \mathbf{J}_n \cdots \mathbf{J}_2 \mathbf{J}_1$ .

The Jones matrix for a birefringent material that induces a phase retardation  $\eta$  between electric field components parallel and orthogonal to a polarization state characterized by an orientation angle  $\theta$  and a circularity related to  $\phi$  is given by<sup>80</sup>

$$\mathbf{J}_b = \begin{bmatrix} e^{i\eta/2} C_{\theta}^2 + e^{-i\eta/2} S_{\theta}^2 & (e^{i\eta/2} - e^{-i\eta/2}) C_{\theta} S_{\theta} e^{-i\phi} \\ (e^{i\eta/2} - e^{-i\eta/2}) C_{\theta} S_{\theta} e^{i\phi} & e^{i\eta/2} S_{\theta}^2 + e^{-i\eta/2} C_{\theta}^2 \end{bmatrix} \quad (29)$$

where  $C_{\theta} = \cos\theta$  and  $S_{\theta} = \sin\theta$ . The Jones matrix of a dichroic material with attenuation ratios of  $P_1$  and  $P_2$  for electric field components parallel and orthogonal, respectively, to a polarization state given by an orientation angle  $\Theta$  and a circularity  $\Phi$  has the form<sup>80</sup>

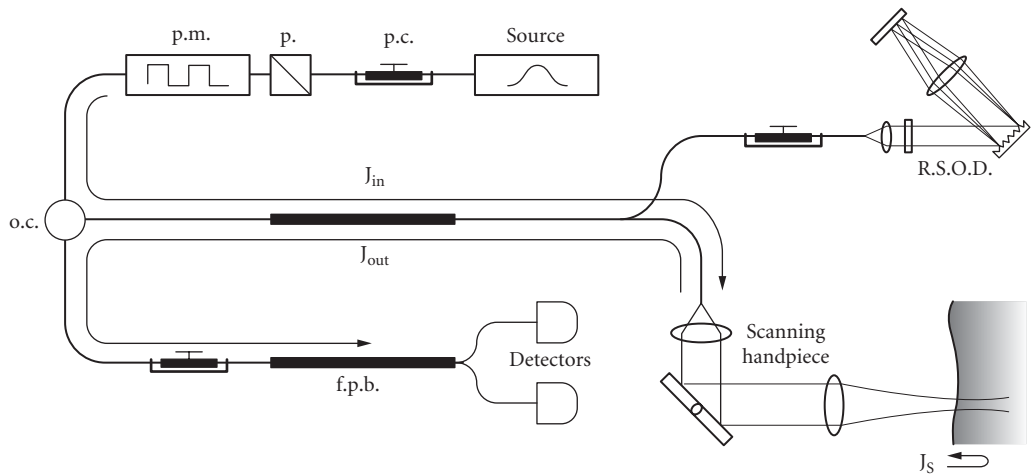
$$\mathbf{J}_d = \begin{bmatrix} P_1 C_{\Theta}^2 + P_2 S_{\Theta}^2 & (P_1 - P_2) C_{\Theta} S_{\Theta} e^{-i\Phi} \\ (P_1 - P_2) C_{\Theta} S_{\Theta} e^{i\Phi} & P_1 S_{\Theta}^2 + P_2 C_{\Theta}^2 \end{bmatrix} \quad (30)$$

As birefringence seems to be the primary polarization property exhibited by biological tissue, most PS-OCT analysis methods concentrate on determination of phase retardation. Schoenenberger et al.<sup>81</sup> analyzed system errors introduced by the extinction ratio of polarizing optics and chromatic dependence of wave retarders, and errors due to dichroism. System errors can be kept small by careful design of the system with achromatic elements, but can never be completely eliminated. In principle, dichroism is a more serious problem when interpreting results as solely due to birefringence. However, Mueller matrix ellipsometry measurements have shown that the error due to dichroism in the eye is relatively small,<sup>82,83</sup> and earlier PS-OCT work shows that dichroism is of minor importance in rodent muscle.<sup>52</sup> Despite this, a method for simultaneous determination of sample birefringence and dichroism is desirable, especially one that can be applied to systems with the unrestricted use of optical fiber and fiber components.

The nondepolarizing polarization properties of an optical system can be completely described by its complex Jones matrix,  $\mathbf{J}$ , which transforms an incident polarization state, described by a complex electric field vector,  $\mathbf{E} = [H, V]^T$ , to a transmitted state,  $\mathbf{E}' = [H', V']^T$ . A Jones matrix can be decomposed in the form  $\mathbf{J} = \mathbf{J}_R \mathbf{J}_p = \mathbf{J}_p \mathbf{J}_R'$ .<sup>80</sup> Birefringence, described by  $\mathbf{J}_R$ , can be parameterized by three variables: a degree of phase retardation  $\eta$  about an axis defined by two angles,  $\gamma$  and  $\delta$ .

Diattenuation, described by  $J_p$ , is defined as  $d = (P_1^2 - P_2^2) / (P_1^2 + P_2^2)$  and can be parameterized by four variables, where  $P_1$  and  $P_2$  are the attenuation coefficients parallel and orthogonal, respectively, to an axis defined by angles  $\Gamma$  and  $\Delta$ . These seven independent parameters, along with an overall common phase  $e^{i\psi}$ , account for all four complex elements of a general Jones matrix  $\mathbf{J}$ . Assuming that birefringence and diattenuation arise from the same fibrous structures in biological tissue and thus share a common axis  $\delta = \Delta$  and  $\gamma = \Gamma$ ,<sup>72</sup> the number of independent parameters is reduced by two to five. In order to determine these five parameters, the sample needs to be probed by two unique polarization states. One incident and reflected polarization state yield three relations involving the two orthogonal amplitudes and the relative phase between them.<sup>52</sup> Therefore, it is possible to use the six relationships defined by two unique pairs of incident and reflected polarization states to exactly solve for the Jones matrix of a sample.

In general terms, a PS-OCT system sends polarized light from a broadband source into the sample and reference arms of an interferometer, and reflected light from both arms is recombined and detected. Figure 17 shows an example of a fiber-based PS-OCT system. The source light is sent to a fiber-based polarization controller and a polarizer. The polarization of the incident state on the sample is modulate by a polarization modulator that allows switching between two polarization states. Light passes a circulator, which sends the light to a fiber-based beam splitter that splits the light into a sample and a reference arm. Upon reflection, the light is recombined in the beam splitter and the circulator sends the light to the detection arm where the light is split into two orthogonal polarization states before detection by two separate detectors. The optical path from source to detector can be described by three Jones matrices. Define  $J_{in}$  as the Jones matrix representing the optical path from the polarized light source (the polarization modulator) to the sample surface,  $J_{out}$  as that going from the sample surface to the detectors, and  $J_s$  as the round-trip Jones matrix for light propagation through a sample.<sup>74</sup> This nomenclature can be applied to all PS-OCT systems, ranging from bulk-optic systems<sup>48,49,51–53,65</sup> to those with fibers placed such that they are traversed in a round-trip manner,<sup>73</sup> to time-domain<sup>66, 68</sup> and spectral-domain<sup>69</sup> PS-OCT systems with the unrestricted use of optical fiber and nondiattenuating fiber components, and even for retinal systems,<sup>76</sup> where the polarization effects of the cornea can be included in  $J_{in}$  and  $J_{out}$ . The electric field of light reflected from the



**FIGURE 17** Schematic of the fiber-based PS-OCT system (p.c., polarization controller; p, polarizer; pm, polarization modulator; oc, optical circulator; RSOD, rapid scanning optical delay; and fpb, fiber polarizing beamsplitter).  $J_{in}$ ,  $J_{out}$ , and  $J_s$  are the Jones matrix representations for the one-way optical path from the polarization modulator to the scanning handpiece, the one-way optical path back from the scanning handpiece to the detectors, and the round-trip path through some depth in the sample, respectively. (Reprinted from Ref. 74 with permission from the Optical Society of America.)



sample surface,  $\mathbf{E}$ , can be expressed as  $\mathbf{E} = e^{i\psi} \mathbf{J}_{\text{out}} \mathbf{J}_{\text{in}} \mathbf{E}_{\text{source}}$ , where  $\psi$  represents a common phase and  $\mathbf{E}_{\text{source}}$  represents the electric field of light coming from the polarized source. Likewise, the electric field of light reflected from some depth within the tissue may be described by  $\mathbf{E}' = e^{i\psi'} \mathbf{J}_{\text{out}} \mathbf{J}_{\text{in}} \mathbf{E}_{\text{source}}$ . These two measurable polarization states can be related to each other such that  $\mathbf{E}' = e^{i\Delta\psi} \mathbf{J}_T \mathbf{E}$ , where  $\mathbf{J}_T = \mathbf{J}_{\text{out}} \mathbf{J}_S \mathbf{J}_{\text{out}}^{-1}$  and  $\Delta\psi = \psi' - \psi$ . Thus,  $\mathbf{J}_T$  can be determined by comparing the polarization state reflected from the surface and from a certain depth in the sample. Now we need to determine how to extract  $\mathbf{J}_S$  from  $\mathbf{J}_T$ .

If the optical system representing  $\mathbf{J}_{\text{out}}$  is nondiattenuating,  $\mathbf{J}_{\text{out}}$  can be treated as a unitary matrix with unit determinant after separating out a common attenuation factor.  $\mathbf{J}_{\text{out}}$  is assumed to be nondiattenuating since optical fibers are virtually lossless.  $\mathbf{J}_S$  can be decomposed into a diagonal matrix  $\mathbf{J}_C = \begin{bmatrix} P_1 e^{i\eta/2} & 0 \\ 0 & P_2 e^{-i\eta/2} \end{bmatrix}$ , containing complete information regarding the amount of sample diattenuation and phase retardation, surrounded by unitary matrices  $\mathbf{J}_A$  with unit determinant that define the sample optic axis,  $\mathbf{J}_S = \mathbf{J}_A \mathbf{J}_C \mathbf{J}_A^{-1}$ .  $\mathbf{J}_T$  can be rewritten such that  $\mathbf{J}_T = \mathbf{J}_{\text{out}} \mathbf{J}_S \mathbf{J}_{\text{out}}^{-1} = \mathbf{J}_{\text{out}} (\mathbf{J}_A \mathbf{J}_C \mathbf{J}_A^{-1}) \mathbf{J}_{\text{out}}^{-1} = \mathbf{J}_U \mathbf{J}_C \mathbf{J}_U^{-1}$ , where  $\mathbf{J}_U = \mathbf{J}_{\text{out}} \mathbf{J}_A$ . Since unitary matrices with unit determinant form the special unitary group  $\text{SU}(2)$ ,<sup>84</sup>  $\mathbf{J}_U$  must also be a unitary matrix with unit determinant by closure and can be expressed in the form

$$\mathbf{J}_U = e^{i\beta} \begin{bmatrix} C_\theta e^{i(\phi-\varphi)} & -S_\theta e^{i(\phi+\varphi)} \\ S_\theta e^{-i(\phi+\varphi)} & C_\theta e^{-i(\phi-\varphi)} \end{bmatrix} \quad (31)$$

$\mathbf{J}_T$  can be obtained from the measurements by combining information from two unique incident polarization states,  $[H'_1, H'_2; V'_1, V'_2] = e^{i\Delta\psi_1} \mathbf{J}_T [H_1, e^{i\alpha} H_2; V_1, e^{i\alpha} V_2]$ , where  $\alpha = \Delta\psi_2 - \Delta\psi_1$ . The polarization properties of interest can be obtained by equating the two expressions for  $\mathbf{J}_T$  to yield

$$e^{i\Delta\psi_1} \begin{bmatrix} P_1 e^{i\eta/2} & 0 \\ 0 & P_2 e^{-i\eta/2} \end{bmatrix} = \begin{bmatrix} C_\theta & S_\theta \\ -S_\theta & C_\theta \end{bmatrix} \begin{bmatrix} e^{-i\phi} & 0 \\ 0 & e^{i\phi} \end{bmatrix} \begin{bmatrix} H'_1 & H'_2 \\ V'_1 & V'_2 \end{bmatrix} \quad (32)$$

$$\begin{bmatrix} H_1 & e^{i\alpha} H_2 \\ V_1 & e^{i\alpha} V_2 \end{bmatrix}^{-1} \begin{bmatrix} e^{i\phi} & 0 \\ 0 & e^{-i\phi} \end{bmatrix} \begin{bmatrix} C_\theta & -S_\theta \\ S_\theta & C_\theta \end{bmatrix}$$

In principle, parameters  $\theta$ ,  $\phi$ , and  $\alpha$  can be solved for with the condition that the off-diagonal elements of the matrix product on the right-hand side of Eq. (32) are equal to zero. In practice, real solution cannot always be found, as measurement noise can induce nonphysical transformations between incident and transmitted polarization states. To account for this, Eq. (32) can be solved by optimizing parameters  $\theta$ ,  $\phi$ , and  $\alpha$  to minimize the sum of the magnitudes of the off-diagonal elements. In principle, this can be achieved using two unique incident polarization states to probe the same volume of a sample. However, when two orthogonal incident polarization states are used,<sup>73</sup> birefringence cannot be retrieved under all circumstances.<sup>85</sup> A better choice is to use two incident polarization states perpendicular in a Poincaré sphere representation to guarantee that polarization information can always be extracted.<sup>66-69, 76, 86</sup> The degree of phase retardation can easily be extracted through the phase difference of the resulting diagonal elements, and the diattenuation by their magnitudes. It should be noted that these phase retardation values range from  $-\pi$  to  $\pi$ , and can therefore be unwrapped to yield overall phase retardations in excess of  $2\pi$ .

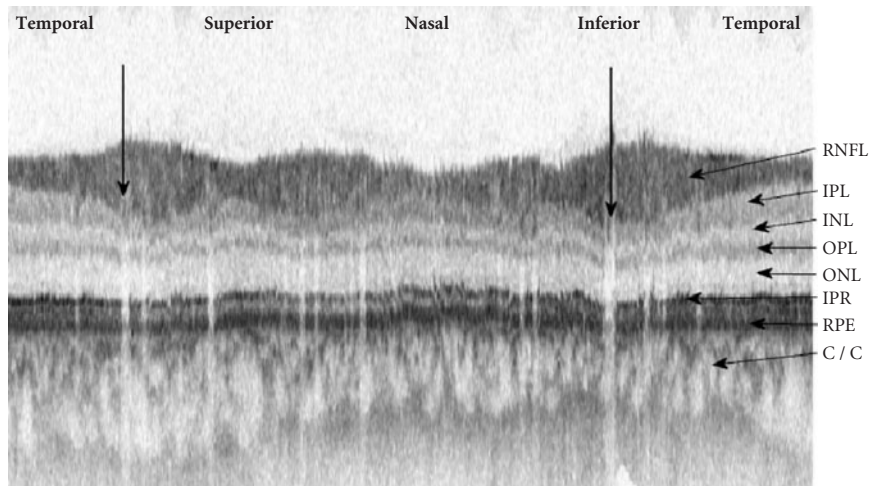
## 18.17 PS-OCT IN OPHTHALMOLOGY

Ophthalmological application of OCT has arguably driven a great deal of its development, and probably represents the most researched clinical application of the technology to date. PS-OCT in particular has been used to measure the birefringence of the human retinal nerve fiber layer in vivo<sup>48,76,77,87-89</sup> for potential early detection of glaucoma, the world's second leading cause of blindness.

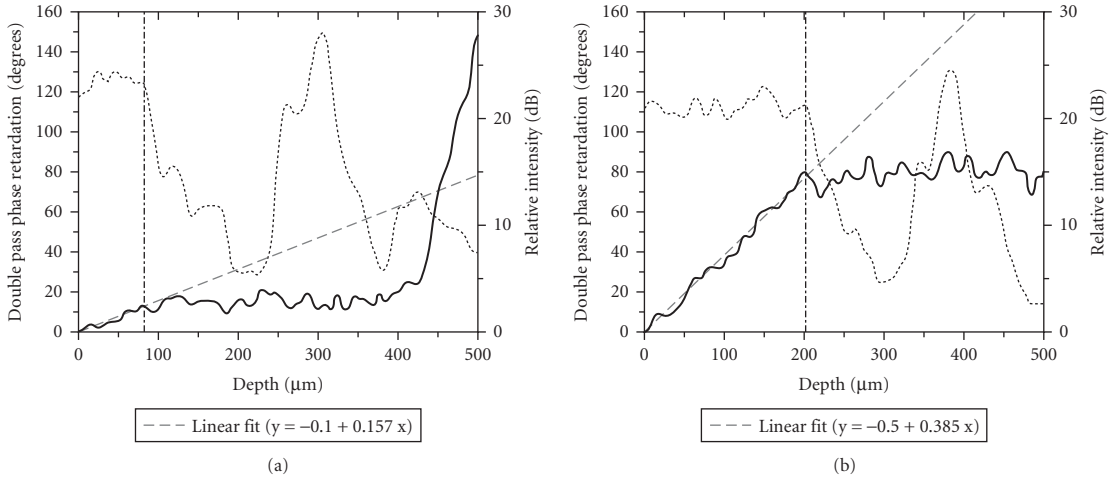
Glaucoma causes damage to the retinal ganglion cells, resulting in a thinning of the retinal nerve fiber layer (RNFL). In addition, nerve fiber layer tissue loss may be preceded by changes in birefringence as ganglion cells become necrotic and axons in the RNFL are replaced by a less organized and amorphous tissue composed of glial cells. When glaucoma is detected at an early stage, further loss of vision can be prevented by treatment. The visual field test is the current standard method of detecting loss of peripheral vision in glaucoma. However, measurements show that up to 40 percent of nerves are irreversibly damaged before loss of peripheral vision can be clinically detected. PS-OCT has the potential to detect changes to the RNFL at an earlier time point through changes in its birefringence and thickness.

Ophthalmic studies can be performed using systems similar to that used by Cense et al.,<sup>76</sup> in which a slit lamp has been adapted for use with PS-OCT. Figure 18 is a typical example of a structural-intensity time-domain OCT image of the retina in the left eye of a healthy volunteer obtained with a circular scan with a radius of 2.1 mm around the optic nerve head (ONH). The image measures 13.3 mm wide and 0.9 mm deep and is shown at an expanded aspect ratio in depth for clarity. Structural layers such as the RNFL, the interface between the inner and outer segments of the photoreceptors, and the retinal pigmented epithelium can be seen.

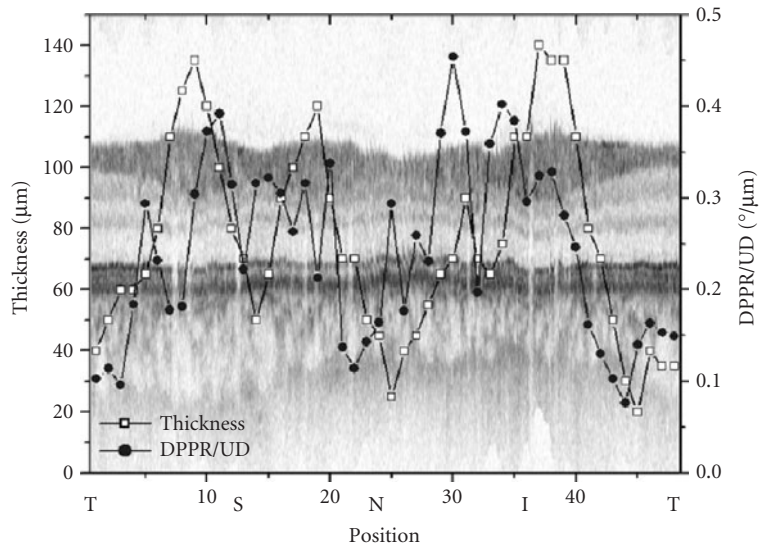
The addition of polarization sensitivity allows for localized quantitative assessment of the thickness and birefringence of the RNFL. Figure 19 shows two examples of combined thickness and birefringence measurements, one of a region temporal to the ONH, the other of a region superior to the ONH. The depth of the RNFL can be determined by a decrease in backscattered intensity from the RNFL to the inner plexiform layer. The birefringence of the RNFL can then be estimated from a linear least-square fit of the measured double-pass phase retardation through the determined depth. Two main observations can be drawn from such graphs: the retinal layers directly below the RNFL are minimally birefringent, and that the thickness and birefringence of the RNFL are not constant. These observations can also be seen in Fig. 20, which overlays the thickness and birefringence determined



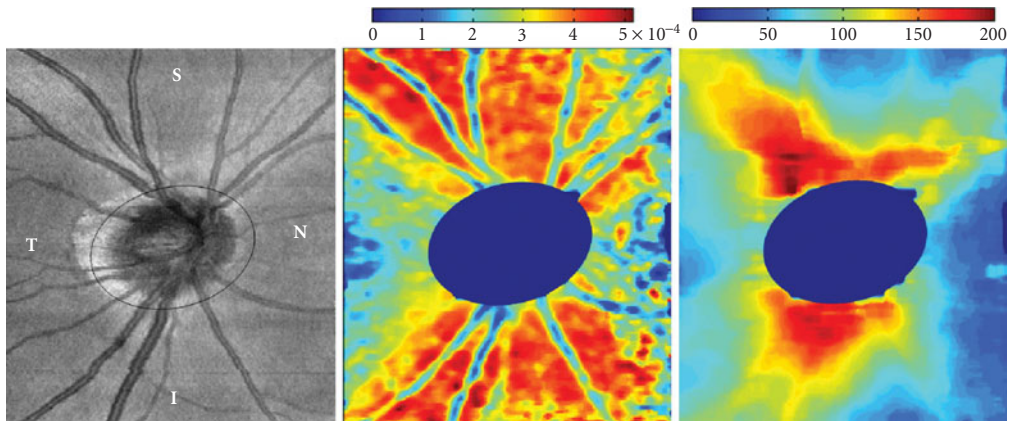
**FIGURE 18** A realigned OCT intensity image created with a 2.1-mm radius circular scan around the ONH. The dynamic range of the image is  $-36$  dB. Black pixels represent strong reflections. The image measures 13.3 mm wide and 0.9 mm deep. Visible structures: retinal nerve fiber layer (RNFL); inner plexiform layer (IPL); inner nuclear layer (INL); outer plexiform layer (OPL); outer nuclear layer (ONL); interface between the inner and outer segments of the photoreceptor layer (IPR); retinal pigmented epithelium (RPE); and choriocapillaris and choroid (C/C). Vertical arrows: locations of the two largest blood vessels. Other smaller blood vessels appear as vertical white areas in the image. (Reprinted from Ref. 77 with permission from the Association for Research in Vision and Ophthalmology.)



**FIGURE 19** Thickness (dotted line) and birefringence (solid line) plots of an area temporal (a) and superior (b) to the ONH. DPPR data belonging to the RNFL is fit with a least-square linear fit. The slope in the equation represents the DPPR/UD or birefringence. The vertical line indicates the boundary of the RNFL, as determined from the intensity and DPPR data. (a) The increase in DPPR at a depth beyond 450  $\mu\text{m}$  is caused by either a relatively low signal-to-noise ratio, or by the presence of a highly birefringent material—for instance, collagen in the sclera. (Reprinted from Ref. 77 with permission from the Association for Research in Vision and Ophthalmology.)



**FIGURE 20** A typical example of combined RNFL thickness and birefringence measurements along a circular scan around the ONH. The intensity image is plotted in the background. The RNFL is relatively thicker superiorly (S) and inferiorly (I). A similar development can be seen in the birefringence plot. The birefringence is relatively higher in the thicker areas, whereas it is lower in the thinner temporal (T) and nasal (N) areas. (Reprinted from Ref. 77 with permission from the Association for Research in Vision and Ophthalmology.)



**FIGURE 21** OCT scan ( $4.24 \times 5.29 \text{ mm}^2$ ) of the retina of a normal volunteer, centered on the ONH. (a) Integrated reflectance map showing a normal temporal crescent (white area temporal to the ONH); (b) birefringence map; and (c) RNFL thickness map (color bar scaled in microns). The circle on the left indicates the excluded area in the birefringence and thickness maps as corresponding to ONH. (S = superior, N = nasal, I = inferior, T = temporal). (Reprinted from Ref. 90 with permission from the International Society for Optical Engineering.)

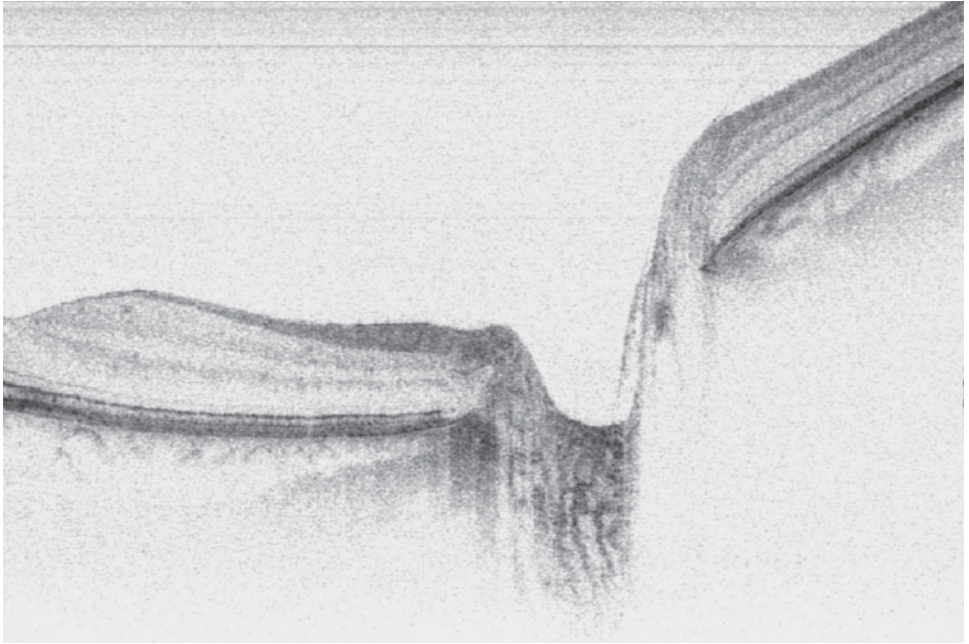
as in Fig. 19 on a circular scan across the ONH. The plots indicate that the RNFL is thickest and most birefringent superiorly and inferiorly to the ONH.

En face maps of RNFL thickness and birefringence can be generated from data obtained with recently developed spectral-domain ophthalmic PS-OCT systems.<sup>90</sup> A three-dimensional volume ( $4.24 \times 5.29 \times 1.57 \text{ mm}^3$ ) of the retina of a normal volunteer (right eye) was scanned at a rate of 29 fps with 1000 A-lines/frame, and contains 190 frames (B-scans) acquired in 6.5 s. The integrated reflectance, birefringence, and retinal nerve fiber layer (RNFL) thickness maps are shown in Fig. 21, confirming previous findings that the RNFL birefringence is not uniform across the retina. Superior, nasal, inferior, and temporal areas of the retina around the ONH are indicated by the letters S, N, I, and T. The integrated reflectance map, obtained by simply integrating the logarithmic depth profiles, illustrates the blood vessel structure around the ONH. The RNFL thickness map is scaled in microns (color bar on the top of the image) indicating a RNFL thickness of up to 200  $\mu\text{m}$ . The central dark-blue area corresponds to the position of the ONH that was excluded from both the thickness and the birefringence maps. A typical bow-tie pattern can be seen for the distribution of the RNFL thickness around the ONH, showing a thicker RNFL superior and inferior to the ONH. The birefringence map illustrates a variation of the birefringence values between 0 and  $5.16 \times 10^{-4}$ , and it clearly demonstrates that the RNFL birefringence is not uniform across the retina; it is smaller nasal and temporal and larger superior and inferior to the ONH.

Given that measurements of the thickness and birefringence of the RNFL can be acquired with the speed and accuracy demonstrated, further research into changes in these parameters with glaucoma can be performed. Experiments—for instance, a longitudinal study with PS-OCT on patients at high risk for development of glaucoma—will either confirm or reject the hypothesis. In addition, PS-OCT can enhance the specificity in determining RNFL thickness in structural OCT images by using changes in tissue birefringence to determine the border between the RNFL and the ganglion cell layer.

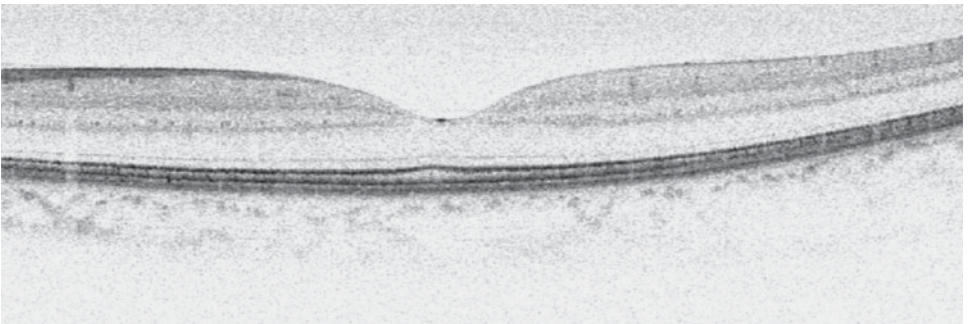
## 18.18 RETINAL IMAGING WITH SD-OCT

Many examples of retinal imaging with SD-OCT are available in the literature. Commercial systems are being introduced into the market at this point. Below, two typical examples of high



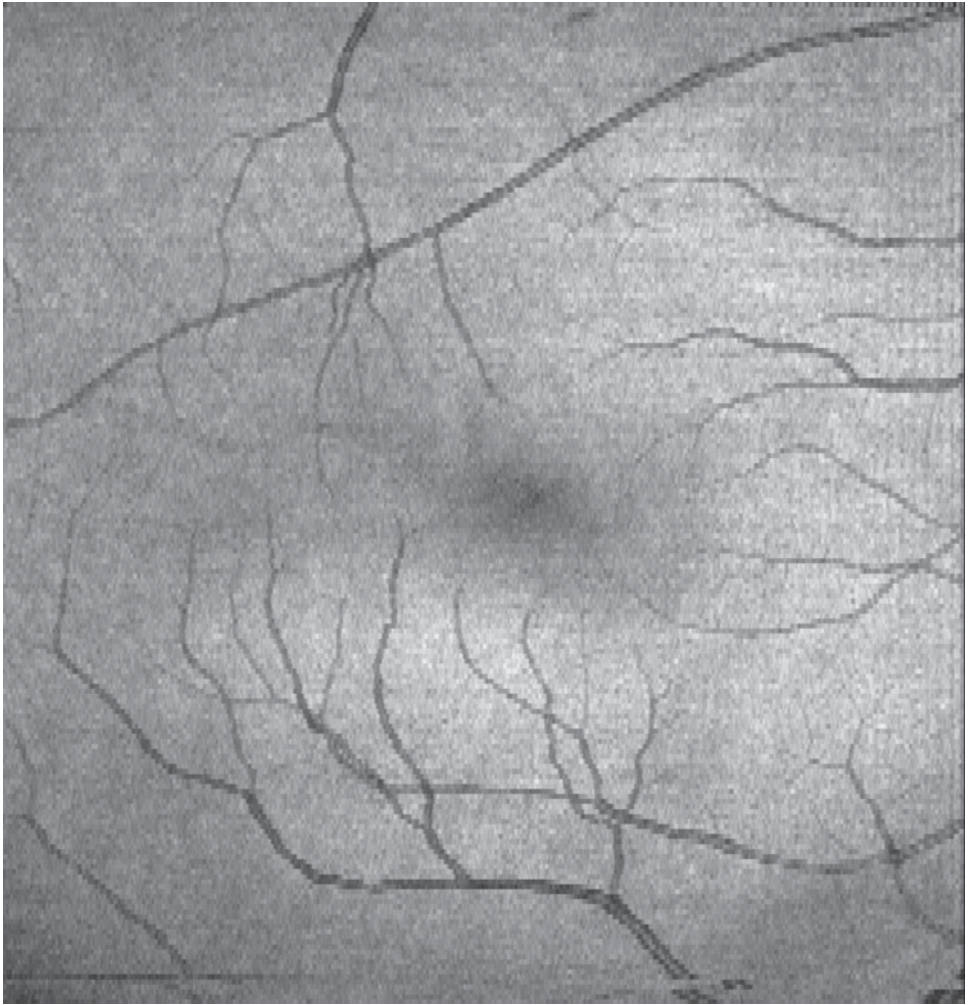
**FIGURE 22** High resolution SD-OCT image of a human retina in vivo, centered on the optic nerve head. The image is magnified in depth by a factor of 2. Image size:  $4.662 \times 1.541$  mm.

quality SD-OCT images are presented, acquired with a system providing a depth resolution of  $3 \mu\text{m}$  in tissue. Both figures consist of approximately 1000 A-lines or depth profiles. Figure 22 shows a cross-section centered on the optic nerve head. Figure 23 shows an image centered on the fovea and the corresponding en face image (Fig. 24) generated from the three-dimensional data set.



**FIGURE 23** High resolution SD-OCT image of a human retina in vivo, centered on the fovea. The image is magnified in depth by a factor of 2. Image size:  $4.973 \times 0.837$  mm.





**FIGURE 24** En face reconstruction of the fovea region of the retina from a three-dimensional volumetric SD-OCT data set of 200 images. Image size: 4.97 mm  $\times$  5.18 mm.

## 18.19 CONCLUSION

Spectral domain or frequency domain OCT (SD/FD-OCT) has become the preferred method for retinal imaging owing to its high imaging speed,<sup>10,19</sup> enhanced signal to noise ratio (SNR),<sup>7,15–17</sup> and the availability of broadband sources permitting ultrahigh resolution retinal imaging.<sup>20,21</sup> However, the state-of-the-art spectrometers are hindering further improvements (1) with limited detection efficiency ( $\sim 25$  percent)<sup>19</sup> and (2) the obtainable spectral resolution causes approximately a 6-dB sensitivity drop over a 1-mm depth range.<sup>19</sup> Furthermore, rapid scanning of the probe beam in SD-OCT has the adverse effect of fringe washout, which causes SNR to decrease.<sup>36</sup> Fringe washout can be addressed by pulsed illumination.<sup>38</sup> A competing technique such as optical frequency domain imaging

(OFDI), the dominant implementation of Fourier domain OCT technologies at 1.3  $\mu\text{m}$ ,<sup>9</sup> has the advantage of larger depth range and better immunity to motion artifacts. OFDI has recently been demonstrated in the 800 and 1050 nm range,<sup>26,27,91</sup> but has not reached the superior resolution of SD-OCT.<sup>20,21</sup>

## 18.20 ACKNOWLEDGMENTS

This research was supported in part by research grants from the National Institutes of Health (1R24 EY12877, R01 EY014975, and RR19768), Department of Defense (F4 9620-01-1-0014), CIMIT, and a gift from Dr. and Mrs. J. S. Chen to the optical diagnostics program of the Wellman Center of Photomedicine. The author would like to thank a number of graduate students and post doctoral fellows who have contributed to the results presented in this chapter, Barry Cense, Nader Nassif, Brian White, Hyle Park, Jang Woo You, Mircea Mujat, Hyungsik Lim, Martijn de Bruin, Daina Burnes, and Yueli Chen. Special thanks to Teresa Chen, MD, my invaluable collaborator at the Massachusetts Eye and Ear Infirmary, without whom all this work would not have been possible.

## 18.21 REFERENCES

1. A. F. Fercher, K. Menedoht, and W. Werner, "Eye-Length Measurement by Interferometry with Partially Coherent-Light," *Optics Letters*, 1988, **13**(3):186–188.
2. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, et al., "Optical Coherence Tomography," *Science*, 1991, **254**(5035):1178–1181.
3. W. Drexler, H. Sattmann, B. Hermann, T. H. Ko, M. Stur, A. Unterhuber, C. Scholda, et al., "Enhanced Visualization of Macular Pathology with the Use of Ultrahigh-Resolution Optical Coherence Tomography," *Archives of Ophthalmology*, 2003, **121**(5):695–706.
4. American National Standards Institute, *American National Standard for Safe Use of Lasers Z136.1*, 2000, Orlando.
5. A. F. Fercher, C. K. Hitzenberger, G. Kamp, and S. Y. Elzaiat, "Measurement of Intraocular Distances by Backscattering Spectral Interferometry," *Optics Communications*, 1995, **117**(1–2):43–48.
6. B. Golubovic, B. E. Bouma, G. J. Tearney, and J. G. Fujimoto, "Optical Frequency-Domain Reflectometry Using Rapid Wavelength Tuning of a Cr4+:Forsterite Laser," *Optics Letters*, 1997, **22**(22):1704–1706.
7. M. A. Choma, M. V. Sarunic, C. H. Yang, and J. A. Izatt, "Sensitivity Advantage of Swept Source and Fourier Domain Optical Coherence Tomography," *Optics Express*, 2003, **11**(18):2183–2189.
8. S. H. Yun, G. J. Tearney, J. F. de Boer, N. Iftimia, and B. E. Bouma, "High-Speed Optical Frequency-Domain Imaging," *Optics Express*, 2003, **11**(22):2953–2963.
9. S. H. Yun, G. J. Tearney, B. J. Vakoc, M. Shishkov, W. Y. Oh, A. E. Desjardins, M. J. Suter, et al., "Comprehensive Volumetric Optical Microscopy In Vivo," *Nature Medicine*, 2006, **12**(12):1429–1433.
10. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, 1995, Cambridge, England: Cambridge University Press.
11. N. Nassif, B. Cense, B. H. Park, S. H. Yun, T. C. Chen, B. E. Bouma, G. J. Tearney, and J. F. de Boer, "In Vivo Human Retinal Imaging by Ultrahigh-Speed Spectral Domain Optical Coherence Tomography," *Optics Letters*, 2004, **29**(5):480–482.
12. W. Drexler, U. Morgner, R. K. Ghanta, F. X. Kartner, J. S. Schuman, and J. G. Fujimoto, "Ultrahigh-Resolution Ophthalmic Optical Coherence Tomography," *Nature Medicine*, 2001, **7**(4):502–507.
13. T. Akkin, C. Joo, and J. F. de Boer, "Depth-Resolved Measurement of Transient Structural Changes during Action Potential Propagation," *Biophysical Journal*, 2007, **93**(4):1347–1353.
14. G. Hausler and M. W. Lindner, "Coherence Radar and Spectral Radar—New Tools for Dermatological Diagnosis," *Journal of Biomedical Optics*, 1998, **3**(1):21–31.
15. T. Mitsui, "Dynamic Range of Optical Reflectometry with Spectral Interferometry," *Japanese Journal of Applied Physics Part 1-Regular Papers Short Notes & Review Papers*, 1999, **38**(10):6133–6137.

16. R. Leitgeb, C. K. Hitzenberger, and A. F. Fercher, "Performance of Fourier Domain vs. Time Domain Optical Coherence Tomography," *Optics Express*, 2003, **11**(8):889–894.
17. J. F. de Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, "Improved Signal-to-Noise Ratio in Spectral-Domain Compared with Time-Domain Optical Coherence Tomography," *Optics Letters*, 2003, **28**(21):2067–2069.
18. M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszewski, and A. F. Fercher, "In Vivo Human Retinal Imaging by Fourier Domain Optical Coherence Tomography," *Journal of Biomedical Optics*, 2002, **7**(3):457–463.
19. N. A. Nassif, B. Cense, B. H. Park, M. C. Pierce, S. H. Yun, B. E. Bouma, G. J. Tearney, T. C. Chen, and J. F. de Boer, "In Vivo High-Resolution Video-Rate Spectral-Domain Optical Coherence Tomography of the Human Retina and Optic Nerve," *Optics Express*, 2004, **12**(3):367–376.
20. B. Cense, N. Nassif, M. C. Pierce, S. H. Yun, B. H. Park, B. E. Bouma, G. J. Tearney, and J. F. de Boer, "Ultrahigh-Resolution High-Speed Retinal Imaging Using Spectral-Domain Optical Coherence Tomography," *Optics Express*, 2004, **12**(11):2435–2447.
21. M. Wojtkowski, V. J. Srinivasan, T. H. Ko, J. G. Fujimoto, A. Kowalczyk, and J. S. Duker, "Ultrahigh-resolution, Highspeed, Fourier Domain Optical Coherence Tomography and Methods for Dispersion Compensation," *Optics Express*, 2004, **12**(11):2404–2422.
22. R. A. Leitgeb, L. Schmetterer, W. Drexler, A. F. Fercher, R. J. Zawadzki, and T. Bajraszewski, "Real-time Assessment of Retinal Blood Flow with Ultrafast Acquisition by Color Doppler Fourier Domain Optical Coherence Tomography," *Optics Express*, 2003, **11**(23):3116–3121.
23. B. R. White, M. C. Pierce, N. Nassif, B. Cense, B. H. Park, G. J. Tearney, B. E. Bouma, T. C. Chen, and J. F. de Boer, "In Vivo Dynamic Human Retinal Blood Flow Imaging Using Ultra-Highspeed Spectral Domain Optical Doppler Tomography," *Optics Express*, 2003, **11**(25):3490–3497.
24. M. Wojtkowski, T. Bajraszewski, I. Gorczynska, P. Targowski, A. Kowalczyk, W. Wasilewski, and C. Radzewicz, "Ophthalmic Imaging by Spectral Optical Coherence Tomography," *American Journal of Ophthalmology*, 2004, **138**(3):412–419.
25. T. C. Chen, B. Cense, M. C. Pierce, N. Nassif, B. H. Park, S. H. Yun, B. R. White, B. E. Bouma, G. J. Tearney, and J. F. de Boer, "Spectral Domain Optical Coherence Tomography—Ultrahigh Speed, Ultra-High Resolution Ophthalmic Imaging," *Archives of Ophthalmology*, 2005, **123**(12):1715–1720.
26. E. C. W. Lee, J. F. de Boer, M. Mujat, H. Lim, and S. H. Yun, "In Vivo Optical Frequency Domain Imaging of Human Retina and Choroid," *Optics Express*, 2006, **14**(10):4403–4411.
27. H. Lim, M. Mujat, C. Kerbage, E. C. W. Lee, Y. Chen, T. C. Chen, and J. F. de Boer, "High-Speed Imaging of Human Retina In Vivo with Swept-Source Optical Coherence Tomography," *Optics Express*, 2006, **14**(26):12902–12908.
28. A. Unterhuber, B. Povazay, B. Hermann, H. Sattmann, A. Chavez-Pirson, and W. Drexler, "In Vivo Retinal Optical Coherence Tomography at 1040 nm—Enhanced Penetration into the Choroid," *Optics Express*, 2005, **13**(9):3252–3258.
29. D. M. de Bruin, D. L. Burnes, J. Loewenstein, Y. Chen, S. Chang, T. C. Chen, D. D. Esmaili, and J. F. de Boer, "In Vivo Three-Dimensional Imaging of Neovascular Age-Related Macular Degeneration Using Optical Frequency Domain Imaging at 1050 nm," *Investigative Ophthalmology and Visual Science*, 2008, **49**:4545–4552.
30. A. F. Fercher, W. Drexler, C. K. Hitzenberger, and T. Lasser, "Optical Coherence Tomography—Principles and Applications," *Reports on Progress in Physics*, 2003, **66**(2):239–303.
31. A. B. Vakhtin, K. A. Peterson, W. R. Wood, and D. J. Kane, "Differential Spectral Interferometry: An Imaging Technique for Biomedical Applications," *Optics Letters*, 2003, **28**(15):1332–1334.
32. W. V. Sorin and D. M. Baney, "A Simple Intensity Noise-Reduction Technique for Optical Low-Coherence Reflectometry," *IEEE Photonics Technology Letters*, 1992, **4**(12):1404–1406.
33. L. Mandel and E. Wolf, "Measures of Bandwidth and Coherence Time in Optics," *Proceedings of the Physical Society of London*, 1962, **80**(516):894–897.
34. S. H. Yun, G. J. Tearney, B. E. Bouma, B. H. Park, and J. F. de Boer, "High-Speed Spectral-Domain Optical Coherence Tomography at 1.3  $\mu\text{m}$  Wavelength," *Optics Express*, 2003, **11**(26):3598–3604.
35. S. H. Yun, G. J. Tearney, J. F. de Boer, and B. E. Bouma, "Motion Artifacts in Optical Coherence Tomography with Frequency-Domain Ranging," *Optics Express*, 2004, **12**(13):2977–2998.
36. S. H. Yun, G. J. Tearney, J. F. de Boer, and B. E. Bouma, "Pulsed-Source and Swept-Source Spectral-Domain Optical Coherence Tomography with Reduced Motion Artifacts," *Optics Express*, 2004, **12**(23):5614–5624.



37. G. Moneron, A. C. Boccara, and A. Dubois, "Stroboscopic Ultrahigh-Resolution Full-Field Optical Coherence Tomography," *Optics Letters*, 2005, **30**(11):1351–1353.
38. J. W. You, T. C. Chen, M. Mujat, B. H. Park, and J. F. de Boer, "Pulsed Illumination Spectral-Domain Optical Coherence Tomography for Human Retinal Imaging," *Optics Express*, 2006, **14**(15):6739–6748.
39. S. Bourquin, A. Aguirre, I. Hartl, P. Hsiung, T. Ko, J. Fujimoto, T. Birks, W. Wadsworth, U. Bting, and D. Kopf, "Ultrahigh Resolution Real Time OCT Imaging Using a Compact Femtosecond Nd:Glass Laser and Nonlinear Fiber," *Optics Express*, 2003, **11**(24):3290–3297.
40. M. E. J. van Velthoven, M. H. van der Linden, M. D. de Smet, D. J. Faber, and F. D. Verbraak, "Influence of Cataract on Optical Coherence Tomography Image Quality and Retinal Thickness," *British Journal of Ophthalmology*, 2006, **90**(10):1259–1262.
41. S. H. Yun, C. Boudoux, G. J. Tearney, and B. E. Bouma, "Highspeed Wavelength-Swept Semiconductor Laser with a Polygonscanner-Based Wavelength Filter," *Optics letters*, 2003, **28**:1981–1983.
42. S. L. Jiao, R. Knighton, X. R. Huang, G. Gregori, and C. A. Puliafito, "Simultaneous Acquisition of Sectional and Fundus Ophthalmic Images with Spectral-Domain Optical Coherence Tomography," *Optics Express*, 2005, **13**(2):444–452.
43. M. Mujat, R. C. Chan, B. Cense, B. H. Park, C. Joo, T. Akkin, T. C. Chen, and J. F. de Boer, "Retinal Nerve Fiber Layer Thickness Map Determined from Optical Coherence Tomography Images," *Optics Express*, 2005, **13**(23):9480–9491.
44. X. J. Wang, T. E. Milner, and J. S. Nelson, "Characterization of Fluid-Flow Velocity by Optical Doppler Tomography," *Optics Letters*, 1995, **20**(11):1337–1339.
45. Y. H. Zhao, Z. P. Chen, C. Saxer, S. H. Xiang, J. F. de Boer, and J. S. Nelson, "Phase-Resolved Optical Coherence Tomography and Optical Doppler Tomography for Imaging Blood Flow in Human Skin with Fast Scanning Speed and High Velocity Sensitivity," *Optics Letters*, 2000, **25**(2):114–116.
46. Y. H. Zhao, Z. P. Chen, C. Saxer, Q. M. Shen, S. H. Xiang, J. F. de Boer, and J. S. Nelson, "Doppler Standard Deviation Imaging for Clinical Monitoring of In Vivo Human Skin Blood Flow," *Optics Letters*, 2000, **25**(18):1358–1360.
47. V. Westphal, S. Yazdanfar, A. M. Rollins, and J. A. Izatt, "Realtime, High Velocity-Resolution Color Doppler Optical Coherence Tomography," *Optics Letters*, 2002, **27**(1):34–36.
48. M. R. Hee, D. Huang, E. A. Swanson, and J. G. Fujimoto, "Polarization-Sensitive Low-Coherence Reflectometer for Birefringence Characterization and Ranging," *Journal of the Optical Society of America B-Optical Physics*, 1992, **9**(6):903–908.
49. J. F. de Boer, T. E. Milner, M. J. C. van Gemert, and J. S. Nelson, "Two-Dimensional Birefringence Imaging in Biological Tissue by Polarization-Sensitive Optical Coherence Tomography," *Optics Letters*, 1997, **22**(12):934–936.
50. J. F. de Boer, S. M. Srinivas, A. Malekafzali, Z. P. Chen, and J. S. Nelson, "Imaging Thermally Damaged Tissue by Polarization Sensitive Optical Coherence Tomography," *Optics Express*, 1998, **3**(6):212–218.
51. M. J. Everett, K. Schoenenberger, B. W. Colston, and L. B. Da Silva, "Birefringence Characterization of Biological Tissue by Use of Optical Coherence Tomography," *Optics Letters*, 1998, **23**(3):228–230.
52. M. G. Ducros, J. F. de Boer, H. E. Huang, L. C. Chao, Z. P. Chen, J. S. Nelson, T. E. Milner, and H. G. Rylander, "Polarization Sensitive Optical Coherence Tomography of the Rabbit Eye," *IEEE Journal of Selected Topics in Quantum Electronics*, 1999, **5**(4):1159–1167.
53. G. Yao and L. V. Wang, "Two-Dimensional Depth-Resolved Mueller Matrix Characterization of Biological Tissue by Optical Coherence Tomography," *Optics Letters*, 1999, **24**(8):537–539.
54. X. J. Wang, T. E. Milner, Z. P. Chen, and J. S. Nelson, "Measurement of Fluid-Flow-Velocity Profile in Turbid Media by the Use of Optical Doppler Tomography," *Applied Optics*, 1997, **36**(1):144–149.
55. Z. P. Chen, T. E. Milner, D. Dave, and J. S. Nelson, "Optical Doppler Tomographic Imaging of Fluid Flow Velocity in Highly Scattering Media," *Optics Letters*, 1997, **22**(1):64–66.
56. J. A. Izatt, M. D. Kulkarni, S. Yazdanfar, J. K. Barton, and A. J. Welch, "In Vivo Bidirectional Color Doppler Flow Imaging of Picoliter Blood Volumes Using Optical Coherence Tomography," *Optics Letters*, 1997, **22**(18):1439–1441.
57. A. M. Rollins, S. Yazdanfar, J. K. Barton, and J. A. Izatt, "Realtime In Vivo Color Doppler Optical Coherence Tomography," *Journal of Biomedical Optics*, 2002, **7**(1):123–129.
58. Z. H. Ding, Y. H. Zhao, H. W. Ren, J. S. Nelson, and Z. P. Chen, "Real-Time Phase-Resolved Optical Coherence Tomography and Optical Doppler Tomography," *Optics Express*, 2002, **10**(5):236–245.

59. V. X. D. Yang, M. L. Gordon, B. Qi, J. Pekar, S. Lo, E. Seng-Yue, A. Mok, B. C. Wilson, and I. A. Vitkin, "High Speed, Wide Velocity Dynamic Range Doppler Optical Coherence Tomography (Part I): System Design, Signal Processing, and Performance," *Optics Express*, 2003, **11**(7):794–809.
60. S. Yazdanfar, A. M. Rollins, and J. A. Izatt, "Imaging and Velocimetry of the Human Retinal Circulation with Color Doppler Optical Coherence Tomography," *Optics Letters*, 2000, **25**(19):1448–1450.
61. S. Yazdanfar, A. M. Rollins, and J. A. Izatt, "In Vivo Imaging of Human Retinal Flow Dynamics by Color Doppler Optical Coherence Tomography," *Archives of Ophthalmology*, 2003, **121**(2):235–239.
62. M. D. Kulkarni, T. G. van Leeuwen, S. Yazdanfar, and J. A. Izatt, "Velocity-Estimation Accuracy and Frame-Rate Limitations in Color Doppler Optical Coherence Tomography," *Optics Letters*, 1998, **23**(13):1057–1059.
63. R. Leitgeb, L. F. Schmetterer, M. Wojtkowski, C. K. Hitzenberger, M. Sticker, and A. F. Fercher, "Flow Velocity Measurements by Frequency Domain Short Coherence Interferometry," *Proceedings of SPIE*, 2002, 4619.
64. J. F. de Boer, C. E. Saxer, and J. S. Nelson, "Stable Carrier Generation and Phase-Resolved Digital Data Processing in Optical Coherence Tomography," *Applied Optics*, 2001, **40**(31):5787–5790.
65. C. K. Hitzenberger, E. Gotzinger, M. Sticker, M. Pircher, and A. F. Fercher, "Measurement and Imaging of Birefringence and Optic Axis Orientation by Phase Resolved Polarization Sensitive Optical Coherence Tomography," *Optics Express*, 2001, **9**(13):780–790.
66. C. E. Saxer, J. F. de Boer, B. H. Park, Y. H. Zhao, Z. P. Chen, and J. S. Nelson, "High-Speed Fiber-Based Polarization-Sensitive Optical Coherence Tomography of In Vivo Human Skin," *Optics Letters*, 2000, **25**(18):1355–1357.
67. B. H. Park, C. Saxer, S. M. Srinivas, J. S. Nelson, and J. F. de Boer, "In Vivo Burn Depth Determination by High-Speed Fiberbased Polarization Sensitive Optical Coherence Tomography," *Journal of Biomedical Optics*, 2001, **6**(4):474–479.
68. M. C. Pierce, B. H. Park, B. Cense, and J. F. de Boer, "Simultaneous Intensity, Birefringence, and Flow Measurements with High-Speed Fiber-Based Optical Coherence Tomography," *Optics Letters*, 2002, **27**(17):1534–1536.
69. B. H. Park, M. C. Pierce, B. Cense, S. H. Yun, M. Mujat, G. J. Tearney, B. E. Bouma, and J. F. de Boer, "Real-time Fiberbased Multifunctional Spectral-Domain Optical Coherence Tomography at 1.3  $\mu\text{m}$ ," *Optics Express*, 2005, **13**(11):3931–3944.
70. S. L. Jiao, G. Yao, and L. H. V. Wang, "Depth-Resolved two-dimensional Stokes Vectors of Backscattered Light and Mueller Matrices of Biological Tissue Measured with Optical Coherence Tomography," *Applied Optics*, 2000, **39**(34):6318–6324.
71. S. L. Jiao and L. H. V. Wang, "Jones-matrix Imaging of Biological Tissues with Quadruple-Channel Optical Coherence Tomography," *Journal of Biomedical Optics*, 2002, **7**(3):350–358.
72. S. L. Jiao and L. H. V. Wang, "Two-dimensional Depth-Resolved Mueller Matrix of Biological Tissue Measured with Double-Beam Polarization-Sensitive Optical Coherence Tomography," *Optics Letters*, 2002, **27**(2):101–103.
73. S. L. Jiao, W. R. Yu, G. Stoica, and L. H. V. Wang, "Optical-fiber-Based Mueller Optical Coherence Tomography," *Optics Letters*, 2003, **28**(14):1206–1208.
74. B. H. Park, M. C. Pierce, B. Cense, and J. F. de Boer, "Jones Matrix Analysis for a Polarization-Sensitive Optical Coherence Tomography System Using Fiber-Optic Components," *Optics Letters*, 2004, **29**(21):2512–2514.
75. J. Strasswimmer, M. C. Pierce, B. H. Park, V. Neel, and J. F. de Boer, "Polarization-Sensitive Optical Coherence Tomography of Invasive Basal Cell Carcinoma," *Journal of Biomedical Optics*, 2004, **9**(2):292–298.
76. B. Cense, T. C. Chen, B. H. Park, M. C. Pierce, and J. F. de Boer, "In Vivo Depth-Resolved Birefringence Measurements of the Human Retinal Nerve Fiber Layer by Polarization-Sensitive Optical Coherence Tomography," *Optics Letters*, 2002, **27**(18):1610–1612.
77. B. Cense, T. C. Chen, B. H. Park, M. C. Pierce, and J. F. de Boer, "Thickness and Birefringence of Healthy Retinal Nerve Fiber Layer Tissue Measured with Polarization-Sensitive Optical Coherence Tomography," *Investigative Ophthalmology and Visual Science*, 2004, **45**(8):2606–2612.
78. D. Fried, J. Xie, S. Shafi, J. D. B. Featherstone, T. M. Breunig, and C. Le, "Imaging Caries Lesions and Lesion Progression with Polarization Sensitive Optical Coherence Tomography," *Journal of Biomedical Optics*, 2002, **7**(4):618–627.
79. R. C. Jones, "A New Calculus for the Treatment of Optical Systems I. Description and Discussion of the Calculus," *Journal of the Optical Society of America A*, 1941, **31**(7):488–493.

80. J. J. Gil, and E. Bernabeu, "Obtainment of the Polarizing and Retardation Parameters of a Nondepolarizing Optical System from the Polar Decomposition of its Mueller Matrix," *Optik*, 1987, **76**(2):67–71.
81. K. Schoenberger, B. W. Colston, D. J. Maitland, L. B. Da Silva, and M. J. Everett, "Mapping of Birefringence and Thermal Damage in Tissue by Use of Polarization-Sensitive Optical Coherence Tomography," *Applied Optics*, 1998, **37**(25):6026–6036.
82. G. J. van Blokland, "Ellipsometry of the Human Retina In Vivo: Preservation of Polarization," *Journal of the Optical Society of America A*, 1985, **2**:72–75.
83. H. B. K. Brink and G. J. van Blokland, "Birefringence of the Human Foveal Area Assessed In Vivo with Mueller-Matrix Ellipsometry," *Journal of the Optical Society of America A*, 1988, **5**(1):49–57.
84. W. K. Tung, *Group Theory in Physics*, 1985, Singapore World Scientific.
85. B. H. Park, M. C. Pierce, and J. F. de Boer, "Comment on Optical-Fiber-Based Mueller Optical Coherence Tomography," *Optics Letters*, 2004, **29**(24):2873–2874.
86. B. H. Park, M. C. Pierce, B. Cense, and J. F. de Boer, "Realtime Multifunctional Optical Coherence Tomography," *Optics Express*, 2003, **11**(7):782–793.
87. B. Cense, H. C. Chen, B. H. Park, M. C. Pierce, and J. F. de Boer, "In Vivo Birefringence and Thickness Measurements of the Human Retinal Nerve Fiber Layer Using Polarization-Sensitive Optical Coherence Tomography," *Journal of Biomedical Optics*, 2004, **9**(1):121–125.
88. M. Pircher, E. Gotzinger, R. Leitgeb, H. Sattmann, O. Findl, and C. K. Hitzenberger, "Imaging of Polarization Properties of Human Retina In Vivo with Phase Resolved Transversal PS-OCT," *Optics Express*, 2004, **12**(24):5940–5951.
89. E. Gotzinger, M. Pircher, and C. K. Hitzenberger, "High Speed Spectral Domain Polarization Sensitive Optical Coherence Tomography of the Human Retina," *Optics Express*, 2005, **13**(25):10217–10229.
90. M. Mujat, B. H. Park, B. Cense, T. C. Chen, and J. F. de Boer, "Autocalibration of Spectral-Domain Optical Coherence Tomography Spectrometers for In Vivo Quantitative Retinal Nerve Fiber Layer Birefringence Determination," *Journal of Biomedical Optics*, 2007, **12**(4).
91. H. Lim, J. F. de Boer, B. H. Park, E. C. W. Lee, R. Yelin, and S. H. Yun, "Optical Frequency Domain Imaging with a Rapidly Swept Laser in the 815-870 nm Range," *Optics Express*, 2006, **14**(13):5937–5944.

---

# GRADIENT INDEX OPTICS IN THE EYE

---

Barbara K. Pierscionek

*Department of Biomedical Sciences  
University of Ulster  
Coleraine, United Kingdom*

---

## 19.1 GLOSSARY

---

**Cataract.** Opacification in the eye lens caused by changes in protein shape, structure, or interaction that causes light to be scattered or absorbed.

**Crystallin proteins.** The major structural proteins in the eye lens.

**Equatorial plane.** The plane of the eye lens that contains the lens equator.

**Gradient index or GRIN.** The property of a lens or medium that has a gradually varying refractive index.

**Homogenous index.** The property of a lens or medium that has a constant refractive index.

**Isoindicial contours.** Contours of constant refractive index.

**Magnetic resonance imaging (MRI).** Noninvasive method used for imaging body organs that utilizes the interaction of a static magnetic field, radio frequency pulses, and magnetic gradients to form images.

**Myopia.** Shortsightedness.

**Optic fiber.** A fiber made of glass or plastic that guides light along its length by total internal reflection.

**Optical fiber perform.** The precursor to an optical fiber: a rod with the refractive index distribution of the desired optical fiber that is stretched to form the fiber.

**Raman microspectroscopy.** A technique used to detect constituents of a system from the spectrum of molecular vibrations created by specific interactions of monochromatic light with matter. These interactions cause scattering of photons at lower frequencies than that of the incident light.

**Ray tracing.** Monitoring the path of rays as they traverse a lens or optical system. This is often used to measure a particular property of the lens or system.

**Reflectometric sensor.** A device that measures optical or material properties of a medium from the proportion of light that the medium reflects.

**Refractive index.** The property of a medium that contributes to its capacity to bend light and that is related to the density of the material.

**Refractive power.** The capacity of a lens or optical system to bend and focus light.

**Sagittal plane.** The vertical plane of the eye lens that contains the optic axis.

## 19.2 INTRODUCTION

This chapter considers the common forms of gradient index in optical media. The lens of the eye has a gradient of index and this is examined in detail; the form of the index profile as well as the magnitude of refractive index are described. The refractive index distribution has been measured or derived from a number of animal eyes as well as from the human eye. The different types of gradients are presented and compared.

## 19.3 THE NATURE OF AN INDEX GRADIENT

The refractive power of a lens is determined by its shape, the refractive index of its medium, and the refractive index of the medium that surrounds the lens. In terms of shape, the more curved the lens the greater is the refraction of light. The relationship between refractive index and refractive power is not as direct: light is not necessarily refracted to a greater degree in a higher-index medium. A ray of light will travel in a straight line whether it travels in air (refractive index = 1), in water (refractive index = 1.33), or in glass (refractive index around 1.5 to 1.6). The effect of refractive index on refraction occurs when there is a change in refractive index at the interface between two media or a gradual variation in index within a medium. The greater the difference between refractive indices at an interface or the steeper the gradient of refractive index in a medium, the greater the degree of refraction.

Gradient index (GRIN) deals with media which have a varying refractive index. The effect of varying or gradient index occurs in nature; changes in temperature or pressure can induce fluctuations or gradations in the refractive index of air. One of the most commonly cited examples of this is seen in the shiny reflective marks on a road that can be seen on a hot day. The heat of the road warms the air closest to it and gradients of temperature and of refractive index are formed. The higher the temperature, the less dense the air and the lower the refractive index. The index gradient has the effect of deflecting light that hits the surface of the road at certain angles so that it appears to have been reflected from the surface of the road.

The steepness of the index gradient controls the degree of bending of light in the GRIN medium and this is used in the creation of GRIN lenses and devices. A refractive contribution from the medium is particularly effective when the surface of the given structure is flat and cannot contribute to the refractive power.

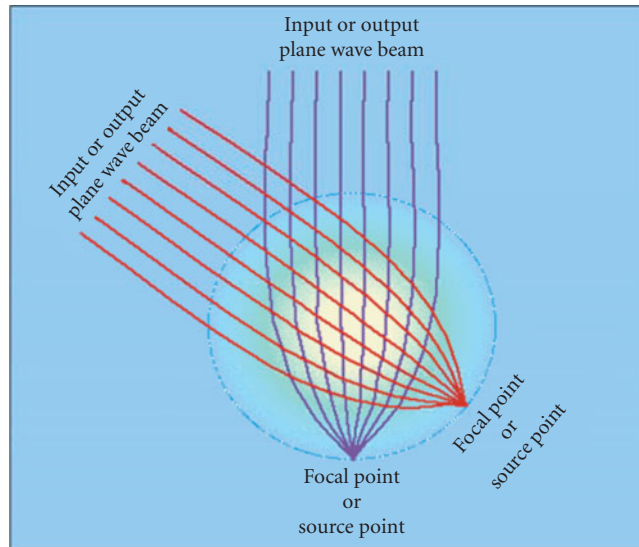
## 19.4 SPHERICAL GRADIENTS

In a lens with a spherical refractive index distribution, the index varies with distance from the center of the lens. The most famous early GRIN lens was Maxwell's fish-eye lens<sup>1</sup> that has a refractive index distribution with spherical symmetry about a point and can be described as having the form

$$n(r) = \frac{n_0}{1 + (r/a)^2} \quad (1)$$

where  $r$  is the distance from the center,  $n(r)$  is the refractive index at  $r$ ,  $n_0$  is the refractive index at the center of the lens, and  $a$  is a constant.

The special property of a fish-eye lens is that all rays emanating from a given object point pass through the same image point.<sup>2</sup> This produces sharp imagery but only for those points on the surface and within the lens. There is no evidence that fish possess such lenses. The name possibly derives from the challenge posed by the Irish Academy of Sciences: to produce a medium with a refractive index distribution that was capable of forming images with the least depth possible. As the eye of the fish is



**FIGURE 1** Light paths through a Luneberg lens in surround media index matched to the lens surface.

known to be relatively flat, Maxwell's solution to the problem was called the fish-eye.<sup>3</sup> The limitation of this lens is that sharp imagery of object points only occurs for those points that lie within or on the surface of the lens.<sup>2,4</sup>

This restriction does not apply to the Luneberg lens,<sup>5</sup> which, like the Maxwell fish-eye lens, has a refractive index distribution that is spherically symmetrical. This lens is a sphere with a refractive index distribution described by

$$n(r) = (2 - r^2)^{1/2} \quad (2)$$

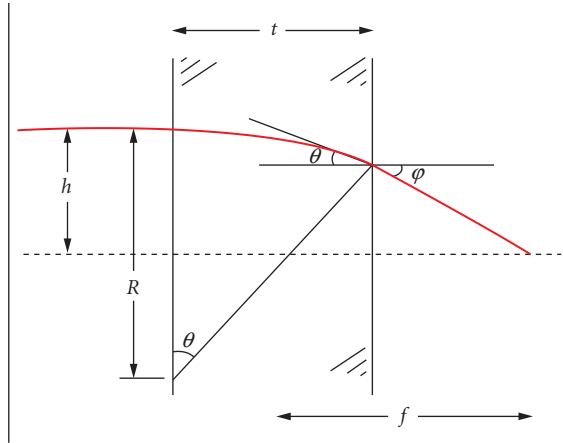
where  $r$  is the distance from the center of the lens and  $n(r)$  is the refractive index at  $r$ .

It has no intrinsic optical axis but focuses parallel rays, incident on any part of its surface, to a point on the opposite surface of the lens (Fig. 1).

The restriction of the Luneberg lens is that with this lens sharp images are only created for objects at infinity. In order to ensure that rays enter in a parallel bundle, the index of the surrounding media must match that of the lens surface. In air this would require an outer surface refractive index of 1, which is difficult to achieve for visible light but can be utilized for frequencies in the microwave range.<sup>2</sup> It has wide ranging applications in microwave antennas and as a radar device. A very useful modern application is in the design of broadband airborne antennas. These allow passengers on an aircraft to access the Internet, email, satellite television, and video.

## 19.5 RADIAL GRADIENTS

When the refractive index distribution is cylindrically symmetrical, that is, the refractive index varies with distance from a fixed line, it is described as a radial index gradient. A simple type of radial index lens is the Wood lens, named after its designer.<sup>6</sup> This lens has a radial refractive index gradient and flat surfaces.<sup>2</sup> In such a lens, parallel rays from objects at infinity are refracted purely by the



**FIGURE 2** Refraction of a ray through a Wood lens.  $R, f \gg h$ ,  $t$ . [From N. Morton, "Gradient Refractive Index Lenses," *Phys. Educ.* **19**:86–90 (1984).]

index distribution within the lens (Fig. 2). When the refractive index decreases with distance from the axis, the lens acts as a converging lens; when the refractive index increases with distance from the axis, the lens diverges light. With a Wood lens, the object and image planes are external to the lens.<sup>2</sup>

Optical fiber cores with radial index gradients are, cross-sectionally, akin to a Wood lens. However, because optical fibers are elongated in the axial direction and have very small diameters, there is an additional periodicity to the ray path. In a graded index optical fiber core, the index distribution follows a general form<sup>7</sup>

$$n(r) = n_0[1 - 2(r/a)^\alpha \Delta]^{1/2} \tag{3}$$

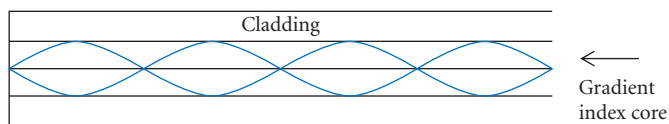
where  $r$  is the radial distance from the center,  $n(r)$  is the refractive index at  $r$ ,  $n_0$  is the refractive index along the axis of the fiber,  $a$  is the fiber radius, and  $\Delta$  is the fractional refractive index difference between the center and edge of the fiber core.

For  $\alpha = 2$  and  $\Delta \ll 1$ , the profile is parabolic and approximated as

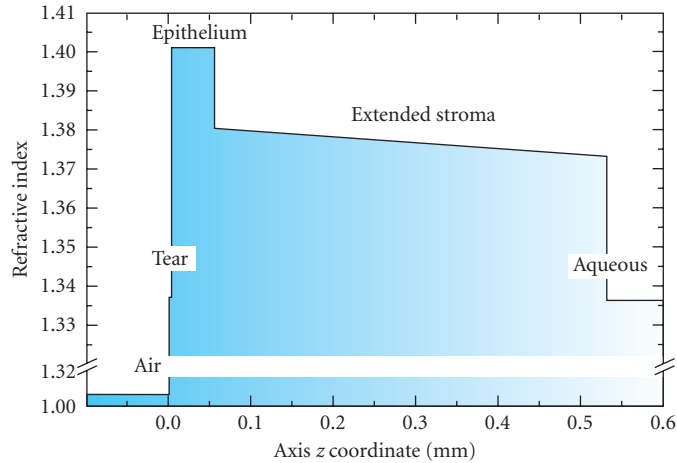
$$n(r) = n_0[1 - (r/a)^2 \Delta] \tag{4}$$

In an optical fiber, the core is surrounded by a single index cladding to minimize light losses along the length of the fiber and to protect the quality of the fiber core. Within the core, the rays follow a sinusoidal path along the axis of the fiber (as seen in Fig. 3).

Optical fibers have many uses particularly when the light needs to be sent some distance from the source or an image needs to be collected from a remote or hard to access location. Optical fibers have replaced wires in communication systems. They are now used routinely in medical instruments and applications, such as endoscopes and sensors.



**FIGURE 3** Light rays follow a sinusoidal path when traveling in the gradient index core of a GRIN fiber.



**FIGURE 4** Variations in corneal refractive index along the optical axis. (From Ref. 10.)

## 19.6 AXIAL GRADIENTS

Refractive index can also vary as a function of distance from a reference plane. This is most often the plane perpendicular to the optic axis and in such case the refractive index varies along the optic axis. Such a gradient can be formed by thin films and has important use in the design of multilayer antireflection coatings.<sup>8,9</sup> Southwell<sup>9</sup> has shown that shape of the index gradient along the axis is a very important determinant that can be used to minimize adverse reflections. Measurement of axial gradients has application in the study of biological tissues, many of which are layered. The lamellar stroma, which constitutes the bulk of the cornea of the eye and has been considered to have a homogenous refractive index, has recently been modeled as an axial index structure<sup>10</sup>. The axial index model has been found to produce more spherical aberration, for larger pupil sizes, than homogenous index models. Any refractive index differences along the axis of the cornea are likely to be very small and the model introduced very slight variations (Fig. 4).

## 19.7 THE EYE LENS

The refractive index of the eye lens has a radial as well as an axial gradient. This is because of the three-dimensional structure as well as the nature of the growth mode of the lens. The lens grows by tissue accretion in layers on the lens surface, with no concomitant loss of existing cells, resulting in a layered structure. The refractive index varies across the layers resulting in a radial gradient, in the equatorial plane and an axial gradient along the optic axis. In spherical lenses (seen in some fish), the index gradient varies along a radial distance from the center in all directions.

The shapes and sizes of eye lenses, like any other part of the eye, vary widely across species. This variation bears no relationship to the size of the animal nor to the size of the eye. For example, the lens of a rabbit is larger than that of a human, while the rat's lens is highly curved and fills the eye almost entirely. With such a variety of lenticular shapes and sizes, it is not surprising that the refractive index distribution also varies between species and is not related to the size of the animal or to its eye but to its visual needs.



The refractive index gradients in human and several types of animal lenses have been measured or derived using various techniques. These have included ray tracing, optic fiber sensing, or calculation from protein and water concentration measurements and MRI scanning. Earlier studies on dispersion have provided refractive index measurements on inner and outer sections of lenses from a number of vertebrates.<sup>11,12</sup>

## 19.8 FISH

The fish lens contributes more to the refractive power of the eye than lenses from land living animals because the corneal power in the fish eye is significantly reduced by the very small difference in refractive index between the cornea and water. Fish lenses are spherical or nearly spherical with steep refractive index gradients and high maximal index values to provide the necessary refractive power to the eye whilst minimizing spherical aberration.<sup>13–20</sup> The refractive index gradients in fish lenses have been fitted with elliptical,<sup>14</sup> parabolic,<sup>16–19</sup> and sixth-order polynomial functions (Table 1).<sup>20</sup>

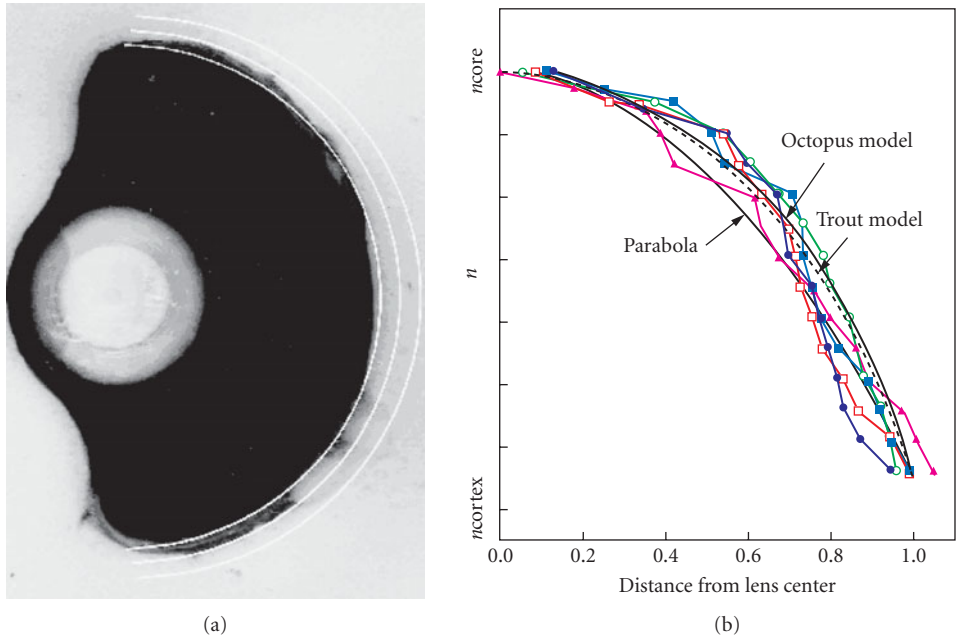
In spite of the differences in the size of fish and their lenses, the maxima of refractive index in the lens centers are very close in value and, with the exception of the Blue Eye Trevally lens, similar values are also found at the lens edge (Table 1). The edge index value in the trevally lens is substantially higher than that found in the other fish lenses. It is also notable that the trevally has an enormous eye relative to its size and its lens is the largest of any lens studied: a diameter of 26.6 mm and a weight of over 9.5 g.<sup>17</sup> It is a deep-sea fish, unlike the other fish listed in Table 1. From a structural aspect, almost a third of the trevally lens protein comprises  $\gamma$ -crystallin.<sup>17</sup> This crystallin protein contributes more to the refractive index of the lens than the other crystallins<sup>21</sup> and is capable of closer protein packing with less surrounding water.<sup>22</sup> Although the proportion of  $\gamma$ -crystallin is comparatively high, it is not high enough to alone account for the refractive index magnitude. This may also be attributed to the very high proportion of insoluble protein that was found on extraction.<sup>17</sup> This does not necessarily indicate that the proteins are in an insoluble state in the intact trevally lens but may suggest a propensity for proteins in this lens to aggregate with no loss of transparency. This will depend on protein types and relative distributions as well as protein-protein and protein-water interactions. It is not known whether the high proportions of  $\gamma$ -crystallin and insoluble protein found in the trevally lens apply to the lenses of other fish.

Ageing changes in the fish lens have not been studied but the effect on the index gradient with size and therefore growth has been investigated. Kroger et al.<sup>19</sup> reported that the refractive index in the center of the African Cichlid fish lens increases with lens size. The resultant steepening of the index gradient with growth causes a decrease in focal length. There is also a concomitant decrease in spherical aberration that is attributed to a smoothing of the index gradient.<sup>19</sup> It has been suggested that depressions in the index profile may have a discrete function in producing multiple focal points.<sup>19</sup>

Jagger<sup>15</sup> compared models of fish lenses and found that a higher-order polynomial function to describe the index gradient provided results that were most akin to measurements for fish lenses in terms of image quality, spherical, and chromatic aberration. It has also been pointed out that there may be internal asymmetries in the index distribution in fish lenses that would render erroneous measurements based on ray tracing assumptions.<sup>23</sup>

**TABLE 1** Details of Refractive Index Profiles in a Variety of Fish Lenses

Fish	Refractive Index (max)	Refractive Index (min)	Wavelength (nm)	Gradient Shape
Goldfish <sup>14</sup>	1.55–7	1.35–8	633	Elliptical
African Cichlid <sup>16</sup>	1.540	1.361	633	Parabolic
Blue Eye Trevally <sup>17</sup>	1.546	1.408	633	Parabolic
Trout <sup>18</sup>	1.538	1.3716	550	Parabolic
Black Dory <sup>20</sup>	1.543	1.368	590	Sixth-order polynomial



**FIGURE 5** (a) Octopus eye showing lens and (b) refractive index gradient for five octopus lenses as measured in the equatorial plane. Model and experimental index profiles are shown together with the index profile for the trout lens. (From Ref. 24.)

## 19.9 OCTOPUS

The eye of the octopus has no cornea; the refractive power of the eye is provided by the lens (Fig. 5a).<sup>24</sup> It was found to have concentric isoindicial contours and an index gradient that could be fitted to a 10th order polynomial (Fig. 5b). The maximum refractive index was 1.509 decreasing to 1.357 for 550 nm.

## 19.10 RAT

The refractive index distribution in the rat lens was measured using ray tracing<sup>25,26</sup> following the method for determining the refractive index in optical fiber performs devised by Chu.<sup>27</sup> This method enabled the calculation of the gradient refractive index profile using, as parameters, the distance of the parallel incident rays from the optic axis and the corresponding angles formed by the exiting rays and the optic axis.

The studies of Campbell and Hughes<sup>25</sup> and Campbell<sup>26</sup> showed that the rat has the steepest refractive index gradient of all species yet measured, rising from around 1.38 at the edge of the lens to a maximum of around 1.5 at the lens center in a profile that can be fitted to a parabolic form. This rise occurs over a very short distance of 2 mm.<sup>26</sup> A refractive index value of 1.50 is approaching the refractive index range for glass and, in terms of protein concentration, is close to the refractive index of pure protein.<sup>28</sup> With such a high protein concentration, the rat lens cannot be deformed, and hence the rat eye has no accommodative capacity. The high magnitude of refractive index in the rat lens is

partially caused by the relatively high proportion of the protein  $\gamma$ -crystallin. In terms of shape, the rat lens is not quite spherical but tends to have a longer equatorial diameter than sagittal radius with little variation between the anterior and posterior curvatures.<sup>26</sup> The radial and axial gradient profiles will not be the same, the latter slightly less steep than the former. The curved nature of the rat lens, its high magnitude of refractive index, and the steep index gradient gives the lens a very high refractive power rendering the rat eye highly myopic. This accords with the animal's visual needs which are for very close distances and do not require a range of focus.

## 19.11 GUINEA PIG

The refractive index distribution of the guinea pig lens, measured with an optical fiber reflectometric sensor, has been found to be parabolic.<sup>29</sup> This was for a single lens, and the age of the animal was not known. Hence magnitudes of refractive index: a maximum value of 1.429 at the lens center and a minimum value of 1.389 at the edge can only be taken as applicable for that particular lens. Calculated protein content for the guinea pig lens was 48 percent<sup>29</sup> and the lens shape is almost spherical with a sagittal width of around 5 mm and an equatorial radius of 5.4 mm. The refractive index distribution will therefore be similar in the axial and radial directions. Because the guinea pig lens is relatively small, its refractive index gradient is quite steep and this adds to its refractive power. The steep distribution of refractive index may be linked to the structural proteins and the uncommon crystallin protein ( $\zeta$ -crystallin) that comprises about 10 percent of the total soluble proteins.<sup>30</sup> This protein differs from the other major crystallins in structure<sup>30,31</sup> and in its response to the effects of heat and the protective action of  $\alpha$ -crystallin.<sup>32</sup>  $\zeta$ -crystallin requires a tenfold greater amount (w/w) of  $\alpha$ -crystallin than the other crystallin proteins to prevent it forming aggregates and disturbing the smoothness of the refractive index gradient.<sup>32</sup> The greater level of protection required by  $\zeta$ -crystallin could indicate a propensity to exist in regions of relatively higher protein concentration. The relatively steep refractive index gradient and the high curvature of the guinea pig lens are needed to meet the refractive requirements of the guinea pig: clear vision at close distances.

## 19.12 RABBIT

The refractive index gradient in the rabbit lens was measured in the equatorial and horizontal planes on thin slices.<sup>33</sup> Isoindicial curves were found as concentric circles in the equatorial plane and as eccentric ellipses in the horizontal plane, the eccentricity arising because of differences in curvature between the anterior and posterior lens surfaces. The three-dimensional refractive index distribution was represented by an elliptical paraboloid

$$n = n_0 - [Z^2/A^2 + (X^2 + Y^2)/B^2] \quad (5)$$

where  $n_0$  is the refractive index at the lens center,  $X$  is the lateral axis (through the middle of the horizontal plane),  $Y$  is the vertical axis (through the middle of the equatorial plane),  $Z$  is the optic axis,  $A$  and  $B$  are constants ( $A$  differs in the anterior and posterior sections of the lens).

The refractive index for 589 nm was reported to vary from 1.464 to 1.431 at the lens center and 1.393 to 1.384 at the surface, which is not very different from the magnitudes of refractive index in the guinea pig lens. However, the rabbit lens is larger than that of the guinea pig (around 10 mm thickness in the horizontal and sagittal planes) and so the index gradient in the rabbit lens would not be as steep providing a lower contribution to refractive power than does the index gradient in the guinea pig lens.

### 19.13 CAT

The lens of the cat is shaped as a biconvex ellipsoid<sup>34</sup> and its refractive index follows a parabolic shape with a maximum of 1.4813 at the center and minimum of 1.3877 in the periphery (values averaged from five cat lenses) for a wavelength 550 nm. The parabolic shape was found in both the sagittal (axial) and equatorial planes and the index variations follow isoindicial contours that retain the geometric shape of the lens.<sup>34</sup>

Jagger<sup>34</sup> cites Matthiessen (1880),<sup>35</sup> who showed the index in a homogenous lens would give equivalent power to one with a variation of index (based on a parabolic gradient). This rule states that the difference between the total index and the core (maximum) index in a grin lens is equal to the difference between the core index and cortical (edge) index

$$n(\text{total}) - n(\text{core}) = n(\text{core}) - n(\text{cortex})$$

so

$$n(\text{total}) = 2 \times n(\text{core}) - n(\text{cortex})$$

According to the rule of Matthiessen,<sup>35</sup> the equivalent to a lens with a parabolic gradient index, with a maximum value of 1.48 and an edge value of 1.38, would require a homogenous index lens to have a refractive index of 1.58. This is higher than the refractive index of pure protein.<sup>28</sup>

### 19.14 BOVINE

The optical properties of the bovine lens have been investigated in a number of studies and the refractive index has been measured using ray tracing<sup>36,37</sup> and fiber optic reflectometric sensing methods.<sup>38</sup> In addition, the refractive index in the bovine lens has been measured in both equatorial and sagittal planes and in lenses from a wide age range from foetal<sup>39</sup> to approximately 15 years of age.<sup>36</sup>

The refractive index, measured in the equatorial plane using a ray tracing technique for a wavelength of 633 nm<sup>36,37</sup> was found to have some variations depending on the age of the lens. In all lenses the gradient followed the same general shape: that of a second-order polynomial, but the index magnitude increased across the gradient with age. The size and therefore the radius of the lens in the equatorial plane had to be incorporated into the equation to account for size and age differences. The general form of the refractive index gradient covering the wide range of lenses investigated is

$$n(r) = a_r - N_r/\rho^2 + C \quad (6)$$

where  $r$  is the radial distance from the center of the lens,  $n(r)$  is the refractive index at  $r$  (mm),  $a_r$  is the refractive index at the center of the oldest lens investigated (approximately 15 years of age),<sup>36</sup>  $N_r$  is a non-normalized gradient function,  $\rho$  is the equatorial radius of the lens (mm), and  $C$  is a correction factor for age-related differences in the refractive index at the lens center.

With

$$N_r = 0.702 - 0.471r + 0.186r^2 \quad (7)$$

and

$$C = -0.0025 - 0.00583r + 0.00131r^2 \quad (8)$$

the final form of the equation is

$$n(r) = 1.4575 - 0.702/\rho^2 + 0.471r/\rho^2 - 0.186r^2/\rho^2 - 0.00583r + 0.00131r^2 \quad (9)$$

In a subsequent study the refractive index along the optic axis of the bovine lens was measured using a fiber optic sensor and a wavelength of 670 nm.<sup>38</sup> The lens age range was not as great as in the ray tracing study of the equatorial plane index: weights varied from 1.676 to 2.646 g corresponding to ages of around 1 to 8 years, respectively.<sup>40</sup> A second-order polynomial was again found to fit the profiles but here slight differences were found in the anterior and posterior sections of the lens. The anterior axial profile was found to fit the following equation

$$n(x) = 1.3790 + 0.018484x - 0.0014925x^2 \quad (10)$$

and the posterior axial profile could be described by

$$n(x) = 1.4102 + 0.009647x - 0.0010741x^2 \quad (11)$$

where  $x$  is the distance from the anterior pole (mm) and  $n(x)$  is the refractive index at  $r$ .

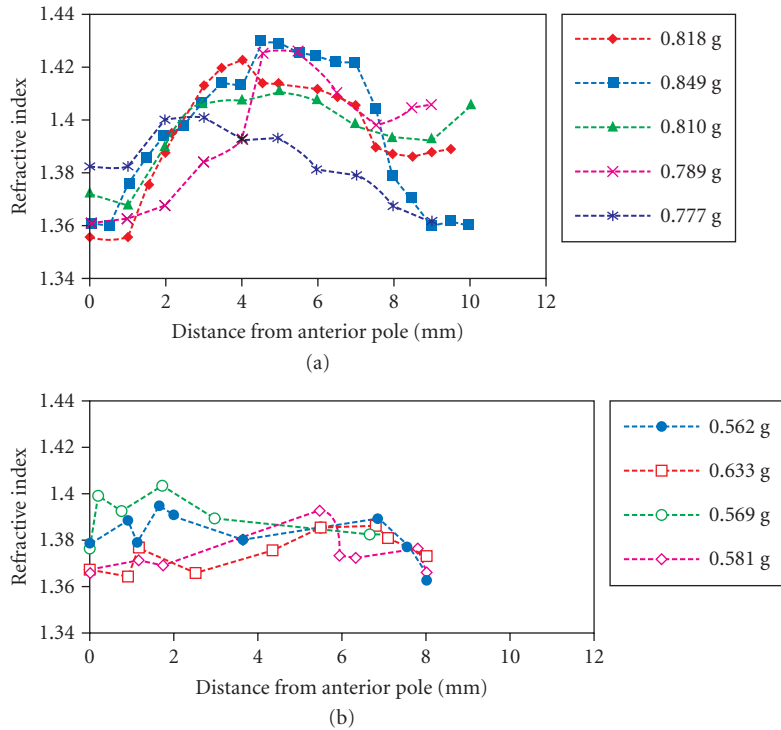
It is not possible to definitively state whether the bovine lens does have contours of refractive index that follow the external shape of the lens but the forms of the gradients in the sagittal and equatorial planes are similar and using dimensional data from the lens it is possible to do approximate calculations to see if the magnitudes of refractive index following a given contour are close. If so, the assumption of concentric contours of refractive index may apply to the bovine lens. From the dimensions of bovine lenses<sup>41</sup> a lens of about 1.5 g weight has an equatorial radius of 7.5 mm, an anterior sagittal thickness of 5 mm and a posterior sagittal thickness of 6 mm.

Taking a contour at 2 mm from the anterior pole, this will be at a radial distance of 4.5 along the equatorial plane (2 mm is 40 percent of the anterior sagittal thickness and 4.5 mm is 40 percent of the equatorial radius) and 2.4 mm from the posterior pole (8.6 mm from the anterior pole). Putting these values into Eqs. (9), (10), and (11) gives refractive index values of 1.4099 (~1.41) along the anterior sagittal thickness, 1.416 along the equatorial plane, and 1.414 along the posterior sagittal thickness, respectively. Taking into account differences in measurement technique, the fact that measurements of radial and axial gradients were conducted on different lenses, and the recognition that in biological tissues there will always be individual variations in lens shape and size, the above calculations can only provide an estimate. This notwithstanding, the values are close and suggest that in the bovine lens isoindicial contours of refractive index may follow the surface shape.

The bovine lens is one of few in which age-related changes to the GRIN optics have been investigated. The changes with age to the index distribution should take into account the fact that the lens continues to grow, and are more correctly described as changes with time and not exclusively occurring as a result of senescence. The lens grows by successive tissue layering and, within each layer, the cells stretch from anterior to posterior pole. As there is an asymmetry between the anterior and posterior surface curvatures, the cells must stretch more and more asymmetrically with growth, and a greater component of the cellular length is found in the posterior part of the lens. How the cellular contents, mainly the proteins, are distributed along the cell length is not known. Nevertheless, as the anterior and posterior refractive index gradients meet in the equatorial plane, the refractive index at that point must be the same. Similarly, refractive indices at the anterior and posterior poles should be close in value, as these are from the same cell layers, and large difference in refractive index between the poles would not meet the refractive needs of the eye. Fiber optic sensing has shown that there is, indeed, very little difference in the refractive index at the anterior and posterior poles of lenses that have a relatively lower weight and were therefore likely to be younger.<sup>38</sup> However, whilst the refractive index does not appear to alter with age at the anterior pole, at the posterior pole a statistically significant correlation with age was found.<sup>38</sup> The consequence of this is that in very old bovine lenses, some differences in polar refractive index may occur. This is most likely to be a consequence of the growth mode and increasing asymmetry of shape that may affect and result in greater differences in anterior and posterior refractive index gradients in advanced age.

As both axial and radial gradients were measured with similar wavelength sources (633 nm for the radial gradient and 670 nm for the axial) wavelength differences would have a negligible impact on the results. Comparison of ray tracing and fiber optic sensing have shown consistent findings.<sup>36,38</sup>

A single study on development of the foetal bovine lens<sup>39</sup> has shown that the refractive index gradient starts to take on its parabolic form around the middle of the gestational period of 9.3 months. Prior to this, the refractive index distribution has no defined shape (Fig. 6).



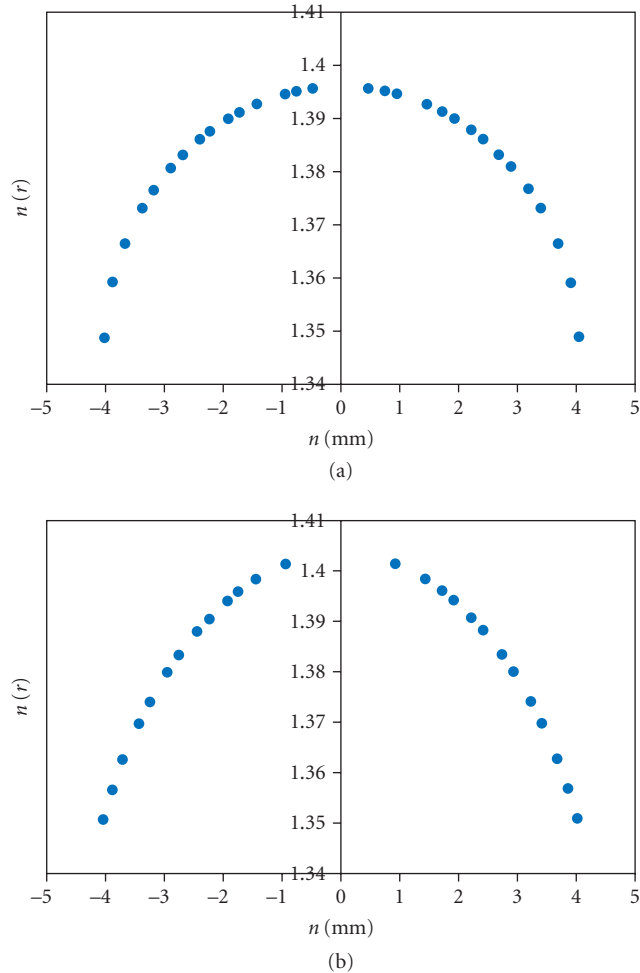
**FIGURE 6** Refractive index profiles along optic axes of foetal bovine lenses. (a) Greater than 0.7 g and (b) lesser than 0.7 g wet weight. (From Ref. 39.)

## 19.15 PIG

The refractive index in the porcine lens is similar in shape to that found in the bovine lens.<sup>42,43</sup> It has been fitted with a second-order polynomial, a shape that has been found from ray tracing measurements of refractive index<sup>42</sup> (as shown as Fig. 7) and from protein concentration profiles determined using Raman microspectroscopy.<sup>44</sup>

Jones and Pope<sup>43</sup> using magnetic resonance imaging (MRI), found that parabolic, power function, and higher-order polynomial fits could be used to provide a very good description of the index profile. They reported that the power function provided the best fit but this was only by a very slight margin. Comparing the results of Pierscionek et al.<sup>42</sup> who measured the refractive index profile in the equatorial plane with those of Jones and Pope,<sup>43</sup> whose measurements were in the sagittal plane, there is very little difference in profile shapes. This suggests that there may be isoindicial contours of refractive index in the porcine lens that follow the surface shape. The porcine lens index gradient has been measured for wavelengths 633 and 532<sup>42</sup> and 589 nm (converted to 543.5 nm).<sup>43</sup> The index magnitude was reported to vary from an average of 1.354 at the edge of the lens up to 1.396 at the center for 633 nm<sup>42</sup> from 1.365 to 1.368 at the edge up to an average of 1.4374 at the center for 543.5 nm,<sup>43</sup> and from 1.358 at the edge to 1.404 at the center for 532 nm. The central index value found by Jones and Pope<sup>43</sup> concurs with measurements made by Sivak and Mandelman.<sup>12</sup>

The higher values of refractive index found by Jones and Pope<sup>43</sup> compared to Pierscionek et al.<sup>42</sup> could have reflected differences in ages of animals; it is not possible to determine whether the two studies<sup>42,43</sup> had used lenses of similar ages. Age-related variations in refractive index have not been studied in the porcine lens.



**FIGURE 7** Refractive index profiles measured in the equatorial plane of porcine lenses shown against radial distance from the center of the lens for (a) 633 nm and (b) 532 nm light. (From Ref. 42.)

## 19.16 HUMAN/PRIMATE

The refractive index gradient of the human lens differs from index gradients in the lenses of most other species. The gradient of refractive index is only found in the cortex of the lens, which is generally taken to represent the outer one-third of the lens, approximately, in terms of radial distance. In the inner, nuclear region, the refractive index is relatively constant around a value of 1.39 to 1.41<sup>12,45,46</sup> (depending on wavelength) and remains so with age as long as there are no opacifications or cataractous changes.<sup>47</sup> This supports findings on protein<sup>48</sup> and water gradients.<sup>49,50</sup> Changes with age occur in the steepness of the cortical index gradient and have been proposed as an answer to why the human lens, in spite of continued growth and increased curvature<sup>51,52</sup> does not become myopic.<sup>29,53</sup> A recent study using MRI has suggested that there is a decrease, with age, in the nuclear

refractive index<sup>54</sup> and that this may explain the fact that eye does not become myopic. However, the results reported are not consistent with what is known about the protein/total water proportions in the human lens and their constancy with age.<sup>29</sup> It is most likely that Moffat et al.<sup>54</sup> were measuring the proportions of free water, which increases in the human lens with age.<sup>55</sup>

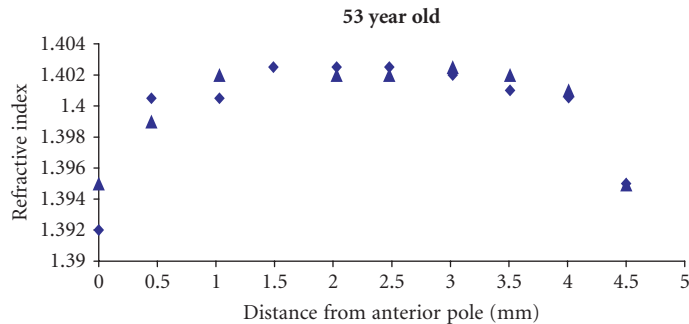
An overall decrease in the nuclear refractive index, would be physiologically unsound, as it would require an imbibing of water or a leakage of protein from the center of the lens. Given that the nuclear refractive index in the human lens is not a part of the gradient index but is constant, an overall decrease in magnitude would have a negligible effect on refractive power of the eye.

The constancy of the nuclear refractive index and the index gradient in the cortex require a higher-order polynomial to fit the index distribution in the human lens. A fourth-order polynomial was found to give a highly significant fit to the refractive index gradient in the equatorial plane of the human lens (Fig. 8).<sup>29</sup>

This form for the radial refractive index gradient (i.e., in the equatorial plane) has been found in studies that used ray tracing (for 633 nm)<sup>45</sup> and for measurements taken with a fiber optic sensor (for 670 nm).<sup>46</sup> The two techniques have also shown results with similar magnitudes for the nuclear refractive index.<sup>45,46</sup>

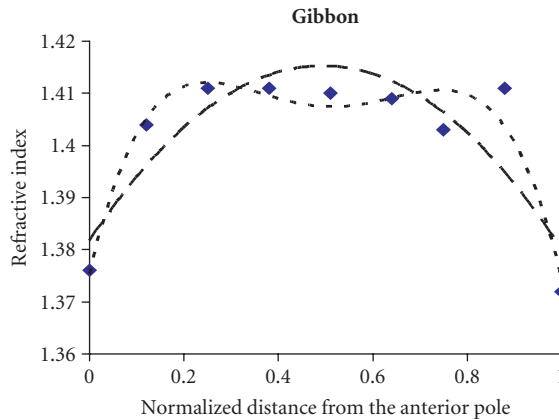
The refractive index along the optic axis similarly has a gradient in the cortical region with a constant index nucleus.<sup>46,47</sup> However, the assumptions of concentric, isoindicial contours that follow the surface shape of the lens, may not apply to younger lenses (from the third decade of life).<sup>46</sup> As with the bovine lens, and consistent with the asymmetry of shape in the sagittal plane, there is a slight difference in the refractive index gradients in the anterior and posterior sections of the lens and this is also affected by lens age.<sup>29</sup>

The difficulty in ascribing a single formulaic description to the refractive index gradient in the human lens is not only complicated by changes to the cortical gradient with age, and by differences in the axial and radial gradients. There is also the confounding factor of accommodation. The human lens, in individuals under the age of around 50, is capable of altering its shape to meet the focusing requirements of the eye. This capacity is greater in younger eyes and it is not certain, what accommodative state a young lens is in when it is removed from the eye. It may not be unreasonable to expect that when the lens is released from the tension applied by the ciliary muscle, in situ, it takes on its most accommodated, that is, most curved shape. This, however, assumes that the lens is an elastic body, with the fully accommodated state: the one to which it returns when applied stretching forces mediated by the ciliary muscle are removed. This assumption has and continues to court controversy and so it remains uncertain what accommodative state the lens is in when removed from the eye. Moreover, given that the capacity for accommodation diminishes with age, the accommodative state of the in vitro lens will vary depending on the age of the lens. How the refractive index gradient alters with accommodation is not known, although some studies have shown that the accommodative process is caused by a change



**FIGURE 8** Variations in refractive index (for 670 nm) along the optic axis of a 53-year-old human lens plotted against the normalized distance from the anterior pole. (From Ref. 29.)





**FIGURE 9** Variations in refractive index (for 670 nm) along the optic axis of the gibbon lens plotted against the normalized distance from the anterior pole. (From Ref. 29.)

in the length of the nucleus<sup>51,56–58</sup> with no change in the thickness of the cortex. This would suggest that the cortical refractive index gradient remains unchanged, the only alteration to the profile being an elongation to the constant index nuclear section.

There is very little information about the refractive index gradients in the lenses of lower-order primates. Results from a single gibbon lens show the same type of profile as for the human lens with a similar index magnitude (Fig. 9).<sup>29</sup>

It is not certain whether the same confounding factor of accommodation applies to the optics and refractive index gradient in the gibbon lens. However, studies on the rhesus monkey have shown that the species does accommodate.<sup>59</sup> As the gibbon belongs to a primate family that is more closely related to humans and its visual needs, from close range feeding to swinging from trees suggests that this species is likely to have an accommodating lens.

From the refractive index gradients, the primate lenses are the only ones that do not have a parabolic gradient but the refractive index gradient flattens to a region of constant refractive index in the nucleus. This constant, maximum value is lower than for other species thus far studied rendering these lenses softer and more pliable. This is consistent with the need for accommodation but also means that the refractive index gradient in a primate lens contributes less to the refractive power than the parabolic gradients, with higher-index magnitudes, found in other species.

## 19.17 FUNCTIONAL CONSIDERATIONS

Gradient index media are found in nature and the concept is exploited in the manufacture of lenses and optical devices for a wide range of applications. The fundamental difference between the biological and the manufactured lenses is that the former is not static in its optical qualities. Within any given species, the refractive index gradient will change with age because the lens continues to grow throughout life and it may change its shape in those species that possess accommodating lenses. Moreover, when considering the eye lens and its optical properties, it should not be considered in isolation. The eye lens functions as part of the optical system of the eye. The rays incident on its anterior surface will have been refracted by the cornea and traversed the anterior chamber. The shapes of the cornea and anterior chamber depths vary from species to species and individual variations within a species can be found. The humours that surround the lens in the eye will also impact on the contribution to refraction by the refractive index gradient. The more

closely the edge of the gradient is matched to the surrounding media, the more contribution is made by the refractive index gradient. No biological lens has yet been shown to have an edge index value matched to the refractive index of the aqueous and vitreous (around 1.336) and, with the techniques of ray tracing and optic fiber sensing, this is the most difficult part of the refractive index gradient to measure.

## 19.18 SUMMARY

The major types of index gradient found in lenses, optical devices, and created by nature have been described. These can be broadly categorized into gradients that vary parallel to the path of light, perpendicular to the path of light, or with radial distance from the center of the structure (found in spherical lenses). The eye lens is an excellent example of a GRIN structure with an index that varies both along and perpendicular to the optic axis. The refractive index distribution of the eye lens is related to its biological organization and to the properties of the major structural entities, the crystallin proteins. The concentration of proteins affects the magnitude of refractive index and the distributions of the different classes influence the shape of the index gradient. In spite of the complex structure of proteins and the changing interactions between protein subclasses and between proteins and water, it has been possible to measure the GRIN structure of the eye lens and in most species examined this appears to have a similar form. Transposing this form to an intraocular implant that cannot only mimic the GRIN structure but also provide some accommodative capacity is the subject of further research.

## 19.19 REFERENCES

1. J. C. Maxwell, *Cambridge Dublin Math Journal* **8**:188 (1854).
2. E. W. Marchand, *Gradient Index Optics*, Academic Press, New York, 1978.
3. A. Frank, F. Leyvraz, and K. B. Wolf, "Hidden Symmetry and Potential Group on the Maxwell Fish-eye," *J. Math. Phys.* **31**:2757–2768 (1990).
4. M. Born and E. Wolf, *Principles of Optics*, Pergamon, New York, 1959.
5. R. K. Luneberg, *Mathematical Theory of Optics*, Brown University, Providence, Rhode Island, 1944, pp. 189–213.
6. R. W. Wood, *Physical Optics*, Macmillan, New York, 1905, p. 72.
7. J. C. Palais, *Optic Fiber Communications*, Prentice Hall, Englewood Cliffs, NJ, 1988, p. 95.
8. P. H. Berning, "Use of Equivalent Films in the Design of Infrared Multilayer Antireflection Coatings," *J. Opt. Soc. Am.* **52**:431–436 (1962).
9. W. H. Southwell, "Gradient-index Antireflection Coatings," *Opt. Lett.* **8**:584 (1983).
10. S. Barbero, "Refractive Power of a Multilayer Rotationally Symmetric Model of the Human Cornea and Tear Film," *J. Opt. Soc. Am. A* **23**:1578–1585 (2006).
11. D. A. Palmer and J. Sivak, "Crystalline Lens Dispersion," *J. Opt. Soc. Am.* **71**:780–782 (1981).
12. J. G. Sivak and T. Mandelman, "Chromatic Dispersion of the Ocular Media," *Vis. Res.* **22**:997–1003 (1982).
13. R. D. Fernald and S. E. Wright, "Maintenance of Optical Quality during Crystalline Lens Growth," *Nature* **301**:618–620 (1983).
14. D. Axelrod, D. Lerner, and P. J. Sands, "Refractive Index within the Lens of a Goldfish Eye Determined from the Paths of Thin Laser Beams," *Vis. Res.* **28**:57–65 (1988).
15. W. S. Jagger, "The Optics of the Spherical Fish Lens," *Vis. Res.* **32**:1271–1284 (1992).
16. R. H. H. Kroger, M. C. W. Campbell, R. Munger, and R. D. Fernald, "Refractive Index Distribution and Spherical Aberration in the Crystalline Lens of the African Cichlid Fish *Haplochromis Burtoni*," *Vis. Res.* **34**:1815–1822 (1994).
17. B. K. Pierscionek and R. C. Augusteyn, "Refractive Index and Protein Distributions in the Blue Eye Trevally Lens," *J. Am. Optom. Assoc.* **66**:739–743 (1995).

18. W. S. Jagger and P. J. Sands, "A Wide-angle Gradient Index Optical Model of the Crystalline Lens and Eye of the Rainbow Trout," *Vis. Res.* **36**:2623–2639 (1996).
19. R. H. H. Kroger, M. C. W. Campbell, and R. D. Fernald, "The Development of the Crystalline Lens is Sensitive to Visual Input in the African Cichlid Fish, *Haplochromis Burtoni*," *Vis. Res.* **41**:549–559 (2001).
20. L. E. Garner, G. Smith, S. Yao, and R. C. Augusteyn, "Gradient Refractive Index of the Crystalline Lens of the Black Oreo Dory (*Alloctytus Niger*): Comparison of Magnetic Resonance Imaging (MRI) and Laser Ray-trace Methods," *Vis. Res.* **41**:973–979 (2001).
21. B. K. Pierscionek, G. Smith, and R. C. Augusteyn, "The Refractive Increments of Bovine  $\alpha$ ,  $\beta$ , and  $\gamma$ -Crystallins," *Vis. Res.* **27**:1539–1541 (1987).
22. C. Slingsby, B. Norledge, A. Simpson, O. A. Bateman, G. Wright, H. P. C. Driessen, P. F. Lindley, D. S. Moss, and B. Bax, "X-ray Diffraction and Structure of Crystallins," *Prog. Retin Eye Res.* **16**:3–29 (1996).
23. P. E. Malkki and R. H. H. Kroger, "Visualization of Chromatic Correction of Fish Lenses by Multiple Focal Lengths," *J. Opt. A; Pure Appl. Opt.* **7**:691–700 (2005).
24. W. S. Jagger and P. J. Sands, "A Wide-angle Gradient Index Optical Model of the Crystalline Lens and Eye of the Octopus," *Vis. Res.* **39**:2841–2852 (1999).
25. M. C. W. Campbell and A. Hughes, "An Analytic Gradient Index Schematic Lens and Eye for the Rat that Predicts Aberrations for Finite Pupil," *Vis. Res.* **21**:1129–1148 (1981).
26. M. C. W. Campbell, "Measurement of Refractive Index in an Intact Crystalline Lens," *Vis. Res.* **24**:409–415 (1984).
27. P. L. Chu, "Non Destructive Measurements of Index Profile of an Optical-fiber Perform," *Electronic Letts.* **13**:736–738 (1977).
28. R. Barer and S. Joseph, "Refractometry of Living Cells. Part 1. Basic Principles," *Quart. J. Microscop. Sci.* **95**:399–423 (1954).
29. B. K. Pierscionek, "Species Variations in the Refractive Index of the Eye Lens and Patterns of Change with Ageing," in O. R. Ioseliani, (ed.), *Focus on Eye Research*, Nova Science Publishers Inc., New York, 2005, pp. 91–116.
30. Q. L. Huang, P. Russell, S. H. Stone, and J. S. Zigler Jr., "Zeta-crystallin a Novel Lens Protein from the Guinea Pig," *Curr. Eye Res.* **6**:725–732 (1987).
31. T. Borrás and A. Rodokanaki, "Primary Structure of Zeta-crystallin Protein from Guinea Pig. Its Similarity to the Enzyme Alcohol Dehydrogenase," *Lens Eye Toxic. Res.* **6**:795–805 (1989).
32. P. V. Rao, J. Horwitz, and J. S. Zigler Jr., "Chaperone-like Activity of  $\alpha$ -Crystallin. The Effect of NADPH on its Interaction with Zeta-crystallin," *J. Biol. Chemistry* **269**:13266–13272 (1994).
33. S. Nakao, S. Fujimoto, R. Nagata, and K. Iwata, "Model of Refractive-Index Distribution in the Rabbit Crystalline Lens," *Opt. Soc. Am.* **58**:1125–1130 (1968).
34. W. S. Jagger, "The Refractive Structure and Optical Properties of the Isolated Crystalline Lens of the Cat," *Vis. Res.* **30**:723–738 (1990).
35. L. Mattheissen, "Untersuchungen über den Aplanatismus und die Periscopie der Kristalllinse in den Augen der Fische," *Pfluegers Archiv.* **21**:287–307 (1880).
36. B. K. Pierscionek, "Growth and Ageing Effects on the Refractive Index in the Equatorial Plane of the Bovine Lens," *Vis. Res.* **29**:1759–1766 (1989).
37. B. K. Pierscionek, D. Y. C. Chan, J. P. Ennis, G. Smith, and R. C. Augusteyn, "A Non-destructive Method of Constructing Three-dimensional Gradient Index Models for Crystalline Lenses: I. Theory and Experiment," *Am. J. Optom. Physiol. Opt.* **65**:481–491 (1988).
38. B. K. Pierscionek, "The Refractive Index along the Optic Axis of the Bovine Lens," *Eye* **9**:776–782 (1995).
39. B. K. Pierscionek, A. Belaidi, and H. H. Bruun, "Optical Development in the Foetal Bovine Lens," *Exp. Eye Res.* **77**:639–641 (2003).
40. O. Hockwin, "Age Changes of Lens Metabolism," *Altern Entwickl* **1**:95–129 (1971).
41. B. K. Pierscionek and R. C. Augusteyn, "Growth Related Changes to Functional Parameters in the Bovine Lens," *Biochim. Biophys. Acta* **1116**:283–290 (1992).
42. B. K. Pierscionek, A. Belaidi, and H. H. Bruun, "Refractive Index Gradient in the Porcine Lens for 532 and 633 nm Light," *Eye* **19**:375–381 (2005).
43. C. E. Jones and J. M. Pope, "Measuring Optical Properties of an Eye Lens Using Magnetic Resonance Imaging," *Magn. Reson. Imaging* **22**:211–220 (2004).

44. C. L. de Korte, A. F. van der Steen, J. M. Thijssen, J. J. Duindam, C. Otto, and G. J. Puppels, "Relation Between Local Acoustic Parameters and Protein Distribution in Human and Porcine Eye Lenses," *Exp. Eye Res.* **59**:617–627 (1994).
45. B. K. Pierscionek and D. Y. C. Chan, "The Refractive Index Gradient of the Human Lens," *Optom. Vis. Sci.* **66**:822–829 (1989).
46. B. K. Pierscionek, "Refractive Index Contours in the Human Lens," *Exp. Eye Res.* **64**:887–893 (1997).
47. B. K. Pierscionek, "Variations in Refractive Index and Absorbance of 670 nm Light with Age and Cataract Formation in Human Lenses," *Exp. Eye Res.* **60**:407–414 (1995).
48. P. P. Fagerholm, B. T. Philipson, and B. Lindstrom, "Normal Human Lens, the Distribution of Protein," *Exp. Eye Res.* **33**:615–620 (1981).
49. A. Huizinga, A. C. C. Bot, F. F. M. de Mul, G. F. J. M. Vrensen, and J. Greve, "Local Variation in Absolute Water Content of Human and Rabbit Eye Lenses Measured by Raman Microspectroscopy," *Exp. Eye Res.* **48**:487–496 (1989).
50. I. Siebinga, G. F. J. M. Vrensen, F. F. M. de Mul, and J. Greve, "Age-related Changes in Local Water and Protein Content of Human Eye Lenses Measured by Raman Microspectroscopy," *Exp. Eye Res.* **53**:233–239 (1991).
51. N. Brown, "The Change in Shape and Internal Form of the Lens of the Eye on Accommodation," *Exp. Eye Res.* **15**:441–459 (1973).
52. N. Brown, "The Change in Lens Curvature with Age," *Exp. Eye Res.* **19**:175–183 (1974).
53. B. K. Pierscionek, "Presbyopia and the Effect of Refractive Index," *Clin. Exp. Optom.* **73**:26–36 (1990).
54. B. A. Moffat, D. A. Atchison, and J. M. Pope, "Age-related Changes in Refractive Index Distribution and Power of the Human Lens as Measured by Magnetic Resonance Micro-imaging in vitro," *Vis. Res.* **42**:1683–1693 (2002).
55. D. Lahm, L. K. Lee, and F. A. Bettelheim, "Age Dependence of Freezable and Nonfreezable Water Content of Normal Human Lenses," *Invest. Ophthalmol. Vis. Sci.* **26**:1162–1165 (1985).
56. B. Patnaik, "A Photographic Study of Accommodative Mechanisms: Changes in Lens Nucleus during Accommodation," *Invest. Ophthalmol.* **6**:601–611 (1963).
57. J. F. Koretz, C. A. Cook, and P. L. Kaufman, "Accommodation and Presbyopia in the Human Eye. Changes in the Anterior Segment and Crystalline Lens with Focus," *Invest. Ophthalmol. Vis. Sci.* **38**:569–578 (1997).
58. M. Dubbelman, G. L. van der Heijde, H. A. Weeber, and G. F. Vrensen, "Changes in the Internal Structure of the Human Crystalline Lens with Age and Accommodation," *Vis. Res.* **43**:2363–2375 (2003).
59. J. F. Koretz, A. M. Bertasso, M. W. Neider, B. True-Gabelt, and P. L. Kaufman, "Slit-lamp Studies of the Rhesus Monkey Eye. II. Changes in Crystalline Lens Shape, Thickness and Position during Accommodation and Aging," *Exp. Eye Res.* **45**:317–326 (1987).

*This page intentionally left blank*

## Edward S. Bennett

*College of Optometry  
University of Missouri  
St. Louis, Missouri*

---

### 20.1 GLOSSARY

---

**Accommodative demand.** This represents the difference between distance and near refractive corrections.

**Aspheric lens design.** An aspherical surface, typically elliptical, often used on the back surface to align with an aspheric anterior cornea. In some presbyopic designs, either a back surface with a hyperbolic or high rate of flattening is used to provide plus power or an aspheric front surface is used for the same purpose.

**Axial ametropia.** Ametropia resulting from the axial elongation of the globe.

**Axial edge lift.** Pertains to the distance between the extension of the tangent of the base curve radius and the absolute edge; it is measured parallel to the lens optical axis or the vertical distance from the lens edge to an extension of the base curve radius.

**Base curve radius (BCR).** The radius of curvature of the central posterior optical section in millimeters.

**Back vertex power (BVP).** Commonly used to express the power of a contact lens, the power is determined from a fixed position with the concave surface of the lens against the lensometer lens stop.

**Bitoric design.** Gas permeable (GP) lens designs incorporating two base curve radii and multiple powers for correction of high corneal astigmatism, typically 2.50 D or greater.

**Contact lens power.** The dioptric power of the contact lens, usually expressed as back vertex power.

**Edge clearance.** The actual distance from the lens edge to the cornea. This value is less than the calculated edge lift due to the asphericity of the anterior cornea.

**Effective power.** The power of the contact lens at the corneal plane.

**Fluoro-silicone/acrylate (F-S/A).** The current generation of rigid gas-permeable contact lenses consisting primarily of fluorine, silicone, and methyl methacrylate.

**Front vertex power (FVP).** The power is determined from a fixed position with the convex surface of the lens against the lensometer lens stop.

**High-order aberrations.** Aberrations induced by the optics of the eye, notably spherical aberration and coma.

**“K.”** This refers to the flatter keratometry reading. GP lenses are generally selected with a base curve radius slightly steeper or flatter than “K.”

**Lacrimal lens.** The tear lens between a GP contact lens and the anterior corneal surface. It is a plus power if the lens exhibits apical clearance; it is minus if the lens exhibits apical bearing.

**Multifocal contact lenses.** Contact lenses that provide more than one correcting power within the lens; these are often used for the correction of presbyopia.

**Optical zone diameter (OZD).** This represents the linear equivalent of the base curve radius and provides the visual optics for both GP and soft lenses.

**Overall diameter.** This represents the linear, chord, edge-to-edge measurement of the lens in millimeters.

**Oxygen permeability (Dk).** The current term to describe the potential of a contact lens material to transmit oxygen which considers both, solubility and diffusion.

**Prism (in contact lenses).** Prism is produced by varying the thickness from the superior to the inferior region of a contact lens while maintaining the same front and back surface curvatures.

**Radial edge lift.** Pertains to the distance between the extension of the tangent of the base curve radius and the absolute edge; it is measured from the lens edge perpendicular to an extension of the base curve radius.

**Refractive ametropia.** Ametropia resulting from abnormal refractive component(s) of the eye.

**Relative spectacle magnification (RSM).** This compares the corrected ametropic retinal image to that of a standard emmetropic schematic eye. Ametropia can be purely axial, resulting from the axial elongation of the globe, refractive, resulting from abnormal refractive component(s) of the eye or a combination of both factors.

**Residual astigmatism.** With spherical GP lenses this pertains to the total ocular astigmatism (defined as the refractive cylinder) minus the corneal (often keratometric) cylinder. With spherical soft lenses, this pertains to the refractive cylinder.

**Rigid GP contact lenses.** Rigid lens materials, smaller in diameter with higher optical quality than soft lenses, which typically incorporate methyl methacrylate, silicone, and fluorine within the polymer matrix.

**Silicone-hydrogel.** The current generation of soft lens materials incorporating silicone within the matrix for higher oxygen permeability.

**Soft toric lenses.** Soft lens designs that correct for refractive astigmatism.

**Spectacle magnification.** This pertains to the ratio of retinal image size of the corrected ametropic eye to the retinal image size of the same eye uncorrected.

---

## 20.2 INTRODUCTION

The optics pertaining to contact lenses have several similarities to those of spectacle lenses. However, the primary differences pertain to such factors as the lesser thickness of contact lenses, the fact they are in direct contact with the tear film—not 11 to 15 mm away from the cornea—the differences present between soft and rigid gas permeable (GP) lens materials, as well as differences in accommodation and convergence effects on the eye. It is evident that, as a result of their optical properties, contact lenses are unique among the forms of correction available to wearers as well as, in many cases, a modality that exceeds other visual correction options in both the quality of vision and the visual freedom present to the wearer. The goal of this chapter is to emphasize the basic contact lens optics commonly used in eye care practices.

## 20.3 CONTACT LENS MATERIAL, COMPOSITION, AND DESIGN PARAMETERS

### Contact Lens Materials

Contact lens materials can be considered either rigid or soft. The first modern rigid lens was the corneal polymethylmethacrylate (PMMA) lens which was popular from the late 1940s to the early 1980s. This material was essentially limited to the PMMA plastic and was, therefore, not oxygen permeable. GP lenses incorporating silicone (for oxygen permeability) as well as stability agents and wetting agents (the latter to help offset the hydrophobic nature of silicone) was introduced in 1979. However, the introduction of fluoro-silicone/acrylate (F-S/A) lenses in the mid 1980s was a major breakthrough and almost all GP lenses in use today are F-S/A materials. The addition of fluorine both assisted in surface wettability via encouraging the tear film mucin to form an even layer on the lens surface but also increased oxygen permeability—through increased oxygen solubility in fluorine—while allowing the silicone content to be reduced. GP materials have many benefits including high quality of vision resulting from the excellent optical quality of these materials and their ability to mold the front surface of the cornea and automatically correct for astigmatism.

Soft (or hydrogel) contact lenses were introduced in 1971. Soft lenses differ from GP lenses as a result of their ability to bind substantial amounts of water.<sup>1</sup> A loose or less tight cross-linking structure allows these materials to range in water content from 25 to greater than 70 percent when equilibrated in normal (0.9 percent) saline.<sup>2</sup> The water content can be expressed by the following equation:<sup>3</sup>

$$\text{Water content} = \frac{\text{Weight of lens water}}{\text{Total weight of hydrated lens}} \times 100$$

Soft lens materials are quite porous relative to GP lenses resulting in a reduction in optical quality which can impact vision. This pore size has been estimated at about 8 Å for low water content materials and 20 to 30 Å for high water content lenses.<sup>4</sup> However, as a result of their larger overall diameter resulting in less movement with the blink, soft lenses have the advantage of being more initially comfortable and easier to adapt to for lens wearers. As soft lenses allow very little oxygen via tear exchange with the blink and the oxygen transmission is achieved primarily through diffusion through the lens, they have not represented a highly successful long-term option for extended wear. However, recently silicone hydrogel lenses have been introduced with oxygen permeability values as much as 5 to 10 times greater than hydrogel lenses and have increased from 0 percent of new fits in the United States in 2002 to over 35 percent in 2006.<sup>5</sup> Silicone hydrogel lens materials combine the high oxygen permeability of silicone with other material characteristics of conventional hydrogel materials that facilitate fluid transport and lens movement.<sup>6</sup> The surfaces of these lenses are treated to make them more hydrophilic and, therefore, more biocompatible. These materials are FDA approved for up to 30 days of continuous lens wear.

### Curvature, Diameter, and Thickness

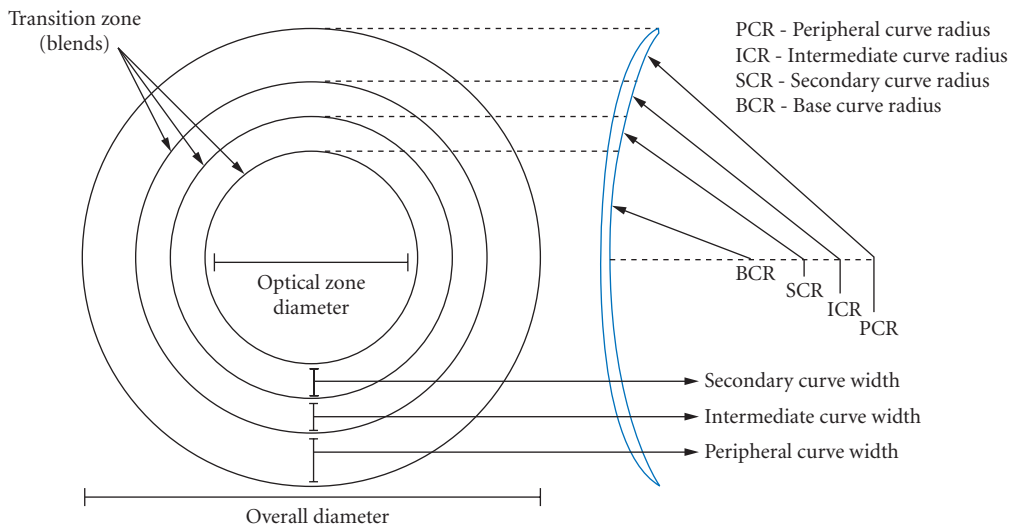
To understand the optics associated with contact lenses, it is important to have a good working knowledge of contact lens design. A brief description of each parameter is discussed below and summarized in Table 1.

**Base Curve Radius (BCR)** The base curve radius of a contact lens is the radius of curvature of the central posterior optical section in millimeters<sup>7</sup> (Fig. 1). This has also been termed the back central optical radius (BCOR). For rigid GP lens materials the base curve radius is typically selected to be approximately the same curvature as the anterior cornea such that the lens will “align” with the anterior corneal surface. With soft or hydrogel lenses, the base curve radius is typically about 1.0 mm flatter than the anterior corneal curvature as they are much larger in diameter and simply drape over



**TABLE 1** Contact Lens Design Parameters

Base Curve Radius	The radius of curvature of the central posterior optical section in millimeters.
Secondary Curve Radius/Width	In a tricurve or tetracurve GP lens design, the secondary curve radius is the curve adjacent to the central base curve radius and is typically flatter than the base curve radius. The width of this curve (often 0.2 to 0.5 mm) represents the linear size of this curve.
Intermediate Curve Radius/Width	In a tetracurve GP lens design, the intermediate curve radius is the curve adjacent to the secondary curve radius and is flatter than the secondary curve radius. The width of this curve (often 0.2 to 0.5 mm) represents the linear size of this curve.
Peripheral Curve Radius/Width	In a tricurve or tetracurve GP lens design, the peripheral curve radius is the outermost curve radius. It is the flattest curve and its parameter is most responsible for the resultant edge clearance (distance from edge to cornea). The width of this curve (often 0.2 to 0.5 mm) represents the linear size of this curve.
Transition Zone Blend	As the junction between two peripheral curves can be rather sharp, laboratories typically blend or smooth these junctions.
Overall Diameter	This represents the linear, chord, edge-to-edge measurement of the lens in millimeters.
Optical Zone Diameter	This represents the linear equivalent of the base curve radius and provides the visual optics for both GP and soft lenses.
Center Thickness	This is the thickness as measured at the exact center of the lens.
Edge Thickness	This is the thickness as measured at the edge of the lens. This value should be similar to the center thickness. As high minus lenses have thick edges, a plus lenticular is often manufactured to thin the anterior edge; likewise for low minus and plus power lenses, a minus lenticular is generated to increase the edge thickness.
Power	Lens power is the dioptric power of the contact lens, usually expressed as back vertex power.



**FIGURE 1** Cross-section of a tetracurve rigid GP contact lens showing overall diameter, optical zone diameter, secondary curve, intermediate curve, and peripheral curve radii and widths and transition zone blends. (From Ref. 30.)

the cornea and align with the adjacent sclera.<sup>8</sup> For GP lenses base curve radius is often specified in millimeters; however, it has also been denoted by its dioptric equivalent. The conversion formula is

$$\text{BCR (in diopters)} = \frac{n' - n}{r}$$

where  $n'$  = index surrounding the lens (although the tear layer is 1.336, the keratometer is calibrated for 1.3375 and this value is used)

$n$  = refractive index of air or 1.0

$r$  = base curve radius in meters

For example, if the base curve radius was verified to be 7.50 mm with the radiuscope, the dioptric equivalent would be  $1.3375 - 1.0/0.0075$  or 45 D. BCR values in diopters and their converted values in millimeters are provided in Table 2.

**Overall Diameter (OAD)/Optical Zone Diameter (OZD)** The linear, chord, edge-to-edge measurement of the lens in millimeters is the overall diameter. GP lenses typically have OAD values between 9.0 to 10.0 mm. Soft or hydrogel lenses typically have OAD values of approximately 14.0 mm. The optical zone diameter is the linear equivalent of the base curve radius and is the central section of the lens which provides usable optics for vision; the lens periphery has a primary function of optimizing tear exchange and the fitting relationship. By convention, the lens OZD for GP lenses is the overall diameter minus the posterior peripheral curve widths (often the OAD—approximately 1.0 to 1.5 mm). The OZD for soft lenses is the overall lens diameter minus twice the anterior peripheral curve width in millimeters.

**Posterior Peripheral Curve Systems** GP lenses have a standard curvature (i.e., the base curve radius) with typically two-to-three peripheral curve radii. These peripheral curve radii are progressively

**TABLE 2** Keratometer Diopter Conversion to Millimeters

BCR (D)	BCR (mm)	BCR (D)	BCR (mm)
40.00	8.44	46.25	7.30
40.25	8.39	46.50	7.26
40.50	8.33	46.75	7.22
40.75	8.28	47.00	7.18
41.00	8.23	47.25	7.14
41.25	8.18	47.50	7.11
41.50	8.13	47.75	7.07
41.75	8.08	48.00	7.03
42.00	8.04	48.25	7.00
42.25	7.99	48.50	6.96
42.50	7.94	48.75	6.92
42.75	7.89	49.00	6.89
43.00	7.85	49.25	6.85
43.25	7.80	49.50	6.82
43.50	7.76	49.75	6.78
43.75	7.72	50.00	6.75
44.00	7.67	50.25	6.72
44.25	7.63	50.50	6.68
44.50	7.58	50.75	6.65
44.75	7.54	51.00	6.62
45.00	7.50	51.25	6.59
45.25	7.46	51.50	6.55
45.50	7.42	51.75	6.52
45.75	7.38	52.00	6.49
46.00	7.34		

flatter from the base curve to the outermost curve. As the cornea is aspheric and flattens from center to periphery, a moncurve GP lens would likely be too steep in the periphery and not allow tear exchange—important for oxygen exchange and debris removal—to occur. Therefore, if the base curve radius was 7.80 mm, one common philosophy recommends a secondary curve radius (SCR) equal to 1 mm flatter than the base curve radius, and a peripheral curve radius 2 mm flatter than the SCR.<sup>9</sup> Both of these curves would be 0.3 mm wide. This is termed a tricurve design. A tetracurve design would include a secondary curve, an intermediate curve, and a peripheral curve as diagrammed in Fig. 1. Some designs (to be discussed) are manufactured such that they have a gradual increase in flattening in the periphery, termed a peripheral aspheric design. Soft lens posterior curvature is either moncurve, bicurve (typically one peripheral curve on average about 0.5 mm wide and as flat as 12 mm in curvature), or aspheric. The nonmonocurve designs are intended to align the lens periphery with the peripheral limbal or scleral topography.<sup>7</sup>

**Center Thickness** Contact lenses are relatively thin lens designs (relative to spectacles) with lenses as thin as 0.035 mm for a high minus power ultrathin soft lens to as high as 0.5 to 0.6 mm for high plus power soft and GP lenses, especially if prism ballast is incorporated into the lens to stabilize it on the eye. Minus power GP lenses tend to vary from 0.09 mm (ultrathin) to 0.19 mm for low minus standard thickness designs. Low plus power lenses typically have center thicknesses in the 0.20 to 0.25 mm range and high plus power lenses have center thicknesses in the 0.26 to 0.60 mm range. Center thickness impacts both lens mass and oxygen transmission. Plus power lenses have greater mass and an anterior center of gravity making centration on the eye more challenging. In addition, oxygen transmission (i.e.,  $Dk/t$  where “Dk” is the oxygen permeability of the material and “t” is the thickness—either  $t_{\text{avg}}$  for average thickness or  $t_c$  for center thickness) reduces as the center thickness increases. However, a minus power GP lens that is too thin will bend or flex on the eye and will negate one of the benefits of this modality, the ability to mold the anterior corneal surface into a sphere and correct astigmatism.

**Edge Thickness** Edge thickness varies for GP and soft lenses but values of 0.06 to 0.12 mm are not uncommon. With most GP designs, the apex is tapered and rounded toward the posterior surface. As high minus power (typically  $\geq 5$  D) lenses tend to have thick edges and both low minus power (typically  $\leq 1.50$  D) and all plus power lenses have thin edges, special manufacturing procedures are utilized to maintain a uniform edge thickness. With GP lenses this is typically a plus lenticular (a manufacturing process that thins the edge) for high minus power lenses and a minus lenticular (a manufacturing process that results in a thicker edge) for low minus and plus power lenses. This uniform edge thickness is important for allowing the upper lid to lift up the lens with the blink while not being so thick that the lid would push the lens inferiorly.

## 20.4 CONTACT LENS POWER

Contact lens power typically differs from spectacle lens power. For example, a spectacle lens power of  $-2.00 -1.00 \times 180$  would indicate that the horizontal meridian of the spectacles has a power of  $-2.00$  D and the vertical meridian has a power of  $-3.00$  D. If this patient has no residual astigmatism (i.e., the anterior corneal astigmatism is also equal to  $-1.00 \times 180$ ), the back surface of a spherical power GP lens will mold the front surface of the cornea into a sphere; therefore, the astigmatism is corrected. A spherical soft lens, however, simply drapes over the eye and does not correct for refractive astigmatism and, in theory, the  $-1.00 \times 180$  correction in the spectacles would not be corrected by such a soft lens and a special design, a soft toric lens (to be discussed) would be indicated.

### The Contact Lens as a Thick Lens

Contact lenses are manufactured at a much lesser center thickness (typically 0.08 to 0.5 mm) than spectacle lenses. Nevertheless, they still need to be considered as thick lenses when determining the overall power. Therefore, instead of simply adding the refractive power of the front surface to the

back surface power (i.e.,  $F_T = F_1 + F_2$ ), the curvatures of both surfaces, center thickness and index of refraction of the lens material need to be considered. The thick lens formula used for spectacle lenses could be considered.<sup>10</sup> It is as follows:

$$F_T = F_1 + F_2 - (t/n)F_1F_2$$

where  $F_T$  = equivalent or true refractive power in diopters

$F_1 = (n' - n)/r_1$  = refractive power of the anterior lens surface

$F_2 = (n - n')/r_2$  = refractive power of the posterior lens surface

$t$  = center thickness of the lens in meters

$n'$  = refractive index of the lens material

$n$  = refractive index of the medium surrounding the lens

$r_1$  = radius of curvature of posterior surface, in meters

$r_2$  = radius of curvature of anterior surface, in meters

However, as the positions of the principal planes vary with the design of the lens and its two surface powers, it is impractical to use the equivalent, or true, power of a contact lens as the absolute definition of refractive power. It is, instead, important to use a fixed position to measure refractive power.

### Back Vertex Power and Front Vertex Power

A fixed position from which refractive power can be determined is via the use of either back vertex power (BVP) or front vertex power (FVP). The equations of these powers are as follows:

$$\text{BVP} = F_1 + \left[ \frac{F_2}{1 - (t/n')F_2} \right]$$

$$\text{FVP} = \frac{F_1}{1 - (t/n')F_1} + F_2$$

The measurement of BVP and FVP is very straightforward on a lensometer. BVP is obtained with the back surface of the lens (concave surface) against the lens stop (see Fig. 2). Front surface



**FIGURE 2** Determination of back surface power via the use of a lensometer.

**TABLE 3** Differences between Back Vertex Power (BVP) and Front Vertex Power (FVP) in Air for GP Lenses ( $N = 1.46$ )

BVP (D)	FVP (D)	Center Thickness (mm)	Power Difference (D)
-15.00	-14.86	0.12	-0.14
-12.00	-11.88	0.12	-0.12
-9.00	-8.91	0.12	-0.09
-6.00	-5.94	0.12	-0.06
-3.00	-2.96	0.14	-0.04
Plano	Plano	0.18	0
+3.00	+2.95	0.21	+0.05
+6.00	+5.87	0.24	+0.13
+9.00	+8.75	0.30	+0.25
+12.00	+11.59	0.36	+0.41
+15.00	+14.41	0.41	+0.59
+20.00	+19.01	0.50	+0.99

power is obtained with the front surface of the lens (i.e., convex surface) against the lens stop. The focal lengths (in meters) are also easy to determine as they are simply the reciprocals of the vertex powers in diopters. For example, a BVP of  $-10$  D would have a focal length of  $1/10$  or  $0.1$  m (or  $10$  cm).

The difference between BVP and FVP is often negligible, especially in all minus and low plus power lenses. This is the result of the smaller center thicknesses that are present in these lens powers. However, with moderate-to-high plus power lenses—notably aphakic powers—this difference becomes significant (see Table 3).

As it can be shown, BVP is always greater than FVP. This has been a source of confusion in the past when it was not uncommon to specify (and/or receive from the fabricating laboratory) aphakic lenses via FVP. However, if the practitioner ordered the lenses based on BVP, the amount of resulting error could be considerable.<sup>11</sup> Fortunately, the contact lens industry has—with rare exception—utilized back surface power for all contact lenses in recent years.

## Effective Power

A primary difference between spectacle power and contact lens power—notably in high ametropia—pertains to the effective power at the corneal plane. A spectacle lens is often positioned  $10$  to  $15$  mm in front of the cornea. This distance from the spectacle lens to the corneal apex is termed “vertex distance.” A contact lens, however, by virtue of being positioned adjacent to the cornea has a vertex distance of zero.

As a rule, as a plus power lens is brought closer to the cornea, its effective power decreases, such that its refractive power must be increased in order to maintain a constant amount of power relative to the eye. Likewise, when a minus power lens is brought closer to the eye, its effective power increases, such that its refractive power must be decreased in order to maintain a constant amount of power relative to the eye.<sup>10</sup> When correcting an individual with a myopic refractive error, a contact lens will require a longer focal length, therefore resulting in less minus power than a spectacle lens. Likewise, when correcting an individual with a hyperopic refractive error, a contact lens will require a smaller focal length, therefore resulting in a higher plus refractive power than with a spectacle lens. Effective power differences between spectacle and contact lens corrections become clinically significant at about  $\pm 4$  D (see Table 4).

**TABLE 4** Effective Spectacle Lens Power at the Corneal Plane (12 mm Vertex Distance)

Minus Lenses (in Diopters)			
Spectacle Lens Power	Effective Lens Power	Spectacle Lens Power	Effective Lens Power
-4.00	-3.75	+4.00	+4.25
-4.50	-4.25	+4.50	+4.75
-5.00	-4.75	+5.00	+5.25
-5.50	-5.25	+5.50	+6.00
-6.00	-5.50	+6.00	+6.50
-6.50	-6.00	+6.50	+7.00
-7.00	-6.50	+7.00	+7.75
-7.50	-7.00	+7.50	+8.25
-8.00	-7.25	+8.00	+8.75
-8.50	-7.75	+8.50	+9.50
-9.00	-8.25	+9.00	+10.00
-9.50	-8.50	+9.50	+10.75
-10.00	-9.00	+10.00	+11.25
-10.50	-9.50	+10.50	+12.00
-11.00	-9.75	+11.00	+12.75
-11.50	-10.00	+11.50	+13.25
-12.00	-10.50	+12.00	+14.00
-12.50	-10.75	+12.50	+14.75
-13.00	-11.25	+13.00	+15.50
-13.50	-11.50	+13.50	+16.00
-14.00	-12.00	+14.00	+16.75
-14.50	-12.50	+14.50	+17.50
-15.00	-12.75	+15.00	+18.25
-15.50	-13.00	+15.50	+19.00
-16.00	-13.50	+16.00	+19.75
-16.50	-13.75	+16.50	+20.50
-17.00	-14.00	+17.00	+21.25
-17.50	-14.50	+17.50	+22.25
-18.00	-14.75	+18.00	+23.00

There are two methods of determining the power at the corneal plane. The first is simply to use the following equation

$$F(c) = \frac{F(\text{sp})}{1 - dF(\text{sp})}$$

where  $F(c)$  = power at the corneal plane  
 $F(\text{sp})$  = power at the spectacle plane  
 $d$  = vertex distance (in m)

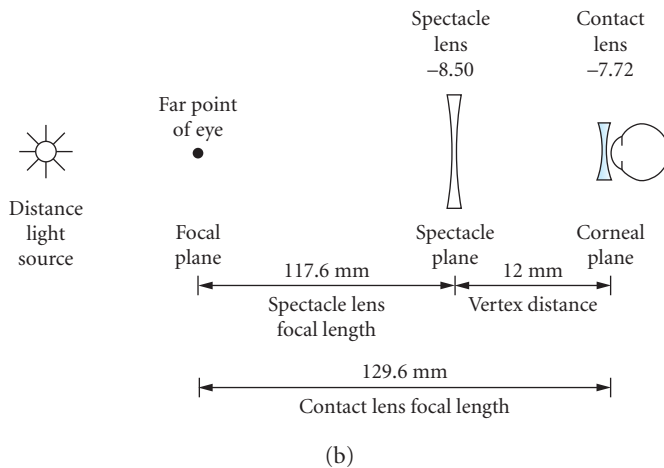
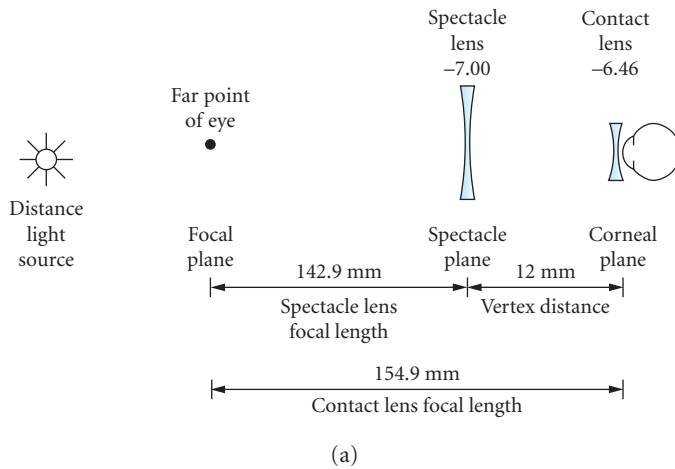
Whereas, the vertex distance should be carefully measured in high ametropia patients, an average value of 12 mm is commonly used and will be used in the calculations presented in this chapter. If a patient has a refractive power of  $-7.00 - 1.50 \times 180$ , and the power in each meridian (i.e.,  $-7.00$  at  $180$  and  $-8.50$  at  $090$ ) is “vertexed” to the corneal plane via this formula, the resulting powers would be  $-6.46$  D and  $-7.71$  D or  $-6.46 - 1.25 \times 180$ , which can be rounded off to  $-6.50 - 1.25 \times 180$ . If the individual was hyperopic and had a refractive error of  $+7.00 - 1.50 \times 180$ , and the power in each meridian (i.e.,  $+7.00$  at  $180$  and  $+5.50$  D at  $090$ ) was vertexed to the corneal plane, the resulting powers would be  $+7.64$  D and  $+5.89$  D or  $+7.64 - 1.75 \times 180$  or rounded off to  $+7.75 - 1.75 \times 180$ .

It can be observed that in compound myopic astigmatism, the spherical and cylinder power components decrease when referred to the corneal plane. In compound hyperopic astigmatism, the spherical and cylindrical power components increase or become greater when referred to the corneal plane.

The second method of effective power calculation would refer back to the diagrams in Figs. 3 and 4 via the use of focal lengths. If the individual's refractive error is  $-7.00 - 150 \times 180$ , the focal lengths of the two primary meridians are as follows:

$$\text{Horizontal meridian: } f_{180} = \frac{1}{F_{180}} = \frac{1}{7 \text{ D}} = 0.1429 \text{ m or } 142.9 \text{ mm (Fig. 3a)}$$

$$\text{Vertical meridian: } f_{090} = \frac{1}{F_{090}} = \frac{1}{8.50 \text{ D}} = 0.1176 \text{ m or } 117.6 \text{ mm (Fig. 3b)}$$



**FIGURE 3** Determination of the effective power for the myopic patient in both the (a) horizontal meridian and (b) vertical meridian.

The focal lengths of the correction at the corneal plane will be 12 mm more; therefore, the final power per meridian is

$$\text{Horizontal meridian: } F_{180} = \frac{1}{0.1549 \text{ m}} = -6.46 \text{ D}$$

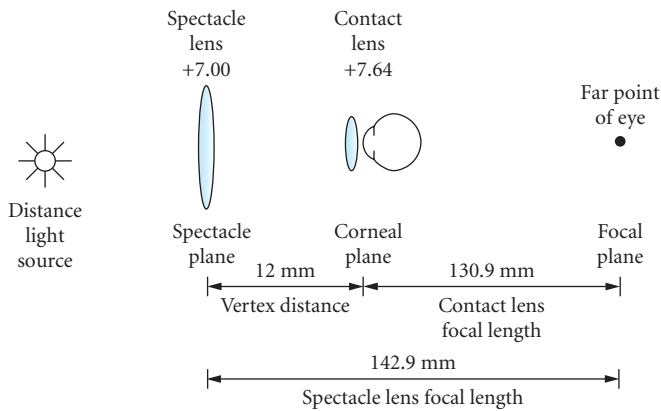
$$\text{Vertical meridian: } F_{090} = \frac{1}{0.1296 \text{ m}} = -7.72 \text{ D}$$

The corneal plane refraction is, therefore,  $-6.46 - 1.26 \times 180$  or rounded off to  $-6.50 - 1.25 \times 180$ .

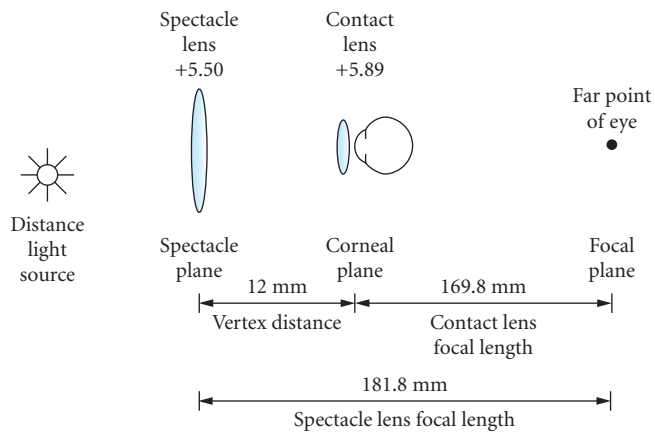
If the individual's refractive error is  $+7.00 - 1.50 \times 180$ , the focal lengths of the two primary meridians are as follows:

$$\text{Horizontal meridian: } f_{180} = \frac{1}{F_{180}} = \frac{1}{-7 \text{ D}} = 0.1429 \text{ m or } 142.9 \text{ mm (Fig. 4a)}$$

$$\text{Vertical meridian: } f_{090} = \frac{1}{F_{090}} = \frac{1}{-5.5 \text{ D}} = 0.1818 \text{ m or } 181.8 \text{ mm (Fig. 4b)}$$



(a)



(b)

**FIGURE 4** Determination of the effective power for the hyperopic patient in both the (a) horizontal meridian and (b) vertical meridian.



The focal lengths of the correction at the corneal plane will be 12 mm less; therefore, the final power per meridian is

$$\text{Horizontal meridian: } F_{180} = \frac{1}{0.1309 \text{ m}} = +7.64 \text{ D}$$

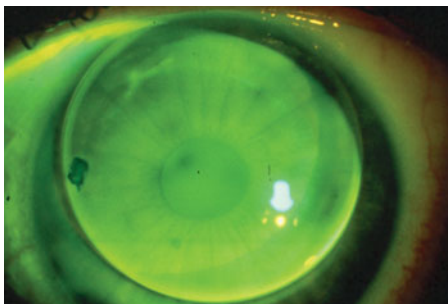
$$\text{Vertical meridian: } F_{090} = \frac{1}{0.1698 \text{ m}} = +5.89 \text{ D}$$

The corneal plane refraction is, therefore,  $+7.64 - 1.75 \times 180$  or rounded off to  $+7.75 - 1.75 \times 180$ .

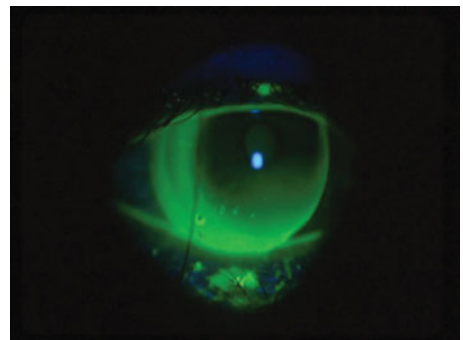
## Lacrimal Lens Considerations

Whereas effective power impacts both soft and GP lenses with patients who have a spherical refractive value equal to 4 D or greater, as soft lenses merely drape over the cornea, no compensation for the lacrimal or tear lens between the cornea and the contact lens is necessary. With GP lenses, however, a lacrimal or tear lens between the cornea and contact lens often exists and needs to be factored into the final contact lens power determination. The baseline values to be considered when determining GP power pertain to the flatter keratometry reading and the spherical refractive value (at the corneal plane). The base curve radius selected for most GP wearers is very similar in curvature to the flatter keratometry reading; therefore, very little—if any—lacrimal lens is present. If fitted in alignment with the flatter keratometry reading, this is typically referred to as an “On *K*” lens-to-cornea fitting relationship. If the lens is fitted more curved or steeper in base curve radius than the baseline corneal curvature value it is termed “steeper than *K*.” If the lens is fitted less curved or flatter than the baseline corneal curvature value, it is termed “flatter than *K*.” The ideal GP lens-to-cornea fitting relationship would consist of an alignment or even tear film between lens and cornea as easily observed when fluorescein dye is applied to the tear film. Such a pattern is shown in Fig. 5. A base curve radius selected excessively steeper than “*K*” resulting in direct contact of the lens against the central cornea is shown in Fig. 6. This apical bearing relationship can ultimately result in central corneal distortion as well as excessive lens awareness due to the large amount of edge clearance present. A base curve radius selected excessively steeper than “*K*” can result in very little tear exchange peripherally, possibly resulting in edge sealoff and corneal edema. (See Fig. 7.)

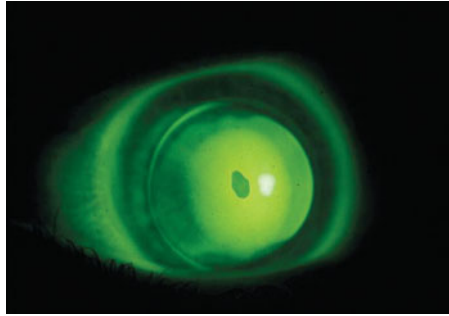
In theory, as a result of the fact that the cornea is aspheric and gradually flattens in curvature from center to periphery, a slightly flatter than “*K*” base curve radius is often recommended. Of course the rate of flattening or corneal eccentricity can vary considerably and this will impact the fitting relationship and the base curve radius that will ultimately result in success. It is also important to remember



**FIGURE 5** An alignment GP lens-to-cornea fitting relationship as demonstrated with fluorescein application.



**FIGURE 6** An apical bearing or flat GP lens-to-cornea fitting relationship as demonstrated with fluorescein application.



**FIGURE 7** An apical clearance or steep GP lens-to-cornea fitting relationship as demonstrated with fluorescein application.

that a lens that is fitted flatter than “*K*” will induce a minus power lacrimal lens; for example, if a lens was fitted 0.50 D flatter than “*K*,” a  $-0.50$  D lacrimal lens would be present and  $+0.50$  D would need to be incorporated into the contact lens power to compensate for the lacrimal lens. Likewise, a lens that is fitted steeper than “*K*” will induce a plus power lacrimal lens; for example, if a lens was fitted 0.50 D steeper than “*K*,” a  $+0.50$  D lacrimal lens would be present and  $-0.50$  D would need to be incorporated into the contact lens power to compensate for the lacrimal lens.

The actual power can be derived in several ways. For example, the GP lenses were to be ordered empirically (i.e., from calculations only, not via the application of a diagnostic lens), the following formula can be used for predicting the power to be ordered

$$F_{cl} = F_{cp} - LLP$$

where  $F_{cl}$  = power of the contact lens  
 $F_{cp}$  = power of the refraction at the corneal plane  
 LLP = lacrimal lens power

Using this formula, if the patient had the following refraction and keratometry values

Refraction:  $-3.00 - 0.75 \times 180$

Keratometry: 43.50 at 180; 44.25 at 090

and a base curve radius was selected 0.75 D flatter than “*K*” or 42.75 (7.89 mm), the following GP contact lens power would be predicted:

$$F_{cl} = -3.00 - (-)0.75 = -2.25 \text{ D}$$

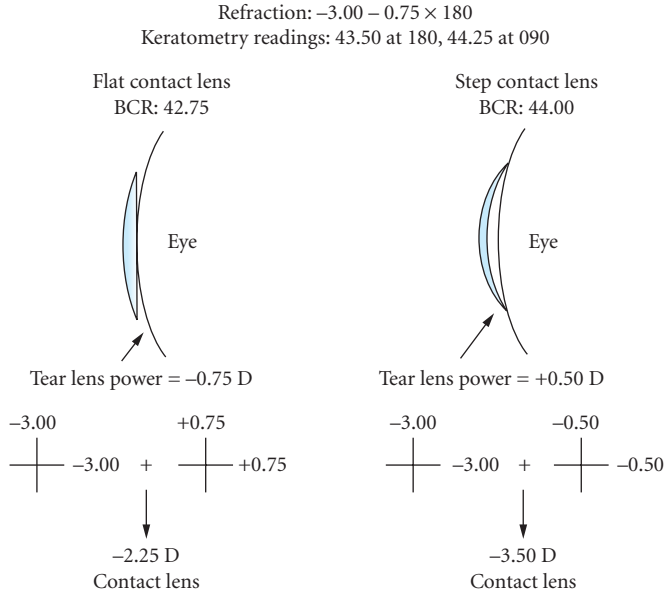
If the base curve radius was selected to be 0.50 D steeper than “*K*,” the predicted contact lens power would be

$$F_{cl} = -3.00 - 0.50 = -3.50$$

This example is shown in Fig. 8. An easy acronym to remember when factoring the tear lens into the contact lens power is “SAM-FAP” (i.e., “steep-add-minus” and “flat-add-plus”).

The most accurate method of determining contact lens power is via the use of a diagnostic lens. The formula for deriving the power when a diagnostic lens is placed on the eye is very simple

$$F_{cl} = F_{dl} + OR$$



**FIGURE 8** The “SAM (steep-add-minus)-FAP (flat-add-plus)” acronym as it applies to power determination based on the tear lens power.

where  $F_{cl}$  = final contact lens power  
 $F_{dl}$  = power of the diagnostic lens  
 OR = spherical overrefraction

For example, if a patient has the following spectacle refraction and keratometry values

Refraction:  $-6.00 - 0.50 \times 180$   
 Keratometry: 43.50 at 180; 44.00 at 090

If a diagnostic GP lens with a power of  $-3.00$  D and a base curve radius equal to 43.25 (7.80 mm) was selected, the predicted contact lens power would equal

$$\begin{aligned}
 F_{cl} &= F_{cp} - LLP \\
 &= -5.50 \text{ (refraction at corneal plane)} - (-)0.25 \\
 &= -5.25 \text{ D}
 \end{aligned}$$

Once the contact lens has been evaluated and an overrefraction performed, it can be determined if the actual contact lens power to be ordered is equal to the predicted. If the spherical overrefraction was  $-2.25$  D, this would be the case as

$$\begin{aligned}
 F_{cl} &= F_{dl} + \text{OR} \quad \text{or} \\
 &= -3.00 + (-)2.25 \\
 &= -5.25 \text{ D}
 \end{aligned}$$

As this example clearly demonstrates, the two most important factors when determining the power of a GP lens are effective power (if the ametropia is  $\geq 4$  D) and lacrimal lens power. If the contact lens power, as determined by overrefraction, deviates from the predicted power, the former is typically more accurate as inaccuracies in refraction and keratometry can impact the predicted value. The final lens order for this patient could be as follows:

Power:  $-5.25$  D  
 OAD/OZD: 9.40/8.00 mm  
 BCR: 7.80 mm  
 SCR/W: 8.80/0.3 mm  
 PCR/W: 10.80/0.3 mm  
 CT: 0.12 mm  
 Material: Boston ES (fluoro-silicone/acrylate)

It is also important to understand how the lens design impacts the sagittal depth relationship between contact lens and cornea. As the optical zone diameter represents the linear equivalent to the base curve radius, if an alignment lens-to-cornea fitting relationship exists but the patient's pupil diameter is quite large, necessitating a larger optical zone diameter to optimize vision, the new lens should exhibit an apical clearance or steep fitting relationship due to the increase in sagittal depth. Essentially the steepest section of the contact lens, the base curve radius, has become larger. Therefore, the base curve radius should be designed to be flatter to compensate for this change in optical zone diameter; typically a 0.25 D change for every 0.5 mm change in OZD should maintain an alignment lens-to-cornea fitting relationship.

## Residual Astigmatism

A factor that has to be considered when determining whether a patient should be fit into GP lenses is their residual astigmatism. This is defined as<sup>12</sup>

Total ocular astigmatism – astigmatism as measured by keratometry

In the previous examples, the refractive and anterior corneal astigmatism (as measured with a keratometer) were the same; therefore, no residual astigmatism would be predicted. If, however, a patient had the following refractive information:

Refraction:  $-2.00 - 1.50 \times 180$   
 Keratometry: 42.50 at 180; 43.00 at 090

the predicted residual astigmatism would equal the refractive (i.e., total ocular) astigmatism minus the keratometric astigmatism or

$$-1.50 \times 180 - (-) 0.50 \times 180 = -1.00 \times 180$$

As a GP lens only corrects anterior corneal astigmatism, this residual astigmatism would not be corrected and could result in symptoms of blurred vision and eyestrain. Typically, when there is  $-0.75$  D or greater of residual astigmatism, a spherical GP lens is not recommended. Fortunately, this is quite uncommon.

## Soft Lens Power

For spherical soft lenses, the primary factor to consider would pertain to effective power considerations when the ametropia equals or exceeds 4 D. As the back surface of a soft lens essentially conforms to—or drapes over—the anterior corneal surface, the power of the lacrimal lens is approximately

plano or afocal.<sup>13</sup> This conformation of lens to the topography of the cornea has been termed “flexure.” Changes in base curve radius, diameter, or any other parameters do not appear to induce any lacrimal lens power. If the lens dehydrates during wear, the refractive index of the material will increase which will increase the magnitude of both plus and minus corrections, and the base curve radius will steepen and the lens will increase in minus power.<sup>14</sup> In addition, it has been reported that as the cornea steepens, the amount of back vertex power is predicted to increase for both plus and minus lenses.<sup>10</sup> Total flexure or conformation to the cornea may not be realistic with soft lenses, notably for plus power lenses.<sup>15</sup>

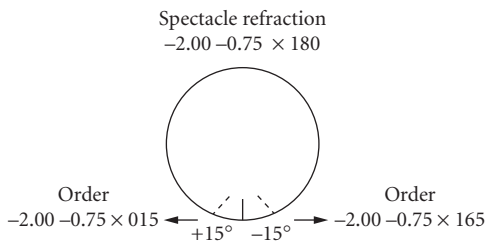
Another factor that impacts soft lens power as well as other design parameters pertains to the amount of expansion that occurs when the lens is hydrated (i.e., from the dry to wet state). According to Douthwaite<sup>13</sup> a  $-3.25$  D 59 percent water content soft lens with a 13 mm overall diameter, 8.60 mm base curve radius, and center thickness of 0.085 mm will have—prior to hydration—a  $-5.82$  D power, 6.10 mm base curve radius, 9.22 mm overall diameter, and 0.060 mm center thickness. Therefore, as a result of all of these factors (dehydration, incomplete flexure, and the parameter changes when hydrated), it cannot be assumed that the labeled lens power of a soft contact lens will represent the true power of the lens on the eye.

## Toric Lens Powers

**Soft Toric Lenses** As spherical soft lenses do not correct for the refractive (ocular) astigmatism, lenses incorporating this correction on the front surface are termed “soft toric lenses.” As it has been estimated that about 45 percent of interested contact lens patients have refractive astigmatism greater than or equal to 0.75 D<sup>16</sup> if the criterion for significant astigmatism is a minimum of 0.75 D, then close to half of the spectacle-wearing population are potential candidates for soft toric lenses.<sup>17</sup>

The astigmatism is typically corrected onto the vertical meridian of the front surface and the lenses use several types of stabilizing techniques that will minimize any rotation of the lens on the eye. Traditionally these designs were prism ballasted inferiorly to weight the lens. Several newer techniques have resulted in the necessary thickness differential between the central, superior, and inferior sections of the lens without prism use.

The cylinder axis of the soft toric lens should align with the axis of the refractive astigmatism of the eye. For example, if a patient has a refractive error of  $-2.00 - 0.75 \times 180$ , this should be the powers of the soft toric lens fit to this patient. The cylinder axis is typically identified by a laser mark; often marks are also present  $20^\circ$  to the left and to the right of the axis. If the axis shifts or rotates with the blink, a new lens should be attempted which compensates for this rotation. The acronym LARS (left-add; right subtract) is valuable in these cases (Fig. 9). If the laser mark shifts  $15^\circ$  to the observer’s left, a new lens with an axis at  $15^\circ$  should be attempted; likewise if the lens shifts  $15^\circ$  to the observer’s right, a lens with an axis of  $165^\circ$  should be attempted.



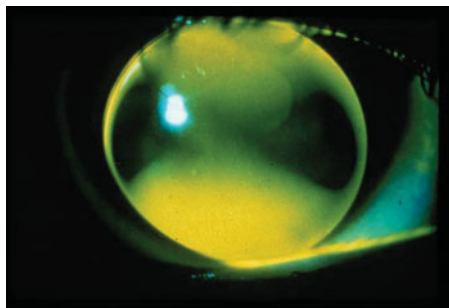
**FIGURE 9** The “LARS” (i.e., left-add; right-subtract) concept as it pertains to determining axis for a soft toric lens.

If lens rotation is minimal, particularly with high spherical refractive errors (i.e.,  $\geq 5$  D) and low astigmatic corrections (i.e.,  $\leq 1.25$  D), little-to-no impact on vision should result. On average, relatively stable soft toric lenses usually average 0 to  $10^\circ$  of nasal rotation.<sup>18</sup> However, when the rotation is significant, a “crossed cylinder” problem is presented. When the axis of cylinder of a correcting contact lens does not align with the best refractive correction, an overrefraction will result in cylinder at an axis oblique to that of the axis of best correction.<sup>10</sup> Residual cylinder will be present in the overrefraction and this value will increase as axis rotation increases and with higher refractive cylinder corrections. The extent of the shift in position of the resultant plus cylinder axis for two cylinders of equal power but opposite sign has been calculated by Pascal.<sup>19</sup> The formula  $(90 + a)/2$  is used, where “ $a$ ” is the angular discrepancy between the two combined cylinders. Therefore, the resultant plus cylinder axis is  $45^\circ$  from the midpoint between the axes of the two combined cylinders. The axis of the resultant plus cylinder of the crossed cylinder combination will appear on the side of the axis of the original plus cylinder opposite to that of the original minus cylinder. If the axis of ocular astigmatism is, for example,  $90^\circ$  and a minus cylinder of equal power is placed at axis  $70^\circ$ , or  $20^\circ$  away, the resultant plus cylinder will appear  $55^\circ$  [i.e.,  $(90 + 20)/2$ ] from 070 on the opposite side, or “x125.” The resultant plus cylinder will require a minus cylinder correction.

**GP Bitoric Lens Powers** Spherical GP wearing patients who exhibit a high amount of corneal astigmatism (typically  $\geq 2.50$  D) will not show the alignment fluorescein pattern demonstrated in Fig. 5 but will, instead, show a “dumbbell-shaped” pattern as shown in Fig. 10. In a high with-the-rule astigmatic patient, excessive edge clearance will exist in the steeper vertical meridian and bearing along the flatter horizontal meridian. As this fitting relationship often results in lens decentration on the eye, more edge awareness and corneal desiccation (dryness-induced epithelial cell loss), a lens that has a toric back surface with curvatures approximating the corneal curvatures in the two principle meridians is indicated. This design—termed bitoric lens design—essentially consists of two curvatures and two corresponding powers. A common method of selecting base curve radii is the Mandell-Moore bitoric lens guide philosophy.<sup>20</sup> This is provided in Table 5. For example, if the keratometry readings were

42.00 at 180      and      45.00 at 090

based on the Mandell-Moore base curve radius selection guide for three diopters of corneal cylinder, the flatter base curve radius would equal 0.25 D flatter than flat “K” or 41.75 D and the steeper base curve radius would equal 0.75 D flatter than steep “K” or 44.25 D. The base curve radii are selected to be flatter than their respective corneal curvatures due to the flattening of the cornea from center to periphery. The greater amount of flattening in the steeper meridian creates a slight amount of



**FIGURE 10** An astigmatic or “dumbbell-shaped” fluorescein pattern of a spherical GP lens on a highly with-the-rule astigmatic patient.

**TABLE 5** Mandell-Moore Base Curve Selection

Corneal Cylinder	Flat Meridian	Steep Meridian
2.0 D	“On K”	0.50 D Flatter than “K”
2.5 D	0.25 D Flatter than “K”	0.50 D Flatter than “K”
3.0 D	0.25 D Flatter than “K”	0.75 D Flatter than “K”
3.5 D	0.25 D Flatter than “K”	0.75 D Flatter than “K”
4.0 D	0.25 D Flatter than “K”	0.75 D Flatter than “K”
5.0 D	0.25 D Flatter than “K”	0.75 D Flatter than “K”

toricity which has been found to enhance tear exchange. Determining the lens powers requires two tear lens calculations (vertical and horizontal meridians), not one as with conventional GP lenses. For example, if the patient’s refraction in the above case was  $Pl - 3.00 \times 180$ , the powers to be ordered for each meridian would consist of the following

$$F_{cph} - LLP \quad \text{and} \quad F_{cpv} - LLP$$

where  $F_{cph}$  = refractive power at the corneal plane in horizontal meridian and  
 $F_{cpv}$  = refractive power at the corneal plane in vertical meridian

$$F_{cph} - LLP = Pl - (-0.25) = +0.25 \text{ D}$$

$$F_{cpv} - LLP = -3.00 - (-0.75) = -2.25 \text{ D}$$

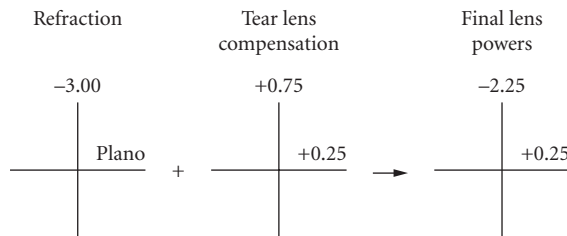
Therefore, the final powers of this bitoric lens would be  $+0.25/-2.25$ . This can also be diagrammed via the use of optical crosses showing the refractive power per meridian in combination with the correction for the lacrimal lens power (i.e., using “SAM-FAP”). This is shown in Fig. 11.

Another method of determining the base curve radii and powers without the use of optical crosses is to use the Mandell-Moore guide. An example is provided in Fig. 12.

An alternative method, which appears to be equally popular and effective, is to fit the patient with bitoric diagnostic lenses.<sup>21</sup> In this case, a spherical overrefraction is typically all that is required and this value is added to the diagnostic lens power in each meridian. If, for example, in the above case, a bitoric diagnostic lens is selected that has base curve radii of 41.50/44.50 and powers of  $Pl/-3.00 \text{ D}$ , and the overrefraction is  $+0.50 \text{ DS}$ , the final powers would be

$$\text{Plano} + (+) 0.50 = +0.50 \text{ D}$$

$$-3.00 + (+) 0.50 = -2.50 \text{ D}$$



**FIGURE 11** The use of optical crosses to show tear lens compensation in determining the final powers of a bitoric lens.

### Mandell-Moore Bitoric Lens Guide - Per Eye

1. Keratometry	42.00 @ 180		45.00 @ 090	
2. Spectacle Rx (Minus Cyl Form)	pl - 3.00 x 180			
	Flattest K	Sphere Power	Steepest K	Sph+Cyl Power
3. Enter K	42.00		45.00	
4. Enter Spectacle Power		plano		-3.00
5. Vertex Adjust Line 4		plano		-3.00
6. Insert Fit Factor	(-) 0.25	(+) 0.25	(-) 0.75	(+) 0.75
Add Lines	3&6	5&6	3&6	5&6
7. Final CL Rx	41.75	+0.25	44.25	-2.25
	Base Curve	Power	Base Curve	Power

### Bitoric Lens Fit Factor

Corneal Cyl	Fit Flat Meridian	Fit Steep Meridian
2.0 Diopters	On K (0 D)	0.50D Flatter
2.5 Diopters	0.25D Flatter	0.50D Flatter
3.0 Diopters	0.25D Flatter	0.75D Flatter
3.5 Diopters	0.25D Flatter	0.75D Flatter
4.0 Diopters	0.25D Flatter	0.75D Flatter
5.0 Diopters	0.25D Flatter	0.75D Flatter

FIGURE 12 The Mandell-Moore bitoric fitting guide.

In this case there is no residual astigmatism. In these examples the bitoric lens has a spherical power effect in that it can rotate on the eye with no impact on vision as the corresponding vergence will be present to correct for the change in corneal curvature. (See Fig. 13.)

If the patient has significant residual astigmatism, a sphero-cylindrical overrefraction will be necessary and the overrefraction in a meridian will be added to the corresponding power in the bitoric diagnostic lens. For example, if a patient had the following refractive information

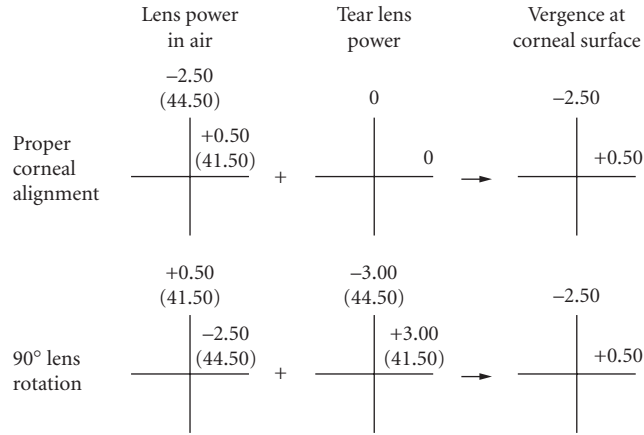
Refraction: +1.00 - 4.00 x 180

Keratometry: 42.00 at 180; 45.00 at 090

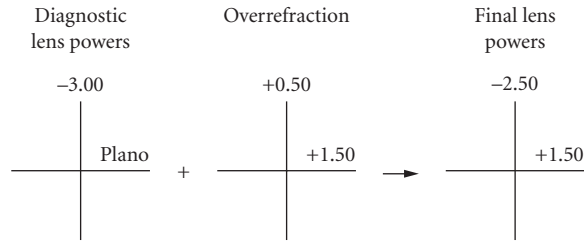
Diagnostic bitoric lens: 41.50/44.50 Pl/-3.00D

Overrefraction: +1.50 - 1.00 x 180





**FIGURE 13** An example of how a bitoric lens can have spherical power effect optics. If the lens rotates 90°, the compensating vergences will be present.



**FIGURE 14** Determining the final powers of a bitoric lens in which a sphero-cylindrical overrefraction simply pertains to adding the overrefraction in a specific meridian to the power of the diagnostic lens in that meridian.

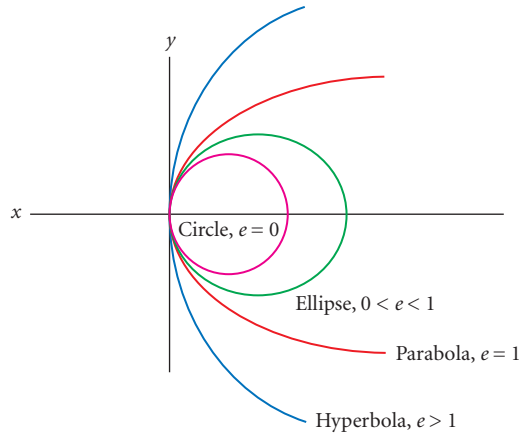
The final powers are determined by adding the overrefraction in the 180 meridian (+1.50) to the diagnostic lens power (Plano) and then adding the overrefraction in the 090 meridian [ $+1.50 + (-)1.00 = +0.50$ ] to the diagnostic lens power in that meridian (-3.00 D). This is shown in Fig. 14.

## 20.5 OTHER DESIGN CONSIDERATIONS

### Aspheric Lens Designs

An aspherical surface has been defined as a surface which is not spherical.<sup>22</sup> In contact lens design terminology, it was originally indicated by Feinbloom<sup>23</sup> that “(a) an ellipsoid represents a better approximation of the form of the surface of the cornea of the human eye, and (b) contact lenses with inner elliptical surfaces represent a marked improvement in the comfort and wearing time by the patient.” This elliptical corneal shape consists of a central spherical zone which progressively flattens as the limbus is approached. The progressive flattening is representative of ellipses—and conic sections—in general.<sup>24</sup>

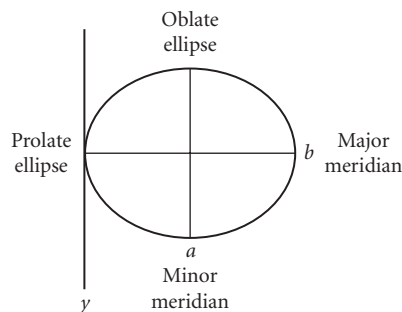
The shapes of conic sections are described mathematically by their eccentricity ( $e$ ). As it pertains to conic sections, eccentricity can be defined as deviating from a circular path. A spherical central



**FIGURE 15** Different conic sections which have the same prolate apical radius but differ in eccentricity or  $e$ -value.

area of a given radius of curvature is combined in a model system with varying rates of paracentral flattening in the same scale as that of the cornea that it represents. The “ $e$ ” increases, as does the rate of paracentral flattening. This can be demonstrated graphically by plotting the conic sections on the same set of axes with the vertex of each figure having the same radius of curvature at the origin of the coordinates (see Fig. 15). A circle has an “ $e$ ” value of 0 with its radius of curvature the same for all points on the curve and the meridians all equal in length. An ellipse has an “ $e$ ” value between 0 and 1; a parabola has an “ $e$ ” value equal to 1; a hyperbola has an “ $e$ ” value greater than 1. As will be discussed, most of the aspheric lens designs in common use mimic the average corneal eccentricity which is a prolate ellipse with an “ $e$ ” value of approximately 0.45.<sup>10,24</sup> A prolate ellipse is produced by elongating the horizontal meridian of a circle (Fig. 16). If “ $a$ ” is the major meridian of the ellipse and “ $b$ ” the minor meridian, the eccentricity is calculated by the following equation:

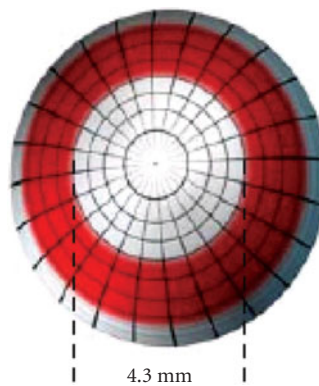
$$e = \frac{\sqrt{a^2 - b^2}}{a}$$



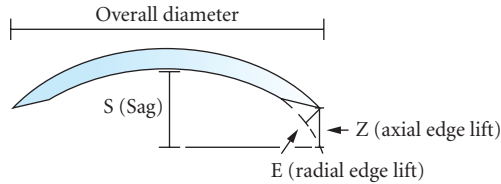
**FIGURE 16** The major meridian of a typical ellipse is the “ $a$ ” axis; rotation around this axis will produce a prolate ellipsoid, characterized by increasing radius of curvature outward from the origin. An oblate ellipsoid is produced by rotation around the minor axis “ $b$ .” This is characterized by a decrease in radius of curvature outward from the origin.

Several GP aspheric lens designs are in common use. Some are manufactured with a spherical base curve radius and an aspheric periphery. In many of these designs, the aspheric section is tangential to the base curve radius, therefore creating an aspheric continuum or continuous curve. Several aspheric designs have a totally aspheric posterior surface; however, the posterior optical zone and the periphery are two different curves. One such design that has enjoyed some success in the United States is the Envision lens (Bausch & Lomb). Originally developed by Hanita Contact Lenses in Israel, this design has an elliptical posterior optical zone of 0.4 and a hyperbolic periphery that is tangential to the posterior optical zone.

Aspheric lens designs have become increasingly popular in the correction of presbyopia. Both center-near and center-distance power soft lenses are currently used which involve continuous change in power from the lens axis to the peripheral section of the lens, therefore creating a multifocal effect. These designs have the benefit of progressive power change resulting in distance, intermediate, and near image correction. This power change is typically on the front surface of the lens and these designs have been termed “simultaneous vision” lenses as the different powers are in front of the pupil at the same time; therefore, some compromise in vision may exist with these designs at any desired viewing distance. This type of presbyopic correction has been found to be preferred by 76 percent of individuals who wore both this correction and monovision (i.e., one eye optimally corrected for distance vision; the other eye optimally corrected for near vision).<sup>25</sup> GP aspheric multifocal designs have enjoyed greater success as a result of the rigidity, optical quality, and ability to provide high add powers. Although some designs are front surface aspherics with an increase in near power from center to periphery, most designs in common use today are center-distance back surface designs although this varies from region to region worldwide. These designs have much greater eccentricity than single vision designs, ranging from high “*e*” value elliptical to hyperbolic back surface designs.<sup>26,27</sup> This type of design has been found to result in similar visual performance to progressive addition spectacle lenses while performing significantly better than both monovision and soft lens multifocals.<sup>28,29</sup> Although the rapid increase in plus power required to provide optimal near vision for patients with high add requirements can compromise distance vision, recently introduced GP multifocal designs have addressed this issue by incorporating some of the add power in a paracentral annular region on the front surface of the lens (see Fig. 17). These designs also rely on some upward shifting of the lens—or translation—when the patient views inferiorly to take advantage of the paracentral/peripheral near power.



**FIGURE 17** The Essentials CSA lens design with the near add power on the front surface shown in red.



**FIGURE 18** Axial versus radial edge lift.

## Posterior Peripheral Curve Design Systems and Calculations

As previously discussed, there are several terms used to describe the peripheral curve radii and width, including secondary curve, intermediate curve, peripheral curve, and aspheric peripheries. Edge lift or edge clearance are terms used to indicate the distance from the GP lens edge to the cornea. Edge lift pertains to the distance between the extension of the tangent of the base curve radius and the absolute edge after the addition of peripheral curves. Radial edge lift (REL) is measured normal to the base curve radius extension (i.e., the distance from the lens edge perpendicular to an extension of the base curve radius.<sup>9</sup> (See Fig. 18.) Axial edge lift is measured parallel to the lens optical axis or the vertical distance from the lens edge to an extension of the base curve radius. Axial edge lift is often used and values of 0.10 to 0.15 mm are typically recommended.<sup>30</sup> In a tricurve design, for example, the peripheral curve often contributes approximately two-thirds of the overall axial edge lift. AEL values can be determined via the tables presented in Musset and Stone.<sup>31</sup>

Table 6 provides some representative AEL values for different designs. A more accurate estimate of the distance between the lens edge and the cornea is the actual edge clearance. As this does not pertain to an extension of the base curve radius, this value would be less than the axial edge lift. An ideal value for axial edge clearance has been found to be 0.08 mm.<sup>32</sup> Although this has traditionally been a difficult parameter to measure, Douthwaite<sup>33</sup> provides a very good overview of his recent efforts in calculating axial edge clearance and his text should be referred to for anyone desiring more information on this topic.

**TABLE 6** Recommended Secondary and Peripheral Curve Radii

Axial Edge Lift = 0.10 mm  
 Secondary Curve Width = 0.3 mm  
 Peripheral Curve Width = 0.3 mm  
 Overall Diameter = 9.0; optical zone diameter = 7.8 mm  
 Base Curve Radius (BCR): varies from 7.5 to 8.3 mm  
 If secondary curve radius (SCR)/width contributes 0.04 and the peripheral curve radius (PCR)/width contributes 0.06 to the overall axial edge lift, the following values would be calculated:

BCR (mm)	SCR (mm)	PCR (mm)
7.50	9.00	10.20
7.70	9.30	10.70
7.90	9.60	11.20
8.10	10.00	11.70
8.30	10.30	12.20

Source: From Ref. 32.

For practical purposes, if the peripheral curve is designed or modified to be flatter or wider, the resulting edge clearance is greater. Flattening the base curve radius will also “lift” the periphery away from the cornea and increase edge clearance. Reducing the optical zone diameter while keeping the overall diameter constant will increase edge clearance as the steepest part of the lens (i.e., the base curve radius) is reduced and the overall peripheral curve width is being increased. Conversely, edge clearance can be reduced by steepening the peripheral curve radius, reducing the peripheral curve width, steepening the base curve radius, or increasing the optical zone diameter.

## Aberrations and Contact Lenses

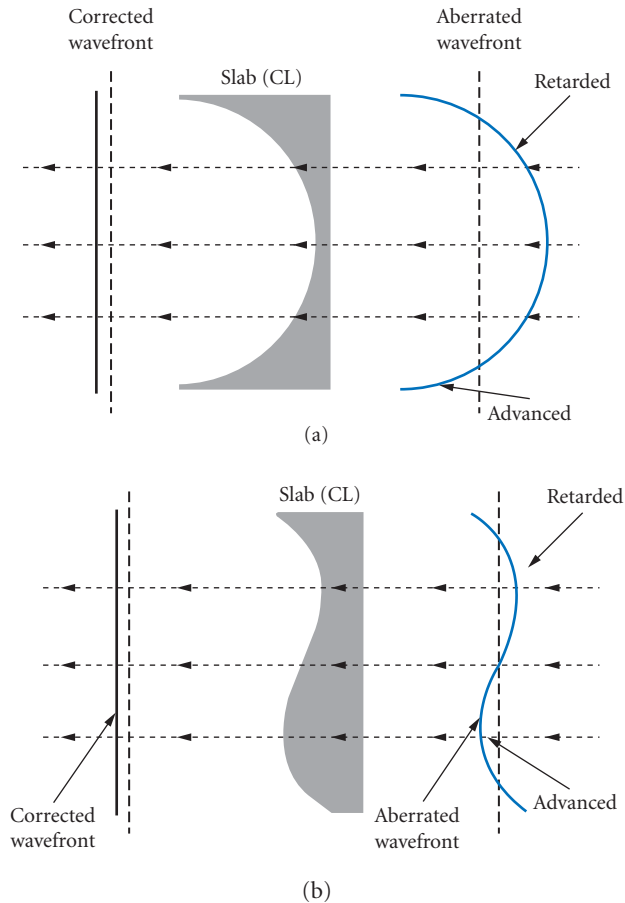
As is well established, there are several aberrations induced by the optics of the eye which can influence the quality of vision. Lower-order aberrations include tilt or prism, astigmatism, and defocus. The most common high-order aberrations found in human eyes are spherical aberration and coma. Whereas in spectacle lens designs, the primary problems pertain to minimizing the effects of oblique astigmatism and curvature of field to optimize quality of vision in peripheral gaze, this is not a significant problem with contact lenses as the lens moves with the eye in all areas of gaze. However, spherical aberration and coma can be problematic with contact lens wear, notably when the lens decenters.<sup>22</sup> The latter can be especially important. If a lens is decentered 0.5 mm, an amount of coma approximately equivalent to the amount of spherical aberration is induced. If this decentration is 1 mm, a magnitude of coma equal to twice the level of spherical aberration is induced.<sup>34</sup> A well-centered lens is especially important as recent research has found that coma is, in fact, more visually compromising than spherical aberration.<sup>35</sup>

The introduction of sensitive, sophisticated aberrometers has been beneficial in evaluating the relationship of contact lenses and aberrations. Several studies have evaluated the optical quality of eyes wearing different types of contact lenses. Certain types of soft lenses (i.e., manufactured via cast-molding or spin-casting) induced more high-order aberrations, such as coma and spherical aberration as measured via aberrometry.<sup>36</sup> In comparing both soft and GP lenses it was found that, whereas both soft and GP lenses induce more aberrations for the eyes that have low wavefront aberrations, soft lens wear tends to induce more higher-order aberrations and GP lens wear tends to reduce higher-order aberrations.<sup>37,38</sup> CRT lenses used to dramatically reshape the cornea to reduce myopia when worn at night have been found to increase higher-order aberrations, especially spherical aberration.<sup>39,40</sup>

To correct for wavefront aberrations it is important to slow the propagation of the wavefront in areas of the pupil where it is advanced and speed up the propagation of the wavefront in areas where it's retarded.<sup>41</sup> This can be accomplished in a contact lens with localized changes in lens thickness. Essentially the lens must be thinner in areas where the eye it is correcting shows the greatest delays in wavefront and thickest where the eye's wavefront demonstrates its greatest advancement. This is shown in Fig. 19 in which the introduction of an aberration-correcting contact lens in both myopic (Fig. 19a) and comatic (Fig. 19b) wavefronts is demonstrated.<sup>41</sup>

Several contact lens designs, especially soft lenses, have been recently introduced with some aberration-correction ability. Keratoconic eyes reportedly have higher-order aberration levels as much as 5.5 times higher than normal eyes;<sup>42</sup> however, custom soft lenses have been found to provide a threefold reduction in aberration on keratoconic eyes.<sup>43,44</sup> A few designs have attempted to introduce a constant level of spherical aberration at all lens powers that is equal in magnitude but opposite in sign to the population mean for human eyes.<sup>42</sup> Some designs achieve this over part of the dioptric power range (i.e., PureVision from Bausch & Lomb); the Biomedics Premier lens from CooperVision achieved this goal over the entire dioptric range.<sup>42,45</sup> The Biomedics XC lens from CooperVision has been introduced with the strategy of having zero spherical aberration at all lens powers and, therefore, will neither add to or correct the eye's own spherical aberration. These lenses have the advantages of not introducing large amounts of spherical aberration often provided by standard spherical soft lenses and as these lenses have no spherical aberration, they will not introduce coma if they decenter.

It's evident that—although documented clinical success has not been definitely established with these designs—the future looks promising for aberration-controlling lenses. Likewise, manufacturers will introduce designs that either center well or decenter by a fixed amount. The introduction of



**FIGURE 19** Schematic showing the correction of a (a) myopic and (b) comatic wavefront by the introduction of an aberration-correcting contact lens. (From Ref. 41.)

contact lens-only aberrometers, such as the ClearWave (AMO Wavefront Sciences), allows for a greater ability to monitor the aberration characteristics of these designs and ultimately introduce designs that are more effective in correcting for the aberrations of the eye.

## 20.6 CONVERGENCE AND ACCOMMODATION EFFECTS

### Convergence

Spectacles which are well centered for distance vision but which are also used for all directions of gaze can induce a prismatic effect when the gaze is not straightahead. When viewing at near, a spectacle-wearing hyperopic patient will experience a base-out effect and a spectacle wearing myopic patient will experience a base-in effect. The fact that the optical center of a contact lens remains (or ideally should remain) centered in front of the eye in different directions of gaze results in several benefits

for the contact lens wearer via reducing or eliminating the following prismatic effects common to spectacle wear including:<sup>10,46</sup>

1. Decreased near convergence demand for bilateral hyperopic patients and increased near convergence demand for bilateral hyperopic patients.
2. Base right prism and base left prism for bilateral myopic patients, and base left prism and base right prism for bilateral hyperopic patients, in right and left gaze, respectively.
3. Vergence demand alterations in anisometropia or antimetropia, required for right and left gaze.
4. Vertical prismatic effects in up and down gaze, and imbalances in down gaze resulting from anisometropia or antimetropia.

There are a few potential problems, however, when changing a patient from spectacles to contact lenses. If the patient requires lateral prism correction to maintain alignment and fusional ability, this is not possible in contact lenses. For the young child this is an important consideration because of the cosmesis and freedom of vision provided by contact lenses. However, only after successful binocular vision training and/or surgical intervention should contact lenses be prescribed. Likewise, patients requiring correction for a vertical deviation will (almost always) not achieve that correction in a contact lens. It is possible to prescribe base down prism in a contact lens; however, this is unlikely to solve the problem unless prescribed binocularly as asthenopic complaints are likely due to the prism differential between the two eyes.

Finally, for a given distance, the contact lens-wearing hyperopic patient exhibits less convergence and the myope exerts more convergence than with spectacles. The binocular myopic patient loses the “base-in” effect received from spectacles and, therefore, if exophoria is present and the patient has an abnormally long or remote near point of convergence, eyestrain and even diplopia may result. This same problem may result for the esophoric hyperopic patient who loses the “base-out” effect provided by spectacles and—if borderline fusional divergence ability was initially present—the decreased convergence demand at near induced by contact lens wear may result in compromised binocularity. Fortunately, as will be discussed, accommodative vergence tends to compensate for the differences in vergence demands between the two modes of correction.

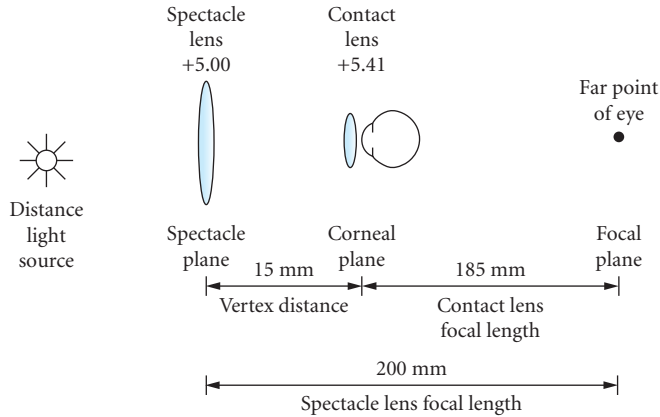
## Accommodative Demands

The review of effective power considerations and, specifically, vertex distance, becomes especially appropriate as it pertains to accommodative demand. As will be discussed, the vergence entering the eye differs with light coming from a distance object versus a near object. Accommodative demand represents the difference between distance and near refractive corrections.<sup>10</sup>

Accommodative demand can best be described via reviewing the differences in demand between a hyperopic and myopic patient, both viewing at a 40 cm distance. Figure 20 shows the vergence of light for a +5 D spectacle-wearing patient, using a vertex distance of 15 mm. A vergence of 0 would arrive at the spectacle lens from a distant object. Via use of the effective power equation, the power at the corneal plane would equal

$$\begin{aligned}
 F(c) &= \frac{F(\text{sp})}{1 - dF(\text{sp})} \\
 &= + \frac{5.00}{1 - 0.015(+5.00)} \\
 &= +5.41 \text{ D}
 \end{aligned}$$

This value can also be obtained by simply subtracting the difference in focal lengths. The spectacle lens has a focal length of  $1/+5.00$  or 0.2 m (or 200 mm). At the corneal plane relative to the far point of the eye, the focal length is reduced by 15 mm or equals 185 mm (0.185 m).  $1/0.185 \text{ m} = +5.41 \text{ D}$

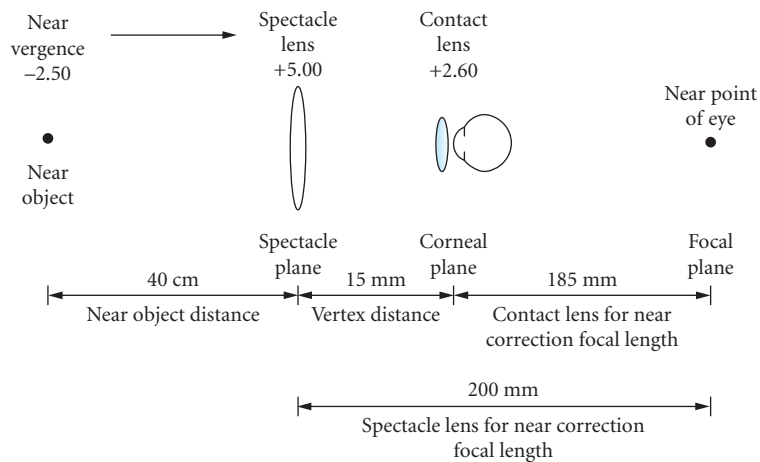


**FIGURE 20** The power and corresponding focal lengths of a +5.00 D spectacle lens at the corneal plane when viewing a distant object.

which would equal the vergence of light at the corneal plane. For the emmetrope, it can be assumed that the vergence of light at the cornea is zero.

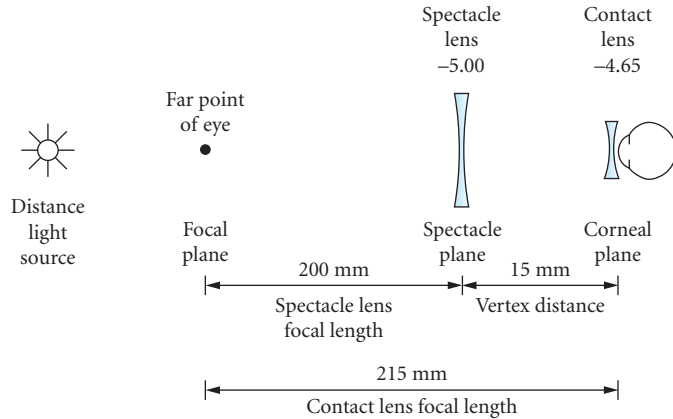
When viewing at a distance of 40 cm, a vergence of  $-2.50$  D (i.e.,  $1/0.40$ ) would enter the spectacle lens (Fig. 21). A vergence of  $+2.50$  D (i.e.,  $-2.50 + (+)5.00$ ) would exit the spectacle lens. To determine the power at the corneal plane, the effective power equation can be used.

$$\begin{aligned}
 F(c) &= \frac{F(\text{sp})}{1 - dF(\text{sp})} \\
 &= + \frac{2.50}{1 - 0.015(+2.50)} \\
 &= +2.60 \text{ D}
 \end{aligned}$$



**FIGURE 21** The power and corresponding focal lengths of a +5.00 D spectacle lens at the corneal plane when viewing a near object.





**FIGURE 22** The power and corresponding focal lengths of a  $-5.00$  D spectacle lens at the corneal plane when viewing a distant object.

Likewise, this value can also be determined by subtracting the difference in focal lengths.  $1/2.50$  or  $0.4$  m ( $400$  mm) provides the focal length of the power exiting the spectacle lens and, considering the  $15$  mm reduction at the corneal plane, the power at the corneal plane would be  $1/0.385$  or  $+2.60$  D. Therefore, the corneal accommodative demand for the  $5$  D hyperope would equal the distance accommodative demand minus the near accommodative demand or  $+5.41 - (+)2.60 = +2.81$  D. This can be compared to the emmetrope who would have a corneal accommodative demand of  $1/0.415$  (i.e.,  $40$  cm plus the  $15$  mm vertex distance) or  $2.41$  D. Therefore, the hyperopic patient requires more accommodation than the emmetropic patients when viewing a near object with spectacles.

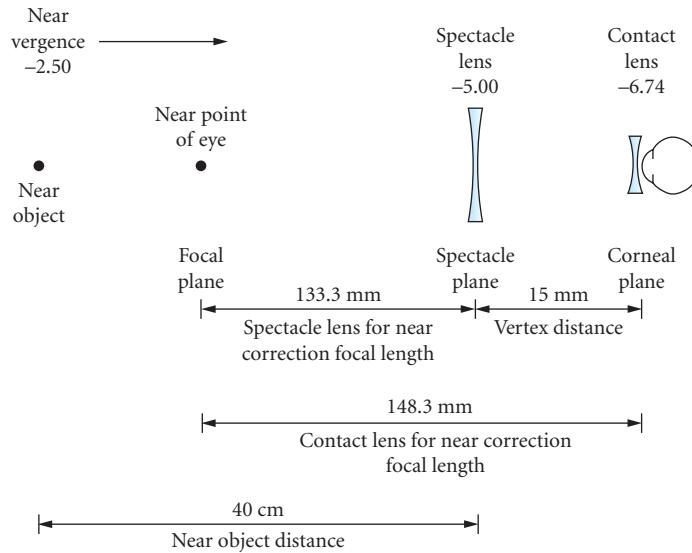
Figure 22 shows the vergence of light for a  $-5$  D spectacle-wearing patient, using a vertex distance of  $15$  mm. A vergence of  $0$  would arrive at the spectacle lens from a distant object. Via use of the effective power equation, the power at the corneal plane would equal

$$\begin{aligned}
 F(c) &= \frac{F(sp)}{1 - dF(sp)} \\
 &= \frac{-5.00}{1 - 0.015(-5.00)} \\
 &= -4.65 \text{ D}
 \end{aligned}$$

This value can also be obtained by simply subtracting the difference in focal lengths. The spectacle lens has a focal length of  $1/-5.00$  or  $-0.2$  m (or  $-200$  mm). At the corneal plane relative to the far point of the eye, the focal length is increased by  $15$  mm or equals  $-215$  mm ( $-0.215$  m).  $1/-0.215$  m =  $-4.65$  D which would equal the vergence of light at the corneal plane.

When viewing at a distance of  $40$  cm, a vergence of  $-2.50$  D would enter the spectacle lens (Fig. 23). A vergence of  $-7.50$  D (i.e.,  $-2.50 + (-)5.00$ ) would exit the spectacle lens. To determine the power at the corneal plane, the effective power equation can be used

$$\begin{aligned}
 F(c) &= \frac{F(sp)}{1 - dF(sp)} \\
 &= \frac{-7.50}{1 - 0.015(-7.50)} \\
 &= -6.74 \text{ D}
 \end{aligned}$$



**FIGURE 23** The power and corresponding focal lengths of a  $-5.00$  D spectacle lens at the corneal plane when viewing a near object.

Likewise, this value can also be determined by subtracting the difference in focal lengths.  $1/-7.50$  or  $-0.1333$  m ( $-133.3$  mm) provides the focal length of the power exiting the spectacle lens and, considering the increase of 15 mm at the corneal plane, the power at the corneal plane would be  $1/-0.1483$  or  $-6.74$  D. Therefore, the corneal accommodative demand for the 5 D myope would equal the distance accommodative demand minus the near accommodative demand or  $-4.60 - (-6.74) = +2.14$  D. This can be compared to the emmetrope who would have a corneal accommodative demand of 2.41 D; therefore, the myopic patient requires less accommodation than the emmetropic patient when viewing a near object with spectacles. The differences in corneal accommodative demand for myopic and hyperopic patients has been well documented.<sup>46-49</sup> Table 7 shows the difference in corneal accommodative demand at the corneal plane as it pertains to myopic and hyperopic spectacle lens powers.<sup>10</sup>

As contact lenses are positioned at the corneal plane, they induce accommodative demands equivalent to that of emmetropia. Therefore, the emerging presbyopic hyperope who changes from spectacles to contact lenses may find that the near symptoms are reduced. However, the emerging

**TABLE 7** Accommodative Demands at the Corneal Plane with Spectacle Lens Wear

Difference in Corneal Plane Accommodative Demand Compared to Emmetropia	Back Vertex Power of Hyperopic Spectacle Lens (D)	Back Vertex Power of Myopic Spectacle Lens (D)
+/- 0.25	+3.25	-3.87
+/- 0.50	+6.00	-8.37
+/- 0.75	+8.62	-13.75
+/- 1.00	+10.87	-20.87

\*Assume a near distance of 40 cm and a vertex distance of 15 m.  
Source: From Ref. 10.

presbyopic highly myopic patient should be advised that a separate near correction (or a contact lens multifocal) may be necessary to satisfy all vision demands.

Overall, it can be concluded that hyperopic-correcting spectacle lenses will require an increase in convergence over that required when uncorrected whereas myopic-correcting lenses reduce the convergence required.<sup>50</sup> However, as indicated in this section, myopic patients accommodate more and hyperopic patients accommodate less when wearing a contact lens correction versus spectacles. Therefore, the overall accommodation-convergence ratio is only minimally impacted.

## 20.7 PRISMATIC EFFECTS

### Prism-Ballasted Contact Lenses

Prism within a contact lens is produced by varying the thickness from the superior to the inferior regions while maintaining the same front and back surface curvatures. There are several types of contact lenses for which prism base down is incorporated within the design. The most popular design, including such prism, pertains to segmented, translating bifocal GP lenses. These lenses include a distance zone in the upper half of the lens, a near zone in the bottom half of the lens and (with some designs) an intermediate power zone between the near and distance segments. Typically, 1.5 to 3.0Δ base down is incorporated within the lens to stabilize it such that the thicker inferior edge will interact with the inferior lid margin on inferior gaze and push the lens up (termed “translation”) such that the wearer will be viewing through the inferior near power zone when reading. Prism has also been used for stabilizing soft toric lenses as well as front toric GP lenses (i.e., used to correct residual astigmatism), although front toric GPs are little used due to the application of soft toric lenses in most of these cases and soft toric lenses tend to use other—reduced mass—options for stabilization.

The amount of prism within a contact lens can be computed using the following formula:<sup>10</sup>

$$P = \frac{100(n-1)(BT-AT)}{BAL}$$

where  $P$  = prismatic power in prism diopters ( $\Delta$ )

$n$  = refractive index of prismatic lens

BT = base thickness of prismatic component of lens

AT = apex thickness of prismatic component of lens

BAL = length of base-apex line or the diameter of the contact lens long the base-apex line of prism

For example, if the diameter of the lens was 9.0 mm, the thickness at the apex was 0.11 mm and the thickness at the base was 0.40 mm, and the refractive index of the material was 1.46, the amount of prism within this lens would be

$$\begin{aligned} P &= \frac{100(n-1)(BT-AT)}{BAL} \\ &= \frac{100(1.45-1)(0.40-0.11)}{9.0} \\ &= 1.48 \text{ or approximately } 1.50\Delta \text{ base down} \end{aligned}$$

Refractive power of a prismatic lens does vary along the base-apex line via the aforementioned formula for deriving back vertex power. This is the result of the fact that lens thickness is increased toward the base with an absence of change in surface curvature. Via Prentice’s rule (to be discussed), back surface power becomes less minus/more plus as thickness increases toward the base of the prism.

## Unintentional (Induced) Prism

Prentice's rule can also be used for determining the prism induced by lens decentration on the eye. This is due to the fact that a contact lens is considered to be separated from the tear lens by a thin layer of air.<sup>51</sup> Prentice's rule is

$$P = Fd$$

where  $P$  = prism induced in prism diopters  
 $F$  = dioptric power of lens  
 $d$  = decentration in centimeters

For example, if a +5.00 D lens decenters 2 mm, the resulting induced prism would be  $+6.00 \text{ D} \times 0.2 = 1.2\Delta$  base down. If the other lens is well centered, this may be sufficient vertical imbalance to cause symptoms of eyestrain. Fortunately, soft lenses typically do not decenter on the eye due to their large size and the fact they drape over the cornea. GP lenses do move with the blink but typically, if decentration is present, it is essentially the same for both eyes. In cases of anisometropia, however, any decentration could result in asthenopic complaints. For example, if the patient was wearing a +2 D lens in the right eye and a +7 D lens in the left eye and both lenses decentered 2 mm inferiorly, the amount of induced prism would be OD:  $+2.00 \times 0.2 = 0.4\Delta$ ; OS:  $+7.00 \times 0.2 = 1.4\Delta$ . The resulting difference would be  $1.0\Delta$ . Typically patients will not be symptomatic with less than  $1\Delta$ ; however, in this case, asthenopic complaints are possible.

## 20.8 MAGNIFICATION

It has been well established that when a spectacle lens is moved toward the eye, the retinal image decreases in size for a plus lens and increases in size for a minus lens. Therefore, with contact lenses both of these effects would tend to be advantageous. The hyperopic patient with a high prescription—and most certainly the aphakic patient—are typically very satisfied to find that objects have returned to nearly a normal size. Myopic patients are pleased to find that everything looks larger with contact lenses. Several methods have been used to assess magnification effects, including magnification of correction and relative spectacle magnification.

### Magnification of Correction (Spectacle Magnification)

Spectacle magnification (SM) is the ratio of retinal image size of the corrected ametropic eye to the retinal image size of the same eye uncorrected.<sup>10</sup> This has been used as an index of how corrective lenses alter retinal image size as compared by the patient before and after corrective lenses are placed on the eye. Spectacle magnification formulas include a power factor and a shape factor<sup>10,47,52-54</sup>

$$\text{SM} = \underbrace{\frac{1}{1} - h(\text{BVP})}_{\text{Power factor}} \times \underbrace{\frac{1}{1} - (t/n')F_1}_{\text{Shape factor}}$$

where SM = spectacle magnification or magnification of correction  
 BVP = back vertex power of correcting lens (D)  
 $h$  = stop distance, from plane of correcting lens to ocular entrance pupil in meters  
 (i.e., vertex distance +3 mm)  
 $t$  = center thickness of correcting lens (m)  
 $n'$  = refractive index of correcting lens  
 $F_1$  = front surface power of correcting lens (D)

With both contact lenses and spectacle lenses the shape factor is nearly always greater than 1.0, representing magnification, due to their respective convex anterior surfaces (i.e.,  $F_1$  is a positive number). However, for contact lenses, the lacrimal lens needs to be considered. Therefore, for a contact lens, the following shape factor should be used:

$$\text{Shape factor} = \frac{1}{1} - (t/n')F_1 \times \frac{1}{1} - (t_L/n_L)F_L$$

where  $t$  = center thickness of correcting lens (m)  
 $n'$  = refractive index of correcting lens  
 $F_1$  = front surface power of correcting lens (D)  
 $t_L$  = center thickness of lacrimal lens in meters  
 $n_L$  = refractive index of lacrimal lens (1.3375)  
 $F_L$  = front surface power of lacrimal lens in keratometric diopters

The power factor for contact lenses is essentially the same as that for spectacle lenses with the exception that, as the vertex distance is 0, the stop distance would be only equal to 3 mm.

In high myopia, a significant minification ( $SM < 1.0$ ) of the retinal image occurs with spectacle correction; this minification is greatly reduced with contact lens wear. Therefore, the highly myopic patient changing from spectacles to contact lenses may comment that their vision is clearer with contact lenses.

The change in spectacle magnification when changing from spectacle to contact lens correction can be compared with the following formula:

$$\frac{\text{Contact lens power factor}}{\text{Spectacle lens power factor}} = 1 - h(\text{BVP})$$

where  $h$  = stop distance, from plane of correcting lens to ocular entrance pupil in meters  
 (i.e., vertex distance +3 mm) and  
 BVP = back vertex distance of spectacle lens (in D)

For the highly myopic (i.e.,  $-10$  D) patient, approximately 15 percent minification will be present with spectacles and only about 5 percent with contact lenses. For the aphakic patient, the magnification will be 25 to 30 percent with spectacles which would be reduced to 5 to 8 percent with contact lenses. Nevertheless, if the patient is a unilateral aphake wearing one contact lens, this image difference may still result in binocular vision problems. Therefore, intraocular lenses, with a stop distance of 0, is optimum in maintaining similar image sizes between the two eyes.

## Relative Spectacle Magnification

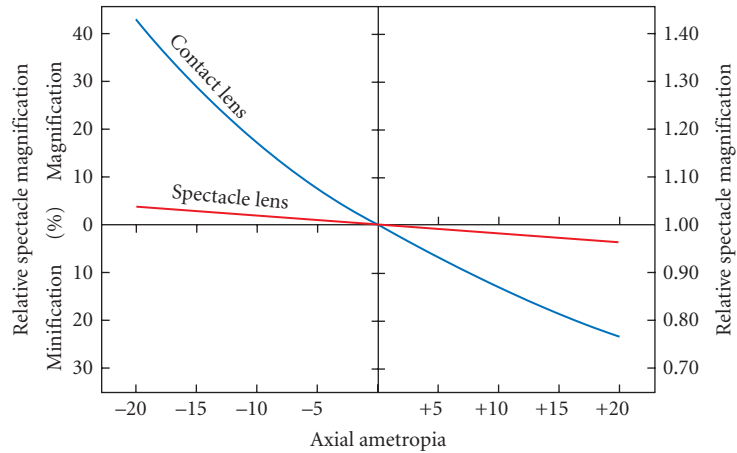
Retinal image size can also be estimated via comparing the corrected ametropic retinal image to that of a standard emmetropic schematic eye. Relative spectacle magnification (RSM) is the ratio derived for spectacle lenses.<sup>10,47,52-54</sup> Ametropia can be purely axial, resulting from the axial elongation of the globe, refractive, resulting from abnormal refractive component(s) of the eye or a combination of both factors.

Relative spectacle magnification for axial ametropia can be derived from the following equation

$$\text{RSM} = \frac{1}{1 + g(\text{BVP})}$$

where RSM = relative spectacle magnification for axial ametropia  
 $g$  = distance in meters from anterior focal point of eye to correcting lens;  $g = 0$  if lens is 15.7 mm in front of the eye  
 BVP = back vertex power of refractive correction (D)

with axial ametropia,  $g$  is equal to 0 or close to it for most spectacle lenses; therefore, the relative spectacle magnification is very close to 0. (See Fig. 24.)<sup>51</sup> In theory, anisometropia of axial origin is best corrected with spectacles and, in fact, anisokonia problems may be encountered with contact lens wear. However, it has been suggested that in axial myopic eyes the retina will stretch to compensate



**FIGURE 24** Relative spectacle magnification for axial ametropia when corrected by spectacles and contact lenses. (From Ref. 51.)

resulting in greater spacing between retinal receptors and receptive fields that effectively reduce the neurological image.<sup>51</sup>

Relative spectacle magnification for refractive ametropia can be derived from the following equation:

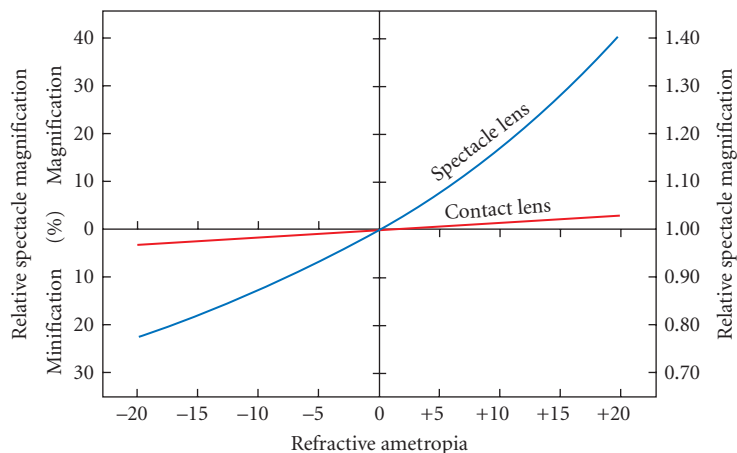
$$\text{RSM} = \frac{1}{1 - d(\text{BVP})}$$

where RSM = relative spectacle magnification for axial ametropia

$d$  = stop distance in meters from correcting lens to entrance pupil (i.e., vertex distance +3 mm)

BVP = back vertex power of refractive correction (D)

This equation is the same for the shape factor for spectacle magnification. It can be observed in Fig. 25 that the RSM for contact lens wear is close to 1.00 but, conversely, it is quite minified for highly myopic lenses and very magnified for high plus lenses.<sup>55</sup>



**FIGURE 25** Relative spectacle magnification for refractive ametropia when corrected by spectacles and contact lenses. (From Ref. 51.)

## 20.9 SUMMARY

This chapter emphasizes the many optical benefits of contact lenses versus spectacles. In particular, the elimination of undesirable aberrations and induced prism when viewing away from the optical center and a lesser magnified (or minified) image size, it is important to emphasize that contact lenses will result in an increase in accommodative demand for the emerging presbyopic myope when changing from spectacles. In addition, children with binocular vision problems who benefit from the lateral prism in spectacles will not benefit from contact lens wear and may, in fact, be very poor candidates. This chapter also emphasizes contact lens optics that are in common use today. Determining the accommodative demand, induced prism and magnification are important in some contact lens wearers. Determination of contact lens power is important in all contact lens wearers.

## 20.10 ACKNOWLEDGMENTS

The author would like to acknowledge the contributions of Teresa Mathew and Jason Bechtoldt.

## 20.11 REFERENCES

1. O. Wichterle and D. Lim, "Hydrophilic Gels for Biological Use," *Nature (London)* **185**:117–118(1960).
2. M. F. Refojo, "The Chemistry of Soft Hydrogel Lens Materials," in M. Ruben, ed., *Soft Contact Lenses*, New York, John Wiley & Sons, 19–39 (1978).
3. M. Refojo, "The Relationship of Linear Expansion to Hydration of Hydrogel Contact Lenses," *Cont. Lens Intraoc. Lens Med. J.* **1**:153–162 (1976).
4. I. Fatt, "Water Flow and Pore Diameter in Extended Wear Gel Materials," *Am. J. Optom. Physiol. Opt.* **55**:294–301 (1978).
5. P. B. Morgan, C. A. Woods, D. Jones, et al., "International Contact Lens Prescribing in 2006," *Contact Lens Spectrum* **22**(1):34–38 (2007).
6. D. F. Sweeney, N. A. Carnt, R. Du Toit, et al., "Silicone Hydrogel Lenses for Continuous Wear," in E. S. Bennett and B. A. Weissman, eds., *Clinical Contact Lens Practice*, Philadelphia, Lippincott Williams & Wilkins, 693–717 (2005).
7. A. Cannella and J. A. Bonafini, "Polymer Chemistry," in E. S. Bennett and B. A. Weissman, eds., *Clinical Contact Lens Practice*, Philadelphia, Lippincott Williams & Wilkins, 233–242 (2005).
8. B. A. Weissman and K. M. Gardner, "Power and Radius Changes Induced in Soft Contact Lens Systems by Flexure," *Am. J. Optom. Physiol. Opt.* **61**:239 (1984).
9. E. S. Bennett, "Silicone/Acrylate Lens Design," *Int. Contact Lens. Clin.* **11**:547 (1984).
10. W. J. Benjamin, "Optical Phenomena of Contact Lenses," in E. S. Bennett and B. A. Weissman, eds., *Clinical Contact Lens Practice*, Philadelphia, Lippincott Williams & Wilkins, 111–163 (2005).
11. R. B. Mandell, "Optics," in R. B. Mandell, ed., *Contact Lens Practice* (4th ed.), Illinois, Springfield, Charles. C. Thomas, 954–980 (1988).
12. M. W. Ford and J. Stone, "Optics and Contact Lens Design," in A. J. Phillips and L. Speedwell, eds., *Contact Lenses* (5th ed.), London, Elsevier, 129–158 (2007).
13. W. A. Douthwaite, "The Contact Lens," in W. A. Douthwaite, ed., *Contact Lens Optics and Lens Design* (3d ed.), London, Elsevier, 27–55 (2006).
14. M. W. Ford, "Computation of the Back Vertex Powers of Hydrophilic Lenses," *Paper read at the interdisciplinary Conference on Contact Lenses, Department of Ophthalmic Optics and Visual Science, City University, London* (1976).
15. B. A. Weissman, "Loss of Power with Flexure of Hydrogel Plus Lenses," *Am. J. Optom. Physiol. Opt.* **63**:166–169 (1986).
16. B. A. Holden, "The Principles and Practice of Correcting Astigmatism with Soft Contact Lenses," *Aust. J. Optom.* **58**:279–299 (1975).

17. E. S. Bennett, P. Blaze, and M. R. Remba, "Correction of Astigmatism," in E. S. Bennett and V. A. Henry, eds., *Clinical Manual of Contact Lenses* (2d ed.), Philadelphia, Lippincott Williams & Wilkins, 351–409 (2000).
18. W. J. Benjamin, "Practical Optics of Contact Lens Prescription," in E. S. Bennett and B. A. Weissman, eds., *Clinical Contact Lens Practice* (2d ed.), Philadelphia, Lippincott Williams & Wilkins, 165–195 (2005).
19. J. L. Pascal, "Cross Cylinder Tests—Meridional Balance Technique," *Opt. J. Rev. Optom.* **87**:31–35 (1950).
20. R. B. Mandell and C.F. Moore, "A Bitoric Guide That is Really Simple," *Contact Lens Spectrum* **3**:83–85 (1988).
21. E. S. Bennett, "Astigmatic Correction," in E. S. Bennett and M. M. Hom, eds., *Manual of Gas Permeable Contact Lenses* (2d ed.), Elsevier, St. Louis, 286–323 (2004).
22. W. A. Douthwaite, "Aspherical Surface," in W. A. Douthwaite, ed., *Contact Lens Optics and Lens Design* (3d ed.), Elsevier, London, 91–125 (2006).
23. W. Feinbloom, "Corneal Contact Lens having Inner Ellipsoidal Surface. US Patent No. 3,227,507, January 4, 1966.
24. G. L. Feldman and E. S. Bennett, "Aspheric Lens Designs," in E. S. Bennett and B. A. Weissman, eds., *Clinical Contact Lens Practice*, Philadelphia, JB Lippincott, 16-1–16-10 (1990).
25. K. Richdale, G. L. Mitchell, and K. Zadnik, "Comparison of Multifocal and Monovision Soft Contact Lens Corrections in Patients with Low-Astigmatic Presbyopia," *Optom. Vis. Sci.* **83**(5):266–273 (2006).
26. T. B. Edrington and J. T. Barr, "Aspheric What?," *Contact Lens Spectrum* **17**(5):50 (2003).
27. E. S. Bennett, "Contact Lens Correction of Presbyopia," *Clin. Exp. Optom.* **91**(3):265–278 (2008).
28. A. S. Rajagopalan, E. S. Bennett, and V. Lakshminarayanan, "Visual Performance of Subjects Wearing Presbyopic Contact Lenses," *Optom. Vis. Sci.* **83**:611–615 (2006).
29. A. S. Rajagopalan, E. S. Bennett, and V. Lakshminarayanan, "Contrast Sensitivity with Presbyopic Contact Lenses," *J. Modern Optics* **54**:7–9:1325–1332 (2007).
30. E. S. Bennett, "Lens Design, Fitting, and Troubleshooting," in E. S. Bennett and R. M. Grohe, eds., *Rigid Gas-Permeable Contact Lenses*, New York, Professional Press, 189–224 (1986).
31. A. Musset and J. Stone, "Contact Lens Design Tables," Butterworths, London, 79–108 (1981).
32. M. Townsley, "New Knowledge of the Corneal Contour," *Contacto* **14**(3):38–43 (1970).
33. W. A. Douthwaite, "Contact Lens Design," in W. A. Douthwaite, ed., *Contact Lens Optics and Lens Design* (3d ed.), Elsevier, London, 165–201 (2006).
34. P. S. Kollbaum and A. Bradley, "Correcting Aberrations with Contact Lenses," *Contact Lens Spectrum* **22**(11):24–32 (2007).
35. A. Bradley, Personal communication, May, 2009.
36. H. Jiang, D. Wang, L. Yang, P. Xie, and J. C. He, "A Comparison of Wavefront Aberrations in Eyes Wearing Different Types of Soft Contact Lenses," *Optom. Vis. Sci.* **83**(10):769–774 (2006).
37. F. Lu, X. Mao, J. Qu, D. Xu, and J. C. He, "Monochromatic Wavefront Aberrations in the Human Eye with Contact Lenses," *Optom. Vis. Sci.* **80**(2):135–141 (2003).
38. X. Hong, N. Himebaugh, and L. N. Thibos, "On-Eye Evaluation of Optical Performance of Rigid and Soft Contact Lenses," *Optom. Vis. Sci.* **78**(12):872–880 (2001).
39. C. E. Joslin, S. M. Wu, T. T. McMahon, and M. Shahidi, "Higher-Order Wavefront Aberrations in Corneal Refractive Therapy," *Optom. Vis. Sci.* **80**(12):805–811 (2003).
40. D. A. Berntsen, J. T. Barr, and G. L. Mitchell, "The Effect of Overnight Contact Lens Corneal Reshaping on Higher-Order Aberrations and Best-Corrected Visual Acuity," *Optom. Vis. Sci.* **82**(6):490–497 (2005).
41. P. S. Kollbaum and A. Bradley, "Correcting Aberrations with Contact Lenses: Part 2," *Contact Lens Spectrum* **22**(12):31–34 (2007).
42. S. Pantanelli and S. MaCrae, "Characterizing the Wave Aberration in Eyes with Keratoconus or Penetrating Keratoplasty Using a High Dynamic Range Wavefront Sensor," *Ophthalmology* **114**(11):2013–2021 (2007).
43. R. Sabesan, M. Jeong, L. Carvalho, I. G. Cox, D. R. Williams, and G. Yoon, "Vision Improvement by Correcting Higher-Order Aberrations with Customized Soft Contact Lenses in Keratoconic Eyes," *Opt. Lett.* **32**(8):1000–1002 (2007).
44. J. D. Marsack, K. E. Parker, K. Pesudovs, W. J. Donnelly, and R. A. Applegate, "Uncorrected Wavefront Error and Visual Performance During RGP Wear in Keratoconus," *Optom. Vis. Sci.* **84**(6):463–470 (2007).



45. P. Kollbaum and A. Bradley, "Aspheric Contact Lenses: Fact and Fiction," *Contact Lens Spectrum* **20**(3):34–38 (2005).
46. M. Alpern, "Accommodation and Convergence with Contact Lenses," *Am. J. Optom.* **26**:379–387 (1949).
47. G. Westheimer, "The Visual World of the New Contact Lens Wearer," *J. Am. Optom. Assoc.* **34**:135–138 (1962).
48. J. Neumueller, "The Effect of the Ametropic Distance Correction Upon the Accommodation and Distance Correction," *Am. J. Optom.* **15**:120–128 (1938).
49. J. S. Hermann and R. Johnson, "The Accommodation Requirement in Myopia," *Arch. Ophthalmol.* **76**:47–51 (1966).
50. W. A. Douthwaite, "Basic Visual Optics," in W. A. Douthwaite, ed., *Contact Lens Optics and Lens Design* (3d ed.), Elsevier, London, 1–26 (2006).
51. R. B. Mandell, "Contact Lens Optics," in R. B. Mandell, ed., *Contact Lens Practice* (4th ed.), Illinois, Springfield, Charles C. Thomas, 954–979 (1988).
52. A. G. Bennett, "Optics of Contact Lenses," (4th ed.), Association of Dispensing Opticians, London, (1966).
53. J. F. Neumiller, "The Optics of Contact Lenses," *Am. J. Optom. Arch. Am. Acad. Optom.* **45**:786–796(1968).
54. S. Duke-Elder and D. Abrams, "Optics," in S. Duke-Elder and D. Abrams, eds., Section I, *Ophthalmic Optics and Refraction*, volume V of Duke-Elder S, ed., *System of Ophthalmology*, C.V. Mosby, Missouri, St. Louis, 25–204 (1970).
55. R. B. Mandell, "Corneal Topography," in R. B. Mandell, ed., *Contact Lens Practice* (4th ed.), Illinois, Springfield, Charles C. Thomas, 107–135 (1988).

---

# INTRAOCULAR LENSES

---

## Jim Schwiegerling

*Department of Ophthalmology  
University of Arizona  
Tucson, Arizona*

---

### 21.1 GLOSSARY

---

**Accommodating intraocular lens.** An artificial lens implanted in the eye that changes power and/or position in response to contraction of the ciliary muscle.

**Accommodation.** The ability to change the power of the eye and focus on objects at different distances.

**Aphakia.** A condition of the eye without a crystalline lens or intraocular lens.

**Aspheric intraocular lens.** An artificial implanted lens that has at least one aspheric surface. Typically used to compensate spherical aberration of the cornea.

**Capsulorhexis.** Removal of the lens capsule. Usually during cataract surgery, the anterior portion of the capsule is removed to allow access to the crystalline lens within the capsule.

**Cataract.** An opacification that occurs within the crystalline lens that reduces the quality of vision.

**Chromophore.** A molecule doped into the lens material that absorbs specific wavelength bands such as ultraviolet or blue light.

**Dysphotopsia.** Stray light effects encountered following implantation of intraocular lenses. Positive dysphotopsia causes streaks or glints of light seen in the peripheral vision. Negative dysphotopsia causes dark bands or shadows to appear in peripheral vision.

**Haptic.** A structure that aids in alignment and support of intraocular lenses within the eye. Typically two or three arms emanate from the edge of the lens. Plate haptics are rectangular flanges that protrude from the sides of the lens.

**Intraocular lens.** An artificial lens that is implanted into the eye to modify the eye's optical power.

**Lens capsule.** An elastic bag that encapsulates the crystalline lens. Most modern cataract procedures leave the capsule in place and only remove the crystalline lens within.

**Limbus.** The boundary between the cornea and the sclera, the white of the eye.

**Multifocal intraocular lens.** An artificial lens that incorporates two or more powers into the same lens. The purpose of these lenses is to allow objects at different distances to be in focus simultaneously.

**Phacoemulsification.** A technique for removing a cataractous crystalline where an ultrasonic probe is used to shatter the lens and then the lens bits are suctioned out of the eye.

**Phakic lens.** An artificial lens implanted into the eye while leaving the crystalline lens intact. Typically this procedure is used to alleviate refractive error and is not a treatment for cataracts.

**Posterior capsule opacification.** A potential side effect of the intraocular lens implant. The lens capsule can adhere to the implant and an opacification develops on the posterior side of the intraocular lens. Laser pulses are typically used to open a hole in the opacified capsule to restore vision.

**Presbyopia.** The state of the eye where accommodation has been completely lost due to stiffening and enlargement of the crystalline lens.

**Pseudophakia.** The state of having an artificial lens implanted within the eye.

**Sclera.** The white of the eye.

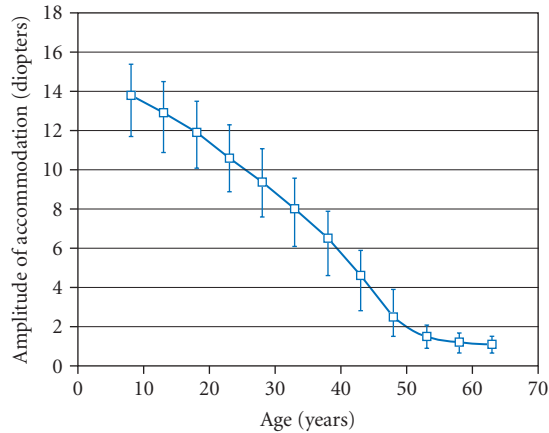
## 21.2 INTRODUCTION

---

The eye forms the optical system of the human visual system. A variety of references provide a good general introduction to the essential components of the eye and their function.<sup>1,2</sup> Here, a brief overview of the main optical elements and their mechanism of action will be provided. The eye consists of two separated lenses that ideally form an image on the retina, the array of photosensitive cells lining the back surface of the eyeball. The eye's first lens is the cornea, which is the clear membrane on the external portion of the eye. The cornea is a meniscus lens and has a power of about 43 diopters (D). The iris resides several millimeters behind the cornea. The iris is heavily pigmented to block light transmission and serves as the aperture stop of the eye. The diameter of the opening in the iris varies with light level, object proximity, and age. The crystalline lens is the second optical element of the eye. It lies immediately behind the iris and provides the focusing mechanism of the visual system. Together, the cornea and iris form images on the external environment onto the retina. At the retina, the light is converted to a neural signal and transmitted to the brain where the signals are interpreted into our perceived image of the surrounding scene.

The crystalline lens has several notable features that provide for a flexible optical system capable of focusing to a wide range of distances. First, the crystalline lens has a gradient index structure with its refractive index varying in both the axial and the radial directions. The maximum index of refractive occurs in the center or nucleus of the lens. The index then gradually decreases toward the lens periphery as well as the front and back surfaces. In addition, the crystalline lens is also flexible, allowing the lens to change shape and consequently power in response to muscular forces. The crystalline lens resides in an elastic bag called the capsule. In the young lens, this capsule causes a contraction of the periphery of the lens and an increase in the central thickness of the lens. As a result of this effect, the radii of curvature of the front and back surfaces of the lens decrease, leading to an increase in the optical power of the lens. In its maximally flexible state, the capsule can form the crystalline lens such that it gains a power up to a maximum 34 D.

The lens and capsule are suspended behind the iris by a series of ligaments called the zonules. One end of a zonule attaches to the perimeter of the lens/capsule and the other end attaches to the ciliary muscle. The ciliary muscle is a ring-shaped muscle that resides behind the iris. In its relaxed state, the ciliary muscle dilates, increasing tension on the zonules and exerting an outward radial force on the perimeter of the lens and capsule. This force causes the crystalline lens shape to flatten. The center thickness of the lens reduces and the radii of curvature of its front and back surfaces increase. The net effect of the lens flattening is a reduction in its power to about 20 D. Constriction of the ciliary muscle causes a reduction in the zonule tension and the elastic capsule causes the lens to thicken and increase its power. Accommodation is the ability to change the power of the crystalline lens through constriction or relaxation of the ciliary muscle. Accommodation allows the eye to bring objects at a variety of distances into focus on the retina. If the eye is focused at infinity when the ciliary muscle is in its relaxed state, then the fully accommodated crystalline lens allows objects at 70 mm from the eye to be brought into proper focus.



**FIGURE 1** Loss of the amplitude of accommodation with age.

The eye is meant to provide a lifetime of vision. However, as with all parts of the body, changes to the eye's performance occur with aging. One effect is the continual decrease in the amplitude of accommodation that occurs almost immediately after birth. Figure 1 shows the accommodative amplitude as a function of age.<sup>3</sup> The preceding description of lens function described a lens that could change its power from 20 to 34 D. With age, this range gradually reduces. The changes are usually not noticeable until the fifth decade of life. At this point, the range of accommodation is usually 20 D in the relaxed state to 23 D in the maximally accommodated state. Under these conditions, if the eye is focused at infinity for a relaxed ciliary muscle, then with maximum exertion, objects at 33 cm from the eye can be brought into proper focus. This distance is a typical distance reading material is held from the eye. Objects closer than 33 cm cannot be brought into focus and ocular fatigue can rapidly set in from continual full exertion of the ciliary muscle. Moving into the sixth decade of life, the range of accommodation continues to reduce, leading to the inability to focus on near objects. This lack of focusing ability is called presbyopia. Reading glasses or bifocal spectacles are often used to aid the presbyope in seeing near objects. These lenses provide additional power to the eye that can no longer be provided by the crystalline lens. Presbyopia is caused by the continual growth of the crystalline lens.<sup>4</sup> The lens consists of a series of concentric layers and new layers are added throughout life. This growth has two effects. First, the lens size and thickness increases with age, leading to a lens that is stiffer and more difficult to deform. Second, the diameter of the lens increases, leading to a reduced tension on the zonules. The combination of these two effects results in the reduction in accommodative range with age and eventually presbyopia.

A second aging effect that occurs in the eye is senile cataract. Cataract is the gradual opacification of the crystalline lens that affects visual function.<sup>5</sup> An increase in lens scatter initially causes a yellowing of the lens. As these scatter centers further develop large opacities occur within or at the surface of the lens. The progression of cataract initially leads to a reduction in visual acuity and losses in contrast sensitivity due to glare and halos. The endpoint of cataract is blindness, with the opacifications completely blocking light from reaching the retina. It is estimated that 17 million people worldwide are blind due to cataract.<sup>6</sup> However, the distribution of these blind individuals is skewed toward less developed countries. The prevalence of blindness due to cataracts is 0.014 percent in developed countries and 2 percent in undeveloped countries. The most common cause of lens opacification is chronic long-term exposure to the UV-B radiation in sunlight. Other less common incidences include traumatic, pharmacological, and congenital cataracts. Risk factors such as smoking and poor diet can lead to earlier emergence of cataract. However, since exposure to sunlight occurs nearly daily, there is high likelihood of developing cataracts within a lifetime. Due to this prevalence, treatments for cataracts have long been sought after.

**TABLE 1** Prevalence of Cataract in the Population of U.S. Residents

Age	Prevalence (%)
43–54	1.6
55–64	7.2
65–74	19.6
75–85	43.1

Currently, there exist no means for preventing cataract formation. Reducing risk factors such as UV-B exposure through limiting sun exposure and using sunglasses, quitting smoking, and healthy diet can delay the onset of cataracts, but still the opacifications are likely to emerge. Table 1 shows the prevalence of cataract in the United States.<sup>7</sup>

Treatment of cataracts is the main means for restoring lost vision caused by the opacifications. Variations in access to these treatments is a reason for the skew in blindness caused by cataracts in developed and undeveloped countries. In developed countries access to cataract treatment is high and usually provided when the cataractous lens has caused relatively mild loss in visual performance. In fact, cataract surgery is the most widely performed surgery in the United States.<sup>6</sup> In undeveloped countries, access may be nonexistent and severely limited leading to a high rate of blindness due to complete opacification of the lens.

### 21.3 CATARACT SURGERY

In restoring vision loss from cataracts, the opacified lens is removed to allow light to enter the eye again. This concept has been known in various evolutionary forms for thousands of years.<sup>8</sup> However, since the crystalline lens forms about one-third of the overall power of the eye, removing the lens leaves the patient severely farsighted. This condition is known as aphakia. Aphakia requires patients to wear high-power spectacle lenses to replace the roughly 20 D of power lost from the crystalline lens. These high-powered lenses are problematic for a variety of reasons. These effects include a 20 to 35 percent magnification of the image and peripheral distortion, since spectacles with spherical surface cannot be corrected for oblique astigmatism at such high positive powers and a ring scotoma due to a gap in the field of view between light entering just inside and just outside the edge of the spectacle lens.<sup>9,10</sup> The advent of intraocular lenses (IOLs) in 1949 served to “complete” cataract surgery.<sup>11,12</sup> IOLs are artificial lenses that are implanted into the eye following the removal of a cataractous lens and are designed to provide the needed optical power to allow normal image formation to occur within the eye. Placement of the IOL within the eye allows aphakic spectacles and their problems can be avoided.

Modern cataract treatments include extracapsular cataract extraction (ECCE) and more recently a variation of this technique known as phacoemulsification. ECCE was first performed over 250 years ago, and has been widely used to remove the opacified lens.<sup>13</sup> In this procedure, a large incision is made in the limbus, which is the boundary of the cornea. A capsulorhexis, which is the removal of the anterior portion of the capsule, is performed and then the crystalline lens is removed, leaving the remaining portion of the capsule intact within the eye. The corneal incision site is then sutured to seal the eye. Phacoemulsification was developed in 1967 by Charles Kelman.<sup>14</sup> In this technique, a small incision is made at the limbus. Tools inserted through this small opening are used to make a capsulorhexis. With the crystalline lens now exposed, a hollow ultrasonic probe is slipped through the opening. The vibrations of the probe tip cause the crystalline lens to shatter and suction is used to aspirate the lens fragments through the hollow center of the probe. The entire lens is removed in this fashion leaving an empty capsule. The incision in phacoemulsification is sufficiently small so that no sutures are needed to seal the wound. Phacoemulsification is similar to modern surgical techniques of orthoscopic and laproscopic surgery. Incisions are kept small to promote rapid healing and limit the risk of infection.

## 21.4 INTRAOCULAR LENS DESIGN

Sir Harold Ridley, a British surgeon specializing in ophthalmology, was dissatisfied with the existing form of cataract surgery since it left patients aphakic and dependent on poorly performing spectacle lenses for visual function.<sup>12</sup> Ridley gained a key insight into “fixing” cataract surgery while treating Royal Air Force pilots from World War II. The canopies of British fighter planes were constructed with a newly developed polymer, polymethyl methacrylate (PMMA), also known as Perspex, plexiglass, or acrylic. In some instances, when the gunfire shattered the canopies, shards of the PMMA would become embedded into the cornea of the pilot. Ridley, in treating these aviators, noted that the material caused little or no inflammatory response of the surrounding ocular tissue. The typical response of the body to a foreign body is to encapsulate and reject the material. However, the PMMA did not elicit this response and sparked the concept of forming a lens from a “biocompatible” material that could be implanted following extraction of the cataractous lens. In 1949, Ridley implanted the first IOL into a 45-year-old woman. Following surgery, the patient was 14 D nearsighted (as opposed to 20 D farsighted in aphakic patients). While the power of the implanted IOL was clearly wrong, Ridley successfully demonstrated that an artificial lens could be implanted into the eye that evokes a marked power change. The error in the power of the IOL stemmed from modeling the implant after the natural crystalline lens. The approximate radii of curvature of the crystalline lens were used for the implant, but the higher-refractive index of the PMMA was not taken into account. Ridley’s second surgery resulted in a similar degree of nearsightedness, but by the third surgery, the IOL power had been suitably refined, leaving the patient mildly nearsighted.

Ridley’s IOL invention is revolutionary, as it restored normal vision following cataract surgery. However, the medical establishment did not immediately accept the value of this invention. Instead, many leading figures in ophthalmology at the time thought the surgery to be reckless and unwarranted. Ridley met with disdain, criticism, and the threat of malpractice. The result of this animosity was a long delay in the evolution of the implant technology as well as a limited use of the lenses by the surgical community. The shift to acceptance of the technology did not begin to occur until the 1970s. In 1981, the Food and Drug Administration finally approved IOLs for implantation following cataract surgery in the United States. Today, phacoemulsification followed by IOL implantation is the most widely performed and successful surgery in the United States. Virtually no aphakic subjects exist anymore and the term “cataract surgery” now implicitly implies implantation of the IOL.

### Intraocular Lens Power

As Ridley’s first two surgeries illustrate, calculation of the IOL power is important to minimize the error in refractive power of the eye following surgery. Ridley ventured into unknown territory in designing the initial IOLs. Accurate determination of the IOL power remains the key component to a successful surgery. Cataract surgeons typically target slight nearsightedness following implantation of the IOL. Conventional IOLs leave the recipient presbyopic because the lens power and position are fixed within the eye. Unlike the crystalline lens, the conventional IOL does not change shape or position and consequently accommodation is lost. Extensive research into accommodating IOLs is currently being pursued and a description of emerging technologies is given further in the chapter. With the fixed-focus IOL, surgeons need to determine the proper power prior to surgery, which can be challenging for several reasons. First, the eye must be tested with the existing crystalline lens in place. The power of the crystalline lens must be deduced from the overall power of the eye and the shape of the cornea. The crystalline lens power must be subtracted from the overall ocular power since the crystalline lens is removed by the surgery. Next, the power of the IOL must be estimated based on its anticipated position within the eye following implantation. Finally, the power of the IOL must also be adjusted for any inherent refractive error in the eye. In other words, the IOL power and the cornea must work in conjunction to give a sharp image on the retina. To further complicate this calculation, the patient is receiving the implant because of cataract, so subjective measurement of refractive error may be confounded by the opacities in the lens.

The IOL power calculation formulas fall into two categories. Some formulas are based on multiple regression analysis of the variables associated with the eye and implant, while other formulas are based on a theoretical prediction of the IOL power. The SRK formula was originally the most widely used regression-type formula to predict IOL power. The SRK formula is given by<sup>15,16</sup>

$$\phi_{\text{IOL}} = A - 0.9 K - 2.5 L \quad (1)$$

where  $\phi_{\text{IOL}}$  is the power of the IOL in diopters,  $A$  is the A-constant of the lens,  $K$  is the corneal keratometry, and  $L$  is the axial length of the eye. The A-constant is a value provided by the IOL manufacturer that is most closely associated with the position of the IOL within the eye. Modern IOLs are about 1 mm thick and about one-fourth the thickness of the natural crystalline lens. Consequently, there is a variation in the positioning of the IOL behind the iris that is dependent on the shape of the implant and surgical technique. The A-constant is empirically derived to account for these factors. Surgeons often customize the A-constants of lenses routinely implanted based on their specific outcomes.

Keratometry is a measure of corneal power. Keratometry is calculated measuring the radius of curvature  $R_a$  in mm of the anterior corneal surface. A power is then calculated as

$$K = 1000 \frac{(n_k - 1)}{R_a} \quad (2)$$

where  $n_k$  is the keratometric index of refraction and  $K$  is in units of diopters. The keratometric index of refraction is an effective corneal index that attempts to incorporate the power associated with the posterior surface of cornea into the keratometry measurement. Values of 1.3315 to 1.3375 are routinely used for  $n_k$  in clinical devices used for measuring keratometry. Measurement of the posterior corneal radius of curvature has been difficult until recently, so historically keratometry measurements have been made based on an estimate of the contribution of this surface to the overall power of the cornea. Note that  $n_k$  is lower than the true refractive index of the cornea, 1.376, to account for the negative power induced by the posterior cornea. As an example, the radii of curvature of the anterior and posterior surfaces of the cornea are approximately 7.8 and 6.5 mm, respectively.<sup>2</sup> The refractive index of the cornea is 1.376 and of the aqueous humor on the back side of the cornea is 1.337. A typical thickness of the cornea is 0.55 mm. With these assumptions, the keratometry of the cornea, as given by Eq. (2), is

$$K = 1000 \frac{(1.3315 - 1)}{7.8} = 42.50 \text{ D} \quad (3)$$

The true power of the cornea is given by

$$\phi_{\text{cornea}} = 1000 \left[ \frac{(1.376 - 1)}{7.8} + \frac{(1.337 - 1.376)}{6.5} - \left( \frac{0.55}{1.376} \right) \left( \frac{(1.376 - 1)}{7.8} \right) \left( \frac{(1.337 - 1.376)}{6.5} \right) \right] = 42.32 \text{ D} \quad (4)$$

Clearly, there are some small differences in the predicted keratometric and the corneal power. This discrepancy is one potential source for error in the calculation of the implant power. As can be seen from Eq. (1), a 0.9 D error in keratometry leads to a 1 D error in the calculated IOL power.

Traditionally, the axial length,  $L$ , of the eye is measured with A-scan ultrasonography. An ultrasonic probe, typically with a frequency of 10 MHz, is placed lightly in contact with the cornea to avoid deforming the surface. Echoes from the crystalline lens surfaces and the retina are then measured. The time of flight of these echoes is directly related to the speed of sound in the various materials of the eye. Table 2 summarizes typical values of these speeds used by commercial A-scan units. The longitudinal resolution of a 10-MHz ultrasound is 200  $\mu\text{m}$ .<sup>17,18</sup> Consequently, Eq. (1) suggests this level of accuracy would lead to a limit of the accuracy of the IOL power of  $\pm 0.25$  D.

**TABLE 2** Speed of Sound in Various Ocular and IOL Materials

Material	Speed of Sound
Cornea	1,640 m/s
Aqueous	1,532 m/s
Lens—Normal	1,640 m/s
Lens—Cataractous	1,629 m/s
Vitreous	1,532 m/s
PMMA IOL	2,760 m/s
Silicone IOL	1,000 m/s

Partial coherence interferometry has recently become a popular technique for measuring axial length.<sup>19</sup> In this technique, a Michelson interferometer is used to measure distances within the eye. A short-coherence-length narrow-band infrared source is shone into the eye. The return beam is interfered with light from the test arm. Since the beam has low-coherence, fringes only appear when the path lengths in the eye and the test arm are nearly identical. By changing the length of the reference arm and observing when fringes appear, the length of the eye can be accurately measured. The axial resolution of this technique is on the order of 12  $\mu\text{m}$ , leading to a resolution of the IOL power of 0.015 D.

The first regression-type formulas are accurate for eyes that fall within “normal” parameters for axial length and corneal power. However, as the more extreme sizes and shapes are approached, the accuracy of formulas degrades markedly. The regression-type formulas were further refined to better handle these extreme cases.<sup>20</sup> While regression analysis is still used for these formulas, a piecewise linear approximation is used to better approximate the full range of variation of human eyes. The SRK II is an example of an evolved formula. The IOL power is defined as

$$\phi_{\text{IOL}} = A_1 - 0.9 K - 2.5 L \quad (5)$$

where

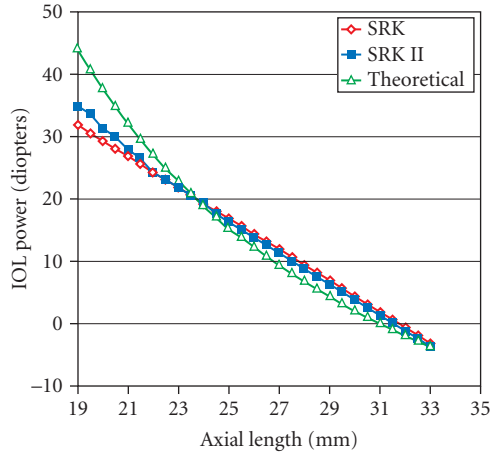
$$\begin{aligned} A_1 &= A + 3 & \text{for} & \quad L < 20.0 \text{ mm} \\ A_1 &= A + 2 & \text{for} & \quad 20.0 \leq L < 21.0 \\ A_1 &= A + 1 & \text{for} & \quad 21.0 \leq L < 22.0 \\ A_1 &= A & \text{for} & \quad 22.0 \leq L < 24.5 \\ A_1 &= A - 0.5 & \text{for} & \quad L \geq 24.5 \end{aligned}$$

Figure 2 compares the SRK and SRK II formulas.

An alternative to regression-based formulas for IOL power prediction are theoretical formulas.<sup>21</sup> Theoretical formulas use geometrical optics to make predictions about the implant power, given knowledge of the axial length, corneal power, and position of the IOL following implantation. Figure 3 shows a simplified model of the eye. The cornea is modeled as a single surface of power  $K$ . In the aphakic eye shown in Fig. 3*a*, the cornea forms an image of a distant object at a distance  $1000 n_{\text{aq}}/K$  behind the corneal vertex for  $K$  in diopters. When the IOL is implanted into the eye, it sits at a location called the effective lens position (ELP). The ELP is deeper into the eye than the iris plane for posterior chamber IOLs. Figure 3*b* shows the IOL implanted into the aphakic of Fig. 3*a*. The image formed by the cornea becomes a virtual object for the IOL located at a distance

$$S = 1000 n_{\text{aq}} / K - \text{ELP} \quad (6)$$





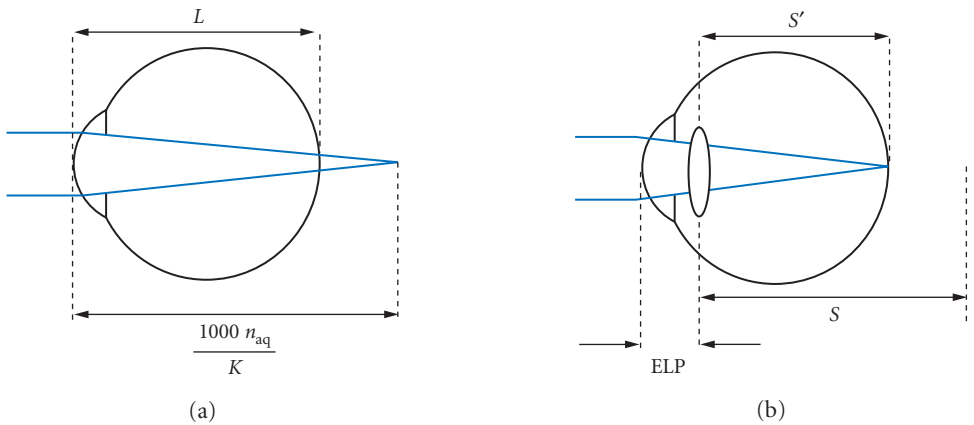
**FIGURE 2** A comparison of the predicted IOL power for the SRK and SRK II formulas assuming  $A = 118$  and  $K = 43$  D. Also shown is a theoretical prediction of the IOL power assuming  $n_{aq} = 1.337$ ,  $K = 43$  D, and  $ELP = 5$  mm.

from the plane of the IOL. The IOL, in turn, needs to image this virtual object onto the retina. If the distance from the IOL to the retina is given by  $S'$ , then

$$\frac{n_{aq}}{S'} - \frac{n_{aq}}{S} = \phi_{IOL} = n_{aq} \left[ \frac{1}{L - ELP} - \frac{1}{1000n_{aq}/K - ELP} \right] \quad (7)$$

which can be rewritten as

$$\phi_{IOL} = \frac{n_{aq}}{K} \left[ \frac{1000n_{aq} - KL}{(L - ELP)(1000n_{aq}/K - ELP)} \right] \quad (8)$$



**FIGURE 3** (a) In the aphakic eye, the image is formed at a distance  $1000n_{aq}/K$ , where  $n_{aq}$  is the refractive index of the aqueous and  $K$  is the power of the cornea. (b) When the IOL is inserted at the effective lens position (ELP), the aphakic image is reformed onto the retina.

Equation (8) is the theoretical IOL power based on thin lens analysis of the pseudophakic eye. Figure 2 also shows the predicted IOL power based on Eq. (8).

A variety of theoretical models exist.<sup>22–25</sup> They are all based on the preceding Eq. (8). Subtle differences exist, however, in the manner each of the variables is calculated. For example, the axial length provided by ultrasonography is slightly short because the echo occurs at the inner limiting membrane of the retina. Light absorption occurs at the photoreceptor layer that is approximately 20  $\mu\text{m}$  deeper into the eye. The ELP is also affected by the anterior chamber depth, which is the distance from the cornea vertex to the iris plane. Eyes with flatter corneas or smaller diameter corneas can have shallower anterior chamber depths, causing a reduction in the ELP. Holladay has worked extensively to determine the location within the eye of available IOLs and to relate the A-constant to the ELP as well as what Holladay calls the surgeon factor (SF).<sup>26–31</sup> The SF is the distance from iris plane to the front principal plane of the IOL. The theoretical models in most cases provide excellent prediction of the IOL power. The typical error in power is less than 1 D in over 80 percent of patients.<sup>20</sup> Surgeons tend to err on the side of nearsightedness to allow the patient to have some plane in their field of vision conjugate to the retina. A farsighted patient would only have a virtual object in focus on their retina and consequently they tend to be less happy since additional refractive correction is required for all object distances. The current generation of IOL power calculations, however, has difficulty in predicting accurate IOL powers in patients who have undergone corneal refractive surgeries such as radial keratotomy (RK) and laser in situ keratomileusis (LASIK). These surgical procedures modify the shape of the cornea to correct for refractive error. However, this shape change can have marked effect on the measurement of the keratometry. The relationship between the shape of the front and back surfaces of the cornea and in the case of LASIK, the corneal thickness has been artificially modified by the refractive procedure. As a result, the keratometric index of refraction is no longer valid and gross errors in IOL prediction can occur as a result. A variety of techniques have been suggested for handling this subset of cataract patients.<sup>32–42</sup> Only a limited number of patients currently fall into this category. However, as the refractive surgery generation ages, an increasing percentage of cataract patients will require specialized techniques for predicting IOL power and much research is still needed in this area to achieve satisfactory results.

## Intraocular Lens Aberrations

Since IOLs are singlets, only limited degrees of freedom exist to perform aberration correction. The human cornea typically has positive spherical aberration.<sup>43</sup> The natural crystalline lenses typically compensates for the cornea with inherent negative spherical aberration leaving the whole eye with low levels of positive spherical aberration.<sup>44</sup> The crystalline lens creates negative spherical aberration through its gradient index structure and aspheric anterior and posterior surfaces. Traditionally, IOLs have been made with spherical surfaces due to the ease of manufacturability. Smith and Lu analyzed the aberrations induced by IOLs using the thin lens Seidel aberration formulas described by Welford.<sup>45,46</sup> Spherical aberration is given by

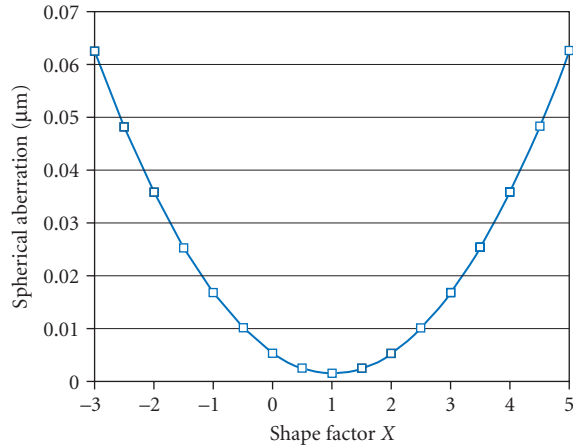
$$\frac{y^4 \phi_{\text{IOL}}^3}{4n_{\text{aq}}^2} \left[ \frac{n_{\text{aq}}^2 (2n_{\text{aq}} + n_{\text{IOL}})}{n_{\text{IOL}} (n_{\text{aq}} - n_{\text{IOL}})^2} X^2 + \frac{4n_{\text{aq}} (n_{\text{aq}} + n_{\text{IOL}})}{n_{\text{IOL}} (n_{\text{aq}} - n_{\text{IOL}})} XY + \frac{2n_{\text{aq}} + 3n_{\text{IOL}}}{n_{\text{IOL}}} Y^2 + \frac{n_{\text{IOL}}^2}{(n_{\text{aq}} - n_{\text{IOL}})^2} \right] \quad (9)$$

where  $y$  is the height of the marginal ray at the IOL,  $\phi_{\text{IOL}}$  is the power of the implant,  $n_{\text{aq}} = 1.337$  is the refractive index of the aqueous, and  $n_{\text{IOL}}$  is the refractive index of the IOL. Typical values for  $n_{\text{IOL}}$  range from 1.41 to 1.55 depending on the material. The shape factor  $X$  is defined as

$$X = \frac{C + C'}{C - C'} \quad (10)$$

where  $C$  is the curvature of the anterior surface of the IOL and  $C'$  is the curvature of the posterior surface. The conjugate factor  $Y$  is defined as

$$Y = \frac{S + S'}{S - S'} \quad (11)$$



**FIGURE 4** The spherical aberration as a function of shape factor for an IOL.

where  $S$  is the distance from the IOL to the virtual object formed by the cornea and  $S'$  is the distance from the IOL to the retina. In the eye, a 43-diopter cornea forms an image of a distant object about 31 mm behind the corneal vertex. The IOL typically resides in a plane 5 mm into the eye. Under these conditions,  $S = 31 - 5 = 26$  mm. If the eye has an axial length of  $L = 24$  mm, then  $S' = 19$  mm. In this example, the conjugate factor  $Y = 6.43$ . Figure 4 shows the spherical aberration as a function of lens shape factor for a 20 D-PMMA lens with the preceding eye parameters and  $y = 2$  mm. Two key features of this plot are apparent. First, the values of spherical aberration are strictly positive. This is true for any positive power lens with spherical surfaces. Since the corneal spherical aberration is typically positive as well, the spherical-surfaced IOL cannot compensate for this aberration, as the natural crystalline lens tends to do. Instead, the total ocular spherical aberration of the pseudophakic eye can only be minimized through the appropriate choice of IOL shape factor  $X$ . In the example in Fig. 4, a shape factor  $X = +1.0$  minimizes the spherical aberration. This shape corresponds to a plano-convex lens with the flat side toward the retina. The shape factor  $X = +1$  is ideal for a perfectly aligned cornea and IOL. However, natural variations in the tilt and decentration of the IOL and the orientation of the visual axis within the eye make this perfect alignment difficult. Typical values for the tilt and decentration of the IOL are less than  $2.6^\circ$  and 0.4 mm, respectively.<sup>47</sup> Atchison has shown that a shape factor of  $X = +0.5$ , double-convex lens with a less curved radius toward the retina, is less sensitive to these errors and modern spherical IOLs tend to target this shape factor.<sup>48,49</sup>

## Aspheric Lenses

Recently, IOLs have shifted toward incorporating an aspheric surface for aberration control.<sup>50</sup> Whereas, conventional spherical-surfaced IOLs seek to minimize the aberrations induced by the implant, aspheric IOLs can be used to compensate spherical aberration induced by the cornea. The corneal spherical aberration stems mainly from its aspheric shape. The cornea over the entrance pupil of the eye is well represented by a conic section of radius  $R$  and conic constant  $Q$ . The sag of the cornea  $z$  is then described as

$$z = \frac{1}{Q+1} \left[ R - \sqrt{R^2 - (Q+1)r^2} \right] \quad (12)$$

**TABLE 3** Radius  $R$  and Conic Constant  $Q$  of the Human Cornea

References	$R$ (mm)	$Q$	Technique
Keily <sup>51</sup>	7.72	-0.26	Photokeratoscopy
Guillon <sup>52</sup>	7.85	-0.15	Photokeratoscopy
Dubbelman <sup>53</sup>	7.87	-0.18	Scheimpflug
Dubbelman <sup>54</sup>	7.87 (male) 7.72 (female)	-0.13	Scheimpflug

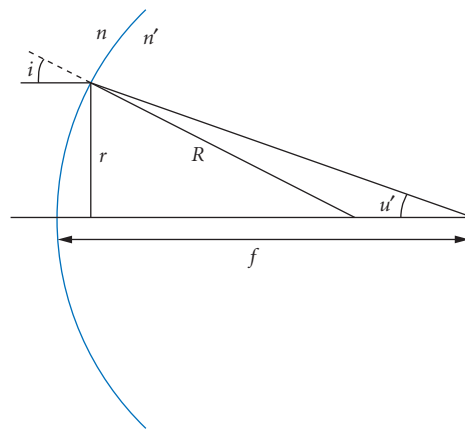
where  $r$  is the radial distance from the optical axis. Table 3 summarizes the values of  $R$  and  $Q$  of the anterior cornea found in several studies of corneal topography.<sup>51-54</sup> There is also a tendency for the cornea to become more spherical with age<sup>54,55</sup> leading to an increase in corneal spherical aberration.<sup>56</sup>

The spherical aberration of a surface can be found from paraxial raytracing values. The spherical aberration wavefront coefficient  $W_{040}$  of a spherical surface is given by<sup>46</sup>

$$W_{040} = -\frac{1}{8} r (ni)^2 \left( \frac{u'}{n'} - \frac{u}{n} \right) \quad (13)$$

where  $n$  is the refractive index,  $i$  is the paraxial angle of incidence of the marginal ray on the surface, and  $u$  the angle the marginal ray makes with respect to the optical axis. The primed terms denote variables following the surface, whereas the unprimed terms denote variable prior to the surface. Figure 5 shows the layout for calculating the spherical aberration of the anterior cornea for a distant object. In this case,  $u = 0$ ,  $n = 1.0$ ,  $u'$  is the marginal ray angle following the surface,  $n' = 1.376$  is the refractive index of the cornea, and  $i$  is the paraxial angle of incidence of the marginal ray. From the figure it is evident that

$$i = \frac{r}{R} \quad (14)$$


**FIGURE 5** Variables in the calculation of the spherical aberration of the anterior cornea.

within the paraxial small angle assumption. Furthermore, the focal length  $f$  of the surface is given by

$$f = \frac{n'}{n'-1}R \quad (15)$$

which leads to a marginal ray angle of

$$u' = -\frac{(n'-1)r}{n'R} \quad (16)$$

Hopkins<sup>57</sup> showed that a correction factor can be added to the spherical surface aberration term to account for the asphericity of a conic surface. This correction factor is given by

$$\Delta W_{040} = \frac{Q(n'-1)r^4}{8R^3} \quad (17)$$

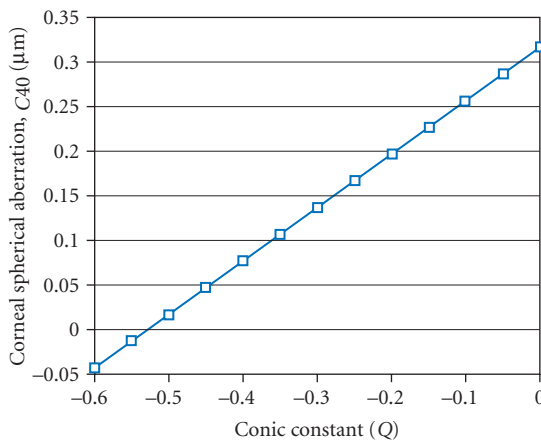
Combining Eq. (13) to Eq. (17), the total spherical aberration of the anterior cornea is given by

$$W_{040} + \Delta W_{040} = \frac{(n'-1)(1+n'^2Q)r^4}{8n'^2R^3} \quad (18)$$

If the preceding technique is repeated for the posterior corneal surface, then its spherical aberration typically is 2 orders of magnitude smaller than the spherical aberration of the anterior surface. Consequently, the contribution of the posterior cornea to the spherical aberration of the aphakic eye can be ignored. While Eq. (18) represents the wavefront coefficient of the corneal spherical aberration, the literature has somewhat adopted a different format. Typically, corneal spherical aberration has been given in terms of the Zernike polynomial expansion coefficient  $c_{40}$  for a normalization radius  $r = 3$  mm.<sup>58</sup> In this case, the corneal spherical aberration is given by

$$c_{40} = W_{040}(r=3) + \Delta W_{040}(r=3) = \frac{1}{6\sqrt{5}} \frac{81(n'-1)(1+n'^2Q)}{8n'^2R^3} = \frac{0.15}{R^3} + \frac{0.284Q}{R^3} \quad (19)$$

The value of  $c_{40}$  is linear with respect to the corneal conic constant, as shown in Fig. 6. Table 4 summarizes the values of  $c_{40}$  from various studies in the literature.<sup>59–62</sup> Since the aberrations of the various elements of the eye add, the implanted aspheric IOL should compensate for the corneal spherical



**FIGURE 6** Corneal spherical aberration as a function of conic constant  $Q$  for a corneal radius  $R = 7.8$  mm.

**TABLE 4** Corneal Spherical Aberration  $c_{40}$ 

References	$c_{40}$
Holladay <sup>50</sup>	$0.270 \pm 0.200 \mu\text{m}$
Wang <sup>59</sup>	$0.280 \pm 0.086 \mu\text{m}$
Belluci <sup>60</sup>	$0.276 \pm 0.036 \mu\text{m}$
Guirao <sup>61</sup>	$0.320 \pm 0.120 \mu\text{m}$

aberration given by Eq. (19). Examples of aspheric IOLs include the Tecnis Z9000 (Advanced Medical Optics, Irvine, CA) designed to correct  $0.27 \mu\text{m}$  of corneal spherical aberration, the AcrySof SN60WF (Alcon Laboratories, Fort Worth, TX) designed to correct for  $0.20 \mu\text{m}$  of corneal spherical aberration, and the SofPort Advanced Optics IOL (Bausch & Lomb, Rochester, NY) designed such that the implant has zero spherical aberration.

## Toric Lenses

Astigmatism is another important consideration following cataract surgery. IOLs have been traditionally rotationally symmetric, again due to manufacturing ease. However, astigmatism may be present in the aphakic eye. This astigmatism is due to a toric shape in the cornea and, since the optical system is not rotationally symmetric, the astigmatism appears along the visual axis. Recently, toric IOLs have become available, but their use has been limited.<sup>63</sup> It is more common to reduce corneal astigmatism at the time of surgery with procedures known as corneal relaxing incisions or limbal relaxing incisions. In these procedures, arcuate incisions are made in the peripheral cornea or limbus, respectively. These incisions have the effect of making the cornea more rotationally symmetric after the incisions heal. Consequently, corneal astigmatism is reduced and a conventional rotationally symmetric IOL can be implanted. Corneal or limbal relaxing incisions are typically performed at the time of cataract surgery. An alternative to relaxing incisions is to perform a refractive surgery procedure such as LASIK to reduce corneal astigmatism. This procedure uses an excimer laser to ablate corneal tissue, leaving the postoperative cornea more rotationally symmetric. Toric IOLs are available and similar power calculations must be taken into account.<sup>64–67</sup> FDA-approved toric IOLs are available in the United States from Staar Surgical (Monrovia, CA) and Alcon Laboratories (Fort Worth, TX).

## Testing IOLs

Clinical testing of implanted IOLs provides insight into the function of the lenses, but can also be difficult due to confounding factors such as cornea clarity, retina function, variations in surgical positioning of the implant, and neural processing. Furthermore, large numbers of subjects and long follow-up times make these types of studies expensive and slow. Several methods, including the Ronchi test and speckle interferometry, have been used to test isolated IOLs.<sup>68,69</sup> Alternatively, model eyes can be used to test the optical performance of IOLs in a configuration similar to their implanted state.<sup>62,70–78</sup> Model eyes allow for rapid evaluation of the lenses that avoids the confounding factors described above. In testing the optical performance of IOLs with a model eye, the geometry of how the lens is used within the eye is taken into account. Aberrations, which affect the IOL optical performance, are dependent on the vergence of light entering the implant. In general, the model eye consists of a lens that represents an artificial cornea and a wet cell, which is a planar-faced vial containing saline into which the IOL is mounted. The wet cell is placed behind the artificial cornea such that the corneal lens modifies incident plane waves and creates a converging beam onto the IOL. The vergence striking the IOL is meant to be representative of the vergence seen by an implanted IOL. An artificial pupil can also be introduced to simulate performance for different pupil sizes. The performance of the lens can then be evaluated at the image plane of the eye model. Several different types of artificial corneas have been used in eye models. The international standard ISO11979-2 recommends using a model cornea that is virtually free of aberrations in conjunction with the light

source used.<sup>79</sup> The standard provides an example model cornea made from a Melles-Griot LAO 034 lens. This lens is a commercially available cemented achromat and consequently is well corrected for both chromatic and spherical aberration. With this model cornea, the spherical aberration of the whole model eye is created solely from the IOL. This model eye specification predated the advent of aspheric IOLs. This type of model eye is therefore only suitable for earlier spherical surface IOLs to evaluate how well these implants minimize their inherent spherical aberration. However, this type of model cornea is not ideal for modern aspheric IOL designs. These newer designs are made to work in conjunction with the cornea and compensate for the corneal spherical aberration. Model corneas with representative levels of corneal spherical and chromatic dispersion have been proposed and used to test these advanced IOL designs.<sup>62</sup>

Measurement of the optical transfer function (OTF) or its modulus the modulation transfer function (MTF) are routine tests for measuring the optical quality of IOLs.<sup>70–78</sup> The MTF of an optical system describes the amount of contrast that is passed through the system. If a high contrast sinusoidal target is imaged by an optical system, then the contrast of the resultant image is reduced. In general, the contrast tends to decrease more severely with higher spatial frequency (i.e., finer spacing between the bars of the sinusoidal target). The MTF of a model eye is conveniently calculated by measuring the line spread function (LSF) of the model eye. The LSF, as its name implies, is simply the image of a narrow slit formed by the model eye. The image  $i(x, y)$  of the vertically oriented slit on the retina is given by

$$i(x, y) = \text{rect}\left(\frac{x}{d}\right) * \text{PSF}(x, y) \quad (20)$$

where  $d$  is the width of the unaberrated slit on the retina and  $\text{PSF}(x, y)$  is the point spread function of the eye. The Fourier transform of Eq. (12) gives

$$\mathfrak{F}\{i(x, y)\} = d \text{sinc}(d\xi)\delta(\eta)\text{OTF}(\xi, \eta) \Rightarrow \text{OTF}(\xi, 0) = \frac{\mathfrak{F}\{i(x, y)\}}{d \text{sinc}(d\xi)} \quad (21)$$

where  $\mathfrak{F}\{\}$  denotes the Fourier transform and  $\text{sinc}(\xi) = \sin(\pi\xi)/(\pi\xi)$ . Equation (21) says that the optical transfer function (OTF) is the ratio of the image spectrum and a sinc function. Note that the denominator of Eq. (13) goes to 0 for  $\xi = 1/d$ . Under this condition, the OTF approaches infinity. Consequently, the size of  $d$  must be made sufficiently small to ensure that the spatial frequencies  $\xi$  fall within a desirable range.

## Multifocal Lenses

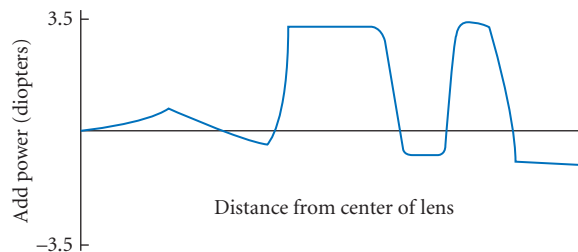
A variety of strategies have been employed to help alleviate presbyopia resulting from cataract surgery. The ideal situation would be to implant an IOL that had the capability to either change its power as the young crystalline does or change its position in response to contraction of the ciliary muscle. The latter case provides accommodation by changing the overall power of the eye through a shift in the separation between the cornea and IOL. These strategies and similar variations are being aggressively pursued as “accommodating” IOLs discussed in more detail in the following section. Clinical results demonstrating the ability of these IOLs to restore some degree of accommodation have been somewhat mixed, but improvement is likely as the accommodative mechanism becomes better understood and the lens designs incorporate these actions.

Multifocal IOLs represent a successful bridging technology between conventional IOLs and a true accommodating IOL. Multifocal optics have two or more distinct powers within their aperture. These lenses take advantage of simultaneous vision. In other words, both in-focus and out-of-focus images are simultaneously presented to the retina. The role of the brain then is to filter out the blurred component and interpret the sharp component providing suitable vision for two distinct distances. For example, suppose the required monofocal (single power) IOL power to correct distance vision in a patient undergoing refractive surgery is 20 D. If a multifocal lens containing dual powers of 20 and 24 D is implanted instead, the patient will have simultaneous vision following surgery. (Note that a 4 D add power in the

IOL plane is approximately equivalent to a 3 D add in the spectacle plane.) If the patient views a distant object, the distance portion of the IOL will form a sharp image on the retina, while the near portion of the lens will create a blurred image of the distant scene on the retina. Similarly, if the patient now views the text of a book, the near portion of the IOL will form a sharp image of the text on the retina, while the distance portion of the IOL will now create a blurred image of the page. In both cases, if the deleterious effects of the blurred portion of the retinal image are sufficiently low, then successful interpretation of the in-focus portion can be made. As a result of simultaneous vision, some contrast is lost in the in-focus image. Multifocal optics therefore represent a trade-off in visual performance. Monofocal lenses provide sharp, high contrast images for distance and horribly blurred low-contrast images for near objects. Multifocal lenses provide reasonably high contrast images for both distance and near vision, but there is some loss in contrast compared to monofocal distance vision. When designing multifocal lenses, control of the out-of-focus portion of the retinal image and understanding the conditions under which the lenses are used are important to optimizing the design. Proper understanding of these issues can provide high-quality multifocal images and lead to happy, spectacle-free patients.

Multifocal IOLs require two or more powers to be simultaneously present in the lens.<sup>80</sup> Zonal refractive lenses introduce multiple powers into the lens by having distinct regions or zones that refract light differently. The width or area of these zones is large compared to the wavelength of light. As a result of these large zones, the bending of the light rays is determined strictly by refraction. The local surface curvature of the lens determines the local lens power, and different powers can be incorporated simply by having regions with different curvatures. A simple example of a zonal refractive lens would be to bore a 2 mm hole in the center of a 20 D lens and then fill the region with a 2 mm region cookie cut from a 24 D lens. These resulting lens would act as a 20 D lens with a 4 D add in the center, providing multifocal optics. One clear problem arises in this design. The junction between the two lens regions will be abrupt and discontinuous leading the stray light effects. Zonal refractive lenses blend the transition between regions. This blend can be rapid, or the transition can be made slowly to introduce regions of intermediate power. Clearly, this concept can be extended to more than two discrete regions, so that multiple annular regions of alternating power can be created. Furthermore, the regions do not even need to be concentric circles. Instead, tear shaped, wedged, and even swirled regions have been demonstrated. The ReZoom IOL and its predecessor the Array IOL (Advanced Medical Optics, Irvine, CA) are examples of zonal refractive lenses. Figure 7 illustrates the power profile of such a lens. As can be expected with simultaneous vision, there exists a degradation in the optical quality of the retinal image with zonal refractive lenses. For example, starburst and halo patterns can result from these types of lenses.<sup>81,82</sup> However, these downsides provide roughly two-thirds of patients with near visual acuity of 20/40 or better, and the differences between zonal refractive and conventional monofocal lenses on complex tasks such as driving is small.<sup>82,83</sup> These lenses provide improved near vision at the expense of the quality of overall vision.

The second optical phenomenon that is exploited to create multifocal optics is diffraction.<sup>84–86</sup> Multifocal optics with diffractive structures are often misunderstood because they move away from the geometrical picture of light rays bending at the surface of the lens. Instead, these lenses take



**FIGURE 7** Power profile of the Array lens. The lens is formed from multiple concentric regions that oscillate between the base lens power and the higher add power.



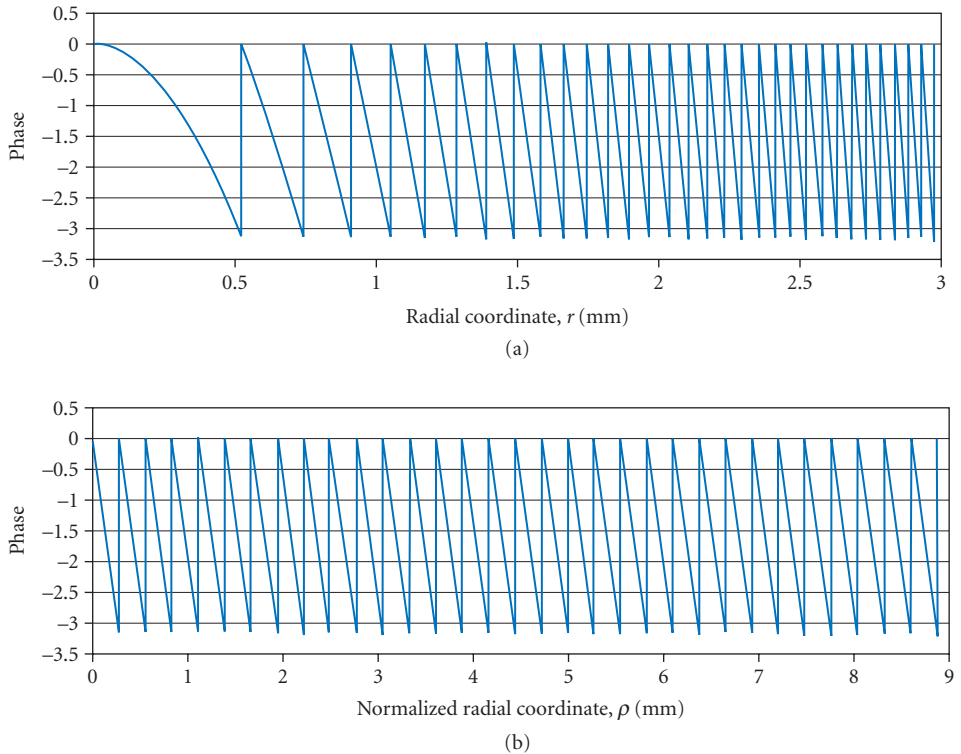
advantage of the wave nature of light. In diffractive IOLs, concentric annular zones are created on the face of the lens. The  $j$ th zone occurs at a radius

$$r_j = \sqrt{2j\lambda_o F} \quad (22)$$

where  $\lambda_o$  is the design wavelength and  $F$  is the focal length of the add power. At the junction of each zone, an abrupt step appears. Both the height of the step and the dimensions of the zones control the degree of multifocality of the lens. The dimensions of the zones are related to the desired add power and in general, the spacing between zones gets progressively smaller from the lens center to its edge. The height of the step at the boundary of each zone determines how much light is put into the add portion. In typical multifocal diffractive lenses, this step height is chosen so that the peaks of one diffractive zone line up with the troughs of the next larger diffractive zone immediately following the lens. As these waves propagate to the retina, the waves from the various diffractive zones mix and there are two distinct regions of constructive interference that correspond to the two main foci of the multifocal lens. The optical phase profile  $\phi(r)$  of a diffractive lens is given by<sup>87</sup>

$$\phi(r) = 2\pi\alpha \left( j - \frac{r^2}{2\lambda_o F} \right) \quad r_j \leq r < r_{j+1} \quad (23)$$

where  $\alpha$  represents a fraction of the  $2\pi$  phase delay. This phase pattern is superimposed onto one of the curved surfaces of the IOL. Note that if the change of variable  $\rho = r^2$  is used, then the phase profile becomes periodic in  $\rho$ . Figure 8 shows the phase profile in both  $r$ - and  $\rho$ -space for a design wavelength of 555 nm,  $\alpha = 0.5$ , and  $F = 250$  mm.



**FIGURE 8** The phase profile of a diffractive lens as a function of  $r$  (a) and  $\rho$  (b). Note that (b) is periodic.

The periodic phase profile can be represented as a Fourier series such that

$$\exp[i\phi(\rho)] = \sum_m c_m \exp\left[-im2\pi\left(\frac{\rho}{2\lambda_o F}\right)\right] \quad (24)$$

where the coefficients  $c_m$  are given by

$$c_m = \frac{1}{2\lambda_o F} \int_0^{2\lambda_o F} \exp\left[-\frac{i\pi\alpha\rho}{\lambda_o F}\right] \exp\left[\frac{im2\pi\rho}{2\lambda_o F}\right] d\rho \quad (25)$$

Carrying out the integration in Eq. (25) gives

$$\exp[i\phi(r)] = \sum_m \exp[i\pi(m-\alpha)] \exp\left[-\frac{i\pi r^2}{\lambda_o(F/m)}\right] \text{sinc}[m-\alpha] \quad (26)$$

Note that each term in the series in Eq. (26) acts like a lens of focal length  $F/m$ . The sinc function dictates how much energy is distributed into each foci. The diffraction efficiency  $\eta_m$  is defined as

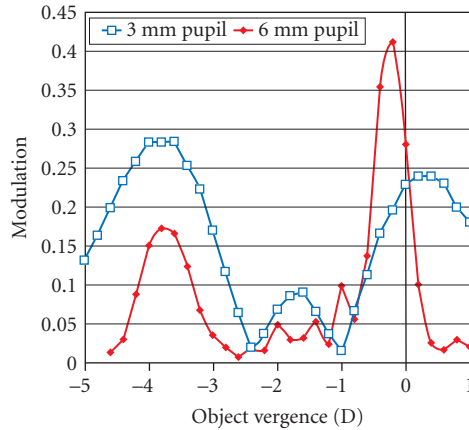
$$\eta_m = \text{sinc}^2[m-\alpha] \quad (27)$$

The diffraction efficiency describes the percent of going into each diffraction order. In the case where  $m = 0$ , the focal length of this order becomes infinite (i.e., zero power). The underlying refractive carrier of the IOL provides all of the power of the implant. Consequently, the refractive power of the IOL is chosen to correct the eye for distance vision. For  $m = 1$ , energy is distributed into the +1 diffraction order. The power of the IOL for this diffraction order is the underlying refractive power of the IOL plus the power  $+1/F$  provided by the diffractive structure. The amount of energy going into each of these diffractive orders is  $\eta_0$  and  $\eta_1$ , respectively. Similarly, energy is sent to higher-diffractive orders that have efficiency  $\eta_m$  and add a power  $m/F$  to the underlying refractive power of the IOL. Table 5 summarizes the add power and the diffraction efficiency for a case where  $\alpha = 0.5$ . Clearly, most of the energy goes into the 0 and +1 diffraction orders.

The Tecnis ZM900 (Advanced Medical Optics, Irvine, CA) is an example of a diffractive lens with  $\alpha = 0.5$  and  $F = 250$  mm. The Acri.LISA IOL (Carl Zeiss Meditec, Oberkochen, Germany) is another example of a diffractive IOL. This IOL is currently only available outside the United States. This lens distributes about twice the energy into the zero-order as it does into the +1 order. This effect biases distance vision and can be achieved by making  $\alpha = 0.414$ . The ReSTOR IOL (Alcon Laboratories, Fort Worth, TX) is another variation of a diffractive lens.<sup>88</sup> These lenses have a diffractive structure over the central 3 mm of the lens and are purely refractive in their periphery. The refractive portion of the lens provides distance vision, while the diffractive portion of the lens provides both near and distance vision. The step heights between annular zones of the lens gradually decrease toward the edge of the diffractive zone. This decrease is called apodization. Traditionally in optics, apodization has referred to a variable transmission of the pupil of an optical system. However, apodization in this regard describes the gradual change in diffraction efficiency that occurs with these types of lenses. The net result of these types of lenses is that the energy sent to the distance and near foci is nearly equal for

**TABLE 5** Add Power and Diffraction Efficiency for  $\alpha = 0.5$  and  $F = 250$  mm

Diffraction Order, $m$	Diffraction Efficiency, $\eta_m$	Add Power
-1	4.5%	-4 D
0	40.5%	0 D
+1	40.5%	4 D
+2	4.5%	8 D
+3	1.6%	12 D



**FIGURE 9** MTF measurements of the ReSTOR apodized diffractive IOL as a function of object vergence: 3- and 6-mm pupil.

small pupils, but is systematically shifted to a distance bias as the pupil size expands. Figure 9 shows the through-focus MTF of the ReSTOR lens for a 3- and 6-mm pupil. The horizontal axis of these plots is in units of vergence or the reciprocal of the object distance in units of inverse meters. As with zonal refractive lenses, there is some loss in visual performance with diffractive lenses.<sup>89-91</sup> This loss is offset by roughly 78 percent of patients achieving near visual acuity of 20/40 or better.<sup>91</sup> Biasing distance vision through modification of the step height or apodization tends to improve performance over conventional equal-split diffractive lenses for large pupil diameters.<sup>92,93</sup> Stray light effects, flares, and halos can be seen with diffractive lenses as well. Apodization of the IOL tends to markedly dampen these stray light effects for large pupils.<sup>93</sup> Finally, diffractive lenses tend to provide superior visual performance when compared to zonal refractive lenses.<sup>94,95</sup>

## Accommodating Lenses

Accommodating IOLs are the natural progression of implant technology. As shown in Fig. 1, accommodation is lost by about the age of 50. Implantation of IOLs may restore vision in cataract patients, but presbyopia remains. Multifocal IOLs have been developed to address the need of presbyopic pseudophakic patients, but there is always a trade-off with these types of lenses. Simultaneous vision by its nature reduces contrast and image quality, and designing multifocal IOLs is always a balance between providing functional near vision and minimizing artifacts attributed to multifocality. Accommodating lenses avoid the issues of simultaneous vision and provide high quality imaging at a variety of object distances. These types of lenses would function in a manner similar to the natural crystalline lens. Several accommodating IOLs are currently available; however, their performance has been marginal at best.<sup>96</sup> Advances in accommodating lens technology are rapidly occurring and this technology will likely replace multifocal IOLs once a suitable solution has been found. Furthermore, cataract surgery may no longer be the requirement for IOL implantation. Once solutions have been developed that offer reasonable levels of accommodation (likely allowing focusing from distance to 50 cm), a procedure known as refractive lens exchange (RLE) or clear lens extraction (CLE) is likely to rise in popularity.<sup>97</sup> In these procedures the noncataractous crystalline lens is removed and replaced with an IOL. These techniques are one option for patients with severe myopia, but surgeons are hesitant to perform these procedures on healthy, albeit presbyopic lenses with low degrees of refractive error. However, there was much resistance to implanting IOLs in the first place and once RLE or CLE demonstrates safe results that restore accommodation, the hesitancy will likely wane.

Two mechanisms are available for creating an accommodating IOL. The trigger for the lens action is the constriction of the ciliary muscle. The first mechanism is an axial shift of the position of the lens. Equation (8) describes the IOL power given the corneal curvature  $K$ , axial length  $L$ , and the effective lens position, ELP. Differentiating Eq. (8) with respect to the ELP gives

$$\frac{d\phi_{\text{IOL}}}{d(\text{ELP})} = \frac{(2K \cdot \text{ELP} - KL - 1000n_{\text{aq}})(KL - 1000n_{\text{aq}})n_{\text{aq}}}{(\text{ELP} - L)^2(K \cdot \text{ELP} - 1000n_{\text{aq}})^2} \quad (28)$$

Assuming a corneal power  $K = 43$  D, and axial length  $L = 24$  mm,  $n_{\text{aq}} = 1.337$ ,  $\text{ELP} = 5$  mm in the unaccommodated state and that the change in ELP is small compared to the total ELP, then the change in IOL power  $\Delta\phi_{\text{IOL}}$  is approximately

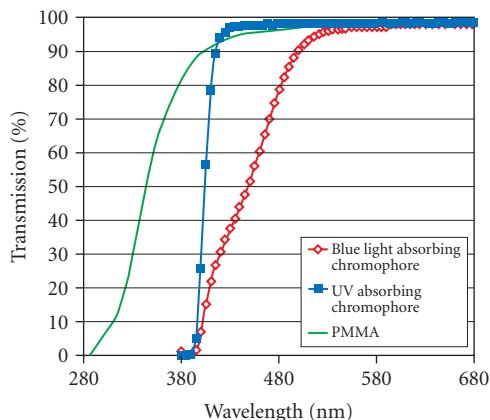
$$\Delta\phi_{\text{IOL}} \approx 1.735 \Delta\text{ELP} \quad (29)$$

Equation (29) shows that under these conditions, the 1 mm of movement of the IOL toward the cornea gives about 1.75 D of accommodation. Examples of IOLs that use axial movement are the FDA-approved Crystalens (Bausch & Lomb, Rochester, NY) and 1CU (HumanOptics AG, Erlangen, Germany) available in Europe. Both lenses have plate-type haptics or flat flanges emanating from the side of the lens. These flanges are hinged such that the lens vaults toward the cornea when the ciliary muscle constricts. Still there is much improvement needed in these technologies. The Synchrony accommodating IOL (Visiogen, Irvine, CA) is a variation on the axial movement concept that uses two lenses that move relative to one another instead of a single lens for the IOL.

The second mechanisms for an accommodating IOL would be a lens that changes its power in response to the constriction of the ciliary muscle. Examples of these types of lenses include the Fluid Vision IOL (Powervision, Belmont, CA) which pumps fluid from a peripheral reservoir into a lens with a deformable membrane surface to change its power, and the AkkoLens (AkkoLens International, Delft, The Netherlands) that uses two cubic phase plates that translate laterally to achieve a power change.<sup>98</sup> These accommodating IOLs represent emerging technologies but clinical demonstration of their capabilities is still needed.

## Phakic Lenses

Phakic IOLs, as their name implies, are lenses that are implanted while the crystalline lens remains clear and intact. These lenses are used for treating nearsightedness instead of cataracts. Since the crystalline lens remains in place, phakic IOLs need to be implanted in the remaining space. There are two suitable locations for this implantation. The first is the anterior chamber, which is the space between the posterior cornea and the iris. Two techniques have been used to support anterior chamber IOLs. The first technique is to wedge the haptics into the angle where the peripheral cornea meets the iris. The second technique is to use specially designed haptics that clip to the peripheral iris to maintain the lens position. The main disadvantage of anterior chamber phakic IOLs is the risk to the corneal endothelium. The corneal endothelium is a thin layer of cells that coats the posterior surface of the cornea. These cells regulate the nutrition and health of the cornea and cannot be replaced if damaged. Mounting an IOL in the anterior chamber risks abrasions of these endothelial cells. Phakic IOLs can also be supported by the sulcus, which means the phakic IOL is placed on the area directly behind the iris, but in front of the crystalline lens. This second type of phakic IOL positioning requires a vaulting of the implant away from the surface of the crystalline lens since contact between the artificial and natural lenses are likely to lead to cataract formation. Examples of phakic IOLs include the Verisyse (Advanced Medical Optics, Santa Ana, CA), which is an iris-supported lens, and the Visian ICL (Staar Surgical, Monrovia, CA), which is a sulcus-supported lens. Excellent results have been achieved with phakic IOLs in extreme nearsightedness.<sup>99</sup> Phakic IOLs for the treatment of farsightedness and astigmatism as well as multifocal lenses are likely to be available in the United States in the near future.



**FIGURE 10** The transmission of a PMMA, and blue-light filtering, and UV-absorbing IOLs for ultraviolet and visible wavelengths.

## Chromophores

The crystalline lens is the main absorber of wavelengths below 400 nm in the eye.<sup>2,100</sup> Removal of the crystalline lens causes a dramatic increase in the level of ultraviolet (UV) radiation that reaches the retina. These highly energetic wavelengths can cause phototoxicity, leading to damage the retinal neural structure and the underlying retinal pigment epithelium. Aphakic individuals are particularly susceptible to this type of damage since they no longer have the appropriate UV filtering. IOLs, in general, do little to mitigate the UV exposure. PMMA has significant transmission for wavelengths between 300 and 400 nm. Mainster recognized the risks to retinal health posed by transparent IOLs<sup>101,102</sup> and most IOLs incorporated a UV-absorbing chromophore by 1986.<sup>103</sup> These UV-absorbing chromophores typically have a cutoff wavelength of 400 nm, leaving high transmission of visible wavelengths, but negligible transmission of UV wavelengths. Blue-green wavelengths can also cause retinal phototoxicity, although much higher dosages are required for damage when compared to UV wavelengths. Several companies have introduced blue-light absorbing IOLs (AF-1 UY, Hoya, Tokyo, Japan and the Acrysof Natural, Alcon laboratories, Fort Worth, TX). These IOLs seek to have a transmission similar to the middle-aged human crystalline lens. Consequently, they partially absorb a portion of the spectrum between 400 and 500 nm. Figure 10 shows the transmission of a 3-mm slab of PMMA immersed in saline, along with the transmissions of IOLs containing UV-absorbing and blue-light absorbing IOLs.

Blue-light absorbing IOLs are somewhat controversial. Critics have suggested that scotopic and color vision are degraded by the lenses.<sup>104,105</sup> However, these claims have not been supported by theoretical analyses and clinical measures of in-eye lens performance.<sup>106–113</sup> Supporters of blue-light filtering lenses advocate their use to promote retinal health. Some laboratory studies have demonstrated protection to cell cultures afforded by the blue-light chromophores, but long-term clinical data remains unavailable.<sup>114</sup> Blue light has been shown to mediate the circadian rhythm<sup>115</sup> and an intermediate violet-absorbing chromophore that balances retinal protection and the circadian clock has been suggested.<sup>116,117</sup> Again, long-term validation of the benefits of such a lens are not available.

## 21.5 INTRAOCULAR LENS SIDE EFFECTS

A variety of optical deficits can arise following the implantation of intraocular lenses (IOLs). While artificial lenses replace cataractous lenses that have degraded markedly in optical quality, these IOLs can also introduce some visual side effects. These side effects include glare, halos, streaks, starbursts,

shadows, and haze. While the vast majority of pseudophakic patients are free from these effects under most circumstances and tolerate the problem in the few situations when they do arise, a small fraction of patients suffer from problems that are comparable or worse than their preoperative state. Understanding the cause of these optical phenomena allows for IOL manufacturers to improve their designs and minimize potential problems in future recipients.

## Dysphotopsia

Dysphotopsia is the introduction of unwanted patterns onto the retina. These unwanted patterns are superimposed over the true retinal image and can cause degradation in visual performance. In pseudophakic dysphotopsia, the design and material of the artificial lens is typically responsible for redirecting the unwanted light to the retina. Dysphotopsia comes in two forms: positive and negative. Positive dysphotopsia is the introduction of bright artifacts onto the retina. These artifacts include arcs, streaks, rings, and halos and may only be present under certain lighting conditions or for certain locations of glare sources in the peripheral field. Negative dysphotopsia, conversely, is the blockage of light from reaching certain portions of the retina. Shadows and dark spots are perceived usually in the temporal field and again this phenomenon is affected by lighting conditions and position. Both positive and negative dysphotopsias affect visual performance because the unwanted images obscure the relevant retinal image formed directly by the IOL. Consequently, understanding their cause and eliminating these undesirable effects will lead to improve performance and satisfaction of patients.

## Posterior Capsule Opacification

Another side effect of IOL implantation is posterior capsule opacification (PCO). Implantation of IOLs into the capsular bag provides a stable platform for the lens. However, in about 25 percent of patients, PCO can result. In PCO, the capsular bag gradually opacifies, and postcataract patients perceive many of the same symptoms seen prior to surgery, namely a gradual loss in acuity, glare, and blurred vision. YAG posterior capsulotomy is typically performed to open a hole in the capsule allowing light to pass through again. The large incidence of PCO causes dissatisfaction in patients and adds expense due to the need for additional equipment, procedures, and time to rectify the situation. Consequently, a reduction in PCO would be beneficial and strong efforts have been made to modify IOL designs and materials to drastically reduce the incidence of PCO. Acrylic materials and square edge designs have both appear to reduce the incidence of PCO.<sup>118,119</sup> However, dysphotopsia increased with introduction of these changes. To gain the benefit of reduced PCO, alterations were needed to acrylic lens edge designs.

Holladay et al.<sup>120</sup> compared the edge glare caused by sharp and rounded edge designs using a nonsequential raytracing technique. Nonsequential raytracing is routinely used for analyzing stray light and illumination effects. Typically, hundreds of thousands of rays are launched from a light source and allowed to refract, reflect, and scatter from various elements. Following the raytracing, the concentration of these rays in a given plane can be evaluated to determine the illumination pattern of the stray light. Holladay et al. found that both square and round edge IOLs produce stray light, but that only the square-edge design concentrated the light into a well-formed arc on the retina. Round-edge designs tended to disperse the stray light over a much larger portion of the retina, suggesting its visual consequences fall below a perceptible threshold. Clinical reports of dysphotopsia support that square-edge designs with a smooth finish are a likely culprit for many of the positive dysphotopsia problems.<sup>121,122</sup> Meacock et al.<sup>123</sup> performed a prospective study of 60 patients split between acrylic lenses with textured and nontextured edges. By prospectively analyzing the two groups, determination about the advantages of textured edges is assessed. One month post-operatively 67 percent of the nontextured IOL patients and 13 percent of textured-IOL patients had glare symptoms. The textured edges provided a statistically significant reduction in glare symptoms. Franchini et al.<sup>124,125</sup> used nonsequential raytracing methods to compare different types of edge design on positive dysphotopsia. They found similar result to Holladay et al.<sup>120</sup> in that the untextured square-edge design produced a

ring pattern on the retina, while a round edge distributed the light over a larger portion of the retina. This group also analyzed an “OptiEdge” design where the edge is beveled in an attempt to minimize stray light effects. In this case, they found an arc pattern is formed on the retina, but this arc in general has intensity far below the smooth edge pattern. Furthermore, Franchini et al. modeled a frosted square-edge design and found that the circular pattern retina is reduced, but the light is distributed over the retina and could possibly reduce contrast in the perceived image. The results of these efforts demonstrate that tailoring the shape and texture of the edge of the IOL can lead to an implant that reduces both PCO and dysphotopsia.

## 21.6 SUMMARY

The addition of IOL implantation following cataract removal revolutionized the procedure and enabled the restoration of high quality vision. While the technique was slow to be accepted by the medical establishment, the benefits of IOLs won over even the harshest critics to become the standard of care. IOLs are entering a postmodern era. Conventional IOLs are reliably fabricated and advances in materials and design have led to small incision surgeries that minimize risks to the patient and allowing outpatient procedures. Power calculations and the devices used to measure ocular dimensions have continued to improve, narrowing the residual error following implantation. Manufacturers are now turning to more subtle improvements to the technology. Aspheric surfaces are being incorporated to minimize ocular spherical aberration. Multifocal IOLs are providing simultaneous distance and near vision, enabling a spectacle-free lifestyle following surgery. Accommodating IOLs are emerging to fully restore accommodation. Furthermore, manufacturers are expanding the role of IOLs into the treatment of refractive error with phakic IOLs. Finally, chromophores are incorporated into IOLs to promote the health of the retina for the remainder of the patient’s life. IOLs have been extraordinarily successful and a true benefit to the dim world of cataracts.

## 21.7 REFERENCES

1. W. J. Smith, *Modern Optical Engineering*, 4th ed., McGraw-Hill, New York, 2000.
2. D. A. Atchison and G. Smith, *Optics of the Human Eye*, Butterworth-Heinemann, Oxford, 2000.
3. A. Duane, “Normal Values of the Accommodation at all Ages,” *Trans. Sec. Ophthalmol. AMA* **53**:383–391 (1912).
4. B. K. Pierscionek, “Reviewing the Optics of the Lens and Presbyopia,” *Proc. SPIE* **3579**:34–39 (1998).
5. National Eye Institute, “Cataract—What You Should Know,” *NIH Publication* 03-201 (2003).
6. J. C. Javitt, F. Wang, and S. K. West, “Blindness due to Cataract: Epidemiology and Prevention,” *Annu. Rev. Public Health* **17**:159–177 (1996).
7. B. E. K. Klein, R. Klein, and K. L. P. Linton. “Prevalence of Age-Related Lens Opacities in a Population,” *Ophthalmology* **99**:546–552 (2003).
8. K. B. Kansupada and J. W. Sassani, “Sushruta: The Father of Indian Surgery and Ophthalmology,” *Doc. Ophthalmology* **93**:159–167 (1997).
9. M. Jalie, *The Principles of Ophthalmic Lenses*, 4th ed., Association of Dispensing Opticians, London, 1984.
10. R. J. Schechter, “Optics of Intraocular Lenses,” in *Duane’s Clinical Ophthalmology*, eds., W. Tasman and E. A. Jaeger, Lippincott, Philadelphia, 2004.
11. D. J. Apple and J. Sims, “Harold Ridley and the Invention of the Intraocular Lens,” *Surv. Ophthalmol.* **40**:279–292 (1996).
12. D. J. Apple, *Sir Harold Ridley and His Fight for Sight*, Slack, New Jersey, 2006.
13. I. Obuchowska and Z. Mariak, “Jacques Daviel—The Inventor of the Extracapsular Cataract Extraction Surgery,” *Klin. Oczna.* **107**:567–571 (2005).

14. C. D. Kelman, "Phacoemulsification and Aspiration. A New Technique of Cataract Removal. A Preliminary Report," *Am. J. Ophthalmol.* **64**:23–35 (1967).
15. J. Retzlaff, "A New Intraocular Lens Calculation Formula," *Am. Intra-Ocular Implant Soc. J.* **6**:148–152 (1980).
16. D. R. Sanders and M. C. Kraff, "Improvement of Intraocular Lens Power Calculation Using Empirical Data," *Am. Intra-Ocular Implant Soc. J.* **6**:268–270 (1980).
17. R. D. Binkhorst, "The Accuracy of Ultrasonic Measurement of Axial Length of the Eye," *Ophthalmic. Surg.* **13**:363–365 (1981).
18. T. Olsen, "The Accuracy of Ultrasonic Determination of Axial Length in Pseudophakic Eyes," *Acta Ophthalmol.* **67**:141–144 (1989).
19. W. Drexler, O. Findl, R. Menapace, G. Rainer, C. Vass, C. K. Hitzenberger, and A. F. Fercher, "Partial Coherence Interferometry: A Novel Approach to Biometry in Cataract Surgery," *Am. J. Ophthalmol.* **126**:524–534 (1998).
20. D. R. Sanders, J. R. Retzlaff, and M. C. Kraff, "Comparison of the SRK II Formula and Other Second Generation Formulas," *J. Cataract Refract. Surg.* **14**:136–141 (1988).
21. T. Olson, "Calculation of Intraocular Lens Power: A Review," *Acta Ophthalmol. Scand.* **85**:472–485 (2007).
22. O. Pomerantzeff, M. M. Pankratov, and G. Wang, "Calculation of an IOL from the Wide-Angle Optical Model of the Eye," *Am. Intra-Ocular Implant Soc. J.* **11**:37–43 (1985).
23. J. T. Holladay, T. C. Prager, T. Y. Chandler, K. H. Musgrove, J. W. Lewis, and R. S. Ruiz, "A Three-Part System for Refining Intraocular Lens Power Calculations," *J. Cataract Refract. Surg.* **14**:17–23 (1988).
24. J. A. Retzlaff, D. R. Sanders, and M. C. Kraff, "Development of the SRK/T Intraocular Lens Implant Power Calculation Formula," *J. Cataract Refract. Surg.* **16**:333–340 (1990).
25. K. J. Hoffer, "The Hoffer Q Formula: A Comparison of Theoretic and Regression Formulas," *J. Cataract Refract. Surg.* **19**:700–712 (1993).
26. J. T. Holladay and K. J. Maverick, "Relationship of the Actual Thick Intraocular Lens Optic to the Thin Lens Equivalent," *Am. J. Ophthalmol.* **126**:339–347 (1998).
27. J. T. Holladay, "International Intraocular Lens and Implant Registry," *J. Cataract Refract. Surg.* **26**:118–134 (2000).
28. J. T. Holladay, "International Intraocular Lens and Implant Registry," *J. Cataract Refract. Surg.* **27**:143–164 (2001).
29. J. T. Holladay, "International Intraocular Lens and Implant Registry," *J. Cataract Refract. Surg.* **28**:152–174 (2002).
30. J. T. Holladay, "International Intraocular Lens and Implant Registry," *J. Cataract Refract. Surg.* **29**:176–197 (2003).
31. J. T. Holladay, "International Intraocular Lens and Implant Registry," *J. Cataract Refract. Surg.* **30**:207–229 (2004).
32. L. C. Celikkol, G. Pavlopoulos, B. Weinstein, G. Celikkol, and S. T. Feldman, "Calculation of Intraocular Lens Power after Radial Keratotomy with Computerized Videokeratography," *Am. J. Ophthalmol.* **120**:739–749 (1995).
33. H. V. Gimbel and R. Sun, "Accuracy and Predictability of Intraocular Lens Power Calculation after Laser in situ Keratomileusis," *J. Cataract Refract. Surg.* **27**:571–576 (2001).
34. J. H. Kim, D. H. Lee, and C. K. Joo, "Measuring Corneal Power for Intraocular Lens Power Calculation after Refractive Surgery," *J. Cataract Refract. Surg.* **28**:1932–1938 (2002).
35. N. Rosa, L. Capasso, and A. Romano, "A New Method of Calculating Intraocular Lens Power after Photorefractive Keratectomy," *J. Refract. Surg.* **18**:720–724 (2002).
36. A. A. Stakheev, "Intraocular Lens Calculation for Cataract after Previous Radial Keratotomy," *Ophthal. Physiol. Opt.* **22**:289–295 (2002).
37. C. Argento, M. J. Cosentino, and D. Badoza, "Intraocular Lens Power Calculation after Refractive Surgery," *J. Refract. Surg.* **29**:1346–1351 (2003).
38. L. Chen, M. H. Mannis, J. J. Salz, F. J. Garcia-Ferrer, and J. Ge, "Analysis of Intraocular Lens Power Calculation in Pose-Radial Keratotomy Eyes," *J. Cataract Refract. Surg.* **29**:65–70 (2003).
39. E. Jarade and K. F. Tabbara, "New Formula for Calculating Intraocular Lens Power after Laser in situ Keratomileusis," *J. Cataract Refract. Surg.* **30**:1711–1715 (2004).



40. G. Ferrara, G. Cennamo, G. Marotta, and E. Loffredo, "New Formula to Calculate Corneal Power after Refractive Surgery," *J. Refract. Surg.* **20**:465–471 (2004).
41. G. Savini, P. Barboni, and M. Zanini, "Intraocular Lens Power Calculation after Myopic Refractive Surgery," *Ophthalmology* **113**:1271–1282 (2006).
42. S. T. Awwad, S. Dwarakanathan, W. Bowman, D. Cavanagh, S. M. Verity, V. V. Mootha, and J. P. McCulley, "Intraocular Lens Power Calculation after Radial Keratotomy: Estimating the Refractive Corneal Power," *J. Refract. Surg.* **33**:1045–1050 (2007).
43. L. Wang, E. Dai, D. D. Koch, and A. Nathoo, "Optical Aberrations of the Human Anterior Cornea," *J. Cataract Refract. Surg.* **29**:1514–1521 (2003).
44. J. Porter, A. Guirao, I. G. Cox, and D. R. Williams, "Monochromatic Aberrations of the Human Eye in a Large Population," *J. Opt. Soc. Am. A* **18**:1793–1803 (2001).
45. G. Smith and C. -W. Lu, "The Spherical Aberration of Intraocular Lenses," *Ophthal. Physiol. Opt.* **8**:287–294 (1988).
46. W. T. Welford, *Aberrations of Optical Systems*, Adam Hilger, Bristol, 1986.
47. A. Castro, P. Rosales, and S. Marcos, "Tilt and Decentration of Intraocular Lenses In Vivo from Purkinje and Scheimplung Imaging," *J. Cataract Refract. Surg.* **33**:418–429 (2007).
48. D. Atchison, "Optical Design of Intraocular Lenses. III. On-Axis Performance in the Presence of Lens Displacement," *Optom. Vis. Sci.* **66**:671–681 (1989).
49. D. Atchison, "Refractive Errors Induced by Displacement of Intraocular Lenses Within the Pseudophakic Eye," *Optom. Vis. Sci.* **66**:146–152 (1989).
50. J. T. Holladay, P. A. Piers, G. Koranyi, M. van der Mooren, and S. Norrby, "A New Intraocular Lens Design to Reduce Spherical Aberration of Pseudophakic Eyes," *J. Refract. Surg.* **18**:683–692 (2002).
51. P. M. Kiely, G. Smith, and L. G. Carney, "The Mean Shape of the Human Cornea," *Optica Acta* **29**:1027–1040 (1982).
52. M. Guillon, D. P. M. Lyndon, and C. Wilson, "Corneal Topography: A Clinical Model," *Ophthal. Physiol. Opt.* **6**:47–56 (1986).
53. M. Dubbelman, H. A. Weeber, and R. G. L. van der Heijde, H. J. Völker-Dieben, "Radius and Asphericity of the Posterior Corneal Surface Determined by Corrected Scheimplung Photography," *Acta Ophthalmol. Scand.* **80**:379–383 (2002).
54. M. Dubbelman, V. A. Sicam, and R. G. L. van der Heijde, "The Shape of the Anterior and Posterior Surface of the Aging Human Cornea," *Vis. Res.* **46**:993–1001 (2006).
55. A. Guirao, M. Redondo, and P. Artal, "Optical Aberrations of the Human Cornea as a Function of Age," *J. Opt. Soc. Am. A* **17**:1697–1702 (2000).
56. T. Oshika, S. D. Klyce, R. A. Applegate, H. Howland, and M. Danasoury, "Comparison of Corneal Wavefront Aberrations after Photorefractive Keratectomy and Laser In Situ Keratomileusis," *Am. J. Ophthalmol.* **127**:1–7 (1999).
57. H. H. Hopkins, *Wave Theory of Aberrations*, Clarendon, Oxford, 1950.
58. J. Schwiegerling, J. G. Greivenkamp, and J. M. Miller, "Representation of Videokeratoscopic Height Data with Zernike Polynomials," *J. Opt. Soc. Am. A* **12**:2105–2113 (1995).
59. L. Wang, E. Dai, D. D. Koch, and A. Nathoo, "Optical Aberrations of the Human Anterior Cornea," *J. Cataract Refract. Surg.* **29**:1514–1521 (2003).
60. R. Bellucci, S. Morselli, and P. Piers, "Comparison of Wavefront Aberrations and Optical Quality of Eyes Implanted with Five Different Intraocular Lenses," *J. Refract. Surg.* **20**:297–306 (2004).
61. A. Guirao, J. Tejedor, and P. Artal, "Corneal Aberrations before and after Small-Incision Cataract Surgery," *Invest. Ophthalmol. Vis. Sci.* **45**:4312–4319 (2004).
62. S. Norrby, P. Piers, C. Campbell, and M. van der Mooren, "Model Eyes for the Evaluation of Intraocular Lenses," *Appl. Opt.* **46**:6595–6605 (2007).
63. D. F. Chang, "When Do I Use LRIs versus Toric IOLs," in *Mastering Refractive IOLs: The Art and Science*, ed., D. F. Chang, Slack, New Jersey, 2008.
64. A. Langenbucher and B. Seitz, "Computerized Calculation Scheme for Bitoric Eikonic Intraocular Lenses," *Ophthal. Physiol. Opt.* **23**:213–220 (2003).

65. A. Langenbucher, S. Reese, T. Sauer, and B. Seitz, "Matrix-Based Calculation Scheme for Toric Intraocular Lenses," *Ophthal. Physiol. Opt.* **24**:511–519 (2004).
66. A. Langenbucher and B. Seitz, "Computerized Calculation Scheme for Toric Intraocular Lenses," *Acta Ophthalmol. Scand.* **82**:270–276 (2004).
67. A. Langenbucher, N. Szentmary, and B. Seitz, "Calculating the Power of Toric Phakic Intraocular Lenses," *Ophthalmic. Physiol. Opt.* **27**:373–380 (2007).
68. L. Carretero, R. Fuentes, and A. Fimia, "Measurement of Spherical Aberration of Intraocular Lenses with the Ronchi Test," *Optom. Vis. Sci.* **69**:190–192 (1992).
69. G. Bos, J. M. Vanzo, P. Maufooy, and J. L. Gutzwiller, "A New Interferometric to Assess IOL Characteristics," *Proc. SPIE* **2127**:56–61 (1994).
70. R. E. Fischer and K. C. Liu, "Advanced Techniques for Optical Performance Characterization of Intraocular Lenses," *Proc. SPIE* **2127**:14–25 (1994).
71. R. Sun and V. Portney, "Multifocal Ophthalmic Lens Testing," *Proc. SPIE* **2127**:82–87 (1994).
72. B. G. Broome, "Basic Considerations for IOL MTF Testing," *Proc. SPIE* **2127**:2–15 (1994).
73. J. S. Chou, L. W. Blake, J. M. Fridge, and D. A. Fridge, "MTF Measurement System that Simulates IOL Performances in the Human Eye," *Proc. SPIE* **2393**:271–279 (1995).
74. E. Keren and A. L. Rotlex, "Measurement of Power, Quality, and MTF of Intraocular and Soft Contact Lenses in Wet Cells," *Proc. SPIE* **2673**:262–273 (1996).
75. D. Tognetto, G. Sanguinetti, P. Sirotti, P. Cecchini, L. Marcucci, E. Ballone, and G. Ravalico, "Analysis of the Optical Quality of Intraocular Lenses," *Invest. Ophthalmol. Vis. Sci.* **45**:2686–2690 (2004).
76. P. A. Piers, N. E. S. Norrby, and U. Mester, "Eye Models for the Prediction of Contrast Vision in Patients with New Intraocular Lens Designs," *Opt. Lett.* **29**:733–735 (2004).
77. R. Rawer, W. Stork, C. W. Spraul, and C. Lingenfelder, "Imaging Quality of Intraocular Lenses," *J. Cataract Refract. Surg.* **31**:1618–1631 (2005).
78. P. G. Gobbi, F. Fasce, S. Bozza, and R. Brancato, "Optomechanical Eye Model with Imaging Capabilities for Objective Evaluation of Intraocular Lenses," *J. Cataract Refract. Surg.* **32**:643–651 (2006).
79. International Standard 11979-2, "Ophthalmic Implants—Intraocular Lenses. Part 2: Optical Properties and Test Methods," ISO, Geneva, 1999.
80. T. Avitabile and F. Marano, "Multifocal Intraocular Lenses," *Curr. Opin. Ophthalmol.* **12**:12–16 (2001).
81. J. D. Hunkeler, T. M. Coffman, J. Paugh, A. Lang, P. Smith, and N. Tarantino, "Characterization of Visual Phenomena with the Array Multifocal Intraocular Lens," *J. Cataract Refract. Surg.* **28**:1195–1204 (2002).
82. H. N. Sen, A. U. Sarikkola, R. J. Uusitalo and L. Llaatikainen, "Quality of Vision after AMO Array Multifocal Intraocular Lens Implantation," *J. Cataract Refract. Surg.* **31**:2483–2493 (2004).
83. K. A. Featherstone, J. R. Bloomfield, A. J. Lang, M. J. Miller-Meeks, G. Woodworth, R. F. Steinert, "Driving Simulation Study: Bilateral Array Multifocal versus Bilateral AMO Monofocal Intraocular Lenses," *J. Cataract Refract. Surg.* **25**:1254–1262 (1999).
84. M. Larsson, C. Beckman, A. Nyström, S. Hård and J. Sjöstrand, "Optical Properties of Diffractive, Bifocal Intraocular lenses," *Proc. SPIE* **1529**:63–70 (1991).
85. A. Issacson, "Global Status of Diffraction Optics as the Basis for an Intraocular Lens," *Proc. SPIE* **1529**: 71–79 (1991).
86. M. J. Simpson, "Diffractive Multifocal Intraocular Lens Image Quality," *Appl. Opt.* **31**:3621–3626 (1992).
87. D. Faklis and G. M. Morris, "Spectral Properties of Multiorder Diffractive Lenses," *Appl. Opt.* **34**:2462–2468 (1995).
88. J. A. Davison and M. J. Simpson, "History and Development of the Apodized Diffractive Intraocular Lens," *J. Cataract Refract. Surg.* **32**:849–858 (2006).
89. H. V. Gimbel, D. R. Sanders, and M. G. Raanan, "Visual and Refractive Results of Multifocal Intraocular Lenses," *Ophthalmology* **98**:881–887 (1991).
90. C. T. Post, "Comparison of Depths of Focus and Low-Contrast Acuties for Monofocal versus Multifocal Intraocular Lens Patients at 1 Year," *Ophthalmology* **99**:1658–1663 (1992).
91. R. L. Lindstrom, "Food and Drug Administration Study Update. One Year Results from 671 Patients with the 3M Multifocal Intraocular Lens," *Ophthalmology* **100**:91–97 (1993).

92. G. Schmidinger, C. Simander, I. Dejaco-Ruhswurm, C. Skorpik, and S. Pieh, "Contrast Sensitivity Function in Eyes with Diffractive Bifocal Intraocular Lenses," *J. Cataract Refract. Surg.* **31**:2076–2083 (2005).
93. J. Choi and J. Schwiegerling, "Optical Performance Measurement and Night Driving Simulation of ReSTOR, ReZoom, and Tecnis Multifocal Intraocular Lenses in a Model Eye," *J. Refract. Surg.* **24**:218–222 (2008).
94. W. W. Hutz, B. Eckhardt, B. Rohrig, and R. Grolmus, "Reading Ability with 3 Multifocal Intraocular Lens Models," *J. Cataract Refract. Surg.* **32**:2015–2021 (2006).
95. U. Mester, W. Hunold, T. Wesendahl, and H. Kaymak, "Functional Outcomes after Implantation of Tecnis ZM900 and Array SA40 Multifocal Intraocular Lenses," *J. Cataract Refract. Surg.* **33**:1033–1040 (2007).
96. O. Findl and C. Leydolt, "Meta-Analysis of Accommodating Intraocular Lenses," *J. Cataract Refract. Surg.* **33**:522–527 (2007).
97. M. Packer, I. H. Fine, and R. S. Hoffman, "The Crystalline Lens as a Target for Refractive Surgery," in *Refractive Lens Surgery*, eds., I. H. Fine, M. Packer, and R. S. Hoffman, Springer-Verlag, Berlin, 2005.
98. A. N. Simonov and G. Vdovin, "Cubic Optical Elements for an Accommodative Intraocular Lens," *Opt. Exp.* **14**:7757–7775 (2006).
99. I. Brunette, J. M. Bueno, M. Harissi-Dagher, M. Parent, M. Podtetenov, and H. Hamam, "Optical Quality of the Eye with the Artisan Phakic Lens for the Correction of High Myopia," *Optom. Vis. Sci.* **80**:167–174 (2003).
100. E. A. Boettner and J. R. Wolter, "Transmission of the Ocular Media," *Invest. Ophthalmol.* **1**:776–783 (1962).
101. M. A. Mainster, "Spectral Transmittance of Intraocular Lenses and Retinal Damage from Intense Light Sources," *Am. J. Ophthalmol.* **85**:167–170 (1978).
102. M. A. Mainster, "Solar Retinitis, Photic Maculopathy and the Pseudophakic Eye," *J. Am. Intraocul. Implant Soc.* **4**:84–86 (1978).
103. M. A. Mainster, "The Spectra, Classification, and Rationale of Ultraviolet-Protective Intraocular Lenses," *Am. J. Ophthalmol.* **102**:727–732 (1986).
104. M. A. Mainster and J. R. Sparrow, "How Much Blue Light Should an IOL Transmit," *Br J Ophthalmol.* **87**:1532–1529 (2003).
105. M. A. Mainster, "Blue-Blocking Intraocular Lenses and Pseudophakic Scotopic Sensitivity," *J. Cataract Refract. Surg.* **32**:1403–1406 (2006).
106. K. Niwa, Y. Yoshino, F. Okuyama, and T. Tokoro, "Effects of Tinted Intraocular Lens on Contrast Sensitivity," *Ophthalm. Physiol. Opt.* **16**:297–302 (1996).
107. S. M. Raj, A. R. Vasavada, and M. A. Nanavaty, "AcrySof Natural SN60AT versus AcrySof SA60AT Intraocular Lens in Patients with Color Vision Defects," *J. Cataract Refract. Surg.* **31**:2324–2328 (2005).
108. A. Rodriguez-Galietero, R. Montes-Mico, G. Munoz, and C. Albarran-Diego, "Blue-Light Filtering Intraocular Lens in Patients with Diabetes: Contrast Sensitivity and Chromatic Discrimination," *J. Cataract Refract. Surg.* **31**:2088–2092 (2005).
109. A. R. Galietero, R. M. Mico, G. Munoz, and C. A. Diego, "Comparison of Contrast Sensitivity and Color Discrimination after Clear and Yellow Intraocular Lens Implantation," *J. Cataract Refract. Surg.* **31**:1736–1740 (2005).
110. J. Schwiegerling, "Blue-Light-Absorbing Lenses and Their Effect on Scotopic Vision," *J. Cataract Refract. Surg.* **32**:141–144 (2006).
111. R. J. Cionni and J. H. Tsai, "Color Perception with AcrySof Natural and AcrySof Single-Piece Intraocular Lenses under Photopic and Mesopic Conditions," *J. Cataract Refract. Surg.* **23**:236–242 (2006).
112. N. Kara-Junior, J. L. Jardim, E. O. Leme, M. Dall'Col, and R. S. Junior, "Effect of the AcrySof Natural Intraocular Lens on Blue-Yellow Perimetry," *J. Cataract Refract. Surg.* **32**:1328–1330 (2006).
113. V. C. Greenstein, P. Chiosi, P. Baker, W. Seiple, K. Holopigian, R. E. Braunstein, and J. R. Sparrow, "Scotopic Sensitivity and Color Vision with a Blue-Light-Absorbing Intraocular Lens," *J. Cataract Refract. Surg.* **33**:667–672 (2007).
114. J. R. Sparrow, A. S. Miller, and J. Zhou, "Blue Light-Absorbing Intraocular Lens and Retinal Pigment Epithelium Protection In Vitro," *J. Cataract Refract. Surg.* **30**:873–878 (2004).
115. G. C. Brainard, J. P. Hanifin, J. M. Greeson, B. Byrne, G. Glickman, E. Gerner, M. D. Rollag, "Action Spectrum for Melatonin Regulation in Humans: Evidence for a Novel Circadian Photoreceptor," *J. Neurosci.* **21**:6405–6412 (2001).

116. M. A. Mainster, "Violet and Blue Light Blocking Intraocular Lenses: Photoprotection versus Photoreception," *Br. J. Ophthalmol.* **90**:784–792 (2006).
117. J. Kraats and D. Norren, "Sharp Cutoff Filters in Intraocular Lenses Optimize the Balance between Light Reception and Light Protection," *J. Cataract Refract. Surg.* **33**:879–887 (2007).
118. E. J. Hollick, D. J. Spalton, P. G. Ursell, M. V. Pande, S. A. Barman, J. F. Boyce, and K. Tilling, "The Effect of Polymethylmethacrylate, Silicone, and Polyacrylic Intraocular Lenses on Posterior Capsular Opacification 3 Years after Cataract Surgery," *Ophthalmology* **106**:49–54 (1999).
119. Q. Peng, N. Visessook, D. J. Apple, S. K. Pandey, L. Werner, M. Escobar-Gomez, R. Schoderbek, K. D. Solomon, and A. Guindi, "Surgical Prevention of Posterior Capsule Opacification. Part 3: Intraocular Lens Optic Barrier Effect as a Second Line of Defense," *J. Cataract Refract. Surg.* **26**:198–213 (2000).
120. J. T. Holladay, A. Lang, and V. Portney, "Analysis of Edge Glare Phenomena in Intraocular Lens Edge Designs," *J. Cataract Refract. Surg.* **25**:748–752 (1999).
121. S. Masket, "Truncated Edge Design, Dysphotopsia, and Inhibition of Posterior Capsule Opacification," *J. Cataract Refract. Surg.* **26**:145–147 (2000).
122. J. A. Davison, "Positive and Negative Dysphotopsia in Patients with Acrylic Intraocular Lenses," *J. Cataract Refract. Surg.* **26**:1346–1355 (2000).
123. W. R. Meacock, D. J. Spalton, and S. Khan, "The Effect of Texturing the Intraocular Lens Edge on Postoperative Glare Symptoms: A Randomized, Prospective, Double-Masked Study," *Arch. Ophthalmol.* **120**:1294–1298 (2002).
124. A. Franchini, B. Z. Gallarati, and E. Vaccari, "Computerized Analysis of the Effects of Intraocular Lens Edge Design on the Quality of Vision in Pseudophakic Patients," *J. Cataract Refract. Surg.* **29**:342–347 (2003).
125. A. Franchini, B. Z. Gallarati, and E. Vaccari, "Analysis of Stray-Light Effects Related to Intraocular Lens Edge Design," *J. Cataract Refract. Surg.* **30**:1531–1536 (2004).

*This page intentionally left blank*

# DISPLAYS FOR VISION RESEARCH

William Cowan

Department of Computer Science  
University of Waterloo  
Waterloo, Ontario, Canada

## 22.1 GLOSSARY

$D(\lambda)$	spectral reflectance of the diffuser
$e_p$	phosphor efficiency
$I_B$	beam current of a CRT
$M_{ij}$	coefficient $ij$ in the regression matrix linking $\Delta X_i$ to $\Delta R_i$
$m_j$	vector indicating the particular sample measured
$N_p$	number of monochrome pixels in a color pixel
$n_j$	vector indicating the particular sample measured
$R(\lambda)$	spectral reflectance of the faceplate of an LCD
$R_j$	input coordinates of the CRT: usually $R_1$ is the voltage sent to the red gun; $R_2$ is the voltage sent to the green gun; and $R_3$ is the voltage sent to the blue gun
$V$	voltage input to a CRT
$V_A$	acceleration voltage of a CRT
$V_0$	maximum voltage input to a CRT
$v_a$	voltage applied to gun $a$ of a CRT; $a = R, G, B$
$v_B$	maximum scanning velocity for the electron beam
$v_h$	horizontal velocity of the beam
$v_v$	vertical velocity of the beam
$X_{ai}$	tristimulus values of light emitted from monochrome pixel of color $a$
$X_j$	tristimulus values of light emitted from a color pixel
$X_i$	tristimulus value $i$ , that is, $X = X_1$ ; $Y = X_2$ ; and $Z = X_3$
$X_{0i}$	tristimulus value $i$ of a reference color
$x_p$	horizontal interpixel spacing, $x_p = v_h \tau_p$
$y_p$	vertical interpixel (interline) spacing

$\gamma$	exponent in power law expressions of gamma correction
$\delta$	interpolation parameter for one-dimensional characterizations
$\Delta R_i$	change in input coordinate $i$
$\Delta X_i$	change in tristimulus value $i$
$\epsilon$	interpolation parameter for inverting one-dimensional characterizations
$V_{\max}$	maximum frequency of CRT input amplifiers
$\mu$	index of color measurements taken along a curve in color space
$\Phi$	power of the emitted light
$\Phi_{a\lambda}^{(AMB)}(V_a)$	spectral power distribution of light emitted from monochrome pixel of color $a$ as a result of ambient light falling on an LCD
$\Phi_{a\lambda}^{(BL)}$	spectral power distribution of light emitted from monochrome pixel of color $a$ as a result of the backlight
$\Phi_{a\lambda}(V_a)$	spectral power distribution of light emitted by monochrome pixel of color $a$ , depending on the voltage with which the pixel is driven
$\Phi_{a\lambda}^{(R)}$	spectral power distribution of light reflected from the faceplate of an LCD
$\Phi(x, y)$	power of the emitted light as a function of screen position
$\Phi_\lambda$	spectral power distribution of light emitted by all monochrome pixels in a color pixel
$\Phi_\lambda$	spectral power of the emitted light
$\Phi_\lambda^{(AMB)}$	light output caused by ambient light falling on the faceplate of an LCD
$\Phi_\lambda^{(BL)}$	light output caused by the backlight of an LCD
$\Phi_0$	maximum light power output from a CRT
$\Phi_{0\lambda}^{(AMB)}$	ambient light falling on the faceplate of an LCD
$\Phi_{0\lambda}^{(BL)}$	light emitted by the backlight-diffuser element of an LCD
$\Phi_\lambda(v_R, v_G, v_B)$	spectral power distribution of the light emitted when voltages $v_R, v_G, v_B$ are applied to its inputs
$\tau_a(\lambda)$	spectral transmittance of the filter on monochrome pixel of color $a$
$\tau_d$	phosphor decay time
$\tau_f$	time spent scanning a complete frame
$\tau_l$	time spent scanning a complete line, including horizontal flyback
$\tau_p$	time the beam spends crossing a single pixel
$\tau(V)$	voltage-dependent transmittance of the light modulating element

## 22.2 INTRODUCTION

Complex images that are colorimetrically calibrated are needed for a variety of applications, from color prepress to psychophysical experimentation. Unfortunately, such images are extremely difficult to produce, especially using traditional image production technologies such as photography or printing. In the last decade technical advances in computer graphics have made digital imaging the dominant technology for all such applications, with the color television monitor the output device of choice. This chapter describes the operational characteristics and colorimetric calibration techniques for color television monitors, with its emphasis on methods that are likely to be useful for psychophysical experimentation. A short final section describes a newer display device, the color liquid crystal display, which is likely to become as important in the decade to come as the color monitor is today.

## 22.3 OPERATIONAL CHARACTERISTICS OF COLOR MONITORS

The color television monitor is currently the most common display device for digital imaging applications, especially when temporally varying images are required. Its advantages include good temporal stability, a large color gamut, well-defined standards, and inexpensive manufacture. A wide variety of different CRT types is now available, but all types are derived from a common technological base. This section describes the characteristics shared by most color monitors, including design, controls, and operation. Of course, a chapter such as this can only skim the surface of such a highly evolved technology. There is a vast literature available for those who wish a greater depth of technical detail. Fink et al.<sup>1</sup> provides a great deal of it, with a useful set of references for those who wish to dig even deeper. Colorimetry is discussed in Chap. 10 in this volume.

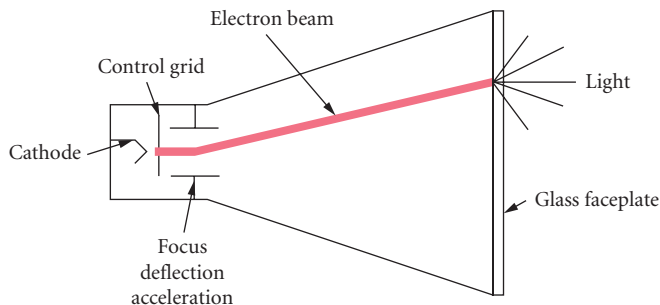
Output devices that go under a variety of names, color television receivers, color video monitors, color computer displays, and so on, all use the same display component. This chapter describes only the display component. Technically, it is known as a color cathode ray tube (CRT), the term that is used exclusively in the remainder of the chapter.

In this chapter several operational characteristics are illustrated by measurements of CRT output. These measurements, which are intended only to be illustrative, were performed using a Tektronix SR690, a now obsolete CRT produced for the broadcast monitor market. While the measurements shown are typical of CRTs I have measured, they are intended to be neither predictors of CRT performance nor ideals to be pursued. In fact, color CRTs vary widely from model to model, and any CRT that is to be used in an application where colorimetry is critical should be carefully characterized before use.

### Color CRT Design and Operation

The color CRT was derived from the monochrome CRT, and shares many features of its design. Thus, this section begins by describing the construction and operation of the monochrome CRT. New features that were incorporated to provide color are then discussed.

**Monochrome CRTs** A schematic diagram of a monochrome CRT is shown in Fig. 1. The envelope is a sealed glass tube from which all the air has been evacuated. Electrons are emitted from the cathode, which is heated red hot. The flux of electrons in the beam, the beam current  $I_B$ , is determined by a control grid. A variety of magnetic and/or electrostatic electrodes then focus, accelerate, and deflect the beam. The beam strikes a layer of phosphor on the inside of the CRT faceplate, depositing



**FIGURE 1** Schematic diagram of a monochrome CRT, showing the path of the electron beam and the location of the phosphor on the inside of the faceplate.



power  $I_B V_A$ , where  $V_A$  is the acceleration voltage. Some of this power is converted to light by the phosphor. The power  $\Phi$  in the emitted light is given by

$$\Phi = \int \Phi_\lambda \cdot \frac{hc}{\lambda} d\lambda$$

$\Phi_\lambda$ , the spectral power distribution of the emitted light is determined, up to a single multiplicative factor, by the chemical composition of the phosphor. The multiplicative factor is usually taken to be a linear function of the beam power. Thus, the efficiency of the phosphor  $e_p$  given by

$$e_p = \Phi / I_B V_A$$

is independent of the beam current. Power not emitted as light becomes heat, with two consequences.

1. If the beam remains in the same spot on the screen long enough, the phosphor coating will heat up and boil the phosphor from the back of the faceplate leaving a hole in any image displayed on the screen.
2. Anything near the phosphor, such as the shadowmask in a color CRT, heats up.

A stationary spot on the screen produces an area of light. The intensity of the light is generally taken to have a gaussian spatial profile. That is, if the beam is centered at  $(x_0, y_0)$ , the spatial profile of the light emitted is given by

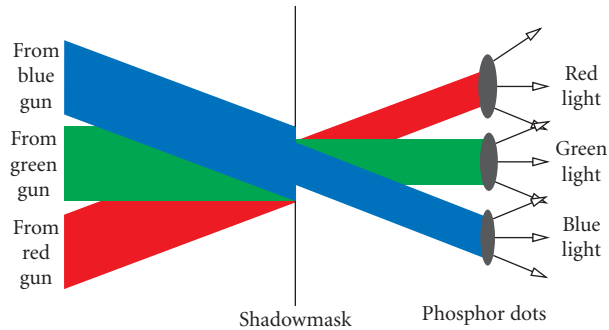
$$\Phi(x, y) \propto \exp \left[ -\frac{1}{2\sigma_x^2}(x - x_0)^2 - \frac{1}{2\sigma_y^2}(y - y_0)^2 \right]$$

The dependence of this spatial profile on the beam current is the subject of active development in the CRT industry. In most applications the beam is scanned around the screen, making a pattern of illuminated areas. The brightness of a given illuminated area depends on the beam current when the electron beam irradiates the area, with the beam current determined by the voltage applied to the control grid.

**Shadowmask Color CRTs** The basic technology of the monochrome CRT is extended to produce color. The standard method for producing color is to take several, usually three, monochrome images that differ in color and mix them additively to form a gamut of colors. (Even very unconventional technologies, such as the Tektronix bichromatic liquid crystal shutter technology, produce color by additive mixture.) This section describes the basic principles of shadowmask CRT technology, which is the dominant technology in color video.

*Geometry of the shadowmask CRT* A color CRT must produce three images to give a full range of color. To do so usually requires a tube with three guns. The three electron beams are scanned in exactly the way that monochrome beams are, but arrive at the screen traveling in slightly different directions. Near the faceplate is a screen called the shadowmask. It is made of metal, with a regular pattern of holes. Electrons in the beams either hit the screen to be conducted away or pass ballistically through the holes. Because the three beams are traveling in different directions, they diverge once they have passed the hole, striking the back of the faceplate in different places. The phosphor on the back of the faceplate is not uniform, but is distributed in discrete areas that radiate red, green, or blue light. The geometry is arranged so that electrons from the red gun hit red-emitting phosphor, electrons from the green gun hit green-emitting phosphor, and electrons from the blue gun hit blue-emitting phosphor. This geometry is illustrated in Fig. 2. Several different geometries of shadowmask tube exist:

1. *Delta guns.* The three guns are arranged at the corners of an equilateral triangle, irradiating phosphor dots arranged in a triad.



**FIGURE 2** Electron beam/shadowmask/phosphor geometry in a shadowmask CRT.

2. *In-line guns.* The three guns are arranged in a line, irradiating phosphor dots side by side. The lines of phosphor dots are offset from line to line, so that the dot pattern is identical to that of the delta gun.
3. *Trinitron.* This is an in-line gun configuration, but the phosphor is arranged in vertical stripes. The holes in the shadowmask are rectangular, oriented vertically.

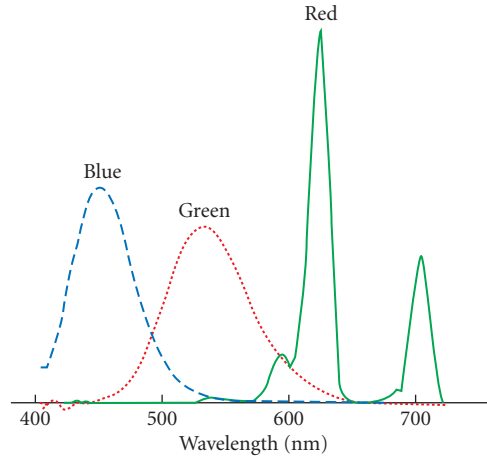
Other types of technology are also possible, though none is in widespread use. Beam index tubes, for example, dispense with the shadowmask, switching among the red, green, and blue signals as the electron beam falls on the different phosphor dots.

The most fundamental colorimetric property determining the colors produced by a shadowmask CRT is the light emitted by the phosphors. Only those colors that are the additive mixture of the phosphor colors (CRT primaries) can be produced. Because the color and efficiency of the phosphors are important determinants of perceived picture quality in broadcast applications phosphor chemistry undergoes continuous improvement, and varies from CRT to CRT. The emission spectra of the phosphors of a “typical” shadowmask CRT are shown in Fig. 3. The chromaticities of these phosphors are:

	$x$	$y$	$z$
Red phosphor	0.652	0.335	0.013
Green phosphor	0.298	0.604	0.098
Blue phosphor	0.149	0.064	0.787

*Common problems in shadowmask CRTs* The shadowmask is a weak point in color CRT design, so that color CRTs suffer from a variety of problems that potentially degrade their performance. The three most common problems are doming, blooming, and shadowmask magnetization. Doming occurs when the shadowmask heats because of the large energy density in the electrons it stops. It then heats up and expands, often not uniformly. The result is a distortion of its shape called “doming.” This geometrical distortion means that the registration of holes and dots is disturbed, and what should be uniform colors become nonuniform. Trinitron tubes have a tensioning wire to reduce this problem. It is often visible as a horizontal hairline running across the whole width of the tube about a third of the way from the top or bottom of the tube, depending on which side of the tube is installed up.

Blooming occurs when too much energy is deposited in the electron beam. Then electrons arrive at the screen in the right position but their entire kinetic energy is not immediately absorbed by the phosphor. They then can move laterally and deposit energy on nearby dots which can be both the wrong color and outside the intended boundary of the bright area. Colors become desaturated, and edges of areas become blurred.



**FIGURE 3** Spectral power distributions of the light output by a typical set of phosphors. The long wavelength peak of the red phosphor is sometimes missed in spectroradiometric measurements.

Magnetic fields build up in the shadowmask when the electromagnetic forces produced by electrons, assisted by the heat build-up, create magnetized areas in the shadowmask. These areas are nonuniform and usually produce large regions of nonuniformity in color or brightness on the screen. CRTs usually have automatic degaussing at power-up to remove these magnetic fields. If this is insufficient to remove the field buildup, inexpensive degaussing tools are available.

**CRT Electronics and Controls** The CRT receives voltage signals at its inputs, one for monochrome operation, three for color operation. The input signals must be suitably amplified to control the beam current using the grid electrode. The amplification process is described first, followed by the controls that are usually available to control the amplification.

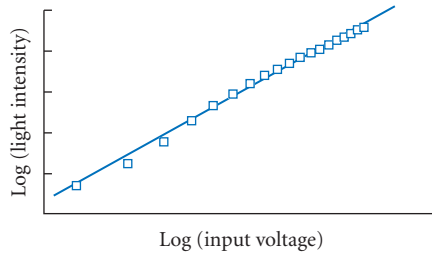
**Amplification** Two properties of the amplification are important to the quality of the displayed image: bandwidth and gamma correction.

The bandwidth of the amplifiers determines the maximum frequency at which the beam current can be modulated. In most applications the beam is moved about the screen to produce the image, so that the maximum frequency translates into a maximum spatial frequency signal. Suppose, unrealistically, that the amplifiers have a sharp cutoff frequency  $v_{\max}$  and that the beam is scanned at velocity  $v_B$ . Then the maximum spatial frequency of which the CRT is capable is  $v_{\max}/v_B$ . In some types of CRT, particularly those designed for vector use, settling time is more important than bandwidth. It can be similarly related to beam velocity to determine the sharpness of spatial transitions in the image.

For colorimetric purposes, the relationship between the voltage applied at the CRT input and the light emitted from the screen is very important. It is determined by the amplification characteristics of the input amplifiers and of the CRT itself, and creating a good relationship is part of the art of CRT design. When the relationship must be known for colorimetric purposes, it is usually determined by measurement and handled tabularly. However, for explanatory purposes it is often written in the form

$$\Phi = \Phi_0 (V/V_0)^\gamma$$

where  $V$  is the voltage input to the CRT, normalized to its maximum value  $V_0$ . The exponent, which is conventionally written as  $\gamma$ , gives this amplification process its name, gamma correction.



**FIGURE 4** Graph of  $\ln$  (light intensity) against  $\ln$  (input voltage), sometimes used to determine the gamma correction exponent. Note the systematic deviation from a straight line.

Figure 4 shows some measured values, and a log-linear regression line drawn through them. Note the following features of this data.

1. The line is close to linear.
2. There are regions of input voltage near the top and bottom where significant deviations from linearity occur.
3. The total dynamic range (the ratio between the highest and lowest outputs) is roughly 100, a typical value. This quantity depends on the setting of brightness (black level) and contrast used during the measurement.

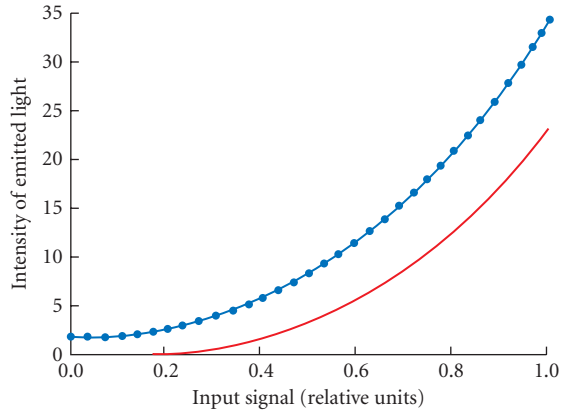
There will be additional discussion of gamma correction in the section on monitor setup.

*Controls that affect amplification in monochrome CRTs* Monochrome monitors have several controls that adjust several properties of the electron beam. Some controls are external to the monitor, some are internal, usually in the form of small potentiometers mounted on the circuit boards. The particular configuration depends on the monitor. In terms of decreasing likelihood of being external they are: brightness (black level), contrast, focus, underscan/overscan, pedestal, gain, horizontal and vertical size. A brief description of the action of each control follows. It is important, however, to remember that while names of controls tend to be constant from CRT to CRT, the action performed by each control often varies.

*Brightness (black level)* This control adjusts the background level of light on the monitor screen. It is designed for viewing conditions where ambient light is reflected from the monitor faceplate. It usually also varies the gamma exponent to a higher value when the background level (black level) is increased. A typical variation of the light intensity/input voltage relationship when brightness is varied is shown in Fig. 5.

*Contrast* This control varies the ratio between the intensity of the lightest possible value and the darkest possible value. High contrast is usually regarded as a desirable attribute of a displayed image, and the contrast control is usually used to produce the highest contrast that is consistent with a sharp image. A typical variation of the light intensity/input voltage relationship when this control is varied is shown in Fig. 6.

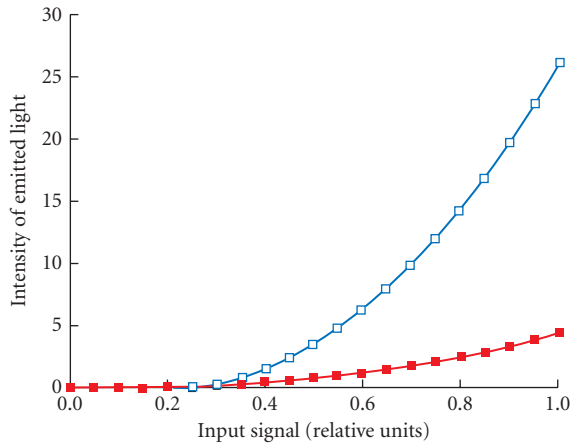
*Focus* This control varies the size of the electron beam. A more tightly focused electron beam produces sharper edges, but the beam can be focused too sharply, so that flat fields show artifactual spatial structure associated with beam motion. Focus is usually set with the beam size just large enough that no intensity minimum is visible between the raster lines on a uniform field.



**FIGURE 5** Variation of the light intensity/input voltage relationship when the brightness (black level) control is varied. The lower curve shows the relationship with brightness set near its minimum; the upper one with brightness set somewhat higher.

*Pedestal and gain* These controls, which are almost always internal, are similar to brightness and contrast, but are more directly connected to the actual amplifiers. Pedestal varies the level of light output when the input voltage is zero. Gain varies the rate at which the light output from the screen increases as the input voltage increases.

*Controls specific to color CRTs* Color monitors have a standard set of controls similar to those of monochrome monitors. Some of these, like brightness and contrast, have a single control applied simultaneously to each of the color components. Others, like gain and pedestal, have three controls, one for each of the color channels. There are several aspects of color that need to be controlled, however, and they are discussed in the paragraphs that follow.



**FIGURE 6** Variation of the light intensity/input voltage relationship when the contrast is varied. The lower curve shows the relationship with contrast set near its minimum; the upper one with contrast near its maximum.

*Purity* Purity is an effect associated with beam/shadowmask geometry. It describes the possibility that the electron beams can cause fluorescence in inappropriate phosphors. There is no standard set of controls for adjusting purity. Generally, there are magnets in the yoke area whose position can be adjusted to control purity, but this control is very difficult to perform. There will be no need to alter them under normal conditions. Purity is most influenced by stray magnetic fields, and can often be improved by moving the CRT.

*White balance* It is important for most monitor applications that when the red, green, and blue guns are turned on equally, calibration white (usually either  $D_{6500}$  or  $D_{9200}$ ) appears on the screen. This should be true at all intensities. Thus, controls that alter the voltage input/light output relationship for each channel should be available. At a minimum, there will be the pedestal and gain for each channel.

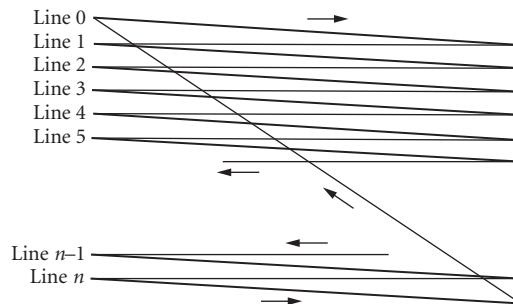
*Degauss* Above we mentioned magnetic fields that build up in the shadowmask. There is generally a set of wires that run around the edge of the faceplate. At power-up a degaussing signal is sent through the wires and the ensuing magnetic field degausses the shadowmask. The degauss control can produce the degauss signal at any time.

**CRT Operation** The CRT forms an image on its screen by scanning the electron beam from place to place, modulating the beam current to change the brightness from one part of the image to another. A variety of scan patterns are possible, divided into two categories. Scan patterns that are determined by the context of the image (in which, for example, the beam moves following lines in the image) are used in vector displays, which were once very common but now less so. Scan patterns that cover the screen in a regular pattern independent of image content are used in raster displays: the scan pattern is called the raster. A variety of different rasters are possible; the one in most common use is a set of horizontal lines, drawn from top to bottom of the screen.

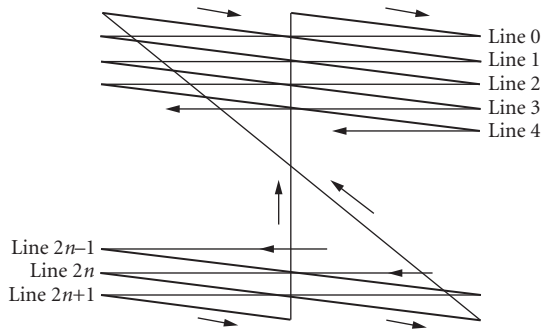
*Raster generation* Almost all raster CRTs respond to a standard input signal that creates a raster consisting of a set of horizontal lines. This section describes the path of the electron beam as it traverses the CRT screen; a later section discusses the signal configurations that provide the synchronization necessary to drive it.

*Frames and fields* The image on the screen is scanned out as a set of lines. Each line is scanned from left to right, as seen from a position facing the screen. Between the end of one line and the beginning of the next the beam returns very quickly to the left side of the screen. This is known as horizontal retrace or flyback. The successive lines are scanned from top to bottom of the screen. One field consists of a scan from top to bottom of the screen. Between the end of one field and the beginning of the next the beam returns very quickly to the top of the screen. This is known as vertical retrace or flyback.

One frame consists of a scan of all the lines in an image. In the simpler type of display a field is identical to a frame, and all the lines of the image are scanned out, from top to bottom of the display. The scan pattern is shown in Fig. 7. It is called noninterlaced.



**FIGURE 7** Scan pattern for a noninterlaced raster.



**FIGURE 8** Scan pattern for an interlaced raster, two fields per frame. Line 3 eventually scans to line  $2n - 1$  by way of the odd lines; line 4 scans to line  $2n$ .

A more complicated type of raster requires more than one field for each frame. The most usual case has two fields per frame, an even field and an odd field. During the even field, the even-numbered lines of the image are scanned out with spaces between them. This is followed by vertical retrace. Then, during the odd field the odd-numbered lines of the image are scanned out into the spaces left during the even-field scan. A second vertical retrace completes the frame. This scan pattern, known as interlaced, is shown in Fig. 8. The purpose of interlace is to decrease the visible flicker within small regions of the screen. It works well for this purpose, provided there are no high-contrast horizontal edges in the display. If they are present they appear to oscillate up and down at the frame rate, which is usually 30 Hz. This artifact can be very visible and objectionable, particularly to peripheral vision. Interlace having more than two fields per frame is possible, but uncommon.

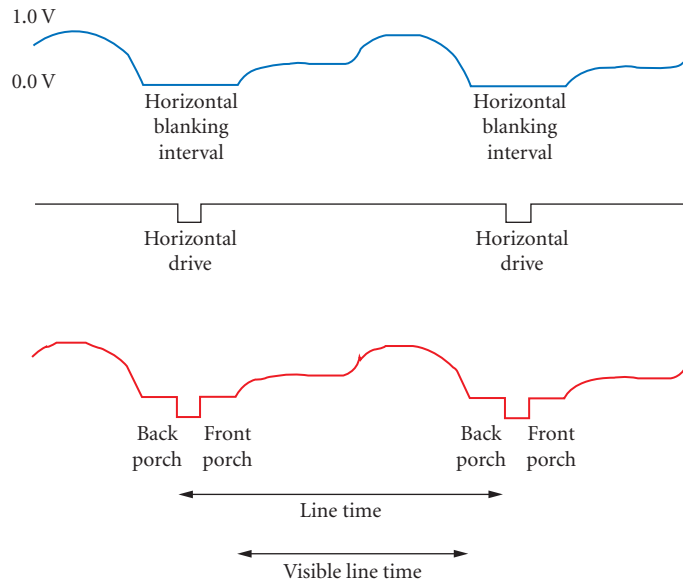
In viewing Figs. 7 and 8 note that the vertical scan is produced by scanning the beam down continuously. Thus the visible lines are actually sloped down from left to right while the retrace, which is much faster than the horizontal scan, is virtually unsloped. The method used to put the odd field of the interlaced scan between the lines of the even field is to scan half a line at the end of the even field followed by half a line at the beginning of the odd field. Thus an interlaced raster has an odd number of lines.

*Relationship of the raster to the CRT input signal* The CRT receives a serial input signal containing a voltage that controls the intensity for each location in the image. This signal must be synchronized with the raster in order to make sure that each pixel is displayed in the right location. This section describes the relationship between the raster and the CRT input.

*Horizontal scanning* The input signal for one line is shown in Fig. 9. It consists of data interspersed with blank periods, called the horizontal blanking intervals. During the horizontal blanking interval the beam is stopped at the end of the line, scanned quickly back for the beginning of the next line, then accelerated before data for the next line begins. In the data portion, 0.0 V indicates black and 1.0 V indicates white. The signal shown would produce a line that is dim on the left, where the line starts, and bright on the right where it ends.

A second signal, the horizontal synchronization signal, has a negative-going pulse once per line, positioned during the horizontal blanking interval. This pulse is the signal for the CRT to begin the horizontal retrace.

When the synchronization signal is combined with the data signal the third waveform in Fig. 9 is produced. The intervals in the horizontal blanking interval before and after the synchronization are known as the front and back porch. They are commonly used to set the voltage level corresponding to the black level, a process known as black clamping. Black clamping reduces the effect of low-frequency noise on the color of the image, but requires good synchronization between the precise timing of the



**FIGURE 9** A schematic input signal that would generate a single line of raster, including the end of the preceding line and the beginning of the succeeding line. The top trace shows the signal that produces the image, including the blank between lines; the middle trace shows the synchronization signal with the horizontal drive pulse located during the blank; the bottom trace shows the synchronization signal and the picture signal in a single waveform.

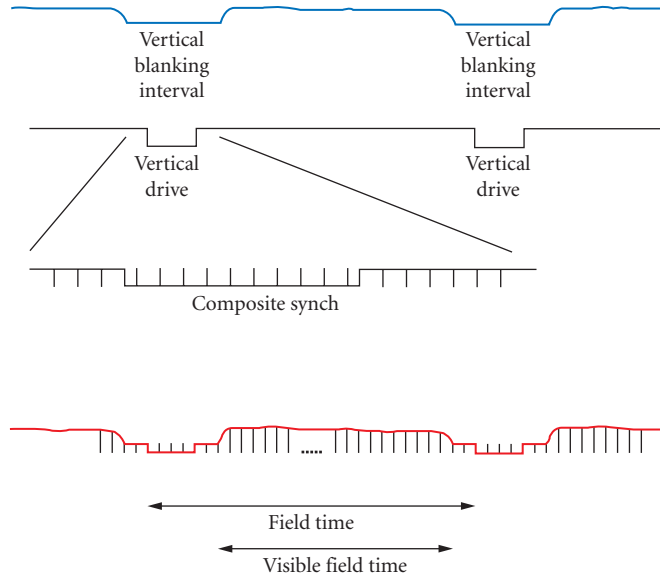
signal and the raster production of the CRT. The ability of the monitor to hold synchronization with particular timing in this signal is an important parameter of the monitor electronics.

*Vertical scanning* Figure 10 shows the input signal for one complete field. There is data for each line, separated by the horizontal blanking period, which is too short to be visible in the figure. Separating the data portion of each field is the vertical blanking interval, during which the beam is scanned back to the top of the screen.

The synchronization signal consists of a vertical drive pulse signaling the time at which the beam should be scanned to the top of the screen. This pulse is long compared to the horizontal drive pulses so that it can easily be separated from them when the synchronization signals are combined into composite synch. In composite synch positive-going pulses in positions conjugate to the positions of the horizontal drive pulses are added during the vertical drive signal. These pulses are designed to keep the phase of the horizontal oscillator from wandering during vertical drive. When they are interpreted incorrectly, as was the case in some monitors early in the history of digital electronics, the result is a small shear in the first few lines at the top of the screen as the horizontal oscillator gets back into phase.

The composite synch signal can be added to the data, as shown in the fourth illustration in Fig. 10. Most real systems have this type of synchronization, which was designed for an era when signals were to be sent long distances by broadcast or wire. Today, in computer graphics applications, we often find that electronic circuitry in the source carefully folds the two synch signals into composite synch and the composite synch signal into the data signal. This signal is carried along a short piece of wire to the receiver, where electronic circuitry strips the signals apart. Bad synchronization is often caused by the malfunction of this superfluous circuitry, but it is unlikely that this situation will change in the immediate future.





**FIGURE 10** A schematic input signal that would generate a single field of a raster, including the end of the preceding field and the beginning of the succeeding field. The top trace shows the picture signal, with the vertical blank shown and the horizontal blank omitted. The second trace shows the vertical synchronization signal with the vertical drive pulse. The third trace shows the horizontal synchronization signal added to the vertical signal to give composite sync. The bottom trace shows composite sync added to the picture signal.

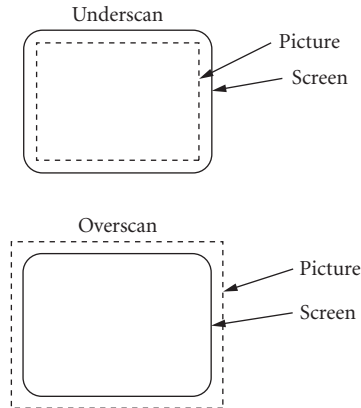
At the input to a color CRT, three or four signals like the ones described above are provided. The four signals are three input signals containing pixel intensities and blanking intervals, plus a fourth signal carrying the composite synchronization signal. To get three signals the synchronization signal is combined with one or more of the pixel signals. When only one pixel signal has synchronization information it is almost always the green one.

*Controls that affect the raster* Several controls affect the placement and size of the raster on the CRT. They are described in this section.

*Horizontal/vertical size and position* These controls provide continuous variation in the horizontal and vertical sizes of the raster, and in the position on the CRT faceplate where the origin of the raster is located.

*Underscan/overscan* Most CRTs provide two standard sizes of raster, as shown in Fig. 11. This control toggles between them. In the underscan position the image is smaller than the cabinet port, so that the whole image is visible, surrounded by a black border. In the overscan position the image is slightly larger than the cabinet port, so that no edges of the image are visible. There is a standard amount of overscan, which is used in home television receivers.

*Convergence* A shadowmask color CRT has, in effect, three rasters, one for each of the primary colors. It is essential that these rasters be positioned and sized exactly the same in all parts of the image; otherwise, spurious colored fringes appear at the edges in the image. To achieve this it is essential



**FIGURE 11** Schematic illustration of overscan and underscan. The rectangle with rounded corners is the cabinet mask that defines the viewing area; the rectangle with square corners is the displayed image.

that all three electron beams should be at the same place at the same time. For example, if the green gun is slightly to the left of the red and blue guns, a white line of a black background will have a green fringe on its left side and a magenta fringe on its right side. Good convergence is relatively easy to obtain with in-line gun configurations, so the usual practice is to adjust convergence at the factory using ring-shaped magnets placed over the yoke of the CRT, then to glue them permanently in place with epoxy cement. Satisfactory readjustment is very difficult to achieve.

Delta gun configurations have, by reputation, less stable convergence. Consequently, there are often controls available to the user for adjustment of convergence. The Tektronix 690SR was an extreme case, with a pullout drawer containing 52 potentiometers for controlling convergence on different areas of the screen. However, these controls, as might be expected, are not independent, so that “fine tuning” the convergence is very difficult, even when the controls are conveniently located.

Magnetic fields are generally the main culprit when convergence is objectionable, since the paths of electrons are curved in magnetic fields. Small fields from power transformers or from other electronic equipment or even the earth’s magnetic field can be the problem. Thus, before considering redoing the convergence on a new monitor, try rotating it and/or moving other equipment around in the room. Another useful trick: some low-cost CRTs have poorly positioned power transformers, and convergence can be improved by moving the power transformer far outside the CRT cabinet.

## Operational Characteristics of Color CRTs

Two factors influence the light output that is produced in response to input signals. One is conventional: the input signal must be sequenced and timed precisely to match what is expected by the CRT. This problem is handled by input signal standards of long standing in the television industry. Synchronization standards are described first, followed by colorimetric standards. The second is technical: the construction of CRTs produces certain variations, even when the input signal is constant. The section ends with a discussion of this variation, specifying the temporal and spatial characteristics of the light output from a CRT and describing reasonable expectations for stability and uniformity in CRT-produced images.

**Timing and Synchronization Standards** Figures 9 and 10 show how the information needed to specify a CRT image are arranged in the input signal. For practical use they need to be specified numerically, thereby allowing CRT manufacturers to ensure that their products will respond appropriately to

the input signals they will encounter in service. Two standards have been created by the Electronic Industries Association (EIA) for this purpose. They are actually standards for television studios, prescribing how signals should be distributed in closed-circuit applications, and specifically between television cameras and studio monitors. In practice they are more widely used. Among other applications, they specify the interface between the tuner and the CRT in television receivers, and the output of digital graphics systems allowing studio monitors to be used as display devices.

The two standards are RS-170,<sup>2</sup> which was designed for lower bandwidth applications, and RS-343,<sup>3</sup> which was designed for higher bandwidth applications. Each gives minimum and maximum timings for each part of the input signal in terms of several parameters that are allowed to vary from application to application. These parameters are then part of CRT input specification, allowing all timing characteristics to be deduced. The most important parameter is the line rate, the number of lines displayed per second. Most CRTs are intolerant of large variations in this parameter. Another important parameter is the field rate, the number of fields displayed per second. Older CRTs were quite intolerant of variations in this parameter, but most modern CRTs can handle signals with a very wide variety of field rates.

RS-170 and RS-343 are monochrome standards. When used for color they are tripled, with the input signal assumed to be in three parallel monochrome signals. As mentioned above, the synchronization signal is either placed on a fourth, or incorporated into a single color signal, usually the green one. However, although this practice is almost universal there is no official color standard for the RGB input to a monitor. This lack can have unfortunate consequences. For example, the NTSC color signal naturally decodes into three signals with peak-to-peak voltages of 0.7 V. Thus RGB monitors were built with inputs expecting this range. RS-170 and RS-343, on the other hand, specify a peak-to-peak voltage of 1.0 V. Early digital graphics systems were built to provide exact RS-170 and RS-343 output. These systems, naturally, overdrove standard monitors badly.

**Colorimetric Standards** In broadcast applications the image transmitter should be able to specify the precise color that will be displayed on the receiver's CRT. This requirement can be supplied by a colorimetric standard. The NTSC color standard was agreed upon for use in the North American broadcast television industry. It is a complete color standard, specifying phosphor chromaticities, color representation on the carrier signal, signal bandwidth, gamma correction, color balance, and so on. Thus, if the NTSC standard were followed in both transmitter and receiver, home television would provide calibrated colors. It is not followed exactly, however, since both television manufacturers and broadcasters have discovered that there are color distortions that viewers prefer to colorimetrically precise color. Furthermore, it is not useful for high-quality imagery since the low bandwidth it allocates to the chromatic channels produces edges that are incompatible with good image quality.

**Spatial and Temporal Characteristics of Emitted Light** The light emitted from a CRT is not precisely uniform, but suffers from small- and large-scale variations. The small-scale variations arise because the image is actually created by a spot that moves over the entire display surface in a time that is intended to be short compared to temporal integration times in the human visual system. In fact, a short enough glimpse of the CRT screen reveals only a single point of light; a longer one reveals a line as the point moves during the viewing time. These patterns are designed to be relatively invisible under normal viewing conditions, but often need to be considered when CRTs are used for vision experimentation, or when radiometric measurements are made. They are controlled largely by the input signal, so they are relatively constant from CRT to CRT.

The small-scale variations are described in this section; the large-scale variations, which occur as spatial nonuniformity and temporal instability in the emitted light, are discussed in the following section.

**Spatial characteristics** The electron beam scans the screen horizontally, making an image that consists of a set of horizontal lines. During the scan its intensity is modulated, so that the line varies in brightness—vertical edges, for example, being created by coordinated modulation in a series of horizontal lines. The sharpness of such a line depends on two factors: the ability of the video amplifiers to produce an abrupt change in intensity and the size of the spot of light on the screen, which is essentially the same as the cross section of the electron beam. Video amplifiers vary greatly in bandwidth

from one model of CRT to another, and adjusting their performance is beyond the reach of virtually all CRT users. Unfortunately, interactions between adjacent horizontal pixels are not even linear,<sup>4-6</sup> and the measurements needed to determine the nonlinearity of a given CRT are extremely demanding. Thus, compensating for amplifier bandwidth or even measuring its effects is beyond the scope of this chapter.

By contrast, the sharpness of a horizontal line is independent of the video amplifiers, but depends only on the spot size and its spatial profile, which is well modeled as a two-dimensional gaussian. The width of the gaussian depends on the focus electrodes, and is user-adjustable on most CRTs. The shrinking raster technique for CRT setup (discussed in subsection “CRT Setup for Image Display” discussed later in this chapter) determines a spot size that has a specific relationship to the interline spacing. Assuming this setup, it is possible to make reasonable assumptions about the contrast of images on a CRT,<sup>7</sup> and these assumptions can be extended to the small-scale spatial structure of arbitrary images. Note that many CRTs used primarily as display terminals are overfocused compared to the shrinking raster criterion because such overfocusing allows the display of smaller text. Overfocusing is easily detected as visible raster lines, usually in the form of closely spaced dark horizontal lines when a uniform field is displayed.

*Temporal characteristics* Because of the scan pattern the light emitted from any portion of the screen has a complicated temporal dependence. The next few paragraphs describe several levels of this dependence, assuming, for simplicity, that the scan is noninterlaced. Similar results for an interlaced scan are easy to derive. The relevant variables are

$\tau_d$	phosphor decay time
$\tau_p$	time the beam spends crossing a single pixel
$v_h$	horizontal velocity of the beam
$x_p$	horizontal interpixel spacing, $x_p = v_h \tau_p$
$\tau_l$	time spent scanning a complete line, including horizontal flyback
$v_v$	vertical velocity of the beam
$y_p$	vertical interpixel (interline) spacing, $y_p = v_v \tau_l$
$\tau_f$	time spent scanning a complete frame (identical to the time spent scanning a complete field)

These temporal factors usually change the colorimetry of the CRT. When one or more is changed, usually because the video source has been reconfigured, the color output for a given input to the CRT usually changes. Thus, a recalibration should be done after any such change. And it is probably practical to recommend that the system be used with the new video timing parameters for a while before the recalibration is done, since tuning or touch-ups will require further recalibration. The intensity of the light emitted from a vanishingly small area of the CRT screen is a function that is zero until the beam traverses the point  $t_s$ , then decays exponentially afterward,

$$\Phi(t) = \sum \Phi_0 \theta(t - t_s) \exp(-(t - t_s) / \tau_d)$$

$\tau_d$  ranges between  $10^{-3} \tau_p$  and  $\tau_f$ , and usually varies from one phosphor to another. Most often the green phosphor has the largest  $\tau_d$ , the blue one the smallest. Broadcast monitors tend to have small  $\tau_d$ 's since they are designed to display moving imagery; data display CRTs tend to have large  $\tau_d$ 's since they are intended to display static imagery. Occasionally a CRT primary (most often red) is a mixture of two phosphors with different decay times. In such cases, the chromaticity of the light emitted by the primary changes over time, though the change is not usually visible.

If a second pixel is  $n_h$  pixels to the right of a given pixel and  $n_v$  lines below it, the light emitted by the second pixel lags behind the light emitted by the first pixel by

$$n_h \tau_p + n_v \tau_l \approx d_h / v_h + d_v / v_v$$

where  $d_h$  and  $d_v$  are the horizontal and vertical screen distances.

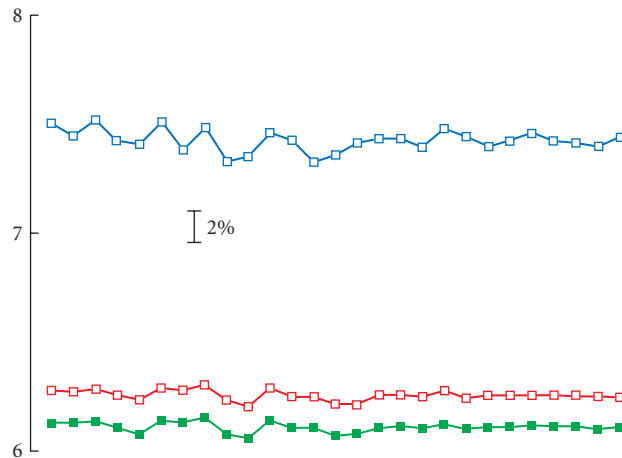
Commonly, a detector measures a finite area of the CRT screen. Then intensity of the detected light is a closely spaced group of exponentials followed by a long gap, then another closely spaced group, and so on. Each peak in the closely spaced group is spaced about  $\tau_p$  from the previous one, occurring when each line of the raster crosses the detection area. Each group is spaced about  $\tau_f$  from the previous one, occurring each time a new field repaints the detected target. The time constant of this composite signal is  $\tau_p$ , the decay time of the phosphor for each individual peak in the signal.

**Stability and Uniformity of CRT Output** The small-scale structure of the emitted light, discussed above, determines the specific form of visual stimuli produced by a CRT. The large-scale structure determines the scope of applicability of measurements made on a particular area of a CRT at a particular time.

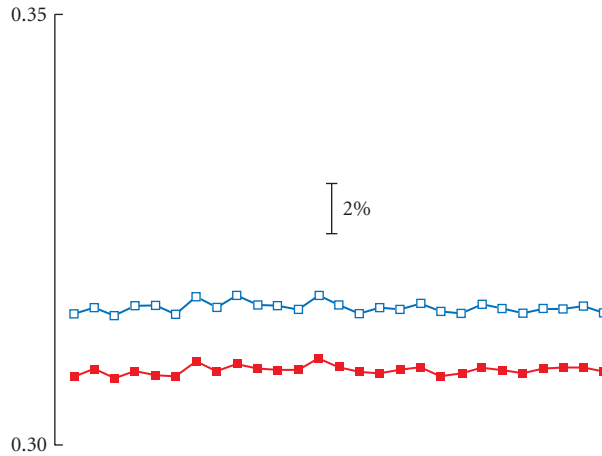
**Temporal stability** How constant is the light emitted by a CRT that receives the same input signal? Figures 12 and 13 show the results of colorimetric measurements taken over a 24-hour period. In each figure the calibration bars shows a 2 percent variation compared to the average value. The variation decreases considerably in the latter part of the graph. It is the period from 5 p.m. until 9 the next morning, showing that variations in the building power are the largest source of temporal instability. Furthermore, this variation affected all guns similarly, as is shown by the smaller variation in Fig. 13, which shows the chromaticity coordinates, thereby factoring out overall changes in intensity.

Measurements like this one are important for establishing the precision at which calibration is sensible: it isn't worthwhile to calibrate to a greater precision than the variation in the colorimetry over the period between calibrations. While this CRT has excellent temporal stability, significantly better than most other light sources, the same is not true for all CRTs. Some vary in output by as much as 20 to 30 percent over times as short as a few minutes.

The measurements shown in Figs. 12 and 13 give an idea of the variation of light output over periods of hours or days, and illustrate the precision it is possible to expect from a recently calibrated CRT. CRTs also vary on a time scale of years, but the effects are not documented. Anecdotal reports suggest the following. First, electronic components change with age, so properties that depend on the



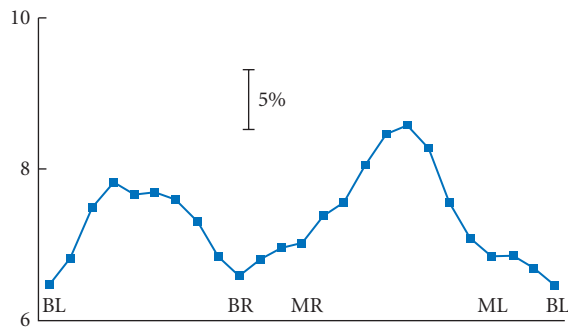
**FIGURE 12** Variation of light output from a color CRT over 24 hours of continuous operation. This graph shows the three tristimulus values when a neutral color is displayed. The latter part of the graph is at night when almost all other equipment in the building is turned off.



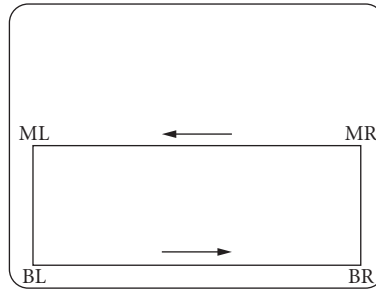
**FIGURE 13** Variation of light output from a color CRT over 24 hours of continuous operation. This graph shows the chromaticity coordinates corresponding to Fig. 12. They show less variation than the tristimulus values which covary considerably.

CRT electronics, such as video amplifier gain and bandwidth, change with age, almost always for the worse. Second, chemical properties do not change with age, so properties such as the spectral power of light emitted by a specific phosphor do not change. One anecdotal report<sup>8</sup> describes no variation in chromaticity coordinates of phosphor emission spectra over several years of CRT operation. It is possible, however, that phosphor concentrations diminish as tubes age, probably because the phosphor slowly evaporates. Such an effect would reduce the intensity of light emitted from the phosphor without changing its chromaticity. The magnitude of this effect is controversial.

*Spatial uniformity* The light emitted by a specific input signal varies a surprising amount from one area of the screen to another. Figure 14 shows the variation of luminance at constant input voltage as



**FIGURE 14** Variation of light output when different parts of a CRT screen are measured from a fixed point. Horizontal lines mark variations of about 5 percent.



**FIGURE 15** The measurement path used for the measurements shown in Fig. 14.

we measure different areas of the screen from a fixed measurement point. (Figure 15 shows the location on the screen of the measurement path.) Note that the light intensity decreases as the measurement point moves away from the center either horizontally or vertically, and is lowest in the corners. Two effects work together to create this effect. As the beam scans away from the center of the tube, it meets the shadowmask at more and more oblique angles, making the holes effectively smaller. Because of the curvature of the tube and the finite distance of the observer, the edges and corners are viewed at angles off the normal to the tube. Light is emitted in a non-Lambertian distribution, preferring directions closer to the normal to the tube face. The effects in Fig. 14 occur in all CRTs. How large they are, however, depends strongly on the type and setup of the monitor. Correcting for this nonuniformity is usually impractical. Doing so requires very extensive measurement.<sup>9</sup>

Closer examination of measured light shows, however, that many experimental situations are not hampered by this nonuniformity. Usually the chromaticity coordinates are very close to being constant, even though the luminance varies greatly. General intuition about color, as well as some recent experiments,<sup>10</sup> shows that humans are quite insensitive to smooth luminance gradients, even when they are as large as 20 percent. This fact, combined with commonsense layout of experimental displays (making them symmetrical with respect to the center of the screen, for example) overcomes spatial nonuniformity without extensive measurement. Important to mention in this respect is the difficulty of creating good stereoscopic viewing conditions on a single monitor. First, there is only a single area of the screen that is ideal for the position of the center of an image. Second, unless one image is horizontally inverted, two images from opposite sides of the screen combined into a single image present drastically different luminance gradients to the two eyes.

## Setup and Viewing Environments for Color CRTs

Many adjustments of the CRT electronics change its performance substantially. Some, such as the purity and convergence, have specific “correct” settings and are not designed to be user-adjustable. Most CRTs with in-line guns, for example, have the ring magnets that adjust the convergence glued into place at the factory. Other controls, such as contrast and brightness, are user-adjustable and should be changed when viewing conditions change if an optimal image is to be produced. These controls are provided in the expectation that the CRT will be displaying an image that is broadcast to a large number of CRTs that are viewed in very different visual environments. There is, therefore, a correct way to adjust these controls; the next few paragraphs describe the basic adjustments to be done. The procedure is based on a technical manual produced by the Canadian Broadcasting Corporation.<sup>11</sup> This manual also provides recommended viewing conditions for critical assessment of displayed images. The procedures and viewing conditions are expected to form the basis of an SMPTE standard for CRT image display.

When CRTs are used for visual experimentation, of course, they are often displaying specifically controlled images in unusual viewing conditions, often total darkness. For such applications the

adjustment procedure described below is unlikely to be interesting, and extreme values of the controls are likely to be desired. For example, an experiment conducted in total darkness is likely to need the black level (brightness) set so that there is no background light emitted from the screen. Or an experiment to measure thresholds is likely to benefit from whatever value of the contrast control minimizes the gain of the CRT at the intensity levels where the measurement is performed. (The objective is to minimize intensity quantization of the stimulus.) In fact, modern CRTs with computer-controllable contrast and black level might be used with different control settings for different trials of a single experiment.

**CRT Setup for Image Display** When a CRT is set up for image display, four adjustments are usually made: focus, brightness, contrast, and color balance and tracking. These adjustments form a rough order, with changes to one often requiring changes to succeeding ones. The following procedures are simplified from Benedikt.<sup>11</sup>

**Focus** Focus should be adjusted by the shrinking raster method. A uniform gray field is displayed on the CRT, and the focus adjustment used to shrink the beam size until raster lines are clearly visible. The beam size is then increased until the raster lines just barely disappear. The smallest beam size for which no horizontal raster lines are visible indicates the correct setting for the focus control.

**Brightness** The brightness or black-level control is set so that zero input to the CRT produces a visual impression of black. This setting *must* be performed in lighting conditions that are exactly those in which the images will be viewed, and with the observer at exactly the viewing distance at which the images will be viewed. With no signal input to the CRT, and the image set to underscan if possible, the brightness control is increased until the image area is noticeably gray. It is then reduced to the highest setting at which the image area looks black.

**Contrast** An input that has all three guns fully on is used for this adjustment. With such a color displayed, the contrast control is adjusted until the luminance of the screen is the maximum luminance desired. This setting should be performed using either a luminance meter or a color comparator, a CRT adjustment device that allows an observer to view the CRT as half of a bipartite field, the other half containing a reference white at the correct luminance. In any case, it is essential that the CRT should not bloom at the contrast setting in use. Blooming occurs when a too-intense electron beam spreads after it has passed through the shadowmask, stimulating more than one phosphor. It reduces the purity of the image, with the visual consequence that bright saturated colors are washed out. Narrow lines of red, green, and blue at full intensity can be used to check for blooming.

**Color balance and tracking** A white with red, green, and blue inputs equal at maximum intensity and a gray with red, green, and blue inputs equal at half intensity are used to set the color balance. In addition, colorimetric capability is needed. Visual colorimetry using a luminous white reference is usual in the broadcast industry, but this measurement can also be made instrumentally. For visual colorimetry the white reference is usually supplied by a color comparator. Adjust the red, green, and blue gain controls until the displayed white matches the reference in chromaticity. Then adjust the red, green, and blue screen controls until the displayed gray matches the reference in chromaticity. It may now be necessary to readjust the brightness and contrast. Do so, then repeat the color balance and tracking adjustment until all the three adjustments are simultaneously satisfactory.

**Viewing Environments** In applications where image quality is critical, it is necessary to control the viewing environment very closely. The following viewing conditions are typical of those used in the broadcast television industry.

1. The luminance of reference white is about 70 cd/m<sup>2</sup>.
2. The observer views the screen from a direction normal to the screen, with the screen-observer distance between 4 and 6 times the screen height.



3. The CRT should be surrounded by a neutral matte area at least 8 times the screen area. The surround should have the chromaticity of the CRT reference white, and a luminance of about  $10 \text{ cd/m}^2$ .
4. A narrow matte-black mask should frame the CRT image.
5. All room lighting should be as close as possible to the chromaticity of the CRT reference white.

## 22.4 COLORIMETRIC CALIBRATION OF VIDEO MONITORS

---

When users of color CRTs talk precisely about colorimetric calibration, a careful distinction between calibration and characterization is usually made. Measuring the input voltage/output color relationship well enough that it is possible to predict the output color from the input voltage, or to discover the input voltage needed to produce a given output color is a CRT characterization. The function mapping voltage to color characterizes the colorimetric performance of the CRT. Adjusting the CRT so that its characterization function matches the characterization function of a standard CRT is a CRT calibration. Talking loosely, however, calibration usually covers both characterization and calibration, and most CRT users are actually more interested in characterizations than in calibrations. Thus, this chapter describes several methods for creating colorimetric characterizations of CRTs. Some such characterization is an essential part of any calibration, but omits the detailed internal adjustment of a CRT, which requires significant electronic expertise if it is to be done safely.

There is no single characterization method that suits all needs. Thus, this chapter provides a variety of methods, each of which does some jobs well and others badly. They can be divided into three basic types.

1. *Exhaustive characterization methods (ECM)* Useful when good precision is needed over the complete monitor gamut. The same or similar algorithms can be used for the characterization of other output devices, such as printers. These methods tend to be both computationally and radiometrically expensive.
2. *Local characterization methods (LCM)* Useful when a precise characterization is needed for only a small portion of the monitor's output range.
3. *Model-dependent characterization methods (MDCM)* Useful when a characterization of moderate precision is needed for the complete monitor gamut. These methods tend to be specific to a given monitor and useful for only a small set of monitor setup methods, but they are computationally and radiometrically inexpensive. In addition they can be done so that the perceptual effects of mischaracterizations remain small, even when the colorimetric effects are large.

Here is a small set of criteria for deciding which type of method is best for a specific application:

1. High precision (better than 1 to 3 percent is impractical by any method discussed in this chapter)—ECM or LCM
2. Complete gamut characterization needed—ECM or MDCM
3. Minimal memory available for characterization tables—LCM or MDCM
4. No or slow floating point available—ECM or LCM
5. Fast inverse needed (we call the transformation from *RGB* to *XYZ* the characterization, the transformation from *XYZ* to *RGB* the inverse)—ECM, LCM, or MDCM
6. Forgiving behavior with out-of-gamut colors—LCM or MDCM
7. Photometry available but not radiometry—MDCM
8. Change of small number of parameters when monitor or monitor setup changes—LCM or MDCM

This list is not comprehensive; each type of characterization is actually a family of methods and there may be an effective method of a particular type even when general considerations seem to rule out that type. Undoubtedly more methods will be discovered as time goes on, leading to a broadening of these rules, and perhaps even a useful splitting of the characterization types into subtypes.

Important considerations when deciding on the characterization method to be used are measurement precision and monitor stability. There is no point in using a characterization method more precise than measurement precision, or more precise than the stability of the CRT over the range of conditions in which the characterization is to be used.

The characterization methods described below make use of a variety of measurements, but measurement methodology is not discussed. For reference, the methods used are (1) spectroradiometry, measurement of spectral power distributions; (2) colorimetry, measurement of tristimulus values; and (3) photometry, measurement of luminance.

The phrase “monitor coordinates” is used throughout this section. It denotes a set of controllable CRT inputs to which the experimenter has immediate access. For example, in applications where the CRT is computer controlled, it indicates the *RGB* values in the color lookup tables of the frame buffer that supplies the video signal to the CRT. The characterization techniques discussed below are independent of the monitor coordinates used. The single exception is the set of model-dependent characterization methods.

## Exhaustive Characterization Methods

These are the most costly characterization methods in terms of time and computer storage. They are also the most precise and the most general. Their utility depends only on the stability of the monitor. (If the monitor doesn’t have enough stability for exhaustive methods to be usable it is unlikely to have enough stability to use any other method.)

**General Description** The idea behind exhaustive characterization is to measure all the colors a monitor can produce and store them in a large table. When a color of given monitor coordinates (*RGB*) is displayed, its tristimulus values are determined by looking them up in the table. When a color of given tristimulus coordinates is desired, the table is searched and that set of monitor coordinates closest to the desired value is chosen for display. (Dithering among nearby colors is also possible if more precision is desired.) This method is also useful for any kind of device, not just for monitors. Thus, software designed for monitors can be reused when, for example, printers must be characterized. The obvious drawback to this method is the number of measurements that must be made. For example, with 24-bit color (8 bits per gun), over 16 million colors can be produced. If each measurement takes 1 s, the measurement process consumes over 4500 h, or almost 200 days, measuring around the clock. The solution to this problem is to sample the set of realizable colors, measure the samples, then interpolate between them. Thus, what we call exhaustive methods are only relatively exhaustive. Handling the practical problems is discussed in the following sections.

**Sampling Densities and Interpolation Algorithms** How many measurements should be made, and which colors should be selected for measurement? The answer depends on the nature of the sampling algorithm to be used, and is an open research question for general sampling. Thus, while it is possible to discuss the issues, practical decisions depend on the experience of the user.

**Sampling procedures** Most sampling algorithms sample the monitor coordinates linearly in the monitor coordinates. For example, if  $512 = 2^9$  samples are to be taken for a color CRT,  $8 = 2^3$  values on each of the red, green, and blue guns would be used, linearly interpolating the range of output voltages. The 512 samples would be the cartesian product of each set of gun values. Thus, if there were 256 possible values for each gun, running from 0 to 255, the 8 values chosen would be 0, 36, 72, 109, 145, 182, 218, and 255. If, as is often the case, saturation exists at the low and/or high ends of the input range, the full range is not used, and these values are scaled to the usable range. Behind this choice is the idea that a well-designed output device should have device coordinates that are close

to perceptually uniform. If so, linear sampling in device coordinates roughly approximates even sampling in a perceptually uniform space. The density of the sampling is related to the interpolation algorithm, which must approximate the exact value to within the desired precision of the characterization. This necessarily entails a trade-off between measurement time and online computational complexity when the characterization is being used: the more dense the sampling, the simpler the interpolation algorithm, and vice versa.

*Interpolation* Most often, linear interpolation is used. In this case, since we are interpolating in a cubic volume the interpolation is trilinear. Thus, given  $R_j$  (RGB), what are the tristimulus values?

1. We assume the existence of a table  $X_i(R_j(n_j))$ , consisting of the three tristimulus values  $X_i$  measured when the three guns are driven by the three voltages  $R_j(n_j)$ . Each three-vector  $n_j$  labels a different sample.
2. Find the sample  $m_j$  that has  $R_j(m_j)$  less than and closest to the values to be estimated  $R_j$ .
3. Take the eight samples  $(m_0, m_1, m_2)$ ,  $(m_0 + 1, m_1, m_2)$ ,  $(m_0, m_1 + 1, m_2)$ , . . . , and  $(m_0 + 1, m_1 + 1, m_2 + 1)$  as the vertices of a polyhedron.
4. Interpolate in  $R_0$  on the four sides running from  $(m_0, m_1, m_2)$  to  $(m_0 + 1, m_1, m_2)$ , from  $(m_0, m_1 + 1, m_2)$  to  $(m_0 + 1, m_1 + 1, m_2)$ , from  $(m_0, m_1, m_2 + 1)$  to  $(m_0 + 1, m_1, m_2 + 1)$ , and from  $(m_0, m_1 + 1, m_2 + 1)$  to  $(m_0 + 1, m_1 + 1, m_2 + 1)$ . The interpolation algorithm for  $X_i(*, l_1, l_2)$  is given by

$$X_i(*, l_1, l_2) = \frac{R_0 - R_0(m_0)}{R_0(m_0 + 1) - R_0(m_0)} X_i(R_0(m_0 + 1), R_1(l_1), R_2(l_2)) \\ + \frac{R_0(m_0 + 1) - R_0}{R_0(m_0 + 1) - R_0(m_0)} X_i(R_0(m_0), R_1(l_1), R_2(l_2))$$

where  $l_j$  is either  $m_j$  or  $m_j + 1$ .

5. Treat the four values as the corners of a polygon. Interpolate in  $R_1$  along the two sides running from  $(m_1, m_2)$  to  $(m_1 + 1, m_2)$  and from  $(m_1, m_2 + 1)$  to  $(m_1 + 1, m_2 + 1)$ . The interpolation algorithm for  $X_i(*, *, l_2)$  is given by

$$X_i(*, *, l_2) = \frac{R_1 - R_1(m_1)}{R_1(m_1 + 1) - R_1(m_1)} X_i(*, R_1(m_1 + 1), R_2(l_2)) \\ + \frac{R_1(m_1 + 1) - R_1}{R_1(m_1 + 1) - R_1(m_1)} X_i(*, R_1(m_1), R_2(l_2))$$

6. Treat these two values as the endpoints of a line segment, and interpolate in  $R_2$ . The final value is given by

$$X_i = \frac{R_2 - R_2(m_2)}{R_2(m_2 + 1) - R_2(m_2)} X_i(*, *, R_2(m_2 + 1)) \\ + \frac{R_2(m_2 + 1) - R_2}{R_2(m_2 + 1) - R_2(m_2)} X_i(*, *, R_2(m_2))$$

The above equations implement trilinear interpolation within a (possibly distorted) cube. It has recently been observed that tetrahedral interpolation has some desirable properties that are lacking in trilinear interpolation. It is accomplished by subdividing the cube into five or six tetrahedra, the corners of which coincide with the corners of the cube. Barycentric coordinates are then used to determine the tetrahedron in which the displayed color lies, and to interpolate within that

tetrahedron. More complex interpolation methods are also possible for nonuniform measurement sampling.<sup>12</sup>

Such interpolation methods implement the characterization directly in terms of the measured values. To test the adequacy of the interpolation it is important to choose regions of color space where curvature of the input/output response is expected to be high and to test the interpolation against exact measurements in that region of color space. If the interpolated values do not match the measured values there are two possible solutions: increase the measurement sampling density or improve the interpolation function. The first is expensive in terms of measurement time and computer memory; the second in terms of online calculation time. (There is a middle way, where a complicated algorithm is used to interpolate among coarsely sampled measurements for the purpose of generating a table large enough that linear interpolation can be done online. This has been used for hardcopy devices but not for monitors.) If the interpolation function is to be generalized it is possible to use higher-order interpolating functions, like splines, or to linearize the measurement space by transforming it before interpolating. For example, it is possible to build a characterization function that provides the logarithms of tristimulus values in terms of the logarithms of input coordinates. Any reasonable powerful generalization is bound to be computationally expensive. Much remains to be done to improve the sampling and interpolation process.

**Inverses** Calculating inverses for multidimensional tabular functions is not straightforward. Speed of computation can be optimized by creating an inverse table, giving values of  $R_i$  for regularly sampled values of  $X_j$ . The easiest way to construct this table is not by measurement, but by calculation (off-line) from the characterization table. Once the table is available, linear interpolation can be done online to provide voltages corresponding to given tristimulus values. Two possible methods exist for calculating inverses from tables.

1. Newton's method of finding zeros using derivatives to calculate an error function is possible, but can have bad numerical instabilities on tabular data that contains measurement error.
2. Examine the table that drives the forward transform to find the cell that contains the desired tristimulus values. Then subdivide the cell, using trilinear interpolation to determine new values for the corners. Determine which subcell contains the desired tristimulus values, and continue subdividing until the solution is sufficiently precise. This method is robust, but too expensive computationally to use online.

A promising alternative, not yet in wide use, takes advantage of barycentric coordinates in a tetrahedral lattice to find the appropriate cell quickly.

Nonlinear interpolation schemes can be used in the forward transformation when producing tables for linear inverse mappings. Presumably tables for nonlinear interpolation schemes in the inverse mapping can also be produced.

Three issues about inverses stand out.

1. Nonlinear schemes have not been sufficiently explored for their potential to be clear.
2. The straightforward application of well-understood computational methodology can provide substantial improvements over present methods.
3. Whatever inversion method is employed, out-of-gamut colors remain a serious unsolved problem.

**Out-of-Gamut Colors** What should be done when a program is asked to display a set of tristimulus values that are outside the color gamut of the monitor on which the color is to be displayed? The answer, of course, depends on the nature of the application. If the application demands drastic action when a request for an out-of-gamut color occurs, the solution is easy. The interpolation algorithm returns an illegal value of  $R_i$ . Then the application can display an error color in that area, or exit with an error. If, instead, the application must display a reasonable color within the monitor gamut, there is no natural solution. Solutions like projecting onto the surface of the monitor gamut have been used with success for some applications. Unfortunately, however, they are frequently computationally expensive and visually unsatisfactory.

## Local Characterization Methods

In many applications only a small region of the color gamut of the monitor must be characterized, but that region must be characterized very precisely. A typical example is a threshold experiment in which there are a small number of reference stimuli. Only colors close to the reference stimuli need be characterized precisely since they are the only colors likely to arise in the experiment. Methods that are specialized for the characterization of small parts of the color gamut, such as the threshold stimulus set, are the subject of this section. They are the methods most often appropriate for vision experimentation.

**General Description** Local characterization methods try to take advantage of simplifications that arise because the set of colors to be characterized is small. For example, a small set of colors all very close to a reference color can be characterized by a linear approximation to the global characterization function. Such simplifications offer many desirable properties such as high precision with minimal measurement (individual colors), linear characterizations and inverses (local regions of color), and simple inverses over extended color ranges (one-dimensional color spaces). To realize this precision, colorimetric measurements of the required precision must be available and easily usable. Detailed descriptions of three limited-gamut characterization schemes follow. Others are certainly possible, and may be worked out by analogy.

*Individual colors* It is often the case that a small number of distinct colors is needed. Under such circumstances the best method available is to perform a colorimetric measurement on each color. This is easy if the colors are arbitrary but need to be known precisely. Then they can be chosen by their *RGB* coordinates, with colorimetric measurements used to establish their tristimulus values. It is less easy if a small number of colors of specified tristimulus values are needed. The best way to deal with the latter problem is to use online colorimetric measurement, adjusting *RGB* values until the required tristimulus values are produced. This should be repeated several times throughout the time during which the characterization is needed to establish the stability of the monitor while the characterization is in use.

This type of characterization leads most directly into the most difficult question that CRT characterization must face: how should a characterization handle spatial and angular variations of light emitted by the CRT? Suppose, for example, an experiment uses only two color stimuli, a red one and a green one. Suppose further that the red one always appears at one point on the screen, while the green one always appears at a different point. Clearly, a characterization method that performs a colorimetric measurement on each stimulus yields maximum precision for minimum effort. Now, should the green color be measured in its particular location and the red one be measured in its location, or should both be measured in a common location, probably the screen center? And how should possible variations in the color of a stimulus when the other is turned off and on be handled? There is no universal best way of resolving such questions. Each must be resolved in a way that suits the particular display application. Note that this problem does not arise only for individual color characterizations. Indeed, it arises with any characterization whatsoever. The fact that every other aspect of individual color characterizations is so simple makes it particularly noticeable in this case.

*Local regions of color* Somewhat more complicated than individual color characterizations are ones in which it is necessary to characterize a small part of the color gamut surrounding a particular color. This might arise, for example, in a matching experiment, where all matches are close to a fixed reference stimulus. In such cases a small region of the color gamut must be characterized very precisely, with an emphasis on exact presentation of differences of color between the reference color and other colors in the region. The procedure is to make a precise colorimetric measurement of the reference color. Call the result  $\mathbf{X}_0$  with components  $X_{0p}$ , which is the result of measuring a color having input voltages  $R_{0p}$ . Then, for small changes in input voltage  $\Delta R_p$ , measure the corresponding changes in tristimulus values  $\Delta X_p$ . When the results are plotted they show a region in which the changes in tristimulus value are linearly related to changes in input coordinate with nonlinear effects growing in importance near the edge of the region. The size of the nonlinear effects, when compared

to the required prevision of the characterization, determines the size of the region that can be characterized linearly. Within this region the tristimulus values are given by

$$X_i = X_{0i} + \sum_{j=1}^3 M_{ij} \Delta R_j$$

The matrix  $M_{ij}$  is determined by a multilinear regression of the data, with each component of the tristimulus values regressed simultaneously against all three input coordinates. Nonlinear effects and interaction terms are not included in the regression. The matrix entry  $M_{ij}$ , which arises as a regression coefficient, is the change of  $X_i$  for a unit change in  $R_j$ . Thus, for example,  $M_{23}$  describes how much the  $Y$  tristimulus value varies with changes in the  $B$  gun of the monitor. This type of characterization is probably the most common in color research. It is most important to remember to determine the limits of such a linear characterization, or at least to determine that the limits are outside the region in which the characterization is to be used.

*One-dimensional color spaces* Often it is necessary to characterize a one-dimensional set of colors, that is, a line, not necessarily straight, in color space. Here is a method for doing so easily. Make a set of colorimetric measurements spaced more or less evenly along the curve. Use the variable  $\mu$  for the measurement number, numbering from one end of the line to the other. Plot the measured tristimulus values and the input voltages as functions of  $\mu$ . The measurement points should be dense enough that intermediate values can be approximated by linear interpolation. Now any set of  $RGB$  values must correspond to a value of  $\mu$ , not necessarily integral. How is this value determined?

1. Each measured  $\mu$  has a corresponding set of  $RGB$  values.
2. Find two consecutive sets such that the desired  $RGB$  values are intermediate between the  $RGB$  values of the sets. Call the  $\mu$  value of the sets  $\mu_0$  and  $\mu_0 + 1$ . Thus, mathematically,

$$R_i(\mu_0) \leq R_i \leq R_i(\mu_0 + 1) \quad \forall i$$

3. Now calculate “how far” between  $\mu_0$  and  $\mu_0 + 1$  the desired  $RGB$  lies. The value for gun  $j$  is  $\delta_j$  where

$$\delta_j = \frac{R_j - R_j(\mu_0)}{R_j(\mu_0 + 1) - R_j(\mu_0)}$$

4. If the three  $\delta$  values are close together this method works. Otherwise, the line of colors must be measured more densely.
5. Use  $\delta = (\delta_1 + \delta_2 + \delta_3)/3$  as the distance between  $\mu_0$  and the desired  $RGB$  value. This distance can then be used to interpolate in the tristimulus values.
6. The interpolated result is the characterized result.

$$X_i = X_i(\mu_0) + \delta(X_i(\mu_0 + 1) - X_i(\mu_0))$$

This method requires no special precautions if the line is relatively straight in  $RGB$  space. It is important to use the interpolation to calculate the tristimulus values of some measured colors to establish the precision of the measurement, or to decide if denser sampling is necessary.

*Inverses* The major reason for using this family of characterization methods is the computational simplicity of the characterizations and their inverses. Neither extensive memory nor expensive computation is needed to realize them.

*Individual colors* Since there is a small number of discrete colors, they can be arranged in a table. To get the characterization, you look up  $RGB$  in the table and read off the tristimulus values. To get the inverse, look up the tristimulus values in the output side of the table and the corresponding input gives  $RGB$  for the desired output. If the tristimulus values to be inverted are not in the table, then the color is not in the characterized set and it cannot be realized.

*Local regions of color* To derive the inverse for local regions of color, write the characterization equation as

$$X_i = X_{0i} = \Delta X_i = \sum_{j=1}^3 M_{ij} \Delta R_j$$

Then, inverting the matrix  $M_{ij}$  to get  $M_{ji}^{-1}$ ,

$$\Delta R_j = \sum_{i=1}^3 M_{ji}^{-1} \Delta X_i$$

which can be written explicitly as a solution for the RGB values needed to generate a color of given tristimulus values. That is,

$$R_j = R_{0j} + \sum_{i=1}^3 M_{ji}^{-1} (X_i - X_{0i})$$

It is important, after doing the above calculation, to check that the *RGB* values so determined lie within the region for which the characterization offers satisfactory precision. If not, however, unlike other methods, the inverse so determined is usually a reasonable approximation to the desired color.

Instead of inverting the matrix off-line, the three linear equations in  $R_j$  can be solved online.

*One-dimensional color spaces* For one-dimensional color sets the interpolation equations provide the inverse provided that *RGB* values and tristimulus values are reversed. To avoid confusion  $\epsilon$  is used in place of  $\delta$ .

1. To calculate  $\epsilon$  find a consecutive pair of tristimulus values that contain the desired values. Call the first of the pair  $\mu_0$ .
2. Then calculate  $\epsilon_i$  using

$$\epsilon_i = \frac{X_i - X_i(\mu_0)}{X_i(\mu_0 + 1) - X_i(\mu_0)}$$

If the three  $\epsilon_i$  values are not close together, then the tristimulus values designate a color that is not in the set, or the wrong interval has been found in a line of colors that must be very convoluted, and the set must be measured more densely.

3. Use  $\epsilon = (\epsilon_1 + \epsilon_2 + \epsilon_3)/3$  as the distance between  $\mu_0$  and the desired tristimulus values.
4. Then the *RGB* values corresponding to the desired tristimulus values are

$$R_j = R_j(\mu_0) + \epsilon(R_j(\mu_0 + 1) - R_j(\mu_0))$$

The results of this calculation should be checked to ensure that the *RGB* values do indeed lie in the color set. This is the second check that is performed, the first being a check for a small spread in the three  $\epsilon_i$  values. The checks are needed if the set is convoluted, and are always passed if the set is simple.

One way of checking whether the measurement density is adequate is to cycle an *RGB* value through the characterization and its inverse. If it fails to meet the required precision, then the measurement density is probably insufficient.

*Out-of-Gamut Colors* Local gamut characterization methods are attractive because of their easy-to-use inverses. Their treatment of out-of-gamut colors is also appealing.

*Individual colors* Colors not in the set of individual colors do not have inverses, and are not found in a search of tristimulus values. Because this characterization method ignores colors near to or between the measured colors, no method is possible for representing colors that are not in the



measured set. However, if the search is done with finite precision, colors close to a measured color are aliased onto it in the search. The precision of the search determines the region of colors that is aliased onto any measured color.

*Local regions of color* When a color is outside the local region to which the characterization applies, this method returns the input coordinates that would apply to a linear characterization. The values are close to those produced by an exact characterization, but outside the bounds of the specified precision. The error increases as the color gets farther away from the characterized region, but slowly in most cases. Thus, this method clearly indicates when a color is outside the characterized region and fails gracefully outside the region, giving values that are close, if not exact.

*One-dimensional color spaces* When a set of *RGB* values or tristimulus values lies outside the one-dimensional space, the fact is indicated during the calculation. The calculation can be carried through anyway; the result is a color in the space that is close to the given color. Exactly how close, and how the characterization defines the “closest” color in the one-dimensional space, depends on the details of how the measurement samples are selected, the curvature of the color set, and the appropriate experimental definition of “close.” If this method comes to have wider utility, investigation of sampling methods might be warranted to find the “best” sampling algorithms. In the meantime, it is possible to say that for reasonably straight sets colors close to the set are mapped onto colors in the set that are colorimetrically near the original color.

## Model-Dependent Characterization Methods

The two types of characterization methods described above are independent of any particular monitor properties. In fact, they can be used for any color-generating device. The third type of characterization makes use of a parametrized model of CRT color production. A characterization is created by measuring the parameters that are appropriate to the CRT being characterized. The next few sections describe the most common model for color CRTs, which is tried and true. Some CRTs may require it to be modified, but any modifications are likely to be small. The emphasis in this description is the set of assumptions on which the model is based, since violations of the assumptions require modifications to the model.

*General Description* The standard model of a color CRT has the following parts.

1. Any displayed color is the additive mixture of three component colors. The component colors are generally taken to be the light emitted by a single phosphor.
2. The spectral power distribution of the light in each component is determined by a single input signal, *R*, *G*, or *B*, and is independent of the other two signals.
3. The relative spectral power distribution of the light in each component is constant. Hence, the chromaticity of the component color is constant.
4. The intensity of the light in each component is a power function of the appropriate input voltage.

Taken together, these parts form a mathematical model of CRT colorimetry.

The standard model has been described many times. For a concise presentation see Cowan;<sup>13</sup> for a historical one see Tannenbaum.<sup>14</sup>

*Gun independence* The light emitted when the input coordinates are the three voltages  $v_R$ ,  $v_G$ , and  $v_B$  (generally called  $v_a$ ) is  $\Phi_\lambda(v_R, v_G, v_B)$ . It turns out to be convenient to form slightly different components than the standard model does, following not the physical description of how a monitor operates, but the logical description of what is done to realize a color on the monitor. The first step when generating a color is to create the *RGB* input from the separate *R*, *G*, and *B* inputs. Imagine turning on each input by itself, and assume that the color when all guns are turned on together is the



additive mixture of the colors produced when the guns are turned on individually. This assumption is called “gun independence.” In terms of tristimulus values it implies the condition

$$X_i = X_{Ri} + X_{Gi} + X_{Bi}$$

(Usually CRTs realize gun independence by exciting different phosphors independently,

$$\Phi_\lambda(\mathbf{v}_R, \mathbf{v}_G, \mathbf{v}_B) = \Phi_\lambda(\mathbf{v}_R, 0, 0) + \Phi_\lambda(0, \mathbf{v}_G, 0) + \Phi_\lambda(0, 0, \mathbf{v}_B)$$

This condition is, in fact, stronger than the gun-independence assumption, and only gun independence is needed for characterization.)

Gun independence was described by Cowan and Rowell,<sup>15</sup> along with a “shotgun” method for testing it. The tristimulus values of many colors were measured, then predictions based on the assumption of gun independence were tested. Specifically, gun independence implies consistency relationships that must hold within the set of measurements. Cowan and Rowell showed that a particular CRT had a certain level of gun independence, but didn’t attempt to make measurements of many CRTs. Clearly, it is worth making measurements on a variety of monitors to determine how widely gun independence occurs, and how large violations of it are when they do occur. For CRTs without gun independence there are two ways to take corrective action:

1. Use a characterization method that is not model-dependent.
2. Modify the monitor to create gun independence. For most CRTs there are good, inexpensive methods for improving gun independence. They will be more widely available in the future as characterization problems caused by failures of gun independence become more widely known.

Extreme settings of the monitor controls can cause violations of gun independence in otherwise sound monitors. The worst culprit is usually turning the brightness and/or contrast up so high that blooming appears at high input levels.

*Phosphor constancy* When gun independence holds, it is necessary to characterize only the colors that arise when a single gun is turned on, since the rest can be derived from them. Usually, it is assumed that the colors produced when a single gun is turned on have constant chromaticity. Thus, for example, the tristimulus values of colors produced by turning on the red gun alone take the form

$$X_{Ri}(\mathbf{v}_R) = E_R(\mathbf{v}_R) \cdot x_{Ri}$$

where  $x_{Ri}$ , the chromaticity coordinates of the emitted light,  $x_R$ ,  $y_R$ , and  $z_R$ , are independent of the voltage input. Thus the tristimulus values of the color produced depend on input voltage only through the intensity  $E_R(\mathbf{v}_R)$ .

The engineering that usually lies behind phosphor constancy is the following. A CRT should be designed so that the beam current in a given electron gun is independent of the beam current in any other gun. The deflection and shadowmask geometries should be designed and adjusted so that the electrons from any gun fall only on a single phosphor type. The physical properties of the phosphor should guarantee that the phosphor emits light of which the chromaticity is independent of the intensity. Meeting these conditions ensures phosphor constancy. It is possible, but unusual, to have phosphor constancy under other conditions, if it happens that effects cancel out appropriately.

The measurement of phosphor constancy is described by Cowan and Rowell.<sup>15</sup> Here is a short description of how to make the required measurements for one of the guns.

1. Turn the gun on at a variety of input voltages, and make a colorimetric measurement at each input voltage.
2. Calculate chromaticity coordinates for each measurement, which are independent of input voltage if phosphor constancy is present.

3. Constancy of chromaticity should also obtain in the presence of varying input to the other guns. To consider this case, turn the two other guns on, leaving the gun to be measured off.
4. Measure the tristimulus values for a baseline value.
5. Then turn the gun to be measured on to a variety of input voltages, making a colorimetric measurement at each.
6. Subtract the baseline tristimulus values from each measurement, then calculate the chromaticity coordinates. These should be constant, independent of the input to the measured gun, and independent of the input to the other two guns as well.

One of the most common monitor characterization problems is a measured lack of phosphor constancy. This is a serious problem, since the functions derived from the CRT model cannot then be easily inverted. Sometimes the measured lack of phosphor constancy is real, in which case there is no solution but to use a different type of characterization procedure. More often, the measurement is caused by poor CRT setup or viewing conditions. This occurs as follows, first for viewing conditions. Ambient light reflected from the screen of the monitor is added to the emitted light. This light is independent of input voltage, and adds equally to all colors. Imagine a phosphor of constant chromaticity  $(x, y)$ , at a variety of input voltages. The tristimulus values of light emitted from the screen are

$$X_i = X_{0i} + E(v) \cdot x_i$$

where  $X_{0i}$  are the tristimulus values of the ambient light reflected from the screen. Note that subtracting the tristimulus values measured when the input voltage is zero from each of the other measurements gives phosphor constancy. (Failure to do the subtraction gives phosphor chromaticities that tend toward white as the input voltage decreases.) This practice should be followed when ambient light is present. Psychophysical experiments usually go out of their way to exclude ambient light, so the above considerations are usually not a problem. But under more normal viewing conditions ambient light is always present.

When the black level/brightness control is set above its minimum there is a background level of light emitted from the screen even when the input voltage is zero. This light should be constant. Thus, it can be handled in exactly the same way as ambient light, and the two can be lumped together in a single normalizing measurement.

Now let's put together this gun independence and phosphor constancy. Gun independence means that the tristimulus values  $X_i(v_R, v_G, v_B)$  can be treated as the sum of three tristimulus values that depend on only a single gun:

$$X_i(v_R, v_G, v_B) = \sum_{a=R,G,B} X_{ai}(v_a)$$

Phosphor constancy means that the tristimulus values from a single gun, once the background has been subtracted away, have constant chromaticity:

$$X_i(v_R, v_G, v_B) = (X_0, Y_0, Z_0) + \sum_{a=R,G,B} E_a(v_a) \cdot x_{ai}$$

To complete the characterization a model is needed for the phosphor output intensity  $E_a(v_a)$ .

*Phosphor output models* Several different methods exist for modeling the phosphor output. The most general, and probably the best, is to build a table for each gun. To do so requires only photometric measurement, and the responsivity of the photometer need not be known. (This property of the model is very convenient. Colorimetric measurements are fairly easy to perform at light levels in the top 70 percent of the light levels produced by a CRT, but hard to perform for dark colors. In this model the top of the output range can be used when doing the colorimetric measurements required to obtain phosphor chromaticities. Then, photometric measurements can be performed easily using photodiodes to calibrate the output response at the low end of the CRT output.) Choose the largest

input voltage to be used, and call it  $v_{a\max}$ . Measure its light output with the photometer. Then measure the light output from a range of smaller values  $v_a$ . The ratio of the light at lower input to the light at maximum input is the relative excitation  $e_a(v_a)$ . Store the values in a table. When excitations for intermediate voltage values are needed, find them by linear interpolation.

$$e_a(v_a) = \mu e_a(v_{a1}) + (1 - \mu)e_a(v_{a2})$$

where the interpolation parameter  $\mu$  is

$$\mu = \frac{v_a - v_{a2}}{v_{a1} - v_{a2}}$$

The memory requirements of the table can be reduced by using more complicated functions, but at the price of complexity of computation in the model.

It is also possible to approximate the measurements using a parameterized function. The most commonly used form is the gamma correction equation

$$e_a(v_a) = e_{a\max}(v_a / v_{a\max})^{\gamma_a}$$

with two parameters  $\gamma_a$  and  $e_{a\max}$ . Here,  $v_{a\max}$  is the maximum input voltage,  $e_{a\max}$  is the maximum relative excitation, usually taken to be 1.0, and  $\gamma_a$  is called the gamma correction exponent. It is determined empirically by regressing the logarithm of the measured excitation against the logarithm of the input voltage.

*Normalization* The excitation that enters the characterization equation  $E_a(v_a)$  is the product of the voltage-dependent relative excitation  $e_a(v_a)$  and a normalization coefficient  $N_a$ :

$$E_a(v_a) = N_a \cdot e_a(v_a)$$

The normalization coefficients complete the characterization model.

*Summary equation* The characterization is summarized by the single equation

$$X_i = X_{0i} + \sum_{a=R,G,B} N_a \cdot e_a(v_a) \cdot x_{ai}$$

which provides the tristimulus values for any set of *RGB* input coordinates.

*Conditions for Use* The assumptions discussed above are the conditions under which the model developed here can be used successfully. The assumptions should be checked carefully before this characterization method is used. The departure of the CRT from the ideal CRT, defined by perfect adherence to these conditions, is a measure of the imprecision in the characterization procedure. Note that when the CRT fails to meet the conditions, which occurs most frequently near the edges of its gamut, the erroneous colors are usually plausible. Thus, the model can be in error by a surprisingly large amount and still produce satisfactory characterizations in noncritical applications.

*Partial models* It is fairly clear that there are partial models meeting some of the conditions, and that they provide useful approaches to characterization. For example, suppose that the phosphor chromaticities vary with input voltage. They can be stored in a table indexed by input voltage (exact values are produced by interpolation) and used in a variant of the characterization equation:

$$X_i = X_{0i} + \sum_{a=R,G,B} N_a \cdot e_a(v_a) \cdot x_{ai}(v_a)$$

Such a generalization works well for the transformation from input voltages to tristimulus values, but the inverse transformation is virtually unusable.

**Measurement of Parameters** The characterization equation requires the measurement of a set of parameters that varies from CRT to CRT. Thus, they must be measured for the CRT to be characterized. Different types of measurements—spectral, colorimetric, and photometric—are needed, as described below.

**Ambient light and black level  $X_{0i}$**  A colorimetric measurement of the screen with all input voltages set to zero produces this value. It should be measured under the exact conditions in which the characterization will be used.

**Phosphor chromaticities  $x_{ai}$**  A colorimetric measurement of the screen with a single gun turned on produces these values. Careful subtraction of the ambient light and black level is important. Measuring at a variety of input voltages produces:

1. A measure of phosphor constancy
2. The range of input voltages over which phosphor constancy holds well enough for use of the characterization model
3. A value for the phosphor chromaticities, produced by averaging over the chromaticities in the characterization range

**Gamma correction functions** Measurement of the relationship between the input voltages and the excitation functions requires photometric measurement at a large variety of input voltages. If interpolation into a table is the method of choice, the sampling of input voltages must be dense enough that the interpolation method chosen, usually linear, is close enough to the exact values to yield the desired accuracy. If a functional relationship, like the power function, is chosen the validity of the function chosen must be inspected very carefully. Whatever method is chosen it is essential that the ambient light/black level be subtracted from the measurements. If it is not, it is counted three times when the contributions from the three guns are added together.

**Normalization coefficients** The normalization coefficients are determined through the characterization equation

$$X_i = \sum_{a=R,G,B} N_a \cdot e_a(v_a) \cdot x_{ai}$$

They may be determined in several ways, falling into two categories: measurement methods and comparison methods. A typical measurement method assumes a single color of known tristimulus values. This knowledge can be produced by a colorimetric measurement or by varying *RGB* to match a known sample such as the reference field of a color comparator. The tristimulus values are substituted into the equation above, giving three equations for the three normalization coefficients. Solving the equations gives the normalization coefficients. Note that the relative excitations and the phosphor chromaticities are both unitless. Thus, the normalization coefficients carry whatever units are used to measure the tristimulus values. The units enter naturally when the linear equations are solved.

The method for solving these linear equations utilizes the inverse transform. A matrix  $\mathbf{M}$ , the entries of which are the phosphor chromaticities

$$\mathbf{M}_{ai} = x_{ai}$$

is defined, and its inverse,  $\mathbf{M}_{ia}^{-1}$  determined. Multiplying it into the characterization equation gives

$$\sum_{i=1}^3 \mathbf{M}_{ia}^{-1} X_i = N_a \cdot e_a(v_a)$$

Then  $N_a$  is calculated from

$$N_a = e_a(v_a) \cdot \sum_{i=1}^3 \mathbf{M}_{ia}^{-1} X_i$$

Note that  $e_a(v_a)$  is known since the input voltage that matches the standard, or of the color that has been measured, is known.

The second method is used when it is possible to match a known standard. For example, it might be possible to find voltages that make colors produced by individual guns equiluminous. Call the voltages  $v_{YR}$ ,  $v_{YG}$ , and  $v_{YB}$ , and substitute them into the  $Y = X_2$  characterization equation

$$N_R e_R(v_{YR}) x_{R2} = N_G e_G(v_{YG}) x_{G2} = N_B e_B(v_{YB}) x_{B2}$$

This equation is easily solved for the ratios  $N_G/N_R$  and  $N_B/N_R$ , leaving a single constant, an overall normalizing coefficient, undetermined. This constant, which carries the units in which the tristimulus values are measured, can be determined only by an absolute measurement (or, equivalently, a color match to an absolute standard). However, because vision is insensitive to variations in overall intensity, the precise value of the overall normalizing coefficient is unimportant in many applications.

**Inverse Transformations** One of the most attractive features of model-dependent characterizations is the simplicity of their inverses. It is simple to invert the characterization equation

$$X_i = \sum_{a=R,G,B} N_a \cdot e_a(v_a) \cdot x_{ai}$$

which determines the tristimulus values of a color in terms of the input voltages that cause it to be displayed, into the inverse equation, which specifies input voltages that create a color of given tristimulus values. Define the matrix of phosphor chromaticities

$$M_{ai} = x_{ai}$$

Then determine its inverse  $M_{ia}^{-1}$  by conventional means. It is used to solve the characterization function

$$\sum_{i=1}^3 M_{ia}^{-1} X_i = N_a \cdot e_a(v_a)$$

Then, after the normalization coefficients are divided out

$$e_a(v_a) = \frac{1}{N_a} \sum_{i=1}^3 M_{ia}^{-1} X_i$$

It is then necessary to find the input voltages that give the appropriate excitations  $e_a$ , which requires inversion of the excitation function. Thus,

$$v_a = e_a^{-1} \left( \frac{1}{N_a} \sum_{i=1}^3 M_{ia}^{-1} X_i \right)$$

If the relative excitation function is specified in a closed form, the inverse can be calculated using only elementary algebra. Otherwise, it is necessary to develop a tabular inverse, which is most easily done by one-dimensional interpolation.

*Violations of phosphor constancy* When phosphor constancy does not hold, the inverse cannot be calculated using the methods given above. Numerical inversion techniques, which are beyond the scope of this section, can be used, but it is probably a better idea to use a different calibration method.

**Out-of-Gamut Colors** Most inversion methods fail internally if they are asked to produce the input coordinates needed to display an out-of-gamut color. For example, tabular inverses fail to find an interval or cell in which to interpolate. Because the functions in model-dependent characterizations are well defined beyond the gamut, these methods generally do not fail, but calculate values that

lie outside the domain of input coordinates. Thus, input coordinate values calculated from inverses of model-dependent characterizations should be checked, because writing out-of-range values to color lookup tables produces unpredictable results.

## Absolute Characterization versus Characterization for Interaction

The characterization methods discussed above are designed to provide absolute characterizations, which are needed for color coordinates defined in terms of an external standard. Probably just as often characterization is needed to support specific interaction techniques. Suppose, for example, an experiment calls for the observer to have three CIELAB controls, one changing  $L^*$ , one changing  $a^*$ , and one changing  $b^*$ . To make this possible a characterization of the CRT in terms of CIE tristimulus values must be done to provide a basis for the transformation from tristimulus values to  $L^*a^*b^*$ . This type of characterization situation has several characteristics that are quite common:

1. At any given time there is a known color on the screen and the problem is to calculate a color slightly different from it. This situation arises because the control is sampled frequently enough that large color changes are produced only cumulatively over many color updates.
2. The new color coordinates must be calculated quickly and frequently.
3. Errors in the calculated color increments are errors in quantities that are already small. Thus, they can be considerably larger (in percentage terms) than would be tolerable for an absolute characterization.
4. The inverse transformation, tristimulus values to  $RGB$  is usually needed.

This combination of requirements is most easily met when the characterization has a linear inverse, and as little calculation as possible in the main loop. Exhaustive characterization methods have the necessary ingredients as they stand, provided an inverse table has been built, as do local methods. But nonlinear transformations, like the one from  $L^*a^*b^*$  to tristimulus values violate the conditions imposed above, and it is necessary to linearize them. They can then be combined with whatever characterization is in use to linearize the entire transformation from input coordinates to color specification.

**Linearization** Consider a transformation  $f$  that takes a vector  $\mathbf{x}$  into a vector  $\mathbf{y}$ . An example is the transformation between  $L^*a^*b^*$  and  $XYZ$ . Suppose two vectors that are known to transform into each other are known:  $\mathbf{x}_0$  ( $L_0^*a_0^*b_0^*$ ) corresponds to the set  $\mathbf{y}_0$  ( $X_0Y_0Z_0$ ). Now, if one vector undergoes a small change, how does the other change? If the transformation is suitably smooth (and the functions of color science are adequately smooth), small changes of one variable can be expressed in terms of small changes of the other by simple matrix multiplication

$$\Delta y_i = \sum_j \mathbf{M}_{ij} \Delta x_j$$

The entries in the matrix  $\mathbf{M}$  are the partial derivatives of  $f$  with respect to  $\mathbf{x}$ , evaluated at  $\mathbf{x}_0$ . That is,

$$\mathbf{M}_{ij} = \left. \frac{\partial f_i}{\partial x_j} \right|_{\mathbf{x}=\mathbf{x}_0}$$

Even though the computation of the derivatives may be expensive, they need to be calculated only infrequently, and the calculation can be scheduled when the system has idle resources available.

## Practical Comments

CRT characterization is essentially a practical matter. It is almost always done because a superordinate goal requires it. Thus, it is appropriate to end this chapter with several practical remarks that make characterization easier and less costly.

Do as little characterization as possible! Characterization is usually done within a fixed budget of time and equipment. Identifying the least characterization that provides all the information needed for the application allows the extra resources to be used to improve precision and to check measurements.

Characterize frequently! CRTs age and drift; colleagues turn knobs unaware that they change a critical characterization. Discovering that a CRT does not match its characterization function invalidates all data collected or experiments run since the characterization was last checked. There are only two defenses. First, measure the CRT as often as is feasible. Second, devise visual checks that can be run frequently and that will catch most characterization errors. It is possible to use luminance-matching techniques like minimally distinct border or minimum motion to check characterizations in a few seconds.

Understand the principles underlying the characterization methods and alter them to fit the specific stimuli needed for the application. In particular, CRTs have many ways of being unstable. The best defense is to characterize the CRT with stimuli that are as close as possible to those that are used in the application: same size, same position on the screen, measurement apparatus at the observer's eye point, same background color and intensity, and so on. Such adherence is the best defense against inadvertently finding a new type of CRT instability that invalidates months of work.

## 22.5 AN INTRODUCTION TO LIQUID CRYSTAL DISPLAYS

The purpose of this section is to provide an introduction to the operational principles that provide color output from liquid crystal displays (LCDs). No actual colorimetric measurements are shown, because LCDs are evolving too fast for them to be of any lasting interest. Instead, schematic illustrations are given, showing features that the author considers likely to be present in future displays.

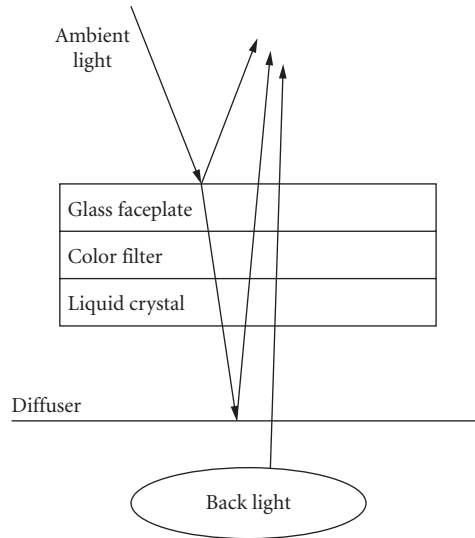
At the time of writing, colorimetric quality of LCDs is uneven. Nonetheless, since LCDs are semiconductor components and since they are now being produced in large quantities for low-cost computers, it is reasonable to expect a rapid decrease in cost combined with an increase in quality over the next decade. As the succeeding sections should make clear, they have a variety of interesting properties which will often make them preferred to CRTs as visual stimulators. Only by actively following their development will it be possible to use them at the earliest opportunity.

Two properties make liquid crystal devices particularly interesting to the visual scientist. First, they are usable as self-luminous, transmissive, or reflective media. They combine this flexibility with the range of temporal and spatial control possible on a CRT. This property is likely to make them particularly valuable for studies of color appearance and color constancy. Second, they have the ability to produce temporal and spatial edges that are much sharper than those producible on a CRT. This property has been called "acutance" in the display literature.<sup>16</sup> High acutance is potentially a serious problem for image display, since the gaussian blur that occurs at edges in CRT images helps to hide artifacts of the CRT raster. But high acutance also presents an opportunity since many small-scale irregularities that cannot be controlled on the CRT can be controlled on the LCD.

### Operational Principles of Monochrome LCDs

This section describes the operation of monochrome LCDs. The components of color LCDs are so similar to those of monochrome ones that they are most easily explained based on an understanding of monochrome LCDs.

**Overview** The important components of a monochrome LCD are shown schematically in Fig. 16. In the dark, the main source of light is the backlight. Its luminous flux passes through a diffuser. The uniformity of the backlight/diffuser pair determines the spatial uniformity of the display surface. Light from the diffuser then passes through a light-modulating element that consists of two crossed



**FIGURE 16** Illustration of the typical components of a single LCD pixel, seen side on. The arrows show the three light paths that contribute to the light emitted from the pixel.

polarizers and an intervening space, usually a few microns thick, filled with the liquid crystal. The liquid crystal rotates the polarization of light passing through it by an amount that depends on the local electric field. Thus, the electric field can be varied to vary the amount of light passing through the light-modulating element. The electric field is produced by a capacitor that covers the area of the pixel. The different varieties of LCD differ mainly in the electronics used to drive the capacitor. At present, the most common type is bilevel, with the capacitor controlled by a transistor that is either on or off, thus defining two light levels, white and black. More interesting are multilevel devices, in which an analog electric field allows more or less continuous variation of light output. Currently available analog devices are capable of 16 levels of gray scale with a pixel that is about 150 microns square.

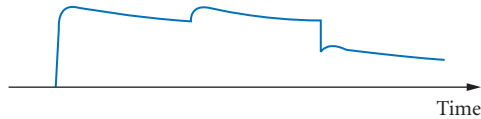
In bilevel devices the array of pixels is very similar to the layout of memory cells in a random access semiconductor memory. Thus, the display itself can be considered a write-only random-access frame buffer. Most LCDs do not, at present, make use of the random-access capability, but accept input via the RS-170 analog signal standard used for CRTs. The multilevel variant is similar, except that an analog value is stored at each location instead of 0 or 1.

An interesting feature of this design is the ability of ambient light to enter the display. It then passes through the light-modulating element and is scattered with random phase by the diffuser. From there it passes back through the light-modulating element to the viewer. Thus, under high-ambient conditions there are two contributions to light emitted from the display: one from the backlight that is linear with respect to the transmittance of the light-modulating element and one from ambient light that is quadratic with respect to the transmittance of the light-modulating element. More specifically, the contribution from the backlight,  $\Phi_{\lambda}^{(BL)}$  is given by

$$\Phi_{\lambda}^{(BL)} = \tau(V)\Phi_{0\lambda}^{(BL)}$$

where  $\Phi_{0\lambda}^{(BL)}$  is the light emitted from the display when the light-modulating element is in its maximally transmissive state and  $\tau(V)$  is the voltage-dependent transmittance of the light-modulating element. Note that the light-modulating process is assumed to be spectrally neutral. (In fact, chemists





**FIGURE 17** Illustration of the time dependence of the light output from an LCD pixel that is turned on for two refreshes, then turned to a lower level for the third refresh.

attempt to make the liquid crystal as spectrally neutral as possible.) The contribution from ambient light  $\Phi_{\lambda}^{(AMB)}$  is

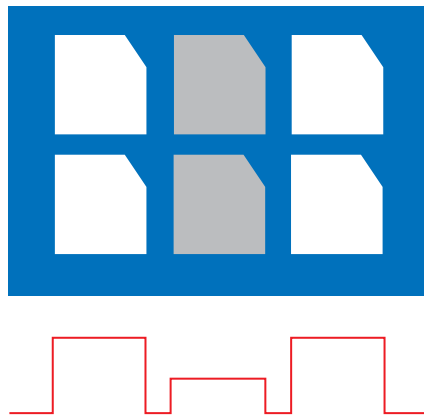
$$\Phi_{\lambda}^{(AMB)} = \tau^2(V)(1 - R(\lambda))D(\lambda)\Phi_{0\lambda}^{(AMB)}$$

where  $\Phi_{0\lambda}^{(AMB)}$  is the ambient light incident on the display,  $R(\lambda)$  is the reflectance of the faceplate of the display, and  $D(\lambda)$  is the proportion of light reflected back by the diffuser. The second contribution is particularly interesting because the display is essentially characterized by a variable reflectance  $\tau^2(V)(1 - R(\lambda))D(\lambda)$ . Thus it is possible to create reflective images that are computer-controlled pixel by pixel. Note the importance of the front surface reflection in this pixel model. Light that is not transmitted at the front surface produces glare; light that is transmitted creates the image. Consequently, it is doubly important that the faceplate be treated to minimize reflection.

**Pixel Structure** The most interesting characteristic of the LCD is the shape and layout of its pixels, which are unlike the CRT pixel both spatially and temporally.

The time dependence of the light output from an LCD pixel is shown schematically in Fig. 17. Each refresh of the pixel produces a sharp rise followed by a slow fall. If the light output decreases from one frame to the next, the decrease occurs abruptly at the beginning of the pixel, without the exponential tail characteristic of CRTs with long persistence phosphors. The turnoff is limited only by the resistance of the circuit that removes charge from the capacitor. The slow fall is produced by loss of charge from the capacitor: the circuit is essentially a sample and hold and the flatness of the pixel profile is determined by the quality of the hold.

The spatial structure of the LCD display surface is shown schematically in Fig. 18. The pixels are nonoverlapping, with dark lines between them because an interpixel mask is used to hide regions



**FIGURE 18** Illustration of the layout and shape of pixels on an LCD. The lower panel shows the spatial dependence of emitted light, on a cross section through the center of a row of pixels. The two middle pixels are turned on to a lower level than the outside ones.

where the electric fields of neighboring capacitors overlap. The light emitted is uniform over the pixel area, so that the intensity profile is uniform. Usually there is a small amount taken off the corner of the pixel to allow space for the electronic circuits that turn the pixel off and on. The pitch of the display, the distance from one pixel to the center of the next pixel, ranges from 100 to 500  $\mu\text{m}$ .

**Problems** Current high resolution, analog, color LCDs suffer from a variety of problems, which may be expected to diminish rapidly in importance as manufacturing technology improves.

*Thickness variation* Uniformity of output depends critically on precise control of the thickness of the liquid crystal from pixel to pixel.

*Heat* Current displays are very temperature-dependent, changing their performance as they warm up and if anything warm, like a human hand, is held in contact with them.

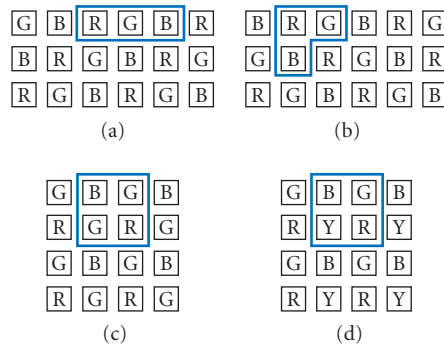
*Gray scale* The present mass-produced LCDs have poor gray-scale capability: each pixel can be only off or on. The manufacturing processes needed to produce gray scale, which produce about 30 levels of gray scale on prototype displays, need considerable development.

*Electronics* A typical display depends on thousands of address wires, connected to the display at the pitch of the pixels. These wires are delicate and subject to failure, which produces a row or column of inactive pixels, called a line-out.

## The Color LCD

The color LCD is a straightforward generalization of the monochrome LCD. Colored filters are placed in the light path to color the light emitted from each pixel. Three differently colored filters, and sometimes four, are used to create several interpenetrating colored images that combine additively to form a full-color image. The repetition pattern of the different images is regular, and it is usually possible to group a set of differently colored pixels into a single-colored pixel with three or four primaries, as shown in Fig. 19. (Terminology is important but nonstandard. In this chapter a single-colored pixel is called a monochrome pixel; a set of differently colored monochrome pixels that combines additively to produce a continuum of color at a given point is called a color pixel.)

**Geometry of the Color Pixel** There are a variety of different geometrical arrangements of monochrome pixels within a color pixel. Several are illustrated in Fig. 19. It is not currently known which



**FIGURE 19** Schematic diagram of a variety of different monochrome pixel geometries: (a) and (b) show two triad geometries; (c) and (d) show two quad geometries. The color pixel is outlined in gray.

arrangement is best for which type of visual information. When the viewer is far enough back from the display the arrangement does not matter, only the relative number of each primary color. One strategy that is being tried is to add a fourth color, white or yellow, in a quad arrangement in order to increase brightness. Such a display offers interesting possibilities for experimentation with mesopic stimuli and for diagnosis and investigation of color-vision defects.

It is obvious that every geometry has directions in which the primaries form unwanted color patterns. Such patterns do not arise in CRT imagery because there is a random correlation between the dot triads that produce color and the pixel locations. Whether there is a regular geometry, as yet untried, that removes the patterns, whether software techniques similar to antialiasing can eliminate their visual effects, or whether a randomized primary arrangement is needed to remove the patterns is a problem as yet unresolved. (It should also be remembered that some stimuli may best be created by graphical techniques that take advantage of the existence of the patterns.) Whatever the solution, however, it must address the interrelationship between spatial and color aspects of the stimulus in the visual system of the perceiver.

**Colorimetry of the Color Pixel** The dominant factor in the colorimetry of LCDs is the interaction between the spectral power distribution of the backlight and the transmittances of the filters. Typically backlight and filter design deals with a trade-off between brightness and color gamut. The higher the excitation purity of the filter the more saturated the primary but the less bright the display. An additional factor considered in this trade-off is heating of the liquid crystal, which is greater when the filters have high excitation purity. Choice of a backlight that has as much as possible of its spectral power in wavelengths passed by the filters will improve performance.

The colorimetric properties of an LCD can be summed up in a few equations. The first is that the color of a pixel is the additive mixture of the colors of its monochrome pixel components. The sum of the spectral powers is

$$\Phi_{\lambda} = \sum_{a=1}^{N_p} \Phi_{a\lambda}(V_a)$$

where  $\Phi_{a\lambda}$  is the spectral power emitted by monochrome pixel  $a$ , which depends on the voltage  $V_a$  applied to it. Similarly, the tristimulus values are the sum of the tristimulus values of the monochrome pixels

$$X_i = \sum_{a=1}^{N_p} X_{ai}$$

where  $X_{ai}$  is the set of tristimulus values for primary  $a$ . Note that “gun independence” is assumed.

The spectral power contributed by monochrome pixel  $a$  is the sum of three components: one from the backlight,  $\Phi_{a\lambda}^{(BL)}(V_a)$ , one from ambient light reflected from the front surface of the LCD,  $\Phi(R)_{a\lambda}$ , and one from ambient light reemitted from the display,  $\Phi_{a\lambda}^{(AMB)}(V_a)$ .

$$\Phi_{a\lambda}(V_a) = \Phi_{a\lambda}^{(BL)}(V_a) + \Phi_{a\lambda}^{(R)} + \Phi_{a\lambda}^{(AMB)}(V_a)$$

The contribution from the backlight is

$$\Phi_{a\lambda}^{(BL)}(V_a) = \tau_a(\lambda)\tau(V_a)\Phi_{0\lambda}^{(BL)}$$

where  $\Phi_{0\lambda}^{(BL)}$  is the spectral power distribution of the backlight,  $\tau(V_a)$  is the voltage-dependent transmittance of the liquid crystal/polarizer sandwich, and  $\tau_a(\lambda)$  is the transmittance of the color filter. The function  $\tau(V)$  is independent of the primary because the display uses the same light-modulating element for each primary. The function  $\tau_a(\lambda)$  is independent of the applied voltage since the light-modulating element is spectrally neutral. In the absence of ambient light, this term is the only contribution to the light emitted by the LCD. The colorimetric properties are then identical to those of a CRT with primaries of the same chromaticity, which is the chromaticity of the product  $\tau_a(\lambda)\Phi_{0\lambda}^{(BL)}$ , though the equivalent of the gamma function  $\tau(V_a)$  is certain to have a different functional form.

The contribution from light reflected from the front surface is

$$\Phi_{a\lambda}^{(R)} = R(\lambda)\Phi_{0\lambda}^{(AMB)}$$

where  $R(\lambda)$  is the reflectance of the front surface and  $\Phi_{0\lambda}^{(AMB)}$  is the ambient light incident on the surface. Usually, the reflectance is spectrally neutral (independent of wavelength) and the reflected light has the same color as the incident ambient light.

The contribution from light reemitted from the display is

$$\Phi_{a\lambda}^{(R)} = (1 - R(\lambda))D(\lambda)\tau_a^2(\lambda)\tau^2(V_a)\Phi_{0\lambda}^{(AMB)}$$

where  $D(\lambda)$  is the reflectance of the diffuser. This contribution can be modeled as a light of the same chromaticity as the spectral power distribution  $(1 - R(\lambda))D(\lambda)\tau_a^2(\lambda)\Phi_{0\lambda}^{(AMB)}$ , with its intensity modulated by the function  $\tau^2(V_a)$ . Note that this chromaticity is not in general the same as the chromaticity produced by light from the backlight. (It will usually be more saturated than the backlight component since it passes through the color filter twice.) Note also that the voltage dependence of the intensity of this light differs from that of the backlight component. Thus, the chromaticity of the primary changes as the voltage changes when ambient light is present, and the effect cannot be subtracted off, as can the ambient component.

**Controls and Input Standards** LCDs will have standard sets of controls and input signals only when the technology is much more mature than it is at present. Currently, the most common input seems to be analog video, as used for CRTs. While this allows a one-for-one substitution of an LCD for a CRT, it seems quite inappropriate for computer applications. Specifically, information is extracted from the frame buffer, a random-access digital device, laboriously sequenced into the serial analog video signal, the reextracted for presentation for the random-access LCD. Thus, it seems likely that digital random-access input standards are likely to supersede video signals for LCD input, and that the display will be more tightly coupled to the image storage system than is common with CRTs.

## Temporal and Spatial Variations in Output

LCD fabrication technology is changing too quickly for quantitative limits to spatial and temporal variability to be predictable. Nonetheless, it seems likely that certain qualitative properties are inherent in all LCD designs. They are discussed in general terms in the next few sections.

**Short-Time Temporal Variation** The light output over a short time (about 100 ms) is shown schematically in Fig. 17. The details shown in that figure, the rise and decay times and the turnoff time, are bound to change as fabrication techniques improve. Currently, the turnoff time is not very good for most LCDs, and images leave shadows that decay visibly for several seconds after they are removed. That defect is not inherent in LCD electronics and it should be expected to disappear as circuitry improves. The other possible defect is the ripple in the light output, which is nonetheless much smaller than the ripple of CRT light output. The size of the ripple is determined by the quality of the sample-and-hold circuit that maintains charge on the capacitor, and is likely to decrease as LCDs improve.

The interaction between ripple and turnoff time is quite different for an LCD than for a CRT. To decrease the ripple on a CRT it is necessary to increase the decay time of the phosphors; degradation of the turnoff time cannot be avoided. For an LCD, on the other hand, the turnoff is active, and independent of the ripple. Thus, it is possible to improve the ripple without degrading the turnoff. As a result, future LCDs are likely to be very useful for generating stimuli that require precise temporal control.

**Long-Time Temporal Variation** Even at the current state of development, the long-time stability of LCDs is quite good, with reports showing variations of about 2 percent on a time scale of several hours. This performance is comparable to good-quality CRTs and incandescent light sources.

The main contributor to instability is heat. LCDs are only stable once they are warmed up, and even small changes in cooling configuration can cause appreciable changes in light output. Thus, good stability of light output requires well-controlled temperature and a long warm-up time.

**Small-Scale Spatial Variation** Small-scale variation is determined by the spatial structure of the pixel, which is illustrated in Fig. 18. The important feature is the sharp edge of the pixel: unlike CRT pixels, there is no blending at the edges of adjacent pixels. This feature makes the creation of some stimulus characteristics easy—sharp vertical and horizontal edges, for example—and makes the creation of other stimulus parameters very difficult—rounded corners, for example. Many graphical techniques and image-quality criteria have evolved to take advantage of display characteristics that are peculiar to the CRT; there is likely to be a substantial investment in devising techniques that are well suited to the very different pixel profile of the LCD.

An interesting future possibility arises because of the similarity of LCD manufacturing technology to that of random-access memories. Consequently, it is likely that future LCDs will be limited in resolution not by pixel size but by input bandwidth. If so, it would be sensible to have logical pixels within the control system that consist of many physical pixels on the display, with enough processing power on the display itself to translate commands referring to logical pixels into drive signals for physical pixels. In fact, a physical pixel could even belong to more than one logical pixel. If such a development occurs, considerable control over the pixel profile will be possible, which may greatly extend the range of spatial variation that images can possess.

**Large-Scale Spatial Variation** Because LCDs have no large-scale structural features, like the beam deflection in the CRT, only manufacturing tolerances should induce spatial variation. The main source of manufacturing variability at the time of writing is the physical size of the pixels and the electronic components—capacitors, especially—that control them. Ideally, such variations would be random, so that spatial variation would add only gaussian noise to the displayed image.

Much more serious is angular variation of emitted light. The light-emitting element uses effects that have strong directional dependence, and it is sometimes even possible to find angles at which an LCD reverses contrast compared to perpendicular viewing. Although this effect is inherent in LC technology, it is reasonable to hope that display improvements will reduce the angular variability below its currently unacceptable level.

---

## 22.6 ACKNOWLEDGMENTS

The author would particularly like to express his gratitude to the late Gunter Wyszecki who provided impetus for his interest in this subject, and who provided intellectual and logistical support during the time in which the author carried out most of the work on which this chapter is based. A large number of other people have provided useful insights and technical assistance over the years, including, but not restricted to: John Beatty, Ian Bell, David Brainard, Pierre Jolicoeur, Nelson Rowell, Marueen Stone, and Brian Wandell. The author would also like to thank the editor, David Williams, for patience above and beyond the call of duty during the very difficult time when I was putting this material into the form in which it appears here.

---

## 22.7 REFERENCES

1. D. G. Fink, K. B. Benson, C. W. Rhodes, M. O. Felix, L. H. Hoke Jr, and G. M. Stamp, "Television and Facsimile Systems," *Electronic Engineers Handbook*, D. G. Fink and D. Christiansen (eds.), 3d ed., McGraw-Hill, New York, 1989, pp. 20-1–20-127.
2. Electronic Industries Association, *Electrical Performance Standards—Monochrome Television Studio Facilities*, RS-170, Washington, D.C., 1957.

3. Electronic Industries Association, *Electrical Performance Standards—Monochrome Television Studio Facilities*, RS-343, Washington, D.C., 1969.
4. N. P. Lyons and J. E. Farrell, "Linear Systems Analysis of CRT Displays," *Society for Information Display Annual Symposium: Digest of Technical Papers* **20**:220–223 (1989).
5. J. B. Mulligan and L. S. Stone, "Halftoning Methods for the Generation of Motion Stimuli," *Journal of the Optical Society of America* **A6**:1217–1227 (1989).
6. A. C. Naiman and W. Makous, "Spatial Non-Linearities of Grayscale CRT Pixels," *Proceedings of SPIE: Human Vision, Visual Processing, and Digital Display III* **1666**:41–56 (1992).
7. B. MacIntyre and W. B. Cowan, "A Practical Approach to Calculating Luminance Contrast on a CRT," *ACM Transaction on Graphics* **11**:336–347 (1992).
8. P. M. Tannenbaum, 1986 (personal communication).
9. D. H. Brainard, "Calibration of a Computer-Controlled Monitor," *Color Research and Application* **14**:23–34 (1989).
10. W. T. Wallace and G. R. Lockhead, "Brightness of Luminance Distributions with Gradient Changes," *Vision Research* **27**:1589–1602 (1987).
11. F. Benedikt, "A Tutorial on Color Monitor Alignment and Television Viewing Conditions," Development Report 6272, Canadian Broadcasting Corporation, Montreal, 1986.
12. I. E. Bell and W. Cowan, "Characterizing Printer Gamuts Using Tetrahedral Interpolation," *Color Imaging Conference: Transforms and Transportability of Color*, Scottsdale, Arizona, 1993.
13. W. B. Cowan, "An Inexpensive Scheme for Calibration of a Color Monitor in Terms of CIE Standard Coordinates," *Computer Graphics* **17**(3):315–321 (1983).
14. P. M. Tannenbaum, "The Colorimetry of Color Displays: 1950 to the Present," *Color Research and Application* **11**:S27–S28 (1986).
15. W. B. Cowan and N. L. Rowell, "On the Gun Independence and Phosphor Constancy of Color Video Monitors," *Color Research and Application* **11**:S34–S38 (1986).
16. L. M. Biberman, "Image Quality," *Perception of Displayed Information*, L. M. Biberman (ed.), Plenum, New York, 1973.

*This page intentionally left blank*

---

# VISION PROBLEMS AT COMPUTERS

---

## Jeffrey Anshel

*Corporate Vision Consulting  
Encinitas, California*

## James E. Sheedy

*College of Optometry  
Pacific University  
Forest Grove, Oregon*

---

### 23.1 GLOSSARY

---

- Accommodation.** In regard to the visual system, accommodation is the focusing ability of the eye.
- Acuity.** A measure of the ability of the eye to resolve fine detail, specifically to distinguish that two points separated in space are distinctly separate.
- Afterimage.** An optical illusion that refers to an image continuing to appear in one's vision after the exposure to the original image has ceased.
- Anisometropia.** A visual condition in which there is a significant refractive difference between the two eyes.
- Astigmatism.** A visual condition in which the light entering the eye is distorted such that it does not focus at one single point in space.
- Binocularity.** The use of two eyes at the same time, where the usable visual areas of each eye overlap to produce a three-dimensional perception.
- Brightness.** The subjective attribute of light to which humans assign a label between very dim and very bright (brilliant). Brightness is perceived, not measured. Brightness is what is perceived when lumens fall on the rods and cones of the eye's retina. The sensitivity of the eye decreases as the magnitude of the light increases, and the rods and cones are sensitive to the luminous energy per unit of time (power) impinging on them.
- Cataracts.** A loss of clarity of the crystalline lens within the eye which causes partial or total blindness.
- Cathode ray tube (CRT).** A glass tube that forms part of most video display terminals. The tube generates a stream of electrons that strike the phosphor coated display screen and cause light to be emitted. The light forms characters on the screen.
- Color convergence.** Alignment of the three electron beams in the CRT that generate the three primary screen colors—red, green, and blue—used to form images on screen. In a misconverged image, edges will have color fringes (e.g., a white area might have a blue fringe on one side).



**Color temperature.** A way of measuring color accuracy. Adjusting a monitor's color-temperature control, for example, may change a bluish white to a whiter white.

**Convergence.** That visual function of realigning the eyes to attend an object closer than optical infinity. The visual axes of the eyes continually point closer to each other as the object of viewing gets closer to the viewer.

**Contrast.** The difference in color and light between parts of an image.

**Diplopia (double vision).** That visual condition where the person experiences two distinct images while looking at one object. This results from the breakdown of the coordination skills of the person.

**Disability glare.** A type of glare that causes objects to appear to have lower contrast and is usually caused by scattering of light within the media in the eye.

**Discomfort glare.** Glare that produces ocular discomfort including eye fatigue, eyestrain, and irritation.

**Dot matrix.** A pattern of dots that forms characters (text) or constructs a display image (graphics) on the VDT screen.

**Dot pitch.** The distance between two phosphor dots of the same color on the screen.

**Electromagnetic radiation.** A form of energy resulting from electric and magnetic effects which travels as invisible waves.

**Ergonomics.** The study of the relationship between humans and their work. The goal of ergonomics is to increase worker's comfort, productivity, and safety.

**Eyesight.** The process of receiving light rays into the eyes and focusing them onto the retina for interpretation.

**Eyestrain (asthenopia).** Descriptive terms for symptoms of visual discomfort. Symptoms include burning, itching, tiredness, aching, watering, blurring, etc.

**Farsightedness (hyperopia).** A visual condition where objects at a distance are more easily focused as opposed to objects up close.

**Font.** A complete set of characters including typeface, style, and size used for screen or printer displays.

**Focal length.** The distance from the eye to the viewed object needed to obtain clear focus.

**Foot-candle.** The amount of illumination inside the surface of an imaginary 1 ft radius sphere would be receiving if there were a uniform point source of 1 cd in the exact center of the sphere. One foot-candle  $\approx$  10.764 lux.

**Glare.** The loss in visual performance or visibility, or the annoyance of discomfort, produced by a luminance in the visual field greater than the illuminance to which the eyes are adapted.

**Hertz (Hz).** Cycles per seconds. Used to express the refresh rate of video displays.

**Illuminance.** The luminous flux incident on a surface per unit area. The unit is the lux or lumen per square meter. The foot-candle (fc) or lumen per square foot is also used. An illuminance photometer measures the luminous flux per unit area at the surface being illuminated without regard to the direction from which the light approaches the sensor.

**Interlaced.** An interlaced monitor scans the odd lines of an image first followed by the even lines. This scanning method does not successfully eliminate flicker on computer screens.

**Lag.** In optometric terms, the measured difference between the viewed object and the actual focusing distance.

**LASIK.** A surgical procedure that alters the curvature of the cornea (front surface of the eye) which reduces the amount of nearsightedness or astigmatism.

**LCD (Liquid crystal display).** A display technology that relies on polarizing filters and liquid-crystal cells rather than phosphors illuminated by electron beams to produce an on-screen image. To control the intensity of the red, green, and blue dots that comprise pixels, an LCD's control circuitry applies varying charges to the liquid-crystal cells through which polarized light passes on its way to the screen.

**Light.** The radiant energy that is capable of exciting the retina and producing a visual sensation. The visible wavelengths of the electromagnetic spectrum extend from about 380 to 770 nm. The unit of light energy is the lumen.

**Luminous flux.** The visible power or light energy per unit of time. It is measured in lumens. Since light is visible energy, the lumen refers only to visible power.

**Luminous intensity.** The luminous flux per solid angle emitted or reflected from a point. The unit of measure is the lumen per steradian, or candela (cd). (The steradian is the unit of measurement of a solid angle.)

**Luminance.** The luminous intensity per unit area projected in a given direction. The unit is the candela per square meter, which is still sometimes called a nit. The foot-lambert (fL) is also in common use. Luminance is the measurable quantity which most closely corresponds to brightness.

**Lux (see foot-candle).** A unit of illuminance and luminous emittance. It is used in photometry as a measure of the intensity of light.

**MHz (MegaHertz).** A measurement of frequency in millions of cycles per second.

**Myopia (nearsightedness).** The ability to see objects clearly only at a close distance.

**Musculoskeletal.** Relating to the muscles and skeleton of the human body.

**Nearpoint.** The nearest point of viewing, usually within arms length.

**Noninterlaced.** A noninterlaced monitor scans the lines of an image sequentially, from top to bottom. This method provides less visible flicker than interlaced scanning.

**Ocular motility.** Relating to the movement abilities of the eyes.

**Parabolic louver.** A type of light fixture that is designed to direct light in a limited and narrowed direction.

**Perception.** The understanding of sensory input (vision, hearing, touch, etc.).

**Pixel.** The smallest element of a display screen that can be independently assigned color and intensity.

**Phosphor.** A substance that emits light when stimulated by electrons.

**Photophobia.** A visual condition of excessive sensitivity to light.

**Presbyopia.** A reduction in the ability to focus on near objects caused by the decreased flexibility in the lens, usually noticed around the age of 40 years old or later.

**Polarity.** The arrangement of the light and dark images on the screen. Normal polarity has light characters against a dark background; reverse polarity has dark characters against a light background.

**Refractive.** Having to do with the bending of light rays, usually in producing a sharp optical image.

**Refresh rate.** The number of times per second that the screen phosphors must be painted to maintain proper character display.

**Resolution.** The number of pixels, horizontally and vertically, that make up a screen image. The higher the resolution, the more detailed the image.

**Resting point of accommodation (RPA).** The point in space where the eyes naturally focus when at rest.

**Specular reflection.** The perfect, mirror-like reflection of light from a surface, in which light from a single incoming direction is reflected into a single outgoing direction.

**Suppression.** The “turning off” of the image of one eye by the brain, most often to avoid double vision or reduce excess stress.

**SVGA (Super video graphics array).** A video adapter capable of higher resolution pixels and/or colors than the  $320 \times 200 \times 256$  and  $640 \times 480 \times 16$  which IBM’s VGA adapter is capable of producing. SVGA enables video adapters to support resolutions of 1024 by 768 pixels and higher with up to 16.7 million simultaneous colors (known as true color).

**VDT (Video display terminal).** An electronic device consisting of a monitor unit (e.g., cathode ray tube) with which to view input into a computer.

**Vision.** A learned awareness and perception of visual experiences (combined with any or all other senses) that results in mental or physical action. Not simply eyesight.

**Visual stress.** The inability of a person to visually process light information in a comfortable, efficient manner.

**Vision therapy.** Treatment (by behavioral optometrists) used to develop and enhance visual abilities.

## 23.2 INTRODUCTION

---

Vision and eye problems are the most commonly reported symptoms among workers at computer monitors. The percentage of computer workers who experience visually related symptoms is generally reported to be from 70 up to almost 90 percent.<sup>1</sup> The current concern for, and popularization of these problems is likely related to the rapid introduction of this new technology into the workplace, the job restructuring associated with it, and a greater societal concern about discomfort associated with work.

The problems are largely symptoms of discomfort. There is little evidence of permanent physiological change or damage associated with extended work at a computer monitor. The vision and eye problems caused by working at a computer monitor have been collectively named “Computer Vision Syndrome (CVS)” by the American Optometric Association. The most common symptoms of CVS obtained from a survey of optometrists in order of frequency are: eyestrain, headaches, blurred vision, dry or irritated eyes, neck and/or backaches, photophobia, double vision, and colored afterimages.

A syndrome is defined as a “group of signs and symptoms that indicate a disease or disease process.”<sup>2</sup> While CVS is not technically a “disease,” it can also be properly characterized as a syndrome due to the fact that it is “a complex of symptoms indicating the existence of undesirable condition or quality.” Thus, CVS is not a single disease entity, as the definition seems to require. However, it is a group of signs and symptoms related to a specific occupational usage of the eyes.

The causes for these symptoms are a combination of individual vision problems, poor office ergonomics, or work habits. Many individuals have marginal vision disorders that do not cause symptoms of less demanding visual tasks. On the other hand, there are numerous aspects of the computer monitor and the computer work environment that make it a more demanding visual task than others—therefore, more individuals are put beyond their threshold for experiencing symptoms.

## 23.3 WORK ENVIRONMENT

---

### Lighting

Improper lighting is likely the largest environmental factor contributing to visual discomfort. The room lighting is a particular problem for computer workers because the horizontal gaze angle of computer work exposes the eyes to numerous glare sources such as overhead lights and windows. Other deskwork is performed with a downward gaze angle and the glare sources are not in the field of view.

Glare is the loss in visual performance or visibility, or the annoyance of discomfort, produced by a luminance in the visual field greater than the illuminance to which the eyes are adapted. There are generally four types of glare: distracting glare, discomfort glare, disabling glare, and blinding glare.

Distracting glare results from light being reflected from the surface of an optical medium and is usually below 3000 lumens. This form of glare most often results in an annoyance to the viewer and leads to eye fatigue. Discomfort glare ranges from 3000 to 10,000 lumens and produces ocular discomfort, including eye fatigue, eyestrain, and irritation. Disabling glare, also known as veiling glare, results when light reaches 10,000 lumens and can actually block visual tasks. This type of glare causes objects to appear to have lower contrast and is caused by scattering of light within the media in the

**TABLE 1** Display and Glare Source Luminances

Visual Object	Luminance (cd/m <sup>2</sup> )
Dark background display	20–25
Light background display	80–120
Reference material with 750 lumens/m <sup>2</sup>	200
Reference material with auxiliary light	400
Blue sky (window)	2,500
Concrete in sun (window)	6,000–1,2000
Fluorescent light (poor design)	1,000–5,000
Auxiliary lamp (direct)	1,500–10,000

eye. Blinding glare results from incident light reflecting from smooth shiny surfaces such as water and snow. It can block vision to the extent that the wearer becomes visually compromised and recovery time is needed to be fully comfortable once again.

The threshold luminance ratios and locations of visual stimuli that cause glare discomfort have been determined,<sup>3</sup> but the physiological basis for glare discomfort is not known. Since large luminance disparities in the field of view can cause glare discomfort, it is best to have a visual environment in which luminances are relatively equal. Primarily because of glare discomfort, the Illumination Engineering Society of North America (IESNA) established maximum luminance ratios that should not be exceeded.<sup>4</sup> The luminance ratio should not exceed 1:3 or 3:1 between the task and visual surrounding within 25°, nor should the ratio exceed 1:10 or 10:1 between the task and more remote visual surroundings. Many objects in the field of view can cause luminance ratios in excess of those recommended by IESNA. Table 1 shows common luminance levels of objects in an office environment. Relative to the display luminance, several objects can greatly exceed the IESNA recommended ratios. A major advantage of using light background displays compared to dark background displays is that it enables better conformity with office luminance levels.

Good lighting design can reduce discomfort glare. Light leaving the fixture can be directed so that it goes straight down and not into the eyes of the room occupants. This is most commonly accomplished with parabolic louvers in the fixture. A better solution is indirect lighting in which the light is bounced off the ceiling—resulting in a large low luminance source of light for the room. Traditional office lighting recommendations have been suggested at about 100 fc (1000 lux) but computerized offices most often require less light, in the range of 50 fc (500 lux) to better balance with the light emanating from the computer display. Proper treatment of light from windows is also important. Shades or blinds should be employed to give flexibility to control outdoor light.

## Screen Reflections

In most cases, the most bothersome source of cathode ray tube (CRT) screen reflections is from the phosphor on the inner surface of the glass. (“Monitor Characteristics” are discussed later.) The phosphor is the material that emits light when struck by the electron beam—it also passively reflects room light. The reflections coming from the phosphor are diffuse. Most computer monitors have a frosted glass surface; therefore, the light reflected from the glass is primarily diffuse with a specular component.

Since most of the screen reflections are diffuse, a portion of any light impinging on the screen, regardless of the direction from which it comes, is reflected into the eyes of the user and causes the screen to appear brighter. This means that black is no longer black, but a shade of gray. The diffuse reflections from the phosphor and from the glass reduce the contrast of the text presented on the screen. Instead of viewing black characters on a white background, gray characters are presented on a white background. Calculations<sup>5</sup> show that these reflections significantly reduce contrast—from 0.96 to 0.53 [contrast =  $(L_t - L_b)/(L_t + L_b)$ , where  $L_t$  and  $L_b$  are luminances of the task and background] under common condition. Decreased contrast increases demand upon the visual system.

A common method of treating reflections is with an antireflection filter placed on the computer monitor. The primary purpose of the filter is to make the blacks blacker, thereby increasing contrast. The luminance of light that is emitted by the screen (the desired image) is decreased by the transmittance factor of the glass filter, whereas light that is reflected from the computer screen (undesired light) is decreased by the square of the filter transmittance, since it must pass through it twice. For a typical filter of 30 percent transmittance, the luminances of black and white are reduced respectively to 9 percent and 30 percent of their values without the filter. This results in an increase in the contrast.

Some antireflection filters have a circular polarized element in them. This feature results in circular polarized lights being transmitted to the computer screen. The resulting reflection on the screen changes the rotation of the polarization of the light so that it is blocked from coming back out through the filter. Since only specular reflected light maintains its polarized properties after reflection, this polarizing feature provides added benefit for only the specular reflections. If significant specular reflections are present, it is beneficial to obtain an antireflection filter with circular polarization.

The influx of liquid crystal displays (LCDs) into the workplace has addressed some of the issues regarding reflections by using a matte surface to diffuse the reflections. However, the matte surface may be disappearing in the future; the trend in high performance LCDs is back to a glossy surface, reintroducing the problem of reflections on the surface of the display. Whichever surface, matte or glossy, users can still experience problems with glare and contrast reduction or image washout. A 2003 survey<sup>6</sup> of LCD users supports this, with 39 percent reported having a glare problem on the LCD and 85 percent of LCD users reacted favorably when an antireflection filter was used with their LCD, stating that they were bothered by glare on their display and preferred working with a glare reduction filter on the display.

Independent, scientific testing of a midtransmission level antireflection computer filter to the same international standard that computer monitors must comply with, ISO 9241-7, has shown that this filter can actually improve a monitor's performance against this standard for reflection reduction and contrast improvement. The significance of this testing is that the quality of the antireflection coatings and the level of absorption technology are important considerations. There are many products on the market today that claim to be antireflection computer filters. These filters offer very low quality, if any, antireflection performance and little to no absorption technology. It is because of these lower performance products that ergonomic specialists, when considering options for reducing reflections and glare on an electronic display, often ridicule antireflection computer filters.

It is important to discuss the counterargument to antireflection computer filters—that monitors today don't need them because they already have antireflection treatments. While this is the case for many computer displays, the amount of glare and reflection reduction can be misleading. A few key points to keep in mind regarding computer displays and antireflection treatments:

- Some simply change the state of the reflection from specular to diffuse through silica coatings or etching the surface of the display (a matte finish).
- Some use a spin coating process that may only reduce first surface reflections down about 1 to 2 percent. Good quality antireflection computer filters reduce first surface reflections to less than 1 percent.
- Flat screens are better, reduce off-axis angle reflections, but still do little for normal incidence reflections.
- Few have absorptive coatings to improve contrast.
- LCDs have many advantages over the older CRT technology, matte surfaces to reduce reflections, but they still have issues with glare and contrast reduction.

Moreover, the trend of going to glossy displays means increased reflections.

## Monitor Characteristics

One obvious method of improving the visual environment for computer workers is to improve the legibility of the display at which they are working. The visual image on computer monitors is compromised in many ways compared to most paper tasks.

A major aspect of computer displays is that the images are composed of pixels—the “picture element” that creates the “dots” of an image. A greater pixel density will be able to display greater detail. Most older computer screens had pixel densities in the range of from 80 to 110 dots/in (dpi: 31 to 43 dots/cm). This pixel density can be compared to impact printers (60 to 100 dpi), laser printers (300 dpi), and advanced laser printers (400 to 1200 dpi). Computer monitors today have dpi specifications that are comparable to laser printers—greater pixel density would result in visually discernible differences, as for printers. Besides pixel density there are many other factors that affect display quality, such as pixel definition (measured by area under the modulation transfer function), font type, font size, letter spacing, line spacing, stroke width, contrast, color, gray scale, and refresh rate.<sup>7</sup> Several studies have shown that increased screen resolution characteristics increase reading speed and/or decrease visual symptoms.<sup>5</sup> Although technological advances have resulted in improved screen resolutions, there is still substantial room for improvement.

To adequately cover all aspects of computer monitor characteristics, it is necessary to discuss both the CRT and the LCD technologies used to generate these images.

The CRT is a light-emitting device. It was first demonstrated over 100 years ago and it is a fairly simple device to describe, but one requiring great precision to manufacture. It consists of a glass bottle under a high vacuum with a layer of phosphorescent material at one end and an electron gun at the other. The electron gun creates a stream of electrons that are accelerated toward the phosphor by a very high voltage. When the electrons strike the phosphor, it glows at the point of impact. A coil of wire called the yoke is wrapped around the neck of the CRT. The yoke actually consists of a horizontal coil and a vertical coil, each of which generates a magnetic field when a voltage is applied. These fields deflect the beam of electrons both horizontally and vertically, thereby illuminating the entire screen rather than a pinpoint. To create an image, the electron beam is modulated (turned on and off) by the signal from the video card as it sweeps across the phosphor from left to right. When it reaches the right side of the tube, the beam is turned off or “blanked” and moved down one line and back to the left side. This process occurs repeatedly until the beam reaches the bottom right-hand side and is blanked and moved to the top left.

The LCD or “flat panel” as it is sometimes called, is a totally different way of generating a computer image. Rather than being light emissive as the CRT is, the LCD is light transmissive. Each dot on the screen acts like an electrically controlled shutter that either blocks or passes a very high intensity backlight (or backlights) that is always on at full brightness. Whereas the dots comprising the image of a CRT-based display vary in size depending on the displayed resolution (if you display VGA resolution, the entire screen is lit with  $640 \times 480 = 307,200$  dots while if you display UXGA, the screen is filled by  $1600 \times 1200 = 1,920,000$  dots), an LCD panel has a fixed number of pixels. Each pixel is made up of three subpixels, a red, a green, and a blue. Since the backlight is white, the colors are achieved by means of tiny red, green, and blue filters.

Although the properties of a liquid crystal material were discovered in the late 1800s, it was not until 1968 before the first LCD display was demonstrated. Simply, light from the backlight is polarized, passed through the liquid crystal material (which twists the light  $90^\circ$ ) and passed through a second polarizer that is aligned with the first one. This results in a structure that blocks the backlight. The liquid crystal material has a unique property, however, that causes its molecules to line up with an applied voltage. The  $90^\circ$  twist goes to 0 and the polarizer sandwich becomes transparent to the backlight. By applying more or less voltage, the light can be varied from 0 (black) to 100 percent (white). Modern displays can make the transition from full off to full on in 256 steps, which is 8 bits ( $2^8 = 256$ ). Each pixel can display 256 shades each of red, green, and blue for a total palette of  $256 \times 256 \times 256 = 16.7$  million colors.

It is interesting that many monitor manufacturers will use that 16.7 million-color number very prominently in their advertising. However, in this chapter we are considering the eyes and visual system. Just how many colors does the human visual system perceive? Well, as near as we can tell (and there are many estimates here), the human visual system perceives between 7 and 8 million colors (in the most generous estimates). So, here we have a display that produces 16.7 million colors, but we can only see about half of those! For this reason, high performance monochrome monitors are often measured in “just noticeable differences,” or JNDs. However, color combinations can be a visual issue. It is important to maintain high contrast (color contrast, as well) between the letters and the background of the display, regardless of which colors are used.

So, which is the better technology? This can be very subjective, but the LCD exhibits no linearity or pincushion (straight lines looking curved) distortion and no flicker, all of which can contribute to computer vision syndrome. Furthermore, the LCD is always perfectly focused as long as it is set to display at its native resolution (the fixed number of pixels that may be addressed by the video card regardless of the resolution being displayed). When displaying a nonnative resolution, the LCD invokes a digital processor called a scaler to add enough pixels to fill the screen without distorting the image. This is a very complex procedure and the final result is a slightly softened image that many people find tiring. A good rule of thumb is to run an LCD monitor at its native resolution unless there is a compelling reason not to.

The electron beam in a CRT display is constantly scanning or drawing the image over and over. When it hits a specific spot on the screen, light is emitted, but as soon as the beam moves on, the spot starts to fade. It is critical that the beam scans all the points on the screen and returns to the first point in a short enough time that the eye does not perceive the dimming. The number of times an image is “painted” on the CRT is called the refresh rate or vertical sync frequency and it must be set to a value  $\geq 75$  Hz to avoid flicker. Some people are hypersensitive to flicker and need an even higher refresh rate. The LCD display does not flicker by design.

## Workstation Arrangement

It is commonly stated that: “the eyes lead the body.” Since working at the computer is a visually intensive task, our body will do what is necessary to get the eyes in the most comfortable position—often at the expense of good posture and causing musculoskeletal ailments such as sore neck and back.

The most common distance at which people view printed material is 40 cm from the eyes. This is the distance at which eye doctors routinely perform near visual testing and for which most bifocal or multifocal glasses are designed. Most commonly, the computer screen is farther from the eyes—from 50 to 70 cm. This distance is largely dictated by other workstation factors, such as desk space and having room for a keyboard. The letter sizes on the screen are commensurately larger compared to printed materials to enable the longer viewing distance.

Because we are now using a vertical, rather than horizontal work surface, the height of the computer screen is a very important aspect of the workstation arrangement. A person can adapt to the vertical location of their task by changing their gaze angle (the elevation of the eyes in the orbit) and/or by changing the extension/flexion of the neck. Typically, users will alter their head position rather than eye position to adjust to a different viewing height. This can cause awkward posture and result in neck and/or backache. The eyes work best with a depression of from 10 to 20°;<sup>5</sup> therefore, this is the preferred viewing angle. As a practical guide, it is often recommended that the computer user set their monitor just below their straight-ahead gaze, with the top of the display tilted back about 10°.

Whatever is viewed most often during daily work should be placed straight in front of the worker when seated at the desk. This applies to the computer and/or the reference documents—whichever is viewed most frequently. Although this seems self-evident, many computer workers situate their work so that they are constantly looking off to one side. If using hard copy on a regular basis, the material should be placed so that it is close in proximity to the monitor. Whether it is below the monitor (an “in-line” setup) or adjacent to it, the closer proximity will allow easier eye movements between the screen and the hard copy. In addition, the hard copy should reside on the same side as the handedness of the user: right side for a right-handed person. This will allow the person to either write or manipulate the hard copy without stretching into an awkward posture.

## Work Habits

Since the computer has essentially taken over as the most essential piece of office equipment, office workers are spending an excessive number of hours viewing their display screens. To simplify a method to break up their viewing habits, one author has devised a method called “the 3 B’s”:



blink, breathe, and break. The rationale for blinking will be discussed in the next section on dry eyes. Breathing is an important aspect of computer use simply because we are sedentary for extended hours and our physical activity is reduced. This reduces blood flow and creates a condition for shallow breathing, which in turn tires the computer user making them lethargic and less productive.

Regarding breaks, this author recommends the “20/20/20” rule: Every 20 min, take just 20 s, and look 20 ft away. This short, more frequent break will allow the accommodative and convergence systems to relax and presumably regain some level of performance.

## 23.4 VISION AND EYE CONDITIONS

Many individuals have marginal vision disorders that do not cause symptoms on less demanding visual work, but will cause symptoms when the individual performs a demanding visual task. Given individual variation in visual systems and work environments, individual assessment is required to solve any given person’s symptoms. Following are the major categories of eye and vision disorders that cause symptoms among computer users.

### Dry Eyes

Computer users commonly experience symptoms related to dry eyes. These symptoms include irritated eyes, dry eyes, excessive tearing, burning eyes, itching eyes, and red eyes. Contact lens wearers also often experience problems with their contact lenses while working at a computer display—related to dry eyes. In response to dry eye and ocular irritation, reflex tearing sometimes occurs and floods the eyes with tears. This is the same reflex that causes tearing when we cry or are exposed to foreign particles in our eye.

Computer workers are at greater risk for experiencing dry eye because the blink rate is significantly decreased and, because of the higher gaze angle, the eyes are wide open with a large exposed ocular surface. Patel et al.<sup>8</sup> measured blink rate by direct observation on a group of 16 subjects. The mean blink rate during conversation was 18.4 blinks/min and during computer use it was 3.6 blinks/min—more than a fivefold decrease. Tsubota and Nakamori<sup>9</sup> measured blink rates on 104 office workers. The mean blink rates were 22 blinks/min under relaxed conditions, 10 blinks/min while reading a book on the table, and 7 blinks/min while viewing text on a computer screen. Although both book reading and computer work result in significantly decreased blink rates, a difference between them is that computer work usually requires a higher gaze angle, resulting in an increased rate of tear evaporation. Tsubota and Nakamori<sup>9</sup> measured a mean exposed ocular surface of 2.2 cm<sup>2</sup> while subjects were relaxed and 1.2 cm<sup>2</sup> while working at a computer. Since the primary route of tear elimination is through evaporation, and the amount of evaporation is a roughly linear function of ocular aperture area, the higher gaze angle when viewing a computer display results in faster tear loss. Even though reading a book and display work both reduce the blink rate, the computer worker is more at risk because of the higher gaze angle.

Other factors may also contribute to dry eye. For example, the office air environment is often low in humidity and can contain contaminants. Working under an air vent also increases evaporation. The higher gaze angle results in a greater percentage of blinks that are incomplete, resulting in poor tear film and higher evaporation rates.

Dry eyes are common in the population, reported as being 15 percent of a clinical population. For example, postmenopausal women, arthritis sufferers, post-LASIK patients, those taking some systemic medications, and contact lens wearers are more prone to dry eyes. All of these conditions should be explored as causative for dry eye symptoms. For many people with marginal dry eye problems, work at a computer will cause their dry eye problem to become clinically significant.

Resolving dry eye symptoms may take many approaches, depending on the source of the problem. The longstanding, traditional approach of simply adding an “artificial tear” to the eyes is losing favor because it has shown to be ineffective. Newer drops are available that create a better quality tear film.



In addition, a prescription drop that incorporates a mild antibiotic to reduce inflammation has shown to be effective in some cases. There is also a swell of attention to oral supplementation to stimulate better tear formation. Regardless of which approach the eye care practitioner takes, resolving dry eyes for the computer user should be a priority in the treatment of computer vision issues.

## Refractive Error

Many patients simply need an accurate correction of refractive error (myopia, hyperopia, or astigmatism). The blur created by the refractive error makes it difficult to easily acquire visual information at the computer display. This reduces work efficiency and induces fatigue. A 2004 study by Daum<sup>10</sup> indicates that even a slightly inaccurate vision prescription at the computer can have a significant negative impact on a worker's productivity. Results indicate that uncorrected vision, even if no symptoms appear, can affect employee productivity and accuracy. In addition, a vision miscorrection of as little as 0.50 D can cause a decrease in productivity of 9 percent and accuracy by 38 percent.

Patients with hyperopia must exert added accommodative effort to see clearly at near distances. Therefore, many hyperopic computer patients require refractive correction that they may not require for a less visually demanding job. Patients with 2.00 to 3.50 D of myopia who habitually read without their glasses often have visual and musculoskeletal difficulties at the computer because, to see clearly without their glasses, they must work too closely to the computer screen. These patients often require a partial correction of their myopia for proper function at the computer display.

Some computer-using patients develop a late-onset myopia from 0.25 to 1.00 D. The preponderance of evidence supports the contention that near workplaces some people are greater risk for the development of myopia. However, there is no evidence that work at a computer monitor causes myopia more than other forms of extended near-point work. The fact is, however, that our society is dictating that we do more and more of our work on the computer, which translates into long hours viewing the display screen. Take for example some school districts that require all of their students to perform all of their work on computers. The consequences of this type of visual stress may take years to realize.

## Accommodation

*Accommodation* is the mechanism by which the eye changes its focus to look at near objects. Accommodation is accomplished by constriction of the ciliary muscle that surrounds the crystalline lens within the eye. The amplitude of accommodation (dioptric change in power) decreases with age. Presbyopia (see later) is the age-related condition that results when the amplitude of accommodation is no longer adequate to meet near visual needs. (See Chaps. 12 and 14.)

Many workers have reduced amplitude of accommodation for their age, or accommodative *infacility* (inability to change the level of accommodation quickly and accurately). These conditions result in a blur at near working distances and/or discomfort. Extended near work also commonly results in *accommodative hysteresis*—that is, the accommodative mechanism becomes “locked” into the near focus and it takes time (a few minutes to hours) to fully relax for distance focus. This is effectively a transient myopia that persists after extended near work. Some theories on myopia development look at this transient myopia as a step in the progression to developed myopia. Accommodative disorders are diagnosed in approximately one-third of the prepresbyopic segment of a clinical population. Accommodative disorders in the younger or prepresbyopic patient are usually treated with near prescription spectacles that enable the worker to relax accommodation. Vision training can also work for some patients.

Most work with display screens are performed with a higher gaze than other near-point work and accommodative amplitude is shown to be reduced with elevation of the eye. Relative to a 40° downward viewing angle, elevations of 20° downward, straight ahead, and 20° upward result in average accommodative decreases of 1.22 D, or 2.05 D, and 2.00 D, respectively, in a group of 80 prepresbyopes

with an average age of 26.4 years.<sup>11</sup> The higher gaze angles at most computer workstations result in a viewing condition for which the amplitude of accommodation is reduced—thus placing a greater strain on accommodation than near tasks performed at lower gaze angles.

## Binocular Vision

Approximately 95 percent of people keep both eyes aligned on the object of regard. Those individuals who habitually do not keep their eyes aligned have *strabismus*. Even though most people can keep their eyes aligned when viewing an object, many individuals have difficulty maintaining this ocular alignment and experience symptoms such as fatigue, headaches, blur, double vision, and general ocular discomfort (see Chap. 13).

Binocular fusion is the sensory process by which the images from each eye are combined to form a single percept. When the sensory feedback loop is opened (e.g., by blocking an eye), the eyes assume their *position of rest* with respect to one another. If the position of rest is outward or diverged, the patient has *exophoria*. If it is inward or converged, the condition is *esophoria*. Clinically, the phoria is measured by occluding one of the eyes and measuring the eye alignment while occluded.

If the patient has a phoric deviation, as is the case for most people, then a constant neuromuscular effort is required to keep the eyes aligned. Whether a person experiences symptoms depends on the amount of the misalignment, the ability of the individual to overcome that misalignment, and the task demands. The symptoms associated with excessive phoria deviations can be eyestrain, double vision, headaches, eye irritation, and general fatigue.

Eye alignment at near viewing distances is more complex than at far viewing distances because of the ocular convergence required to view near objects and because of the interaction between the ocular convergence and accommodative mechanisms. Treatment of these conditions can include refractive or prismatic correction in spectacles, or vision training.

## Anisometropia

One of the possible causes of binocular vision dysfunction is a difference in the correction for each eye. If there is a significant refractive difference between the two eyes, the condition of *anisometropia* exists. While this can happen developmentally and genetically, it can also be induced with surgical intervention of cataracts. If one eye has a cataract procedure performed and surgery on the other eye is not needed, then there may be a difference in the refractive errors, creating an anisometropic condition.

This condition is more significant if the refractive error of each eye is corrected with spectacles. Wearing a correction for two variable prescriptions can create an image-size differential between the eyes. One solution is to fit the patient with contact lenses. Wearing the prescription directly on the eye will reduce the magnification differential and allow for equal-sized images.

## Presbyopia

*Presbyopia* is the condition in which the normal age-related loss of accommodation results in an inability to comfortably maintain focus on near objects. This usually begins at about the age of 40. The usual treatment for presbyopia is to prescribe reading glasses or multifocal lenses that have a distance vision corrective power (if needed) in the top of the lens and the near vision corrective power in the bottom of the lens. The most common lens designs for correcting presbyopia are bifocals and progressive addition lenses (PALs). As usually designed and fitted, these lenses work well for the most common everyday visual tasks and provide clear vision at 30 cm with a downward gaze angles of about 25°. The computer screen is typically farther away (from 50 to 70 cm) and higher (from 10 to 20° of ocular depression). A presbyope who tries to wear their usual multifocal correction at the computer will either not see the screen clearly or will need to assume an awkward

posture, resulting in neck and back strain. This is because the zone of intermediate vision is most often too narrow for full-time computer use. Many, if not most, presbyopic computer workers require a separate pair of spectacles for their computer work. Several newer lenses are designed specifically for people with occupations that require intermediate viewing distances. (See Chaps. 12 and 14.)

## 23.5 REFERENCES

---

1. National Institute for Occupational Safety and Health, "Potential Health Hazards of Video Display Terminals," *DHHS (NIOSH) Publication No. 81-129*, National Institute of Occupational Safety and Health, Cincinnati, 1981.
2. *The American Heritage Dictionary*, 4th ed., Houghton-Mifflin, Boston, MA, 2000.
3. S. K. Guth, "Prentice Memorial Lecture: The Science of Seeing—A Search for Criteria," *American Journal of Optometry and Physiological Optics* **58**:870–885 (1981).
4. "VDT Lighting. IES Recommended Practice for Lighting Offices Containing Computer Visual Display Terminals," Illumination Engineering Society of North America, New York, 1990.
5. J. E. Sheedy, "Vision at Computer Displays," *Vision Analysis*, Walnut Creek, CA, 1995.
6. T. Allan, "Glare Screen Home Usage Test Report," Decision Analysis, Inc. Study 2003–0559 (2005).
7. ANSI/HFS 100, *American National Standard for Human Factors Engineering of Visual Display Terminal Workstations*, Human Factors Society, Santa Monica, CA, 1999.
8. S. Patel, R. Henderson, L. Bradley, B. Galloway, and L. Hunter, "Effect of Visual Display Unit Use on Blink Rate and Tear Stability," *Optometry and Vision Science* **68**(11):888–892 (1991).
9. K. Tsubota and K. Nakmori, "Dry Eyes and Video Display Terminals. Letter to Editor," *New England Journal of Medicine* **328**:524 (1993).
10. Daum KM, Clore KA, Simms SS, Vesely JW, Dwilczek DD, Spittle BM, and Good GW. "Productivity Associated with Visual Status of Computer Users," *Optometry* **75**:33–47 (2004).
11. P. H. Ripple, "Accommodative Amplitude and Direction of Gaze," *American Journal of Ophthalmology* **35**:1630–1634 (1952).

---

# HUMAN VISION AND ELECTRONIC IMAGING

---

## Bernice E. Rogowitz

*IBM T. J. Watson Research Center  
Hawthorne, New York*

## Thrasyvoulos N. Pappas

*Department of Electrical and Computer Engineering  
Northwestern University  
Evanston, Illinois*

## Jan P. Allebach

*Electronic Imaging Systems Laboratory  
School of Electrical and Computer Engineering  
Purdue University  
West Lafayette, Indiana*

---

## 24.1 INTRODUCTION

---

The field of *electronic imaging* has made incredible strides over the past decade as increased computational speed, bandwidth, and storage capacity have made it possible to perform image computations at interactive speeds. This means larger images, with more spatial, temporal, and chromatic resolution, can be captured, compressed, transmitted, stored, rendered, printed, and displayed. It also means that workstations and PCs can accommodate more complex image and data formats, more complex operations for analyzing and visualizing information, more advanced interfaces, and richer image environments, such as virtual reality. This, in turn, means that image technology can now be practically used in an expanding world of applications, including video, home photography, internet catalogues, digital libraries, art, and scientific data analysis. These advances in technology have been greatly influenced by research in human perception and cognition, and in turn, have stimulated new research into the vision, perception, and cognition of the human observer. Some important topics include spatial, temporal, and color vision, attentive and preattentive vision, pattern recognition, visual organization, object perception, language, and memory.

The study of the interaction between human vision and electronic imaging is one of the key growth areas in *imaging science*. Its scope ranges from printing and display technologies to image processing algorithms for image rendering and compression, to applications involving interpretation, analysis, visualization, search, design, and aesthetics. Different electronic imaging applications call on different human capabilities. At the bottom of the visual food chain are the visual phenomena mediated by the threshold sensitivity of low-level spatial, temporal, and color mechanisms. At the next level are perceptual effects, such as color constancy, suprathreshold pattern, and texture analysis.

Moving up the food chain, we find cognitive effects, including memory, semantic categorization, and visual representation, and moving to the next level, we encounter aesthetic and emotional aspects of visual processing.

## Overview of This Chapter

In this chapter, we review several key areas in the two-way interaction between human vision and technology. We show how technology advances in electronic imaging are increasingly driven by methods, models, and applications of vision science. We also show how advances in vision science are increasingly driven by the rapid advances in technologies designed for human interaction. An influential force in the development of this new field has been the Society for Imaging Science and Technology (IS&T)/Society of Photographic and Instrumentation Engineers (SPIE) Conference on Human Vision and Electronic Imaging, which had its origins in 1988 and since 1989 has grown annually as a forum for multidisciplinary research in this area.<sup>1-14</sup> This chapter has been strongly influenced by the body of research that has been presented at these conferences, and reflects their unique perspectives.

A decade ago, the field of *human vision and electronic imaging* focused on the threshold sensitivity of the human visual system (HVS) as it relates to display technology and still-image compression. Today, the scope of human vision and electronic imaging is much larger and keeps expanding with the electronic imaging field. We have organized this chapter to reflect this expanding scope. We begin with early vision approaches in Sec. 24.2, showing how the explicit consideration of human spatial, temporal, and color sensitivity has affected the development of algorithms for compression and rendering, as well as the development of image quality metrics. These approaches have been most successful in algorithmically evaluating the degree to which artifacts introduced by compression or rendering processes will be detected. Section 24.3 considers how image features are detected, processed, and perceived by the human visual system. This work has been influential in the design of novel gaze-dependent compression schemes, new visualization and user interface designs, and perceptually based algorithms for image retrieval systems. Section 24.4 moves higher up the food chain to consider emotional and aesthetic evaluations. This work has influenced the development of virtual environments, high-definition TV, and tools for artistic appreciation and analysis. Section 24.5 concludes the chapter paper, and suggestions for further reading are given in Sec. 24.6.

## 24.2 EARLY VISION APPROACHES: THE PERCEPTION OF IMAGING ARTIFACTS

---

There are many approaches to characterizing image artifacts. Some approaches are based purely on physical measurement. These include measuring key image or system parameters to ensure that they fall within established tolerances, or comparing an original with a rendered, displayed, or compressed version, on a pixel-by-pixel basis, using a metric such as mean square error. These approaches have the advantage of being objective and relatively easy to implement; on the other hand, it is difficult to generalize from these data. Another approach is to use human observers to judge perceived quality, either by employing a panel of trained experts or by running psychological experiments to measure the perception of image characteristics. These approaches have the advantage of considering the human observer explicitly; on the other hand, they are costly to run and the results may not generalize. A third approach is to develop metrics, based on experiments measuring human visual characteristics, that can stand in for the human observer as a means of estimating human judgments. These perceptual models are based on experiments involving the detection, recognition, and identification of carefully controlled experimental stimuli. Since these experiments are designed to reveal the behavior of fundamental mechanisms of human vision, their results are more likely to generalize. This section reviews several models and metrics based on early human vision which have been developed for evaluating the perception of image artifacts.

Early vision models are concerned with the processes mediating the threshold detection of spatial, temporal, and color stimuli. In the early days of television design, Schade<sup>15</sup> and his colleagues introduced the notion of characterizing human threshold sensitivity in terms of the response to spatial-frequency patterns, thereby beginning a long tradition of work to model human spatial and temporal sensitivity using the techniques of linear systems analysis. The simplest model of human vision is a threshold contrast sensitivity function (CSF). A curve representing our sensitivity to spatial modulations in luminance contrast is called a spatial CSF. A curve representing our sensitivity to temporal modulations in luminance is a temporal CSF. These band-pass curves vary depending on many other parameters, including, for example, the luminance level, the size of the field, temporal modulation, and color. In early applications of visual properties to electronic imaging technologies, these simple threshold shapes were used. As the field has progressed, the operational model for early vision has become more sophisticated, incorporating interactions between spatial, temporal, and color vision, and making more sophisticated assumptions about the underlying processes. These include, for example, the representation of the visual system as a set of band-pass spatial-frequency filters,<sup>16,17</sup> the introduction of near-threshold contrast masking effects, and nonlinear processing.

One key area where these models of early vision have been applied is in the evaluation of image quality. This includes the psychophysical evaluation of image quality, perceptual metrics of image distortion, perceptual effects of spatial, temporal, and chromatic sampling, and the experimental comparison of compression, sampling, and halftoning algorithms. This work has progressed hand-in-hand with the development of vision-based algorithms for still image and video compression, image enhancement, restoration and reconstruction, image halftoning and rendering, and image and video quantization and display.

## Image Quality and Compression

The basic premise of the work in perceptual image quality is that electronic imaging processes, such as compression and halftoning, introduce distortions. The more visible these distortions, the greater the impairment in image quality. The human vision model is used to evaluate the degree to which these impairments will be detected. Traditional metrics of image compression do not incorporate any models of human vision and are based on the mean squared error (i.e., the average squared difference between the original and compressed images). Furthermore, they typically fail to include a calibration step, or a model of the display device, thereby providing an inadequate model of the information presented to the eye.

The first perceptually based image quality metrics used the spatial contrast sensitivity function as a model of the human visual system.<sup>18,19</sup> The development of perceptual models based on multiple spatial-frequency channels greatly improved the objective evaluation of image quality. In these models, the visual system is treated as a set of spatial-frequency-tuned channels, or as Gabor filters with limited spatial extent distributed over the visual scene. The envelope of the responses of these channels is the contrast sensitivity function. These multiple channel models provide a more physiologically representative model of the visual system, and more easily model interactions observed in the detection of spatially varying stimuli. One important interaction is contrast masking, where the degree to which a target signal is detected depends on the spatial-frequency composition of the masking signal.

In 1992, Daly introduced the *visual differences predictor*,<sup>6,20</sup> a multiple-channel model for image quality that models the degree to which artifacts in the image will be detected, and thus will impair perceived image quality. This is a spatial-frequency model which incorporates spatial contrast masking and light adaptation. At about the same time, Lubin<sup>21</sup> proposed a similar metric that also accounts for sensitivity variations due to spatial frequency and masking. It also accounts for fixation depth and image eccentricity in the observer's visual field. The output of such metrics is either a map of detection probabilities or a point-by-point measure of the distance between the original and degraded image normalized by the HVS sensitivity to error at each spatial frequency and location. These detection probabilities or distances can be combined into a single number that represents the overall picture quality. While both models were developed for the evaluation of displays and high quality imaging systems, they have been adapted for a wide variety of applications.

Perceptually based image compression techniques were developed in parallel with perceptual models for image quality. This is not surprising since quality metrics and compression algorithms are closely related. The image quality metric is trying to characterize the human response to an image; an image compression algorithm is either trying to minimize some distortion metric for a given bit rate, or trying to minimize the bit rate for a given distortion. In both cases, a perceptually based distortion metric can be used.

In 1989, three important papers introduced the notion of “perceptually lossless” compression. (See Refs. 22–24.) In this view, the criterion of importance in image compression was the degree to which an image could be compressed without the user’s perceiving a difference between the compressed and original images. Safranek and Johnston<sup>25</sup> presented the perceptual subband image coder (PIC) which incorporated a perceptual model and achieved perceptually lossless compression at lower rates than state-of-the-art perceptually lossy schemes. The Safranek-Johnston coder used an empirically derived perceptual masking model that was obtained for a given CRT display and viewing conditions. As with the quality metrics we discussed earlier, the model determines the HVS sensitivity to errors at each spatial frequency and location, which is called the *just noticeable distortion level* (JND). In subsequent years, perceptual models were used to improve the results of traditional approaches to image and video compression, such as those that are based on the discrete cosine transform (DCT). (See Refs. 26 to 28.)

Perceptually based image quality metrics have also been extended to video. In 1996, Van den Branden Lambrecht and Verscheure<sup>29</sup> described a video quality metric that incorporates spatio-temporal contrast sensitivities as well as luminance and contrast masking adjustments. In 1998, Watson extended his DCT-based still image metric, proposing a video quality metric based on the DCT.<sup>30,31</sup> Since all the current video coding standards are based on the DCT, this metric is useful for optimizing and evaluating these coding schemes without significant additional computational overhead.

In recent years, significant energy has been devoted to comparing these methods, models, and results, and to fine-tuning the perceptual models. The Video Quality Experts Group (VQEG), for example, has conducted a cross-laboratory evaluation of perceptually based video compression metrics. They found that different perceptual metrics performed better on different MPEG-compressed image sequences, but that no one model provided a clear advantage, including the nonperceptual measure, peak signal-to-noise ratio (Ref. 32). We believe that the advantages of the perceptual metrics will become apparent when this work is extended to explicitly compare fundamentally different compression schemes over different rates, image content, and channel distortions. An important approach for improving low-level image quality metrics is to provide a common set of psychophysical data to model. Modelfest (Ref. 33) is a collaborative modeling effort where researchers have volunteered to collect detection threshold data on a wide range of visual stimuli, under carefully controlled conditions, in order to provide a basis for comparing the predictions of early vision models. This effort should lead to a converged model for early vision that can be used to develop image quality metrics and perceptually based compression and rendering schemes. (Comprehensive reviews of the use of perceptual criteria for the evaluation of image quality can be found in Refs. 34 and 35.)

## Image Rendering, Halftoning, and Other Applications

Another area where simple models of low-level vision have proven to be quite successful is image halftoning and rendering. All halftoning and rendering algorithms make implicit use of the properties of the HVS; they would not work if it were not for the high spatial-frequency cutoff of the human CSF. In the early 1980s, Allebach introduced the idea of halftoning based on explicit human visual models and models of the display device. However, the development of halftoning techniques that rely on such models came much later. In 1991, Sullivan et al.<sup>36</sup> used a CSF model to design halftoning patterns of minimum visibility and Pappas and Neuhoff<sup>37</sup> incorporated printer models in error diffusion. In 1992, methods that use explicit visual models to minimize perceptual error in image halftoning were independently proposed by Analoui and Allebach,<sup>38</sup> Mulligan and Ahumada,<sup>39</sup> and Pappas and Neuhoff.<sup>40</sup> Allebach et al. extended this model-based approach to color in 1993<sup>41</sup> and to video in 1994.<sup>42</sup> Combining the spatial and the chromatic properties of early vision, Mulligan<sup>43</sup> took advantage of the low spatial-frequency sensitivity of chromatic channels to hide high spatial-frequency halftoning artifacts.



Low-level perceptual models and techniques have also been applied to the problem of target detection in medical images (e.g., Ref. 44) and to the problem of embedding digital watermarks in electronic images (Refs. 45 and 46).

## Early Color Vision and Its Applications

The trichromacy of the human visual system has been studied for over a hundred years, but we are just in the infancy in applying knowledge of early color vision to electronic imaging systems. From an engineering perspective, the CIE 1931 model, which represents human color-matching as a linear combination of three color filters, has been the most influential model of human color vision. This work allows the determination of whether two patches will have matching colors, when viewed under a constant illuminant. The CIE 1931 color space, however, is not perceptually uniform. That is, equal differences in chromaticity do not correspond to equal perceived differences. To remedy this, various transformations of this space have been introduced. Although CIE  $L^*a^*b^*$  and CIE  $L^*u^*v^*$  are commonly used as perceptually uniform spaces, they still depart significantly from this objective. By adding a nonlinear gain control at the receptor level, Guth's ATD model<sup>47,48</sup> provided a more perceptually uniform space and could model a wide range of perceptual phenomena. Adding a spatial component to a simplified version of Guth's model, Granger<sup>49</sup> demonstrated how this approach could be used to significantly increase the perceived similarity of original and displayed images. The observation that every pixel in an image provides input both to mechanisms sensitive to color variations and to mechanisms sensitive to spatial variations has provided other important contributions. For example, Uriegas<sup>50,51</sup> used the multiplexing of color and spatial information by cortical neurons to develop a novel color image compression scheme.

Understanding how to represent color in electronic imaging systems is a very complicated problem, since different devices (e.g., printers, CRT or TFT/LCD displays, film) have different mechanisms for generating color, and produce different ranges, or gamuts, of colors. Several approaches have been explored for producing colors on one device that have the same appearance on another. Engineering research in this field goes under the name of *device independent color*, and has recently been energized by the need for accurate color rendering over the Internet. (See Ref. 52.)

## Limitations of Early Vision Models for Electronic Imaging

The goal of the early vision models is to describe the phenomena of visual perception in terms of simple mechanisms operating at threshold. Pursuing this Ockham's razor approach has motivated the introduction of masking models and nonlinear summation models, and has allowed us to extend these simple models to describe the detection and recognition of higher-level patterns, such as textures and multiple-sinewave plaids. These models, however, eventually run out of steam in their ability to account for the perception of higher-level shapes and patterns. For example, the perception of a simple dot cannot be successfully modeled in terms of the response of a bank of linear spatial-frequency filters.

In image quality, early vision approaches consider an image as a collection of picture elements. They measure and model the perceived fidelity of an image based on the degree to which the pixels, or some transformation of the pixels, have changed in their spatial, temporal, and color properties. But what if two images have identical perceived image fidelity, but one has a lot of blur and little blockiness while the other has little blur and a lot of blockiness? Are they equivalent? Also, since it is not always possible to conceal these artifacts, it is important to understand how their effects combine and interact in order to minimize their objectionable and annoying effects. This has led to the development of new suprathreshold scaling methods for evaluating the perceived quality of images with multiple suprathreshold artifacts (e.g., Refs. 53 and 54), and has led to the development of new dynamic techniques to study how the annoyance of an artifact depends on its temporal position in a video sequence (Ref. 55).

In color vision, a key problem for electronic imaging is that simply capturing and reproducing the physical color of individual color pixels and patches in an image is not sufficient to describe



the perceived colors in the image. Hunt,<sup>56</sup> for example, has shown that surrounding colors affect color appearance. Furthermore, as McCann<sup>57</sup> has demonstrated using spatially complex stimuli, the perception of an image depends, not simply on the local luminance and color values, and not simply on nearby luminance and color values, but on a more global consideration of the image and its geometry.

### 24.3 HIGHER-LEVEL APPROACHES: THE ANALYSIS OF IMAGE FEATURES

Higher-level approaches in vision are dedicated to understanding the mechanisms for perceiving more complex features such as dots, plaids, textures, and faces. One approach is to build up from early spatial-frequency models by positing nonlinear mechanisms that are sensitive to two-dimensional spatial variations, such as t-junctions (e.g., Ref. 58). Interesting new research by Webster and his colleagues<sup>59</sup> identifies higher-level perceptual mechanisms tuned to more complex visual relationships. Using the same types of visual adaptation techniques commonly used in early vision experiments to identify spatial, temporal, or color channel properties, they have demonstrated adaptation to complex visual stimuli, including, for example, the adaptation to complex facial distortions and complex color distributions.

Another interesting approach to understanding the fundamental building blocks of human vision is to consider the physical environment in which it has evolved, with the idea in mind that the statistics of the natural world must somehow have guided, or put constraints on, its development. This concept, originally introduced by Field,<sup>60,61</sup> has led to extensive measurements of the spatial and temporal characteristics of the world's scenes, and to an exploration of the statistics of the world's illuminants and reflective surfaces. One of the major findings in this field is that the spatial amplitude spectra of natural scenes falls off as  $f^{-1.1}$ . More recently, Field et al.<sup>62</sup> have related this function to the sparseness of spatial-frequency mechanisms in human vision. This low-level description of the global frequency content in natural scenes has been tied to higher-level perception by Rogowitz and Voss.<sup>63</sup> They showed that when the slope of the fall-off in the amplitude spectrum is in a certain range (corresponding to a fractal dimension of from 1.2 to 1.4), observers see nameable shapes in images like clouds. This approach has also been applied to understanding how the statistics of illuminants and surfaces in the world constrains the spectral sensitivities of visual mechanisms (Ref. 64), thereby influencing the mechanisms of color constancy (see also Refs. 65 and 66).

The attention literature provides another path for studying image features by asking which features in an image or scene naturally attract attention and structure perception. This approach is epitomized by Triesman's work on "preattentive" vision<sup>67</sup> and by Julesz' exploration of "textons."<sup>68,69</sup> Their work has had a very large impact in the field of electronic imaging by focusing attention on the immediacy with which certain features in the world, or in an image, are processed. Their work both identifies image characteristics that attract visual attention and guide visual search, and also provides a paradigm for studying the visual salience of image features. An important debate in this area is whether features are defined bottom up—that is, generated by successive organizations of low-level elements such as edges, as suggested by Marr<sup>70</sup>—or whether features are perceived immediately, driven by top-down processes, as argued by Stark.<sup>71</sup>

This section explores the application of feature perception and extraction in a wide range of electronic imaging applications, and examines how the demands of these new electronic tasks motivate research in perception. One important emerging area is the incorporation of perceptual attention or "importance" in image compression and coding. The idea here is that if we could identify those aspects of an image that attracted attention, we could encode this image using higher resolution for the areas of importance and save bandwidth in the remaining areas. Another important area is image analysis, where knowing which features are perceptually salient could be used to index, manipulate, analyze, compare, or describe an image. An important emerging application in this area is digital libraries, where the goal is to find objects in a database that have certain features, or that are similar to a target object. Another relevant application area is visualization, where the goal is to develop methods for visually representing structures and features in the data.

## Attention and Region of Interest

A new idea in electronic imaging is to analyze images to identify their “regions of interest,” features that draw our attention and have a special saliency for interpretation. If we could algorithmically decompose an image into its salient features, we could develop compression schemes that would, for example, compress more heavily those regions that did not include perceptual features, and devote extra bandwidth to regions of interest. Several attempts have been made to algorithmically identify regions of interest. For example, Leray et al.<sup>72</sup> developed a neural network model that incorporates both the response of low-level visual mechanisms and an attentional mechanism that differentially encodes areas of interest in the visual image, and used this to develop a compression scheme.

Stelmach and Tam<sup>73</sup> measured the eye movements of people examining video sequences and concluded that there wasn’t enough consistency across users to motivate developing a compression scheme for TV broadcast based on user eye movements. Geissler and Perry,<sup>74</sup> however, demonstrated that by guiding a user’s eye movements to a succession of target locations, and only rendering the image at high resolution at these “foveated” regions, the entire image appeared to have full resolution. If the system knew in advance where the eye movements would be directed, it could adjust the resolution at those locations, on the fly, thus saving bandwidth. Finding those perceptually relevant features, however, is a very difficult problem. Yarbus<sup>75</sup> showed that there is not a single stereotypic way that an image is viewed; the way the user’s saccades are placed on the picture depends on the task. For example, if the goal is to identify how many people there are in the picture, the set of eye movements differs from those where the goal is to examine what the people are wearing.

A most important voice in this discussion is that of Stark.<sup>76,77</sup> Noton and Stark’s classic papers<sup>71,78</sup> set the modern stage for measuring the eye movement paths, or “scanpaths,” of human vision. They used this methodology to explore the role of particular visual stimuli in driving attention (bottom up) versus the role of higher-level hypothesis-testing and scene checking goals (top down) in driving the pattern of visual activity. Stark and his colleagues have contributed both to our basic understanding of the processes that drive these scanpaths, and to integrating this knowledge into the development of better electronic imaging systems. Recent evidence for top-down processing in eye movements has been obtained using an eye tracking device that can be worn while an observer moves about in the world (Ref. 79). As observers perform a variety of everyday tasks, they perform what the authors call “planful” eye movements, eye movements that occur in the middle of a task, in anticipation of the subject’s interaction with an object in an upcoming task.

## Image Features, Similarity, and Digital Libraries Applications

In digital libraries applications, the goal is to organize and retrieve information from a database of images, videos, graphical objects, music, and sounds. In these applications, the better these objects are indexed and organized, the more easily they can be searched. A key problem, therefore, is to identify features and attributes of importance, and to provide methods for indexing and searching for objects based on these features. Understanding which features are meaningful, or what makes objects similar, however, is a difficult problem. Methods from signal processing and computer vision can be used to algorithmically segment images and detect features. Since these databases are being designed for humans to search and navigate, it is important to understand which features are salient and meaningful to human observers, how they are extracted from complex objects, and how object features are used to judge image and object similarity. Therefore, these algorithms often incorporate heuristics gleaned from perceptual experiments regarding human color, shape, and pattern perception. For example, the work by Petkovic and his colleagues<sup>80</sup> includes operators that extract information about color, shape, texture, and composition. More recently, knowledge about human perception and cognition has been incorporated more explicitly. For example, Frese et al.<sup>81</sup> developed criteria for image similarity based on a multiscale model of the human visual system, and used a psychophysical model to weight the parameters of their model. Another method for explicitly incorporating human perception has been to study how humans explicitly judge image similarity and use

this knowledge to build better search algorithms. Rogowitz et al.,<sup>82</sup> for example, asked observers to judge the similarity of a large collection of images and used a multidimensional scaling technique to identify the dimensions along which natural objects were organized perceptually. Mojsilović et al.<sup>83</sup> asked observers to judge the similarity of textured designs, then built a texture retrieval system based on the perceptual results.

In this expanding area, research opportunities include the analysis of human feature perception, the operationalization of these behaviors into algorithms, and the incorporation of these algorithms into digital library systems. This includes the development of perceptual criteria for image retrieval, the creation of perceptual image similarity metrics, methods for specifying perceptual metadata for characterizing these objects, and perceptual cues for navigating through large multimedia databases. An important emerging opportunity lies in extending these methods to accommodate more complex multidimensional objects, such as three-dimensional objects and auditory patterns.

## Visualization

One consequence of the expanding computer age is the creation of terabytes and terabytes of data, and an interest in taking advantage of these data for scientific, medical, and business purposes. These can be data from satellite sensors, medical diagnostic sensors, business applications, simulations, or experiments, and these data come in many different varieties and forms. The goal of visualization is to create visual representations of these data that make it easier for people to see patterns and relationships, identify trends, develop hypotheses, and gain understanding. To do so, the data are mapped onto visual (and sometimes auditory) dimensions, producing maps, bar charts, sonograms, statistical plots, etc. A key perceptual issue in this area is how to map data onto visual dimensions in a way that preserves the structure in the data without creating visual artifacts. Another key perceptual issue is how to map the data onto visual dimensions in a way that takes advantage of the natural feature extraction and pattern-identification capabilities of the human visual system. For example, how can color and texture be used to draw attention to a particular range of data values, or a departure from a model's predictions? (See Ref. 84 for an introduction to these ideas.)

One important theme in this research is the return to Gestalt principles of organization for inspiration about the analysis of complex visual information. The Gestalt psychologists identified principles by which objects in the visual world are organized perceptually. For example, objects near each other (proximity) or similar to each other (similarity) appear to belong together. If these fundamental principles could be operationalized, or if systems could be built that take advantage of these basic rules of organization, it could be of great practical importance. Some recent papers in this area include Kubovy's<sup>85</sup> theoretical and experimental studies, experiments by Hon et al.<sup>86</sup> on the interpolation and segmentation of sampled contours, and an interactive data visualization system based on Gestalt principles of perceptual organization (Ref. 87).

## User Interface Design

As more and more information becomes available, and computer workstations and web browsers support more interactivity and more color, we move into a new era in interface design. With so many choices available, we now have the luxury to ask how best to use color, geometry, and texture to represent information. A key vision paper in this area has been Boynton's paper<sup>88</sup> on "the eleven colors which are almost never confused." Instead of measuring color discriminability, Boynton asked people to sort colors, to get at a higher-level color representation scheme, and found that the responses clustered in 11 categories, each organized by a common color name. That is, although in side-by-side comparison, people can discriminate millions of colors, when the task is to categorize colors, the number of perceptually distinguishable color categories is very small. In a related experiment, Derefeldt and Swartling<sup>89</sup> tried to create as many distinct colors as possible for a digital map application; they were only able to identify 30. More recently, Yendrikovskij<sup>90</sup> extracted thousands of images from the web and used a *K*-means clustering algorithm to group the colors of the pixels. He also found a small

number of distinct color categories. This suggests that the colors of the natural world group into a small number of categories that correspond to a small set of nameable colors. This result is important for many applications that involve selecting colors to represent semantic entities, such as using colors to represent different types of tumors in a medical visualization. It is also important for image compression or digital libraries, suggesting that the colors of an image can be encoded using a minimal number of color categories. Several groups have begun developing user-interface design tools that incorporate perceptually based intelligence, guiding the designer in choices of colors, fonts, and layouts based on knowledge about color discriminability, color deficiency, color communication, and legibility. Some examples include Hedin and Derefeldt<sup>91</sup> and Rogowitz et al.<sup>92</sup>

## 24.4 VERY HIGH-LEVEL APPROACHES: THE REPRESENTATION OF AESTHETIC AND EMOTIONAL CHARACTERISTICS

The higher-level applications discussed above, such as visualization and digital libraries, push the envelope in electronic imaging technology. They drive developers to address the content of the data, not just the pixel representation. As the technology improves and becomes more pervasive, new application areas are drawn into the web, including applications that use visual media to convey artistic and emotional meaning. For these applications, it is not just the faithfulness of the representation, or its perceptual features, that are of concern, but also its naturalness, colorfulness, composition, and appeal. This may seem an enormous leap, until we realize how natural it is to think of the artistic and emotional aspects of photographic imaging. Interest in using electronic media to convey artistic and emotional goals may just reflect the maturation of the electronic imaging field.

### Image Quality

Although the world “quality” itself connotes a certain subjectivity, the major focus of the research in image quality, discussed so far, has been aimed at developing objective, algorithmic methods for characterizing the judgments of human observers. The research in Japan on high-definition TV has played a major role in shaping a more subjective approach to image quality. This approach is based on the observation that the larger format TV, because it stimulates a greater area of the visual field, produces images which appear not only more saturated, but also more lively, realistic, vivid, natural, and compelling (Ref. 93). In related work on the image quality of color images, DeRidder et al.<sup>94</sup> found that, in some cases, observers preferred somewhat more saturated images, even when they clearly perceived them to be somewhat less natural. The observers’ responses were based on aesthetic characteristics, not judgments of image fidelity.

### Virtual Reality and Presence

Virtual reality is a set of technologies whose goal is to immerse the user in a visual representation. Typically, the user will wear a head-tracking device to match the scene to the user’s viewpoint and stereo glasses whose two interlaced views provide a stereoscopic view. A tracked hand device may also be included so that the user can interact with this virtual environment. Although these systems do not provide adequate visual resolution or adequate temporal resolution to produce a realistic visual experience, they do give the sense of “presence,” a sense of being immersed in the environment. This new environment has motivated considerable perceptual research in a diverse set of areas, including the evaluation of depth cues, the visual contributions to the sense of presence, the role of auditory cues in the creation of a realistic virtual environment, and cues for navigation. This new environment also provides an experimental apparatus for studying the effects of large-field visual patterns and the interaction of vision and audition in navigation.

## Color, Art, and Emotion

Since the beginning of experimental vision science, there have been active interactions between artists and psychologists. This is not surprising since both groups are interested in the visual impression produced by physical media. As artists use new media, and experiment with new visual forms, perceptual psychologists seek to understand how visual effects are created. For example, Chevreul made fundamental contributions to color contrast by studying why certain color dyes in artistic textiles didn't appear to have the right colors. The Impressionists and the Cubists read the color vision literature of their day, and in some cases, the artist and vision scientist were the same person. For example, Seurat studied the perception of color and created color halftone paintings. Artists have continued to inspire vision scientists to this day. For example, Papathomas,<sup>95</sup> struck by the impressive visual depth illusions in Hughes's wall sculptures, used this as an opportunity to deepen our knowledge of how the visual system constructs the sensation of depth. Koenderink<sup>96</sup> used three-dimensional sculptures to study how the human observer constructs an impression of three-dimensional shape from the set of two-dimensional views, where there is an infinite number of valid reconstructions. Tyler<sup>97</sup> brought psychophysical testing to the issue of portrait composition, and discovered that portrait painters systematically position one of the subject's eyes directly along the central median of the painting.

Another approach in this area is to understand the artistic process in order to emulate it algorithmically. The implementation, thus, is both a test of the model and a method for generating objects of an artistic process. For example, Burton<sup>98</sup> explored how young children express themselves visually and kinesthetically in drawings, and Brand<sup>99</sup> implemented his model of shape perception by teaching a robot to sculpt. Some of these goals are esoteric, but practical applications can be found. For example, Dalton<sup>100</sup> broadened the dialogue on digital libraries by exploring how algorithms could search for images that are the same, except for their artistic representation.

Also, as more and more artists, fabric designers, and graphic artists exploit the advances of electronic imaging systems, the more important it becomes to give them control over the aesthetic and emotional dimensions of electronic images. The MIT Media Lab has been a major player in exploring the visual, artistic, and emotional parameters of visual stimuli. In 1993, Feldman and Bender<sup>101</sup> conducted an experiment where users judged the affective aspects (e.g., "energy," "expressiveness") of color pairs, demonstrating that even such seemingly vague and subjective dimensions could be studied using psychophysical techniques. They found that color energy depended on the color distance between pairs in a calibrated Munsell color space, where complementary hues, or strongly different luminances or saturations, produced greater emotional energy. These authors and their colleagues have gone on to develop emotion-based color palettes and software tools for artistic design.

## 24.5 CONCLUSIONS

---

The field of human vision and electronic imaging has developed over the last decade, evolving with technology. In the 1980s, the bandwidth available to workstations allowed only the display of green text, and in those days people interested in perceptual issues in electronic imaging studied display flicker, spatial sampling, and gray-scale/resolution trade-offs. As the technology has evolved, almost all desktop displays now provide color, and the bandwidth allows for high-resolution, 24-bit images. These displays, and the faster systems that drive them, allow the development of desktop image applications, desktop digital libraries applications, and desktop virtual reality applications. Application developers interested in art, visualization, data mining, and image analysis can now build interactive image solutions, integrating images and data from multiple sources worldwide. This rapid increase in technology enables new types of imaging solutions, solutions that allow the user to explore, manipulate, and be immersed in images; and these in turn pose new questions to researchers interested in perceptually based imaging systems.

Some of these new questions include, for example: How do we measure the image quality of virtual reality environments? How do we measure presence, and what are the factors (including interactions of multiple media) that contribute to it? How do we model the interactive visualization of

multivariate data? How do we provide environments that allow people to search through a sea of images? How do we create tools that artists and designers can use? How do we use image technology to perform remote surgery? How do we build user interfaces that sense our mood? With each advance in technology, the systems we build grow increasingly richer, and the questions we ask about human observers become more complex. In particular, they increasingly require a deeper understanding of higher levels of human perception and cognition. Low-level models of retinal and striate cortex function are the foundation, but upon this base we need to consider other dimensions of the human experience: color perception, perceptual organization, language, memory, problem solving, aesthetic appreciation, and emotional response.

As electronic imaging becomes ever more prevalent in the home and the workplace, the two-way interaction between human vision and electronic imaging technology will continue to grow in scope and importance. Research in human vision and its interactions with the other senses will continue to shape the solutions that are developed to meet the requirements of the imaging industry, which in turn will continue to motivate our understanding of the complex process of vision that is the window to the world around us.

## 24.6 ADDITIONAL INFORMATION ON HUMAN VISION AND ELECTRONIC IMAGING

A number of archival journals and conferences have provided a venue for dissemination and discussion of research results in the field of human vision or in the field of electronic imaging. For human vision, these include the annual meeting of the Optical Society of America, the annual meeting of the Association for Research in Vision and Ophthalmology, and the European Conference on Visual Perception (ECVP). For electronic imaging, these include the IEEE Signal Processing Society's International Conference on Image Processing, the annual meeting of the Society for Information Display (SID), and several conferences sponsored or cosponsored by the Society for Imaging Science and Technology (IS&T), especially the International Conference on Digital Printing Technologies (NIP), the Color Imaging Conference (CIC), co-sponsored with SID, the Conference on Picture and Image Processing (PICS), and several conferences that are part of the Symposium on Electronic Imaging cosponsored by the Society of Photographic and Instrumentation Engineers (SPIE) and IS&T.

The major forum for human vision *and* electronic imaging is the Conference on Human Vision and Electronic Imaging cosponsored by the International Imaging Society (IS&T) and the Society for Photographic and Imaging Engineers (SPIE).

There are a number of excellent books on various aspects of human perception and electronic imaging, including *Human Color Vision* by Kaiser and Boynton,<sup>102</sup> *Foundations of Vision* by Wandell,<sup>103</sup> *Visual Perception* by Cornsweet,<sup>104</sup> *Spatial Vision* by De Valois and De Valois,<sup>105</sup> *The Senses*, edited by Barlow and Molon,<sup>106</sup> *The Artful Eye*, edited by Gregory, Harris, Rose, and Heard,<sup>107</sup> *How the Mind Works* by Pinker,<sup>108</sup> *Information Visualization: Optimizing Design For Human Perception*, edited by Colin Ware,<sup>109</sup> *Digital Image and Human Vision*, edited by Watson,<sup>110</sup> *Computational Models of Visual Processing*, edited by Landy and Movshon,<sup>111</sup> and *Early Vision and Beyond*, edited by Papathomas et al.<sup>112</sup>

Finally, we cite a number of review articles and book chapters.<sup>34,35,113</sup>

## 24.7 REFERENCES

1. G. W. Hughes, P. E. Mantey, and B. E. Rogowitz (eds.), *Image Processing, Analysis, Measurement, and Quality*, Proc. SPIE, vol. 901, Los Angeles, California, Jan. 13–15, 1988.
2. B. E. Rogowitz (ed.), *Human Vision, Visual Processing, and Digital Display*, Proc. SPIE, vol. 1077, Los Angeles, California, Jan. 18–20, 1989.
3. B. E. Rogowitz and J. P. Allebach (eds.), *Human Vision and Electronic Imaging: Models, Methods, and Applications*, Proc. SPIE, vol. 1249, Santa Clara, California, Feb. 12–14, 1990.



4. M. H. Brill (ed.), *Perceiving, Measuring, and Using Color*, *Proc. SPIE*, vol. 1250, Santa Clara, California, Feb. 15–16, 1990.
5. B. E. Rogowitz, M. H. Brill, and J. P. Allebach (eds.), *Human Vision, Visual Processing, and Digital Display II*, *Proc. SPIE*, vol. 1453, San Jose, California, Feb. 27–Mar. 1, 1991.
6. B. E. Rogowitz (ed.), *Human Vision, Visual Processing, and Digital Display III*, *Proc. SPIE*, vol. 1666, San Jose, California, Feb. 10–13, 1992.
7. J. P. Allebach and B. E. Rogowitz (eds.), *Human Vision, Visual Processing, and Digital Display IV*, *Proc. SPIE*, vol. 1913, San Jose, California, Feb. 1–4, 1993.
8. B. E. Rogowitz and J. P. Allebach (eds.), *Human Vision, Visual Processing, and Digital Display V*, *Proc. SPIE*, vol. 2179, San Jose, California, Feb. 8–10, 1994.
9. B. E. Rogowitz and I. P. Allebach (eds.), *Human Vision, Visual Processing, and Digital Display VI*, *Proc. SPIE*, vol. 2411, San Jose, California, Feb. 6–8, 1995.
10. B. E. Rogowitz and J. P. Allebach (eds.), *Human Vision and Electronic Imaging*, *Proc. SPIE*, vol. 2657, San Jose, California, Jan. 29–Feb. 1, 1996.
11. B. E. Rogowitz and T. N. Pappas (eds.), *Human Vision and Electronic Imaging II*, *Proc. SPIE*, vol. 3016, San Jose, California, Feb. 10–13, 1997.
12. B. E. Rogowitz and T. N. Pappas (eds.), *Human Vision and Electronic Imaging III*, *Proc. SPIE*, vol. 3299, San Jose, California, Jan. 26–29, 1998.
13. B. E. Rogowitz and T. N. Pappas (eds.), *Human Vision and Electronic Imaging TV*, *Proc. SPIE*, vol. 3644, San Jose, California, Jan. 25–28, 1999.
14. B. E. Rogowitz and T. N. Pappas (eds.), *Human Vision and Electronic Imaging V*, *Proc. SPIE*, vol. 3959, San Jose, California, Jan. 24–27, 2000.
15. O. H. Schade, “Optical and Photoelectric Analog of the Eye,” *Journal of the Optical Society of America* **46**:721–739 (1956).
16. F. W. Campbell and J. G. Robson, “Application of Fourier Analysis to the Visibility of Gratings,” *Journal of Physiology* **197**:551–566 (1968).
17. N. Graham and J. Nachmias, “Detection of Grating Patterns Containing Two Spatial Frequencies: A Comparison of Single-Channel and Multiple-Channels Models,” *Vision Research* **11**:252–259 (1971).
18. H. Snyder, “Image Quality and Observer Performance in Perception of Displayed Information,” in *Perception of Displayed Information*, L. M. Bieberman (ed.), Plenum Press, New York, 1973.
19. P. G. J. Barten, “The SQRI Method: A New Method for the Evaluation of Visible Resolution on a Display,” *Proceedings of the Society for Information Display*, **28**:253–262 (1987).
20. S. Daly, “The Visible Differences Predictor: An Algorithm for the Assessment of Image Fidelity,” in A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, MA, pp. 179–206, 1993.
21. J. Lubin, “The Use of Psychophysical Data and Models in the Analysis of Display System Performance,” in A. B. Watson (ed.), *Digital Images and Human Vision*, MIT Press, Cambridge, MA, pp. 163–178, 1993.
22. V. Ramamoorthy and N. S. Jayant, “On Transparent Quality Image Coding Using Visual Models,” *Proc. SPIE*, **1077**:146–154, 1989.
23. S. Daly, “The Visible Difference Predictor: An Algorithm for the Assessment of Image Fidelity,” *Proc. SPIE*, **1077**:209–216, 1989.
24. A. B. Watson, “Receptive Fields and Visual Representations,” *Proc. SPIE*, **1077**:190–197, 1989.
25. R. J. Safranek and J. D. Johnston, “A Perceptually Tuned Sub-Band Image Coder with Image Dependent Quantization and Post-Quantization Data Compression,” in *Proc. ICASSP-89*, vol. 3, Glasgow, Scotland, pp. 1945–1948, May 1989.
26. H. A. Peterson, A. J. Ahumada, and A. B. Watson, “Improved Detection Model for DCT Coefficient Quantization,” *Proc. SPIE*, **1913**:191–201, 1993.
27. A. B. Watson, “DCT Quantization Matrices Visually Optimized for Individual Images,” *Proc. SPIE*, **1913**:202–216, 1993.
28. D. A. Silverstein and S. A. Klein, “DCT Image Fidelity Metric and its Application to a Text Based Scheme for Image Display,” *Proc. SPIE*, **1913**:229–239, 1993.

29. C. J. Van den Branden Lambrecht and O. Verscheure, "Perceptual Quality Measure Using a Spatio-Temporal Model of the Human Visual System," in V. Bhaskaran, F. Sijstermans, and S. Panchanathan (eds.), *Digital Video Compression: Algorithms and Technologies*, Proc. SPIE, vol. 2668, San Jose, California, pp. 450–461, Jan./Feb. 1996.
30. A. B. Watson, "Toward a Perceptual Visual Quality Metric," Proc. SPIE, **3299**:139–147, 1998.
31. A. B. Watson, Q. J. Hu, J. F. McGowan, and J. B. Mulligan, "Design and Performance of a Digital Video Quality Metric," Proc. SPIE, **3644**:168–174, 1999.
32. P. J. Corriveau, A. A. Webster, A. M. Rohaly, and J. M. Liebert, "Video Quality Experts Group: The Quest for Valid and Objective Methods," Proc. SPIE, **3959**:129–139, 2000.
33. T. Carney, C. W. Tyler, A. B. Watson, W. Makous, B. Beutter, C. Chen, A. M. Norcia, and S. A. Klein, "Modelfest: Year One Results and Plans for Future Years," Proc. SPIE, **3959**:140–151, 2000.
34. M. P. Eckert and A. P. Bradley, "Perceptual Quality Metrics Applied to Still Image Compression," *Signal Processing* **70**:177–200 (1998).
35. T. N. Pappas and R. J. Safranek, "Perceptual Criteria for Image Quality Evaluation," in *Handbook of Image and Video Processing*, A. C. Bovik (ed.), Academic Press, New York, pp. 669–684, 2000.
36. J. Sullivan, L. Ray, and R. Miller, "Design of Minimum Visual Modulation Halftone Patterns," *IEEE Transactions on Systems, Man, and Cybernetics* **21**:33–38 (Jan./Feb. 1991).
37. T. N. Pappas and D. L. Neuhoff, "Printer Models and Error Diffusion," *IEEE Trans. Image Process* **4**:66–79 (1995).
38. M. Analoui and J. P. Allebach, "Model Based Halftoning Using Direct Binary Search," Proc. SPIE, **1666**:96–108, 1992.
39. J. B. Mulligan and A. J. Ahumada, "Principled Halftoning Based on Human Vision Models," Proc. SPIE, **1666**:109–121, 1992.
40. T. N. Pappas and D. L. Neuhoff, "Least Squares Model Based Halftoning," Proc. SPIE, **1666**:165–176, 1992.
41. T. J. Flohr, B. W. Kolpatzik, R. Balasubramanian, D. A. Carrara, C. A. Bouman, and J. P. Allebach, "Model Based Color Image Quantization," Proc. SPIE, **1913**:270–281, 1993.
42. C. B. Atkins, T. J. Flohr, D. P. Hilgenberg, C. A. Bouman, and J. P. Allebach, "Model Based Color Image Sequence Quantization," Proc. SPIE, **2179**:318–326, 1994.
43. J. B. Mulligan, "Digital Halftoning Methods for Selectively Partitioning Error into Achromatic and Chromatic Channels," Proc. SPIE, **1249**:261–270, 1990.
44. M. P. Eckstein, A. J. Ahumada, and A. B. Watson, "Image Discrimination Models Predict Signal Detection in Natural Medical Image Backgrounds," Proc. SPIE, **1316**:58–69, 1997.
45. I. J. Cox and M. L. Miller, "Review of Watermarking and the Importance of Perceptual Modeling," Proc. SPIE, **1316**:92–99, 1997.
46. C. I. Podilchuk and W. Zeng, "Digital Image Watermarking Using Visual Models," Proc. SPIE, **1316**:100–111, 1997.
47. S. L. Guth, "Unified Model for Human Color Perception and Visual Adaptation," Proc. SPIE, **1077**:370–390, 1989.
48. S. L. Guth, "Unified Model for Human Color Perception and Visual Adaptation II," Proc. SPIE, **1913**:440–448, 1993.
49. E. M. Granger, "Uniform Color Space as a Function of Spatial Frequency," Proc. SPIE, **1913**:449–461, 1993.
50. E. M. Uriegas, J. D. Peters, and H. D. Crane, "Comparison of Digital Color Images Based the Model of Spatiochromatic Multiplexing of Human Vision," Proc. SPIE, **2179**:400–406, 1994.
51. E. M. Uriegas, J. D. Peters, and H. D. Crane, "Spatiotemporal Multiplexing: A Color Image Representation for Digital Processing and Compression," Proc. SPIE, **2657**:412–420, 1996.
52. J. Gille, J. Luszcz, and J. O. Larimer, "Error Diffusion Using the Web-Safe Colors: How Good Is it Across Platforms?" Proc. SPIE, **3299**:368–375, 1998.
53. H. de Ridder and G. M. Majoor, "Numerical Category Scaling: An Efficient Method for Assessing Digital Image Coding Impairments," Proc. SPIE, **1249**:65–77, 1990.
54. J. A. Roufs and M. C. Boschman, "Methods for Evaluating the Perceptual Quality of VDUs," Proc. SPIE, **1249**:2–11, 1990.
55. D. E. Pearson, "Viewer Response to Time Varying Video Quality," Proc. SPIE, **3299**:2–15, 1998.



56. R. W. G. Hunt, *The Reproduction of Colour in Photography, Printing and Television*. Fountain Press, England, 1987.
57. J. J. McCann, "Color Imaging System and Color Theory: Past, Present and Future," *Proc. SPIE*, **3299**:38–46, 1998.
58. C. Zetzsche and E. Barth, "Image Surface Predicates and the Neural Encoding of Two Dimensional Signal Variations," *Proc. SPIE*, **1249**:160–177, 1990.
59. M. A. Webster and O. H. MacLin, "Visual Adaptation and the Perception of Distortions in Natural Images," *Proc. SPIE*, **3299**:264–273, 1998.
60. D. J. Field, "What the Statistics of Natural Images Tell us about Visual Coding," *Proc. SPIE*, **1077**:269–276, 1989.
61. D. J. Field, "Relations Between the Statistics of Natural Images and the Response Properties of Cortical Cells," *Journal of the Optical Society of America A* **4**:2379–2394 (1987).
62. D. J. Field, B. A. Olshausen, and N. Brady, "Wavelets, Blur and the Sources of Variability in the Amplitude Spectra of Natural Scenes," *Proc. SPIE*, **2657**:108–119, 1996.
63. B. E. Rogowitz and R. Voss, "Shape Perception and Low Dimension Fractal Boundary Contours," *Proc. SPIE*, **1249**:387–394, 1990.
64. L. T. Maloney, "Photoreceptor Spectral Sensitivities and Color Correction," *Proc. SPIE*, **1250**:103–110, 1990.
65. G. D. Finlayson, "Color Constancy and a Changing Illumination," *Proc. SPIE*, **2179**:352–363, 1994.
66. D. H. Brainard and W. T. Freeman, "Bayesian Method for Recovering Surface and Illuminant Properties from Photosensor Responses," *Proc. SPIE*, **2179**:364–376, 1994.
67. A. Treisman and G. Gelade, "A Feature Integration Theory of Attention," *Cognitive Psychology* **12**:97–136 (1980).
68. B. Julesz, "AI and Early Vision II," *Proc. SPIE*, **1077**:246–268, 1989.
69. B. Julesz, "Textons, the Elements of Texture Perception and Their Interactions," *Nature* **290**:91–97 (1981).
70. D. Marr, *Vision*. W. H. Freeman Company, San Francisco, CA, 1982.
71. D. Noton and L. Stark, "Eye Movements and Visual Perception," *Scientific American* **224**(6):34–43 (1971).
72. P. Leray, F. Guyot, P. Marchal, and Y. Burnod, "CUBICORT: Simulation of the Visual Cortical System for Three Dimensional Image Analysis, Synthesis, and Hypercompression for Digital TV, HDTV and Multimedia," *Proc. SPIE*, **2179**:247–258, 1994.
73. L. B. Stelmach and W. J. Tam, "Processing Image Sequences Based on Eye Movements," *Proc. SPIE*, **2179**:90–98, 1994.
74. W. Geisler and J. S. Perry, "Real Time Foveated Multiresolution System for Low Bandwidth Video Communication," *Proc. SPIE*, **3299**:294–305, 1997.
75. A. Yarbus, *Eye Movements and Vision*. Plenum Press, New York, 1967.
76. A. M. Liu, G. K. Tharp, and L. Stark, "Depth Cue Interaction in Telepresence and Simulated Telemanipulation," *Proc. SPIE*, **1666**:541–547, 1992.
77. L. Stark, H. Yang, and M. Azzariti, "Symbolic Binding," *Proc. SPIE*, **3959**:254–267, 2000.
78. D. Noton and L. Stark, "Scanpaths in Saccadic Eye Movements While Viewing and Recognizing Patterns," *Vision Research* **11**:929–942 (1971).
79. J. B. Pelz, R. L. Canosa, D. Kucharczyk, J. S. Babcock, A. Silver, and D. Konno, "Portable Eye Tracking: A Study of Natural Eye Movements," *Proc. SPIE*, **3959**:566–582, 2000.
80. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by Image and Video Content; the QBIC System," *Computer* **28**:23–32 (Sept. 1995).
81. T. Frese, C. A. Bouman, and J. P. Allebach, "Methodology for Designing Image Similarity Metrics Based on Human Visual System Models," *Proc. SPIE*, **3016**:472–483, 1997.
82. B. E. Rogowitz, T. frees, J. R. Smith, C. A. Bouman, and E. B. Kalin, "Perceptual Image Similarity Experiments," *Proc. SPIE*, **3299**:576–590, 1998.
83. A. Mojsilovic, J. Kovacevic, J. Hu, R. J. Safranek, and S. K. Ganapathy, "Retrieval of Color Patterns Based on Perceptual Dimensions of Texture and Human Similarity Rules," *Proc. SPIE*, **3644**:441–452, 2000.
84. B. E. Rogowitz and L. Treinish, "Data Visualization: The End of the Rainbow," *IEEE Spectrum*, 52–59 (Dec. 1998).

85. M. Kubovy, "Gestalt Laws of Grouping Revisited and Quantified," *Proc. SPIE*, **3016**:402–408, 1997.
86. A. K. Hon, L. T. Maloney, and M. S. Landy, "Influence Function for Visual Interpolation," *Proc. SPIE*, **3016**:409–419, 1997.
87. B. E. Rogowitz, D. A. Rabenhorst, J. A. Garth, and E. B. Kalin, "Visual Cues for Data Mining," *Proc. SPIE*, **2657**:275–300, 1996.
88. R. M. Boynton, "Eleven Colors that are Almost Never Confused," *Proc. SPIE*, **1077**:322–331, 1989.
89. G. A. M. Derefeldt and T. Swartling, "How to Identify upto 30 Colors without Training: Color Concept Retrieval by Free Color Naming," *Proc. SPIE*, **2179**:418–428, 1994.
90. S. N. Yendrikhoviski, "Computing Color Categories," *Proc. SPIE*, **3959**:356–364, 2000.
91. C. E. Hedin and G. A. M. Derefeldt, "Palette: A Color Selection Aid for VDU Displays," *Proc. SPIE*, **1250**:165–176, 1990.
92. B. E. Rogowitz, D. A. Rabenhorst, J. A. Garth, and E. B. Kalin, "Visual Cues for Data Mining," *Proc. SPIE*, **2657**:275–300, 1996.
93. H. Kusaka, "Apparent Depth and Size of Stereoscopically Viewed Images," *Proc. SPIE*, **1666**:476–482, 1992.
94. H. de Ridder, F. J. J. Blommaert, and E. A. Fedorovskaya, "Naturalness and Image Quality: Chroma and Hue Variation in Color Images of Natural Images," *Proc. SPIE*, **2411**:51–61, 1995.
95. T. V. Pappathomas, "See How They Turn: False Depth and Motion in Hughes's Reverspectives," *Proc. SPIE*, **3959**:506–517, 2000.
96. J. J. Koenderink, A. J. van Doorn, A. M. L. Kappers, and J. T. Todd, "Directing the Mental Eye in Pictorial Perception," *Proc. SPIE*, **3959**:2–13, 2000.
97. C. W. Tyler, "Eye Placement Principles in Portraits and Figure Studies over the Past Two Millennia," *Proc. SPIE*, **3299**:431–438, 1998.
98. E. Burton, "Seeing and Scribbling: A Computer Representation of the Relationship Between Perception and Action in Young Children's Drawings," *Proc. SPIE*, **3016**:314–323, 1998.
99. M. Brand, "Computer Vision for a Robot Sculptor," *Proc. SPIE*, **3016**:508–516, 1997.
100. J. C. Dalton, "Image Similarity Modes and the Perception of Artistic Representations of Natural Images," *Proc. SPIE*, **3016**:517–525, 1997.
101. U. Feldman, N. Jacobson, and W. R. Bender, "Quantifying the Experience of Color," *Proc. SPIE*, **1913**:537–547, 1993.
102. P. K. Kaiser and R. M. Boynton, *Human Color Vision*. Optical Society of America, Washington, DC.
103. B. A. Wandell, *Foundations of Vision*. Sinauer, Sunderland, MA, 1995.
104. T. N. Cornsweet, *Visual Perception*. Academic Press, New York, 1970.
105. R. L. D. Valois and K. K. D. Valois, *Spatial Vision*. Oxford University Press, New York, 1990.
106. H. B. Barlow and J. D. Molon (eds.), *The Senses*. Cambridge University Press, Cambridge, 1982.
107. R. L. Gregory, J. Harris, D. Rose, and P. Heard (eds.), *The Artful Eye*. Oxford University Press, Oxford, 1995.
108. S. Pinker, *How the Mind Works*. Norton, New York, 1999.
109. C. Ware (ed.), *Information Visualization: Optimizing Design For Human Perception*. Morgan Kaufmann Publishers, San Francisco, CA, 1999.
110. A. B. Watson (ed.), *Digital Image and Human Vision*. MIT Press, Cambridge, MA, 1993.
111. M. S. Landy and J. A. Movshon (eds.), *Computational Models of Visual Processing*. MIT Press, Cambridge, MA, 1991.
112. T. V. Pappathomas, C. Chubb, A. Gorea, and E. Kowler (eds.), *Early Vision and Beyond*. MIT Press, Cambridge, MA, 1995.
113. B. E. Rogowitz, "The Human Visual System: A Guide for the Display Technologist," in *Proceedings of the SID*, vol. 24/3, pp. 235–252, 1983.

*This page intentionally left blank*

---

# VISUAL FACTORS ASSOCIATED WITH HEAD-MOUNTED DISPLAYS

---

**Brian H. Tsou**

*Air Force Research Laboratory  
Wright Patterson AFB, Ohio*

**Martin Shenker**

*Martin Shenker Optical Design, Inc.  
White Plains, New York*

---

## 25.1 GLOSSARY

---

**Aniseikonia.** Unequal right and left retinal image sizes. (Also see Chaps. 12 and 13.)

**C, F lines.** Hydrogen lines at the wavelengths of 656 and 486 nm, respectively.

**D (diopter).** A unit of lens power; the reciprocal of focal distance in m. (Also see Chap. 12.)

**Prism diopter.** A unit of angle; 100 times the tangent of the angle. (Also see Chaps. 12 and 13.)

---

## 25.2 INTRODUCTION

---

Some virtual reality applications require a head-mounted display (HMD). Proper implementation of a binocular and lightweight HMD is quite a challenge; many aspects of human-machine interaction must be considered when designing such complex visual interfaces. We learned that it is essential to couple the knowledge of visual psychophysics with sound human/optical engineering techniques when making HMDs. This chapter highlights some visual considerations necessary for the successful integration of a wide field-of-view HMD.

---

## 25.3 COMMON DESIGN CONSIDERATIONS AMONG ALL HMDS

---

A wide field-of-view HMD coupled with a gimbaled image-intensifying (see Chap. 31 of Vol. II for details) or infrared (see Chap. 33 of Vol. II for details) sensor enables helicopter pilots to fly reconnaissance, attack, or search-and-rescue missions regardless of weather conditions. Before any HMD

can be built to specifications derived from mission requirements, certain necessary display performance factors should be considered. These factors include functionality, comfort, usability, and safety. A partial list of subfactors that affect vision follows:

Functionality	Comfort	Usability	Safety
Field of view	Eye line-of-sight/focus	Eye motion box	Center-of-mass
Image quality	Alignment	Range of adjustment	Weight

### Safety

An analytic approach to considering these factors has been proposed.<sup>1</sup> In general, factors can be weighted according to specific needs. At the same time, safety—especially in the context of a military operation—cannot be compromised.<sup>2</sup> The preliminary weight and center-of-mass safety requirements for helicopters<sup>3</sup> and jets<sup>4</sup> have been published.

It is extremely difficult to achieve the center-of-mass requirement. Hanging eye lenses in front of one’s face necessitates tipping the center-of-mass forward. A recent survey disclosed that pilots routinely resort to using counterweights to balance the center-of-mass of popular night-vision goggles<sup>5</sup> in order to obtain stability and comfort, even at the expense of carrying more weight.

Light yet sturdy optics and mechanical materials have already made the HMD lighter than was previously possible. In the future, miniature flat-panel displays (see Chap. 22 in this volume, Chaps. 17 and 19 in Vol. II, and Chap. 8 in Vol. V for details) and microelectromechanical system (MEMS) technology<sup>6</sup> might further lessen the total head-borne weight. Previous experience<sup>7</sup> shows that both the field-of-view and exit pupil have a significant impact on HMD weight. Although we can calculate<sup>1</sup> the eye motion box [see Eq. (1)] that defines the display exit pupil requirement, we still do not have a quantitative field-of-view model for predicting its optimum value.

### Usability

In order to avoid any vignetting, the display exit pupil size should match the eye motion box. The eye motion box is the sum of lateral translation of the eye and the eye pupil projection for viewing the full field of view (FOV). We assume the radius of eye rotation is 10 mm.

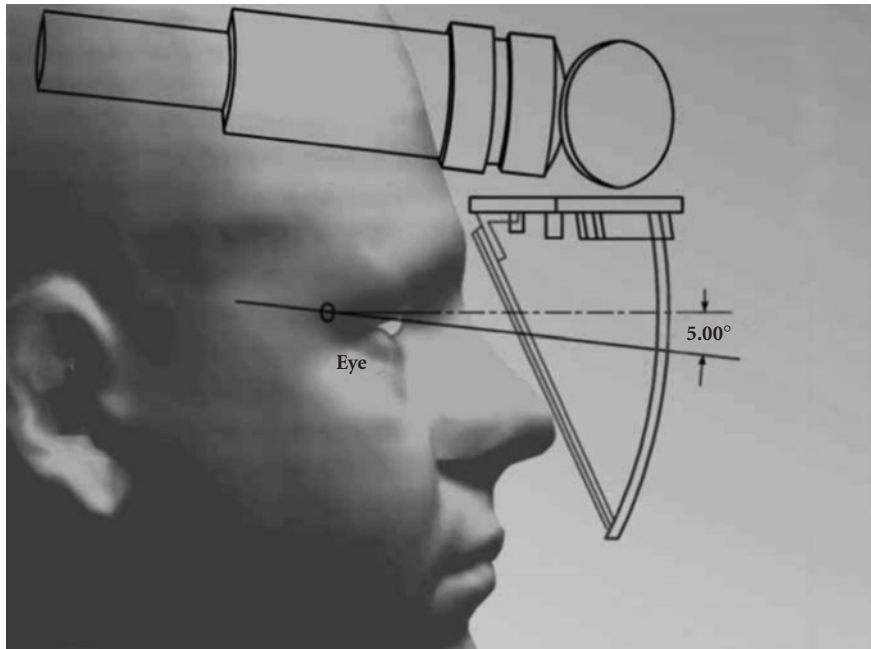
$$\begin{aligned} \text{Eye motion box in mm} \\ = 2 \times 10 \times \sin(\text{FOV}/2) + \text{eye pupil diameter} \times \cos(\text{FOV}/2) + \text{helmet slippage} \end{aligned} \tag{1}$$

The range of pertinent optomechanical HMD adjustments for helicopter aviators<sup>8</sup> is listed as minimum-maximum in millimeters in the following table:

	Interpupillary Distance	Eye to Top of Head	Eye to Back of Head
Female	53–69	107.7–141.7	152.5–190.0
Male	56–75	115.1–148.2	164.4–195.4

### Comfort

Levy<sup>9</sup> shows that, in total darkness, the eye line-of-sight drifts, on average, about 5° down from the horizon. Figure 1 shows the geometric relationship of the visual axis to the horizon. Levy argues that the physiological position of rest of the eye orbit is downward pointing. Menozzi et al.<sup>10</sup> confirm that



**FIGURE 1** Geometric layout of the HMD showing the visual axis being  $5^\circ$  down from the horizon, as discussed in the text. Also shown are the schematic depictions of miniature CRT, folding mirror, beam-splitter, and spherical combiner. Its prescription is given in Table 1.

maximum comfort is in fact achieved when line of sight is pointed downward. The viewing comfort was determined by the perceived muscle exertion during gaze in a given direction for a given time. The median line of sight for the most comfortable gaze at 1 m is around  $10^\circ$  downward.

Our laboratory has also examined how normal ocular-motor behavior might improve the comfort of using any HMD. In a short-term wear study,<sup>11</sup> Gleason found that a range of eyepiece lens power of AN/AVS-6 Aviator Night Vision Imaging Systems (ANVIS) produced comparable visual acuity independent of luminance and contrast. The single best-overall eyepiece lens power produced visual acuity equal to, or better than, that of subject-adjusted eyepiece lens power, producing visual acuity within 2 percent of optimal. Infinity-focused eyepieces made visual acuity worse, reducing it by 10 percent. In his long-term wear (4 hours) study,  $-1.5$  diopter (D) eyepiece lens power caused half of the subjects ( $n = 12$ ) to complain of blurred or uncomfortable vision. These studies indicate that those users, who are optically corrected to a “most plus-best binocular visual acuity” endpoint (see Sec. 25.6, “Appendix”), achieve satisfactory comfort and near optimal binocular visual acuity for extended ANVIS viewing when an eyepiece lens power of approximately  $-0.75$  D is added to the clinical refraction. This result, consistent with an earlier report by Home and Poole,<sup>12</sup> may extend to other binocular visual displays.

Alignment tolerances<sup>13</sup> are the maximum permissible angular deviation between the optical axes of a binocular device that displays separate targets to the two eyes (see Chap. 13, “Binocular Vision Factors That Influence Optical Design,” in this volume for general issues). The ocular targets may be either identical or dichoptic (for stereoscopic presentation) but will be treated differently depending on whether there is a requirement for superimposition of the targets onto a direct view of the environment. If the direct view is either blocked out or too dark to see (e.g., flying at night), then it is classified as a nontransparent or closed HMD. A closed HMD resembles immersive virtual reality simulation where the real world is not needed.

**Closed HMD** Normal subjects have a substantial tolerance for angular deviation between the images in the two eyes for the three degrees of freedom of eye rotation: horizontal, vertical, and cyclorotational. These limits are especially of interest to optometrists and ophthalmologists. The optic axes of spectacle lenses must be aligned with the visual axes of the eyes of the wearer. Displacement of a spectacle lens from a centered position on the visual axis of the eye introduces a prismatic deviation that is entirely analogous to the misalignment of binocular devices. The American National Standards Institute (ANSI) has published ANSI 280.1–1987<sup>14</sup> and permits a maximum deviation of  $\frac{2}{3}$  prism diopter (23 arc minutes) horizontally and  $\frac{1}{3}$  prism diopter (11.5 arc minutes) vertically. This standard can be adopted for closed binocular HMD. These tolerance values are summarized as follows:

Recommended Tolerances for Closed HMD

	Tolerances (arc minutes)
Horizontal	±23
Vertical	±11.5
Cyclorotational	±12

Regarding the cyclorotational tolerances, both ophthalmic lenses and binocular displays can produce rotational deviation of the images about the line of fixation. ANSI spectacle standards do not exist for this alignment axis. Earlier research (dealing with distortion of stereoscopic spatial localization resulting from meridional aniseikonia at oblique axes) shows that there is considerable tolerance for cyclotorsional rotations of the images.<sup>15</sup> However, the effects of cyclorotational misalignment are complicated by the fact that these rotations may result in either an oculomotor (cyclofusional) or sensory (stereoscopic) response. If the eyes do not make complete compensatory cyclorotations, a declination error (rotational disparity) will exist. Declination errors result in an apparent inclination (rotation about a horizontal axis) of the display; that is, the display will appear with an inappropriate pitch orientation.

Sensitivity to declination (stereoscopic response to rotational disparity) is quite acute. Ogle<sup>15</sup> reports normal threshold values of  $\pm 6$  arc minutes, which corresponds to an object inclination of  $5^\circ$  at 3 m distance. If this threshold value were adopted for the cyclotorsion tolerance measurement, it would be overly conservative, as it would deny any reflex cyclofusional capacity to compensate for misalignment errors. A 97 percent threshold<sup>15</sup> ( $\pm 12$  arc minutes) is suggested for adoption as the cyclofusional tolerance.

**Transparent HMD** Transparent HMD optically superimposes a second image upon the directly viewed image using a beam combiner (see Sec. 25.6, “Appendix”). The difference in alignment angles between the direct and superimposed images equals the binocular disparity between target pairs.<sup>16</sup> Horizontal binocular disparity is the stimulus to binocular stereoscopic vision. Consequently, sufficient lateral misalignment will lead to the perception of a separation in depth between the direct and superimposed images.

The exact nature of the sensory experience arising from lateral misalignment depends on the magnitude of the relative disparity. Although the threshold disparity for stereopsis is generally placed at about 10 arc seconds for vertical rods,<sup>17</sup> it can approach 2 arc seconds under optimal conditions using greater than 25 arc-minutes-long vertical line targets.<sup>18</sup> The normative value of threshold stereopsis in the standardized Howard-Dolman testing apparatus is 16 arc seconds. This defines 100 percent stereoacuity for clinical-legal requirements.

This extremely small stereothreshold ( $\sim 2$  arc seconds for optimal conditions) establishes an unpractically small tolerance for the design of displays. However, this precision is only required if the relative depths of the direct and superimposed fields need to correspond and the direct and superimposed fields are dissimilar (nonfusible). An example would be the need to superimpose a fiduciary marker over a particular location in the direct field, which is important to military operations. If this is not a requirement, the binocular disparity between the direct and superimposed fields merely needs to not exceed the range for pyknostereopsis in order to permit sensory averaging of the fields. When two

views are overlaid with small disparity separations ( $<20$  arc minutes), they are depth averaged to form a single depth plane (called *pyknostereopsis*).<sup>19,20</sup> This sets an alignment tolerance of  $\pm 20$  arc minutes (i.e., a maximum binocular disparity of 20 arc minutes). This limit is also within the disparity range required for maximum efficacy of oculomotor alignment of the eyes.<sup>21</sup> Because of the poor sensitivity of the human visual system to absolute disparity,<sup>22</sup> this amount of oculomotor misalignment will not influence the perception of absolute distances of objects in the display.

Vertical disparities do not result in local stereopsis, but they can have an effect on the overall stereoscopic orientation of objects in the binocular field. This response to vertical disparity is known as the *induced effect*. The sensitivity to vertical disparity for extended targets is similar to that of the horizontal disparity.<sup>23</sup> The other factor to consider when establishing vertical tolerances is the dimension of Panum's fusional area. Images falling within the Panum's area will fuse even if these images do not match exactly on the corresponding points between the two eyes. The vertical dimension of Panum's area can be as much as 2.5 times smaller than the horizontal dimension.<sup>24</sup>

The recommended cyclorotational tolerance is based on Ogle's<sup>15</sup> normal threshold value of  $\pm 6$  arc minutes, as discussed in the previous section. Because reflex cyclofusional would be restricted in transparent HMD by the direct field, no relaxation for compensatory eye movements has been included in this tolerance recommendation.

In formulating the final recommendations, the range of *pyknostereopsis* was reduced by half to introduce a margin for individual differences. Summary recommendations for tolerance in transparent HMD are provided as follows:

Recommended Tolerances for Transparent HMD

	Tolerance (arc minutes)
Horizontal	$\pm 10$
Vertical	$\pm 4$
Cyclorotational	$\pm 6$

## Functionality

A larger field of view equates with larger optics and, hence, a heavier piece of equipment. As early as 1962, optical engineers partially overlapped the individual telescopes to achieve a wider field of view without greatly increasing weight.<sup>25</sup> The impact of partial overlap on image quality, however, was not fully addressed until recently.<sup>26,27</sup> These psychophysical studies on binocular summation, rivalry, and stereopsis support a  $40^\circ$  binocular overlap in the middle of the field of view as both necessary and sufficient, with a divergent configuration of the two optical axes preferred over a convergent configuration (please note the visual axes are still parallel or convergent). Quantitative flight data (eye gaze) and pilots' subjective assessment are consistent with laboratory predictions.<sup>28</sup>

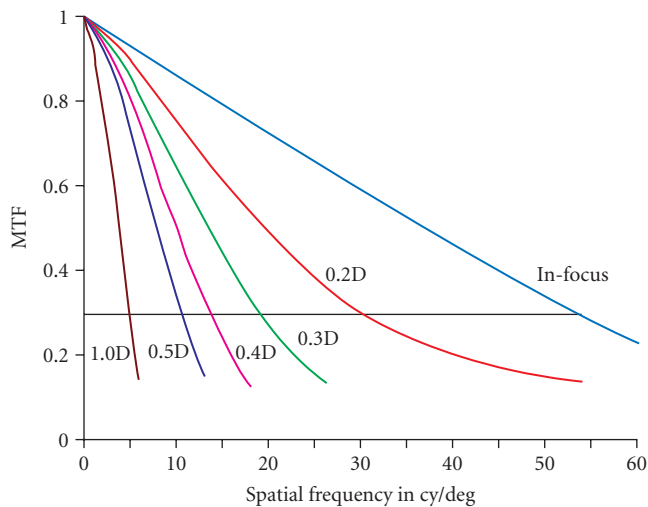
Still, how large should the total field of view be? Everyday experience tells us that it should be as large as our visual field. On the other hand, most studies consistently show there is no significant human performance gain as the field of view extends beyond the  $40^\circ$  to  $60^\circ$  range.<sup>29</sup> Restricting head movements or manipulating field-of-view aspect ratio<sup>30</sup> does not seem to affect the performance except for the smaller field of view (from  $15^\circ$  to  $30^\circ$ ). These studies suggest that (1) we are basically foveated beings,<sup>31–33</sup> and (2) unless compensated by large target size, our peripheral sensitivity is very poor.<sup>34</sup> Nevertheless, quantitative models for the useful field-of-view<sup>35,36</sup> requirement are urgently needed so that HMD field of view is optimized to suit each application and scenario.

Image quality is perhaps the most important parameter for the acceptance of HMD since it is the aspect most noticeable by the users. Various image quality metrics have been devised (see Chap. 1 of this volume for reviews). The fundamental limitation on image quality in any display is either the loss of resolution, or lack thereof. Historically, the poor resolution of the miniature image source has hampered HMD progress; soon, the expected higher pixel density should significantly improve perceived image quality. For example, with high-definition TV's approximate 2000 pixels, a  $50^\circ$  field-of-view

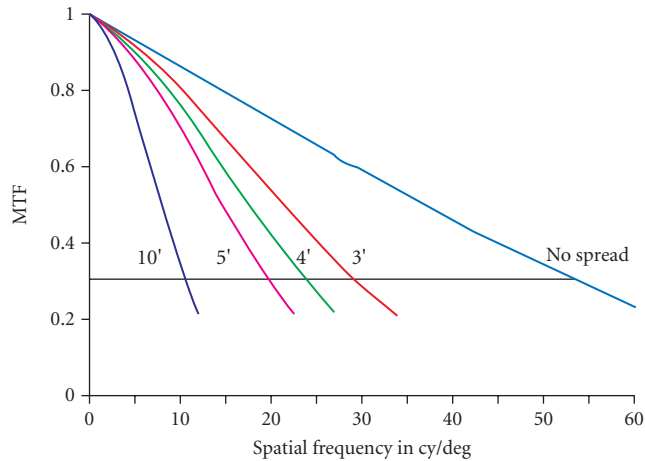


HMD will have 20 cy/deg resolution or a respectable 20/30. The loss of resolution in visual optics is primarily caused by the defocus and lateral chromatic smearing of an image point.<sup>37</sup> When the eye is centered in the exit pupil of a display, defocus is due to residual field curvature and astigmatism. As the eye is decentered (or off aperture) in the exit pupil, residual amounts of uncorrected spherical aberration and coma in the display cause additional astigmatism. Astigmatism is defined as the variation in focus as a function of orientation in the exit pupil. Recently, Mouroulis and Zhang<sup>38</sup> have shown that MTFa (the area under the modulation transfer function over a specified frequency range of from 5 to 24 cy/deg) gives good correlation with subjective sharpness judgment for all target orientations. There is another convenient approximation to use for image quality. Figure 2 shows plots of the variation in modulation transfer function (MTF) (see Chap. 1 of Vol. II for reviews) for a 3-mm aberration-free pupil out to a frequency of 1 arc minute per line pair (or 20/10) as its image is defocused from the best focus position. A 3-mm pupil has been chosen because the eye is generally considered to be diffraction-limited in image quality near this value; we conclude that for “each tenth of diopter defocus 1 arc minute of resolution is lost” (see Fig. 2). We arrived at the same conclusion using Greivendamp’s schematic eye.<sup>39</sup> Greivendamp et al. showed their model correlates quite well with Snellen visual acuity. It should be noted that, for small defocus errors (<0.5 D), a real eye can see slightly better than our calculation predicts.<sup>40</sup> This may be attributable to the residual spherical aberration in the real eye. While this spherical aberration lowers the MTF for the exact in-focus image, it increases the depth-of-focus over which the maximum resolution is visible. Regardless, our rule of thumb is quite useful for the purpose of design evaluation.

Similarly, we approximate the residual chromatic smear as the sum of the lateral color plus the off-aperture manifestation of the residual longitudinal color by using a pair of wedges with zero deviation at a middle wavelength but a specific chromatic spread between the C and F spectral lines. The loss of MTF for a 3-mm, otherwise aberration-free exit pupil is shown in Fig. 3. It can be seen that these defects have a very significant effect on the resolution capabilities at both the high and low spatial frequencies. Therefore, a rule of thumb of “ $\frac{2}{3}$  arc minute of resolution is lost for each arc minute of chromatic spread between C and F lines” is applicable.



**FIGURE 2** Loss of MTF in terms of resolution with defocus. The diopter values are the amount of defocus. At 30 percent modulation, the resolutions are about 1.1 arc minutes per line pair for in-focus, 2 for 0.2 D defocus, 3 for 0.3 D, 4.5 for 0.4 D, 6 for 0.5 D, and 12 for 1.0 D, respectively.



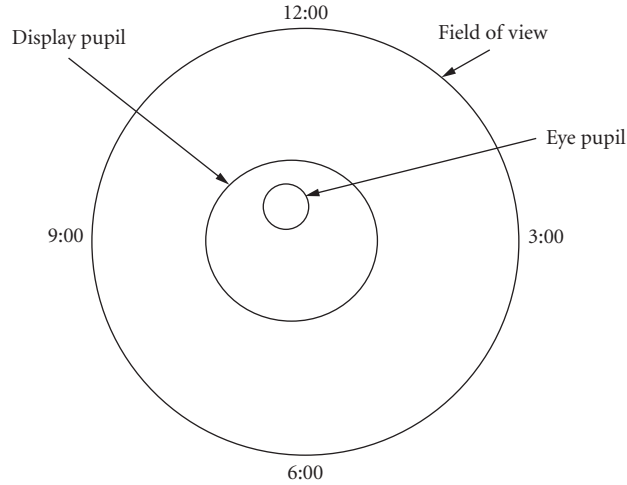
**FIGURE 3** Loss of MTF with chromatic smear. The arc minute values represent smear between C and F lines. At 30 percent modulation, the resolutions are about 1.1 arc minutes per line pair for no chromatic spread, 2 for 3 arc minutes spread, 2.5 for 4', 3.1 for 5', and 6 for 10', respectively.

## 25.4 CHARACTERIZING HMD

Traditionally, we design and evaluate the image at the focal plane of photographic lenses with a plane, aberration-free wavefront, entering the entrance pupil of the objective. This is likely the arrangement in which the photographic lens will be utilized. The design of visual optics is usually performed in a similar manner with a perfect wavefront at the exit pupil of the display (which is defined as the entrance pupil of the eye in the design process). However, the use and the real performance of such a system is best defined, from a visual point of view, by ray tracing from a perfect object and evaluating the nature of the light exiting the system. There are at least two characteristics of this light that are of interest to the display designer.

The first of these is the direction of a ray that is defined as intersecting the center of the eye pupil that is viewing a particular point on the object. This datum is used for evaluation of the system magnification (i.e., mapping and distortion). If the system is supposed to be displaying an image projected to infinity (collimated display), then the apparent direction of a point in the field of view should not change as the eyes move relative to the display. Any deviation of this ray direction across the field of view will bring about an apparent change of the display distortion as well. This variation of the display distortion is often referred to as “dynamic distortion” or “swimming.” Next, if all the rays over the small (3 to 4 mm) entrance pupil of the eye are in focus, a relatively well-corrected display will demonstrate essentially diffraction-limited performance. However, if the eye cannot accommodate for the defocus, then this can cause a significant loss of modulation in the image.

Thus, our approach to evaluating the display image quality is to assess the instantaneous image quality available to the observer as he/she looks at different field angles from a particular point within the eye motion box. We define a small eye pupil at various positions in the larger exit pupil of the display; then we compute the apparent sagittal and tangential foci for various points in the field of view, as seen from this particular eye point. Both the exit pupils and the field of view are defined in terms of a clock face, as shown in Fig. 4. The displacement of the small eye pupil is in the 12:00 direction



**FIGURE 4** Spatial and directional relationships between display exit pupil and eye pupil with respect to field of view.

within the larger exit pupil of the display. This radial direction is the same as the direction of the off-axis point in the field of view. We make the following further definitions:

6:00 is the direction in the field-of-view opposite to the direction in which the eye pupil is displaced or off-aperture within the large exit pupil of the display.

3:00 and 9:00 are directions in the field perpendicular to the direction in which the eye pupil is off-aperture in the large exit pupil of the display. These values are identical.

We simply trace a matrix of object points on the object and calculate the tangential and sagittal foci in diopters for each of these points. We are thus assuming that the aberration variation across the 3- to 4-mm eye pupil is small and that the only residual defect is defocus. The three possibilities are as follows:

If the tangential and sagittal dioptric values are equal and the defocus is such that the observer can accommodate for this apparent distance, then the resolution potential of the display at this point in the field reverts to the diffraction limited potential of the aperture of his/her eye.

If the tangential and sagittal values are unequal but he/she can accommodate to either of the foci, then the image quality loss is defined by a focus halfway between the two foci.

If the tangential and/or sagittal values are such that the observer cannot accommodate for the defocus, then the loss of image quality is defined by this total defocus value.

In this manner, a particular optical design can be evaluated for image quality. An indication that the design is not sufficiently corrected is when focus conditions at various points in the field of view change rapidly for small displacements in the exit pupil. Our method for evaluating the potential display image quality is completely independent of their scale. It is valid for the 1.5-m focal length tilted mirrors used in the collimated displays for flight simulators, the 150- to 200-mm focal length refractive collimators used in heads-up displays, the 50- to 75-mm focal length eyepieces used in binocular magnifiers, and the 20-mm focal length erecting eyepiece systems used in HMD. However, for the HMD, the loss of image quality is likely to be the most severe. We will illustrate this method with a 50° field-of-view full-overlap binocular HMD, using only four elements to minimize weight. To conserve expense, all elements are simple spheres. The display has an overall focal length of 24.6 mm and is

**TABLE 1** Prescription for a 50° Field-of-View Display Optics

Surface	Radius (mm)	Thickness (mm)	Glass
Object	∞	0.519	Air
1	∞	2.54	BK7
2	20.559	5.715	Air
3	∞	8.89	SK16
4	-26.75	33.02	Air
5	119.907	2.54	SF58
6	30.384	11.43	LAF2
7	-48.552	7.62	Air
8	105.5	5.08	SK16
9	-105.5	95.325	Air
10	∞	-38.1	Mirror
11	98.298	76.2	Mirror
Stop	∞	0	
Image	∞		

dimensioned to provide a 19-mm exit pupil with a horizontal field of view of 50° provided by a 23-mm image source format. Our electronic driver bandwidth can only adequately display 1000 video lines, resulting in a resolution of slightly better than 20/60 (10 cy/deg). This resolution is deemed minimal in order to support flying nap-of-the-earth (terrain following) or contour flight.<sup>41</sup> The prescription is given in Table 1.

Table 2 lists the optical defocus values exhibited by this display example. A negative value indicates that the light exiting the display is convergent and corresponds to images focused beyond infinity. Objective measurement of the display system confirms the defocus table. The display focus is then set to 0.87 D (see subsection entitled “Comfort”). Table 3 lists the visual defocus values according to the

**TABLE 2** Optical Defocus versus Pupil Displacements

Field Direction	Field Angle (deg)	Pupil Displaced							
		0 mm		2 mm		4 mm		6 mm	
		12-6	3-9	12-6	3-9	12-6	3-9	12-6	3-9
12:00	25	-0.4*	+0.2 <sup>†</sup>	-0.5	+0.1	-1.0	-0.1	-1.7	-0.5
12:00	20	-0.6	0	-0.6	-0.1	-1.2	-0.3	-2.0	-0.6
12:00	15	-0.4	-0.1	-0.6	-0.1	-1.2	-0.3	-2.0	-0.7
12:00	10	-0.2	0	-0.4	-0.1	-1.1	-0.4	-2.0	-0.7
12:00	5	-0.1	0	-0.3	-0.1	-1.1	-0.4	-2.1	-0.8
	0	0	0	-0.3	-0.1	-1.3	-0.4	-2.3	-0.9
6:00	5	-0.1	0	-0.5	-0.1	-1.5	-0.5	-2.6	-1.0
6:00	10	-0.2	0	-0.7	-0.2	-1.8	-0.6	-2.9	-1.1
6:00	15	-0.4	-0.1	-0.9	-0.2	-2.0	-0.6	-3.1	-1.1
6:00	20	-0.6	0	-1.1	-0.2	-2.1	-0.5	-2.9	-0.9
6:00	25	-0.4	+0.2	-0.7	+0.1	-1.5	-0.2	-1.7	-0.5
3:00	25	+0.2	-0.4	-0.2	-0.5	-1.0	-0.7	-1.8	-1.1
3:00	20	0	-0.6	-0.3	-0.7	-1.2	-1.0	-2.1	-1.4
3:00	15	-0.1	-0.4	-0.4	-0.5	-1.3	-0.9	-2.2	-1.3
3:00	10	0	-0.2	-0.4	-0.3	-1.3	-0.7	-2.3	-1.1
3:00	5	0	-0.1	-0.3	-0.2	-1.3	-0.5	-2.3	-1.0
	0	0	0	-0.3	-0.1	-1.3	-0.4	-2.3	-0.9

\*Negative diopters indicate convergent rays or farther than optical infinity.

<sup>†</sup>Positive diopters indicate divergent rays or closer than optical infinity.

**TABLE 3** Visual Defocus versus Pupil Displacements

Field Direction	Field Angle (deg)	Pupil Displaced							
		0 mm		2 mm		4 mm		6 mm	
		12–6	3–9	12–6	3–9	12–6	3–9	12–6	3–9
12:00	25	-0.3*	+0.3 <sup>†</sup>	-0.3	+0.3	-0.9	0	-1.3	0
12:00	20	-0.3	+0.3	-0.3	+0.3	-0.2	0	-1.3	0
12:00	15	-0.3	+0.3	-0.2	+0.2	-0.2	0	-1.3	0
12:00	10	-0.1	+0.1	-0.2	+0.2	-0.1	0	-1.3	0
12:00	5	0	0	-0.1	+0.1	-0.1	0	-1.3	0
	0	0	0	-0.1	+0.1	-0.3	0	-2.3	-0.9
6:00	5	0	0	-0.2	+0.2	-1.0	0	-2.6	-1.0
6:00	10	-0.1	+0.1	-0.3	+0.3	-1.2	0	-2.9	-1.1
6:00	15	-0.3	+0.3	-0.7	0	-1.5	0	-3.1	-1.1
6:00	20	-0.3	+0.3	-0.9	0	-1.5	0	-2.9	-0.9
6:00	25	-0.3	+0.3	-0.4	+0.4	-1.3	0	-1.2	0
3:00	25	+0.3	-0.3	+0.2	-0.2	-0.2	0	-1.8	-1.1
3:00	20	+0.3	-0.3	+0.2	-0.2	-1.2	-1.0	-2.1	-1.4
3:00	15	+0.2	-0.2	+0.1	-0.1	-0.4	0	-2.2	-1.3
3:00	10	+0.1	-0.1	0	0	-0.6	0	-2.3	-1.1
3:00	5	0	0	-0.1	+0.1	-0.8	0	-2.3	-1.0
	0	0	0	-0.1	+0.1	-0.8	0	-2.3	-0.9

\*Negative diopters indicate convergent rays or farther than where eye accommodates.

<sup>†</sup>Positive diopters indicate divergent rays or closer than where eye accommodates.

eye’s ability to accommodate. Most observers cannot relax their accommodation beyond infinity (or > -0.87 D in our example) and thus suffer a loss in resolution corresponding to the full amount of defocus. The subjective evaluations agree with the predictions. Namely, while performance is satisfactory for the center of the field when the eye is positioned in the center of the display exit pupil, the performance falls off radically as the eye is decentered in the pupil. In addition, there is a significant falloff of performance across the field of view. We suspect the image quality would not be acceptable if the display resolution were much higher than the 20/60 used.<sup>42</sup>

## 25.5 SUMMARY

We have described various visual issues dealing with a 50° binocular HMD breadboard. Both optical axes point downward by 5° and converge and intersect at a point about 1 m in front. Preliminary evaluation of the breadboard in the laboratory has not revealed any problems with respect to usability and comfort. However, the optics are undercorrected for both spherical aberration and astigmatism as shown by the defocus calculations. The astigmatism causes the variation of focus across the field of view when viewed from the center of the eye motion box. The spherical aberration causes the variation of focus when the eye pupil is decentered in the eye motion box. We found that using the defocus to predict display resolution is both valuable and intuitive in assessing HMD image quality during design and initial evaluation.

## 25.6 APPENDIX

There has been a concern<sup>43</sup> with respect to the transparent HMD and heads-up display that virtual images appear closer than the directly viewed image, inducing an undesired accommodation of the viewer. After investigating the issue, we did not find any support for misaccommodation. First, let us define the following terms:

*Refractive error* is defined by the eye's far point, which is the farthest distance to which the eye can focus. Eye focus is the dioptric distance at which the eye is focused, making objects at that distance optically conjugate with the eye's retina. Therefore, eye focus equals the dioptric sum of refractive error plus accommodation.

*Accommodation* is the eye's ability to focus closer than the far point by increasing its optical power through the action of the ciliary muscles. The hyperope's far point is beyond optical infinity, and therefore negative in value. Hyperopes must accommodate to focus to optical infinity. Myopia has a positive refractive error—that is, the eye is optically too strong to focus on distance objects. Do not confuse this with the clinical convention of associating myopia with the negative lens power of the correcting spectacles.

How is refractive error<sup>44</sup> determined? Optometers and clinical refraction optometers (e.g., infrared,<sup>45–48</sup> laser,<sup>49,50</sup> and stigmatoscopy<sup>51</sup>) measure eye focus, not accommodation. Optometers cannot distinguish between nonaccommodating 1-D myopes, emmetropes accommodating 1 D, or 1-D hyperopes accommodating 2 D since all are focused 1 m away. Optometers approximate refractive error (i.e., far point) only when accommodation is minimized under cycloplegia. Cycloplegia is the pharmacologically induced paralysis of accommodation.

Thus, clinical refraction does not determine refractive error in the strict sense. Clinical refraction determines the most-plus lens power resulting in best binocular visual acuity based on a subjective blur criterion (see Chap. 12, "Assessment of Refraction and Refractive Errors and Their Influence on Optical Design" and Chap. 13, "Binocular Vision Factors That Influence Optical Design," in this volume for general issues). Therefore, its theoretical endpoint places the distal end of the eye's depth-of-field (see Chap. 1, "Optics of the Eye," in this volume for reviews) at the eye chart, and does not place eye focus at optical infinity. Since eye charts are usually located about 6 m (+0.17 D) away and the eye's depth-of-field is +0.43 D for detecting blur of small sharp-edged high-contrast targets,<sup>52</sup> the clinical emmetrope is about +0.6 D myopic. This concept is supported by Morgan's<sup>53</sup> report that subjective tests tend to overcorrect by +0.75 D, and Bannon et al.'s<sup>51</sup> report that apparent emmetropes focused from +0.5 to +0.75 D closer than a fixated distant target. The presumed benefit of a myopic endpoint is that the eye's entire depth of field is maximized in real space when distant objects are viewed.

There are two potential causes for misaccommodation with transparent HMD: near-awareness of the virtual image and accommodative trap by the transparent combiner. While proximal accommodation would be the result of the former, the latter may be related to the Mandelbaum's phenomenon. However, Jones<sup>54</sup> found that unless the exit pupil is abnormally constrained (<3 mm), proximal accommodation does not affect the accommodative accuracy for targets in virtual image displays. Therefore, those deleterious proximal effects in virtual image displays may be avoided by utilization of an exit pupil in binocular HMD sufficiently large to encompass a natural pupil (e.g., display exit pupil matches eye motion box).

Contrary to accepted opinion, Gleason<sup>55</sup> found that Mandelbaum's phenomenon is not caused by accommodative capture. Almost fifty years ago, Mandelbaum<sup>56</sup> reported that distant scenery appears blurred when viewed through a window screen. The blur subsided with side-to-side head motion and returned after a delay when head motion ceased. These collective observations are called Mandelbaum's phenomenon. Mandelbaum hypothesized that the eye reflexively focused on the intervening screen instead of the desired distant scenery. Indeed, the assumed disruption of voluntary accommodation by intervening objects is now called the Mandelbaum effect.<sup>57</sup> Gleason investigated Mandelbaum's phenomenon using objective and continuous measurements of accommodation. In each trial, subjects looked directly at the distant E and were instructed to keep the E as clear as possible. After 7 s, a mesh screen was placed in the line of sight at the distance of either causing the greatest Mandelbaum phenomenon or at 40 cm in front. The 40-cm distance was near the subjects' dark focus. At 15 s, subjects were instructed to focus on the screen. At 23 s, subjects were instructed to refocus on the distant E. After the trial, the experimenter shook the screen side to side to verify that Mandelbaum's phenomenon was seen. Subjects were then questioned about what they saw during the trial. After each maximum Mandelbaum's phenomenon trial, subjects reported the distant letter blurred within 1 s of screen interjection; the letter remained blurred until the experimenter shook the screen at trial's end. These observations indicate that Mandelbaum's

phenomenon occurred during the trial. Mandelbaum's phenomenon was weak or nonexistent in 40-cm trials.

A narrow range of screen distances (0.7 to 0.9 D) produced maximum Mandelbaum's phenomenon. These results nearly match Mandelbaum's report of from 0.6 to 1.1 D. The screen distance of maximum Mandelbaum's phenomenon was at 0.8 D. At this distance, the screen strands subtended 0.8 arc minute and the nominal screen gap width was 5.5 arc minutes. Dark focus averaged 1.7 D, matching Leibowitz and Owens<sup>58,59</sup> reports of 1.7 and 1.5 D for 124 and 220 college-age observers, respectively.

Gleason found that eye focus was essentially equal for all viewing tasks requiring fixation of the distant letter. If Mandelbaum's phenomenon were caused by the screen's capture of accommodation, then accommodation would shift eye focus toward the screen. But the eye focus is essentially unchanged before and after the introduction of the intervening screen, indicating that the screen did not influence accommodation. Eye focus shifted to the screen only after the instruction to focus on the screen. Moreover, eye focus returned to its initial level after the instruction to refocus on the distant letter. Clearly, the intervening screen did not disrupt volitional accommodation, either. Even more interesting is the fact that accommodative accuracy was not diminished by the perceptual blur as eye focus shifted from the screen back to the distant letter. In fact, it turns out that clinical emmetropia promotes Mandelbaum's phenomenon. Since clinical emmetropes are actually myopic (+0.6 D nearsighted), when they look at the mesh screen (placed at +0.8 D), the distant scenery (0.0 D) is defocused by 0.6 D, but the screen is only defocused by 0.2 D. Thus, the more blurred distant scenery may be suppressed for the benefit of the screen. This imbalance of retinal blur increases with uncorrected clinical myopia and, consequently, less time is needed to initiate Mandelbaum's phenomenon. On the other hand, an uncorrected hyperope can actually focus to the distant scenery, making the nearby screen blurrier. Now, the screen is suppressed instead of the distant scenery, precluding Mandelbaum's phenomenon. In summary, all five subjects in Gleason's study saw Mandelbaum's phenomenon without a concomitant shift in accommodation, which eliminates the misaccommodation concern over transparent HMD.

## 25.7 ACKNOWLEDGMENTS

The author (BT) thanks Tom Metzler, Mike Richey, Trang Bui, Luc Biberman, and Walter Hollis for the opportunities to be involved with Army HMD programs. We thank Jay Enoch, Neil Charman, and Vince Billock for their critical reviews of the manuscript. We also thank Sarah Tsou for her editorial support. Finally, we acknowledge the significant contributions by Jerry Gleason toward the advancements of HMD designs.

## 25.8 REFERENCES

1. B. H. Tsou, "System Design Considerations for a Visually Coupled System," in S. R. Robinson (ed.), *The Infrared and Electro-Optics System Handbook: Vol. 8. Emerging Systems and Technologies*. SPIE Optical Engineering Press, Bellingham, WA, 1993, pp. 515–536.
2. D. A. Fulghum, "Navy Orders Contractors to Stop Work on Ejectable Night Vision Helmets," *Aviation Week and Space Technology* **133**(23):67–68 (Dec. 3, 1990).
3. B. J. McEntire and D. F. Shanahan, "Mass Requirements for Helicopter Aircrew Helmets," *Advisory Group for Aerospace Research and Development (AGARD) Conference Proceedings* **597**:20.1–20.6 (Nov. 7–9, 1996).
4. C. E. Perry and J. R. Buhrman, "HMD Head and Neck Biomechanics," in J. E. Melzer and K. Moffitt (eds.), *Head Mounted Displays: Designing for the User*, McGraw-Hill, New York, 1997, pp. 147–172.
5. B. McLean, S. Shannon, J. McEntire, and S. Armstrong, "Counterweights Used with ANVIS," Report USAARL-96-30, US Army Aeromedical Research Laboratory, Fort Rucker, AL, 1990.
6. K. Y. Lau, "Microscanner Raster-Scanning Display: A Spyglass for the Future," *Optics and Photonics News* **10**(5):47–50, 84 (May 1999).

7. M. Shenker, "Optical Design Criteria for Binocular Helmet-Mounted Displays," *SPIE* 778:70–78 (1987).
8. S. M. Donelson and C. C. Gordon, "1988 Anthropometric Survey of US Army Personnel: Pilot Summary Statistics," Report Natick-TR-91-040, US Army Natick Research, Development and Engineering Center, Natick, MA, 1991.
9. J. Levy, "Physiological Position of Rest and Phoria," *Am. J. Ophthalmol.* 68(4):706–713 (1968).
10. M. Menozzi, A. Buol, H. Krueger, and Ch. Miede, "Direction of Gaze and Comfort: Discovering the Relation for the Ergonomic Optimization of Visual Tasks," *Ophthalmic Physiol. Opt.* 14:393–399 (1994).
11. G. A. Gleason and J. T. Riegler, "The Effects of Eyepiece Focus on Visual Acuity Through ANVIS Night Vision Goggles During Short- and Long-Term Wear," unpublished report performed for A. F. Workunit 71842604, 1997. (Part of the report was first presented at 1996 Human Factors Society Annual Meeting.)
12. R. Home and J. Poole, "Measurement of the Preferred Binocular Dioptric Settings at a High and Low Light Level," *Opt. Acta* 24(1):97–98 (1977).
13. R. Jones, "Binocular Fusion in Helmet Mounted Displays," unpublished report performed for A. F. Workunit 71842601, 1996. (Part of the report was first presented at 1992 Society for Information Display Annual Meeting.)
14. American National Standards Institute, 1430 Broadway, New York, NY 10018.
15. K. N. Ogle, *Researches in Binocular Vision*, Hafner Publishing Co., New York, 1964, pp. 283–290.
16. G. A. Fry, "Measurement of the Threshold of Stereopsis," *Optometric Weekly* 33:1029–1033 (1942).
17. G. Sigmar, "Observations on Vernier Stereo Acuity with Special Consideration to Their Relationship," *Acta Ophthalmol.* 48:979–998 (1970).
18. E. E. Anderson and F. Weymouth, "Visual Perception and the Retinal Mosaic," *Am. J. Physiol.* 64:561–591 (1923).
19. R. A. Schumer, "Mechanisms in Human Stereopsis," Ph.D. dissertation, Stanford University, Palo Alto, CA, 1979.
20. C. W. Tyler, "Sensory Processing of Binocular Disparity," in Schor and Ciuffreda (eds.), *Vergence Eye Movements: Basic and Clinical Aspects*, Butterworths, Boston, 1983, pp. 199–295.
21. A. M. Norcia, E. E. Sutter, and C. W. Tyler, "Electrophysiological Evidence for the Existence of Coarse and Fine Disparity Mechanisms in Humans," *Vis. Res.* 25:1603 (1985).
22. C. J. Erkelens and H. Collewijn, "Motion Perception During Dichoptic Viewing of Moving Random Dot Stereograms," *Vis. Res.* 25:345–353 (1985).
23. K. N. Ogle, *Researches in Binocular Vision*, Hafner Publishing Co., New York, 1964, pp. 173–199.
24. C. M. Schor and C. W. Tyler, "Spatial Temporal Properties of Panum's Fusional Area," *Vis. Res.* 21:683–692 (1981).
25. M. Shenker, J. LaRussa, P. Yoder, and W. Scidmore, "Overlapping-Monoculars—An Ultrawide-Field Viewing System," *Appl. Opt.* 1:399–402 (1962).
26. S. S. Grigsby and B. H. Tsou, "Grating and Flicker Sensitivity in the Near and Far Periphery: Naso-Temporal Asymmetries and Binocular Summation," *Vis. Res.* 34:2841–2848 (1994).
27. S. S. Grigsby and B. H. Tsou, "Visual Processing and Partial-Overlap Head-Mounted Displays," *J. Soc. for Information Display* 2/2:69–73 (1994).
28. T. H. Bui, R. H. Vollmerhausen, and B. H. Tsou, "Overlap Binocular Field-of-View Flight Experiment," *1994 Society for Information Display International Symposium Digest of Technical Papers* 25:306–308 (1994).
29. M. J. Wells, M. Venturino, and R. K. Osgood, "The Effect of Field-of-View Size on Performance at a Simple Simulated Air-to-Air Mission," *SPIE* 1116:126–137 (1989).
30. B. H. Tsou and B. M. Rogers-Adams, "The Effect of Aspect Ratio on Helmet-Mounted Display Field-of-View," *Report of the 30th Meeting of Air Standardization Coordination Committee (ASCC) Working Party 61: Aerospace Medical and Life Support Systems Symposium: Aeromedical Aspects of Vision* 4:136–146. Defense and Civil Institute of Environmental Medicine, Toronto, Canada, 1990.
31. A. T. Bayhill and D. Adler, "Most Naturally Occurring Human Saccades Have Magnitudes of 15 Degrees or Less," *Invest. Ophthalmol.* 14:468–469 (1975).
32. G. H. Robinson, "Dynamics of the Eye and Head During Movement Between Displays: A Qualitative and Quantitative Guide for Designers," *Human Factors* 21:343–352 (1979).
33. J. Wolfe, P. O'Neil, and S. C. Bennett, "Why Are There Eccentricity Effects in Visual Search? Visual and Attentional Hypotheses," *Percept. Psychophys.* 60:140–156 (1998).



34. S. P. McKee and K. Nakayama, "The Detection of Motion in the Peripheral Visual Field," *Vis. Res.* **24**(1):25–32 (1984).
35. K. K. Ball, B. L. Beard, D. L. Roenker, R. L. Miller, and D. S. Griggs, "Age and Visual Search: Expanding the Useful Field of View," *J. Opt. Soc. Am.* **A5**:2210–2219 (1988).
36. P. Havig and B. H. Tsou, "Is There a Useful Field of Visual Search?" [Abstract]. *Optical Society of America Annual Meeting*, Baltimore, MD, Oct. 4–9, 1998.
37. L. N. Thibos and A. Bradley, "Modeling the Refractive and Neuro-Sensor Systems of the Eye," in P. Mouroulis (ed.), *Visual Instrumentation Handbook*, McGraw-Hill, New York, 1999, pp. 4.19–4.20.
38. P. Mouroulis and H. Zhang, "Visual Instrument Image Quality Metrics and the Effects of Coma and Astigmatism," *J. Opt. Soc. Am.* **A9**(1):34–12 (1992).
39. J. E. Greivenkamp, J. Schwiegerling, J. M. Miller, and M. D. Mellinger, "Visual Acuity Modeling Using Optical Raytracing of Schematic Eyes," *Am. J. Ophthalmol.* **120**(2):227–240 (1995).
40. W. N. Charman and J. A. M. Jennings, "The Optical Quality of the Monochromatic Retinal Image As a Function of Focus," *Br. J. Physiol. Opt.* **31**:119–134 (1976).
41. D. Greene, "Night Vision Pilotage System Field-of-View (FOV)/Resolution Tradeoff Study Flight Experiment," Report NV-1-26, Center for Night Vision and Electro-Optics, Ft. Belvoir, VA, 1988.
42. G. E. Legge, K. T. Mullen, G. C. Woo, and F. W. Campbell, "Tolerance to Visual Defocus," *J. Opt. Soc. Am.* **A4**(5):851–863 (1987).
43. J. H. Iavecchia, H. P. Iavecchia, and S. N. Roscoe, "Eye Accommodation to Head-Up Virtual Images," *Human Factors*, **130**(6):687–702 (1988).
44. G. A. Gleason and J. T. Riegler, "Do Clinical Refraction Result in Best Visual Acuity?" unpublished report performed for A. F. Workunit 71842604, 1997. (Part of the report was first presented at 1996 Association for Research in Vision and Ophthalmology Annual Meeting.)
45. T. N. Cornsweet and H. D. Crane, "Servo-Controlled Infrared Optometer," *J. Opt. Soc. Am.* **60**:548–554 (1970).
46. H. D. Crane and C. M. Steele, "Accurate Three Dimensional Eyetracker," *Appl. Opt.* **17**:691–705 (1978).
47. F. W. Campbell and J. G. Robson, "High-Speed Infra-Red Optometer," *J. Opt. Soc. Am.* **49**:268–272 (1959).
48. N. A. McBrien and M. Millodot, "Clinical Evaluation of the Canon Autorefractometer R-1," *Am. J. Optom. Physiol. Opt.* **62**:786–792 (1985).
49. R. T. Hennessy and H. W. Leibowitz, "Laser Optometer Incorporating the Badal Principle," *Behav. Res. Meth. and Instru.* **4**(5):237–239 (1972).
50. A. Morrell and W. N. Charman, "A Bichromatic Laser Optometer," *Am. J. Optom. Physiol. Opt.* **64**:790–795 (1987).
51. R. E. Bannon, F. H. Cooley, H. M. Fisher, and R. T. Textor, "The Stigmatometry Method of Determining the Binocular Refractive Status," *Am. J. Optom. Arch. Am. Acad. Optom.* **27**:371 (1950). (Cited in I. M. Borish, *Clinical Refraction*, 3d ed., Professional Press, Chicago, 1975, pp. 783–784.)
52. F. W. Campbell, "The Depth of Field of the Human Eye," *Opt. Acta* **4**:157–164 (1957).
53. M. W. Morgan, "The Meaning of Clinical Tests in the Light of Laboratory Measurements," *The Ind. Opt.* **2**:6 (1947). (Cited in I. M. Borish, *Clinical Refraction*, 3d ed., Professional Press, Chicago, 1975, p. 356.)
54. R. Jones, "Proximal Accommodation in Virtual Displays," *Society for Information Display International Symposium Digest of Technical Papers* **24**:195–198 (1998).
55. G. A. Gleason and R. V. Kenyon, "Mandelbaum's Phenomenon Is Not Caused by Accommodative Capture," unpublished report performed for A. F. Workunit 71842604, 1998. (Part of the report was first presented at 1997 Association for Research in Vision and Ophthalmology Annual Meeting.)
56. J. Mandelbaum, "An Accommodative Phenomenon," *AMA Arch. Ophthalmol.* **63**:923–926 (1960).
57. D. A. Owens, "The Mandelbaum Effect: Evidence for an Accommodative Bias Toward Intermediate Viewing Distance," *J. Opt. Soc. Am.* **65**:646–652 (1975).
58. H. W. Leibowitz and D. A. Owens, "Anomalous Myopias and the Dark Focus of Accommodation," *Science* **189**:646–648 (1975).
59. H. W. Leibowitz and D. A. Owens, "New Evidence for the Intermediate Position of Relaxed Accommodation," *Doc. Ophthalmol.* **46**:133–147 (1978).

# INDEX

*Index note:* The *f* after a page number refers to a figure, the *n* to a note, and the *t* to a table.

- Aberrations (in general):
  - and contact lenses, 20.24–20.25, 20.25*f*
  - and forward scattering of light, 1.21
  - in intraocular lenses, 21.9–21.10, 21.10*f*
  - and Stiles-Crawford effect, 8.8
- Aberrations (human eye), 1.3, 16.6
  - absence of, 1.12–1.14, 1.13*f*, 1.14*f*
  - and AO-controlled light delivery
    - to generate aberrations, 15.24
    - longitudinal chromatic aberration, 15.22
    - transverse chromatic aberration, 15.22–15.23
  - astigmatism, 15.2
  - chromatic, 1.19–1.20, 1.19*f*
    - and accommodation, 1.34
    - age-related, 14.14
    - correcting color coordinates for, 10.17
    - correction of, 1.25–1.26, 1.26*f*
    - longitudinal, 15.22
    - and macular pigment, 1.9
    - transverse, 15.22–15.23
    - with visual instruments, 1.28
  - control of, 1.3
  - correction of, 1.25–1.26, 1.26*f*, 10.17
  - defocus, 15.2
  - and depth of focus, 1.28
  - higher-order, 16.6, 16.7
  - and idiosyncratic peculiarities, 1.6
  - image quality
    - for aberration-free eye, 1.12–1.14
    - calculating, 1.21–1.22
  - monochromatic, 1.4, 1.14–1.19
    - age-related, 14.12–14.14, 14.13*f*
    - correction of, 1.25, 1.26, 1.26*f*
    - off-axis, 1.18–1.19, 1.18*f*
    - on the visual axis, 1.15–1.18
  - and observed optical performance, 1.23–1.26
  - properties of, 15.4–15.7, 15.5*f*, 15.6*f*
- Aberrations (human eye) (*Cont.*):
  - and pupil diameter, 1.8
  - and refractive surgery, 1.15
- Aberrometers, 1.15, 1.23, 12.6
- Ablation rate (refractive correction), 16.16–16.18, 16.18*f*
- Abney's law, obedience to, 10.44, 11.37
- Absorption, lenticular, 1.9
- Accommodating intraocular lenses, 14.29–14.30, 21.1, 21.14, 21.18–21.19
- Accommodation, 1.3, 1.29–1.36, 12.3, 21.2
  - accuracy of, 1.32–1.34, 1.33*f*, 1.34*f*
  - in aging eyes, 1.35–1.36, 1.35*f*, 14.4, 21.3
  - application to instrumentation, 1.34–1.35
  - and change in lens, 1.5
  - with computer work, 23.10–23.11
  - defined, 12.1, 13.1, 16.1, 21.1, 23.1
  - dynamics of, 1.31–1.32, 1.31*f*
  - and extraocular muscle movement, 12.16
  - fluctuations in, 1.32
  - with head-mounted displays, 13.31, 25.10–25.12
  - and ocular aberration, 1.17–1.18
  - and presbyopia, 14.8, 14.9*f*, 14.29–14.30
  - and refractive errors, 12.15–12.16
  - and spherical ametropias, 16.5
  - stability of, 1.32
  - vergence input, 1.29, 1.30, 1.34
- Accommodation demand, 13.30
- Accommodative demand:
  - with contact lenses, 20.26–20.30, 20.27*f*–20.29*f*, 20.29*t*
  - defined, 20.1
- Accommodative miosis, 1.30
- Acellular, 16.1
- Achromatic detection:
  - and chromatic adaptation, 11.47–11.49, 11.48*f*
  - measurements favoring, 11.37

- Achromatic mechanism (color vision),  
11.1, 11.37
- Achromatic signals, multiplexing of chromatic signals and (color vision), 11.76–11.79, 11.78*f*
- Action spectra (ocular radiation), 7.2, 7.3, 7.4*f*
- Adaptation:  
in aging eyes, 14.15  
in contrast detection, 2.25–2.27  
multiplicative and subtractive, 2.26  
phoria, 13.21–13.22  
in wearing contact lenses, 12.12
- Adaptive optics (AO) (in retinal microscopy and vision), 1.25, 15.1–15.24
- AO-controlled light delivery to the retina, 15.22–15.24
- alignment, 15.23
- in an AO SLO, 15.23
- conventional AO vision systems, 15.23
- to generate aberrations, 15.24
- longitudinal chromatic aberration, 15.22
- measuring activity of individual cones, 15.24
- transverse chromatic aberration, 15.22–15.23
- uses of, 15.23–15.24
- for correcting monochromatic aberrations, 1.25
- and correction for SCE-1, 9.5
- defined, 15.1, 17.1
- history of, 15.2–15.3
- imaging of the retina, 15.16–15.22
- contrast and resolution, 15.21–15.22
- flood-illuminated AO ophthalmoscope, 15.16–15.17, 15.16*f*, 15.17*f*
- optical coherence tomography, 15.19–15.21, 15.20*f*, 15.21*f*
- scanning laser ophthalmoscope, 15.17–15.19, 15.18*f*, 15.19*f*, 17.9
- implementation of, 15.7–15.15, 15.7*f*
- control system, 15.12–15.15
- wavefront corrector, 15.9–15.12, 15.10*f*, 15.11*f*
- wavefront sensor, 15.8–15.9, 15.8*f*
- properties of ocular aberrations, 15.4–15.7, 15.5*f*, 15.6*f*
- wavefront corrector, 15.9–15.12, 15.10*f*, 15.11*f*
- wavefront sensor, 15.8–15.9, 15.8*f*
- Adaptive optics retina cameras, 15.3, 15.12  
(See also Ophthalmoscopes)
- Additivity:  
in color matching, 10.8
- of color opponency and color appearance, 11.67
- with equilibrium colors, 11.66
- field, 11.51, 11.52*f*
- obedience to Abney's law, 10.44, 11.37
- Stiles-Crawford effect, 9.10*f*–9.11*f*
- Adjustment tasks:  
judgment tasks *vs.*, 3.2
- psychophysical measurement of, 3.4–3.5
- magnitude production, 3.5
- matching, 3.4–3.5
- nulling, 3.4
- threshold, 3.4, 3.5*f*
- Aesthetics and emotional characteristics (electronic imaging), 24.9–24.10
- color, art, and emotion, 24.10
- image quality, 24.9
- virtual reality and presence, 24.9
- After-effect (color vision):  
and McCollough effect, 11.76
- and orientation of contours, 11.75–11.76, 11.75*f*
- and orientation-selectivity, 11.80
- Afterimage, 23.1
- Against-the-rule astigmatism, 1.6
- Age-related changes in vision, 1.5, 1.7, 1.35–1.36, 14.1–14.30
- in accommodation, 1.35–1.36, 1.35*f*, 14.4, 21.3
- in bovine lenses, 19.10
- changes in the eye causing, 21.3–21.4
- in color vision, 14.15, 14.17
- and demographic changes in world, 14.2–14.4, 14.2*f*
- in depth and stereovision, 14.22
- economic/social implications of, 14.4
- and fluorescence, 1.21
- and health implications of living longer, 14.3–14.4
- index of diffusion for, 1.23
- and level of aberration, 1.18
- minimal, 14.22
- ocular diseases, 14.22–14.27
- age-related macular degeneration, 14.24–14.25
- cataract, 14.24
- diabetic retinopathy, 14.25–14.26
- glaucomas, 14.26–14.27
- life-span environmental radiation damage, 14.22–14.23

- Age-related changes in vision (*Cont.*):  
 in optics, 14.4–14.14  
   accommodation and presbyopia, 14.8, 14.9f  
   anterior and posterior chambers, 14.5, 14.6  
   cornea, 14.5, 14.6f  
   eye size, 14.11  
   lens, 14.7–14.8  
   pupil, 14.6–14.7, 14.7f  
   retina, 14.9–14.11  
   tears, 14.4–14.5  
   transparency/cataract, 14.8  
 presbyopic corrections for, 14.27–14.30  
   accommodation restoration, 14.29–14.30  
   contact lenses, 14.27–14.28  
   intraocular lenses, 14.28–14.29  
   noncataract related refractive surgeries, 14.29  
   spectacles, 14.27  
 in pupil diameter, 1.8, 1.18  
 in retinal image quality, 14.11–14.14  
   chromatic aberration, 14.14  
   intraocular scatter, 14.12  
   monochromatic aberrations, 14.12–14.14, 14.13f  
 in retinal reflectance, 1.11  
 and RMS wavefront error, 1.15–1.17, 1.16t  
 in scattering, 1.20  
 in sensitivity, 14.15, 14.16f  
 in spatial vision, 14.17–14.19, 14.18f  
 in temporal vision, 14.19–14.21, 14.20f  
 in transmittance, 1.9  
 in visual acuity, 4.13  
 in visual field, 14.21, 14.21f
- Age-related macular degeneration (AMD), 14.24–14.25  
 defined, 14.1  
 OFDI at 1050 nm in, 18.17, 18.17f  
 OFDI in diagnosis of, 18.8
- Airy diffraction pattern, 1.12
- Airy's disk, 4.1, 4.4f, 4.5
- Alignment:  
   of eye (*see* Eye alignment)  
   of photoreceptors, 8.5–8.8, 8.6f, 8.7f
- A-line, 18.1
- Allelotropia, 13.7
- Amacrine cells, 2.10, 2.11
- Amblyopia, 2.34, 2.35, 12.1, 12.16
- American Conference of Governmental Industrial Hygienists (ACGIH), 7.9
- American National Standard Institute (ANSI), 7.9, 7.11
- Ametropias, 1.6–1.7, 12.4, 16.4–16.5, 16.5f  
 astigmatism, 16.5–16.6  
 correcting, 12.16, 16.8f  
 defined, 12.1, 13.1  
 spherical, 16.5  
 uncorrected, 12.16
- Amplification, in color CRTs, 22.6–22.9, 22.7f
- Angle  $\alpha$ , 1.20
- Angle of incidence, obliquity of, 8.20
- Angular resolution, in diffraction-limited eye, 1.12
- Aniridia, 9.1
- Aniseikonia, 1.41–1.42  
 binocular factors in, 12.16–12.17  
 defined, 9.1, 12.1, 13.1, 25.1  
 distortion from interocular anisomagnification, 13.17–13.18
- Anisometropia, 1.7, 1.41, 1.42  
 binocular factors in, 12.16–12.17  
 defined, 12.1, 23.1  
 interocular blur suppression with, 13.18–13.19  
 as problem with computer work, 23.11
- Anisophoria, optically induced, 13.26
- Anterior chamber, 14.5, 14.6, 16.3
- Aphakia, 12.14, 14.28  
 correction of, 12.14–12.15  
 defined, 9.1, 12.1, 14.1, 21.1
- Apodization, 6.14, 21.17, 21.18  
 defined, 8.1
- Stiles-Crawford effect, 8.8
- Aqueous humor, 1.3f, 14.5, 14.6, 14.26, 16.3
- Aspheric intraocular lenses, 21.1, 21.10–21.12, 21.11f, 21.12t
- Aspheric lens design, 20.1, 20.20–20.23, 20.21f–20.23f
- Asthenopia (eyestrain), 23.2
- Astigmatic dial, 12.7f
- Astigmatism, 1.6, 1.7, 1.7f, 1.18f, 12.4, 16.5–16.6  
 and binocular instrumentation, 1.41–1.42  
 and contact lenses, 20.15

- Astigmatism (*Cont.*):  
 correction of, 16.9  
   cylindrical correction, 13.16, 13.17  
   with hydrogel contact lenses, 12.13  
   with spectacle lenses, 12.9, 15.2  
   with spherical contact lenses, 12.12  
 defined, 12.1, 23.1  
 determining axis of, 12.6  
 following cataract surgery, 21.13  
 irregular, 16.6  
 measuring, 12.7  
 off-axis, 1.18, 1.18*f*
- Asymmetric color matching, 11.27, 11.29, 11.30*f*  
 and changes in overall intensity, 11.68  
 and chromatic adaptation, 11.69  
 and color constancy, 11.71
- Atrophic (dry) age-related macular degeneration, 14.1
- Attention, in human vision, 24.6, 24.7
- Autocorrelation noise, 18.11–18.12
- Autokeratometers, 12.6
- Autorefractors, 12.6
- Axial ametropia, 20.1
- Axial aniseikonia, 13.17, 13.18
- Axial edge lift, 20.1
- Axial gradients, 19.5
- Axial resolution, 18.1
- Back central optical radius (BCOR), 20.3
- Back vertex power (BVP), 12.1, 20.1  
   contact lenses, 20.7–20.8, 20.8*t*  
   spectacle lenses, 12.9
- Base curve:  
   of contact lenses, 12.1, 12.12  
   of spectacle lenses, 12.1, 12.9
- Base curve radius (BCR):  
   for contact lenses, 20.3, 20.4*f*, 20.5  
   defined, 20.1
- Baseline, 13.1
- Beam separators, 5.10
- Beamsplitters, 5.9–5.10, 5.9*f*
- Beer-Lambert law, 2.7
- Bessel functions, 8.1, 8.13, 8.22
- Best-correction, 1.6
- Bezold-Brücke hue shift, 11.1, 11.67–11.68, 11.67*f*
- Bichromatic test mixtures, sensitivity to, 11.39, 11.40
- Bifocal jump, with monocular magnification, 13.15–13.16, 13.15*f*
- Bifocal lenses, 12.10*f*, 14.27  
   contact lenses, 12.13, 14.28  
   intraocular lenses, 14.28  
   for presbyopia, 12.8, 12.10  
   vertical prism introduced by, 13.28–13.29
- Bimorph mirror, 15.1, 15.10, 15.10*f*, 15.11
- Binocular cues, 13.3, 13.7
- Binocular disparities, 2.40, 13.1, 13.4–13.5  
   induced by prisms or lenses, 13.28  
   and perceived direction, 13.7–13.8
- Binocular field, 1.3  
   horizontal angular extent of, 1.38*f*  
   stereopsis in, 1.38–1.42
- Binocular fusion, 13.1, 13.12–13.13
- Binocular instrumentation:  
   and chromostereopsis, 1.20  
   differential focusing for, 1.7  
   tolerances in, 1.41–1.42
- Binocular parallax, 13.1, 13.22
- Binocular rivalry, 13.1, 13.19
- Binocular rivalry suppression, 13.12, 13.33
- Binocular stereoscopic discrimination, 2.40–2.41, 2.41*f*
- Binocular vision factors, 13.1–13.35  
   in computer work, 23.11  
   coordination of eyes, 13.20–13.25  
     binocular parallax, 13.22  
     cross-coupling and direction/distance of gaze, 13.23–13.24, 13.23*f*  
     defocus and efforts to clear vision, 13.22–13.23  
     intrinsic stimuli to vergence, 13.21–13.22  
     perceived distance, 13.24  
     zone of clear and single binocular vision, 13.24–13.25, 13.25*f*
- distortion by monocular magnification, 13.13–13.16  
   bifocal jump, 13.15–13.16, 13.15*f*  
   from convergence responses to prism, 13.19  
   discrepant views of objects/images, 13.16  
   motion parallax, 13.14–13.15  
   perspective distortion, 13.13, 13.14*f*  
   stereopsis, 13.13, 13.14  
   distortion from interocular anisomagnification, 13.16–13.19  
   aniseikonia, 13.17–13.18  
   interocular blur suppression with anisometropia, 13.18–13.19  
   lenses and prisms, 13.16–13.17, 13.17*f*

- Binocular vision factors (*Cont.*):
- eye alignment, 13.20–13.25
    - magnification induced errors of, 13.25–13.27
    - prism induced errors of, 13.27–13.29
  - eye movements, 13.19–13.20
  - focus and responses to distance, 13.30
  - fusion and suppression, 13.12–13.13
  - gaze control, 13.29–13.30
  - and head mounted visual display systems, 13.31–13.35
    - distance conflicts, 13.31–13.32
    - optical errors, 13.34–13.35
    - spatial location conflicts, 13.33, 13.34*f*
    - visual-vestibular conflicts, 13.32–13.33
  - lens effects on vergence and phoria, 13.25–13.27
  - perceived direction, 13.7–13.10
    - corresponding retinal points, 13.8
    - horopter, 13.8–13.9, 13.9*f*, 13.10*f*
    - vertical horopter, 13.9
  - perceived space, 13.3–13.7
    - binocular cues, 13.7
    - extraretinal information for eye movements, 13.7
    - kinetic cues, 13.4–13.7, 13.5*f*, 13.6*f*
    - monocular cues, 13.3–13.7
  - prisms
    - distortion from interocular anisomagnification, 13.16–13.17
    - effects on vergence and phoria, 13.25–13.27
    - errors of alignment with, 13.25–13.29
  - in refractive errors, 12.15–12.17
    - aniseikonia, 12.16–12.17
    - anisometropia, 12.16–12.17
    - convergence and accommodation, 12.15–12.16
  - stereopsis, 13.11–13.12, 13.11*f*
  - visual field, 13.3
- Binocularity, 23.1
- Biological waveguides, 8.1–8.29
- in cochlear hair cells and human hair, 8.24–8.26
  - in fiber-optic plant tissues, 8.26–8.28, 8.27*f*
  - models of, 8.8–8.15, 8.13*f*, 8.14*f*
    - physical assumptions in, 8.22
    - and propagation of light, 8.21, 8.22
  - retinal layer of rods and cones, 8.8–8.9, 8.9*f*, 8.10*f*, 8.12–8.15
  - three-segment model, 8.9, 8.11–8.15
- Biological waveguides (*Cont.*):
- and photoreceptors, 8.3–8.5
    - modal patterns in monkey/human receptors, 8.19–8.24, 8.22*f*
    - orientation and alignment, 8.5–8.8, 8.6*f*, 8.7*f*
    - photoreceptor optics, 8.3
    - quantitative observations of single receptors, 8.15, 8.16*f*, 8.17, 8.18*f*, 8.19*f*
  - in sponges, 8.28–8.29
  - and Stiles-Crawford effect of the first kind, 8.3
- Bipolar cells (retina), 2.9, 2.10
- Bipolar mechanisms (color vision), 11.1, 11.80–11.81
- Birefringence, 1.10, 18.1, 18.20, 18.22, 18.25, 18.27
- Bitoric lenses, 20.1, 20.17–20.20, 20.18*f*–20.20*f*
- Blindness:
- from cataract, 14.24, 21.3
  - from diabetic retinopathy, 14.25
  - economic impact of, 14.4
  - from glaucoma, 14.26
  - and increasing life span, 14.3
  - from uncorrected ametropia, 12.4
- Bloch's law, 2.28
- Blue to red (B/R) ratio, 8.28
- Blue-light photochemical injury, 7.4, 7.10
- Blur, 1.28
  - in correction for SCE-1, 9.4–9.6, 9.14
  - with monovision, 14.28
  - suppression of, 13.18–13.19, 14.28
  - (*See also* Astigmatism; Defocus)
- Born approximation, 8.1, 8.15
- Bovine lenses, 19.8–19.11, 19.11*f*
- Bowman's layer, 16.4
- Bowman's membrane (cornea), 14.5
- Boynton illusion, 11.72, 11.73*f*, 11.78
- Brightness:
- in color CRTs, 22.7, 22.19
  - defined, 11.1, 11.8, 23.1
  - and luminous efficiency, 11.70, 11.70*f*
  - in monochrome CRTs, 22.8*f*
  - in visual acuity tests, 4.8
- Brightness matching, 10.43–10.45
- Bruch's membrane, 8.1, 14.25, 18.1
- Bumper function, 11.59, 11.60*f*
- Bunsen-Roscoe law of photochemistry, 7.1–7.3, 7.7

- C line, 25.1
- Calibration:
- of color CRTs, 22.20–22.21 (*See also* Characterization, of CRTs)
  - errors in, and SCE-1 correction, 9.6–9.13
  - in Maxwellian viewing systems, 5.17–5.18, 5.17*f*
- Cameras:
- adaptive optics retina cameras, 15.3, 15.12 (*See also* Ophthalmoscopes)
  - color, 10.38–10.40
- Capillary, 18.1
- Capsulorhexis, 21.1
- Cardinal directions (color vision), 11.1, 11.56
- Cardinal mechanisms (color vision), 11.1, 11.54, 11.56
- evidence for, 11.79
  - and sensitivity losses, 11.29
- Cat lenses, 19.9
- Cataract, 1.21, 14.8
- in aging eyes, 14.24, 21.3–21.4
  - defined, 14.1, 19.1, 21.1, 23.1
  - infrared, 7.7, 7.10
  - postsurgical correction of, 12.14
  - from radiation, 7.4–7.6, 7.6*f*, 14.23
  - treatment of, 12.14
- Cataract lenses, 12.15
- Cataract surgery, 14.4, 14.24
- intraocular lenses in, 21.4, 21.5 (*See also* Intraocular lenses)
  - renormalization following, 14.17
- Cathode ray tubes (CRTs), 23.1
- color (*see* Color cathode ray tubes)
  - and Computer Vision Syndrome, 23.7, 23.8
  - monochrome, 22.3–22.4, 22.3*f*
    - controls for, 22.7, 22.8*f*
    - design and operation of, 22.3–22.4, 22.3*f*
    - standards for, 22.14
  - in optical systems, 5.16
  - screen reflections with, 23.5
- Cauchy's equation, 1.19, 1.20
- Cavanagh, P., 11.80
- Center of rotation (eye), 1.42, 13.1
- Center thickness, for contact lenses, 20.6
- Central visual processing, 2.12–2.14, 2.13*f*
- Characterization:
- of CRTs, 22.20–22.34
    - absolute *vs.* for interaction, 22.33
    - calibration *vs.*, 22.20
    - choice of method for, 22.20–22.21
  - of head-mounted displays, 25.7–25.10, 25.8*f*, 25.9*t*–25.10*t*
- Characterization, of CRTs (*Cont.*):
- exhaustive, 22.21–22.23
  - local, 22.24–22.27
  - model-dependent, 22.27–22.33
- Chief ray (retinal imaging), 4.3
- Choroid, 18.1
- Chromatic aberration, 1.19–1.20, 1.19*f*
  - and accommodation, 1.34
  - age-related, 14.14
  - correcting color coordinates for, 10.17
  - correction of, 1.25–1.26, 1.26*f*
  - longitudinal, 15.22
  - and macular pigment, 1.9
  - transverse, 15.22–15.23
  - with visual instruments, 1.28
- Chromatic adaptation, 11.35*f*
  - achromatic detection and, 11.47–11.49, 11.48*f*
  - color appearance and, 11.68
  - and luminous efficiency, 11.37
  - second-site, 11.17–11.18
  - and the Sloan notch, 11.49, 11.51
- Chromatic contrast detection, 2.29–2.31, 2.30*f*
- Chromatic CSFs, 2.29–2.31
- Chromatic detection:
- and color appearance, 11.69
  - on neutral fields, 11.34
- Chromatic discrimination, 11.57–11.62
  - and chromatic adaptation, 11.69
  - and color appearance, 11.69
  - defined, 11.1
  - and gap effect, 11.72, 11.74
  - near detection threshold, 11.58–11.59
  - pedestal experiments, 11.59, 11.60*f*, 11.61–11.62
- Chromatic mechanisms, 11.1, 11.80–11.81
- Chromatic signals, multiplexing of achromatic signals and, 11.76–11.79, 11.78*f*
- Chromatic spatial CSFs, 2.29
- Chromatic temporal CSFs, 2.31
- Chromatic valence data, 11.27, 11.28*f*
- Chromaticity coordinates, 10.1, 10.4*t*, 10.20–10.21, 10.22*f*
- Chromaticity diagrams, 10.20–10.21, 10.22*f*
- Chromophores, 21.1, 21.20, 21.20*f*
- Chromostereopsis, 1.20
- CIE 1931 2° color-matching functions, 10.6, 10.12

- CIE 1964 10° color-matching functions, 10.12–10.13, 10.16
- CIE luminous efficiency functions, 10.44–10.45, 11.37, 11.38*f*
- CIELAB, 10.42–10.43
- Cilia, 8.1
- Ciliary body, 1.3*f*
- Ciliary muscle, 21.2
- Ciliary photoreceptors, 8.3
- Ciliary ring, 1.30–1.31
- Circular polarization (OCT), 18.21
- Clear lens extraction (CLE), 21.18
- Closed HMDs, 25.3, 25.4
- Cochlea, 8.1, 8.26
- Cochlear hair cells, light guide effect in, 8.24–8.26
- Coherence length, 18.1
- Coherent coupling, 1.27–1.28
- Coherent illumination, pupil size and, 6.9–6.12, 6.11*t*
- Coleoptile, 8.1, 8.26
- Collagen, 16.1, 16.4
- Color:
- in electronic imaging, 24.5–24.6, 24.8–24.9
  - and visual acuity, 4.10
- Color appearance:
- and chromatic adaptation, 11.68
  - and chromatic detection and discrimination, 11.69
  - and color constancy, 11.71
  - and color opponency, 11.62–11.66
    - color-opponent response or valence functions, 11.63, 11.65*f*
    - hue scaling, 11.63, 11.64*f*
    - opponent-colors theory, 11.62–11.63
    - spectral properties of color-opponent mechanisms, 11.63
    - unique hues and equilibrium colors, 11.63–11.66
  - defined, 11.2
  - and habituation, 11.69–11.70
  - phenomenological aspects of, 11.5
  - and postreceptoral mechanisms, 11.3, 11.4*f*, 11.5
  - and stabilized borders, 11.74–11.75
  - trichromacy *vs.*, 11.5
- Color assimilation, 11.2, 11.4*f*
- Color balance and tracking, in color CRTs, 22.19
- Color cameras, 10.38–10.40
- Color cathode ray tubes (color CRTs), 22.1–22.34
- colorimetric calibration/characterization, 22.20–22.34
- absolute *vs.* characterization for interaction, 22.33
  - choice of method for, 22.20–22.21
  - exhaustive methods, 22.21–22.23
  - local methods, 22.24–22.27
  - model-dependent methods, 22.27–22.33
- design and operation of, 22.3–22.13
- electronics and controls, 22.6–22.13, 22.7*f*–22.13*f*
  - and monochrome CRTs, 22.3–22.4, 22.3*f*
  - shadowmask color CRTs, 22.4–22.6, 22.5*f*, 22.6*f*
- operational characteristics of, 22.13–22.18
- colorimetric standards, 22.14
  - spatial characteristics of emitted light, 22.14–22.15
  - spatial uniformity, 22.17–22.18, 22.17*f*, 22.18*f*
  - stability of output, 22.16–22.17, 22.16*f*, 22.17*f*
  - temporal characteristics of emitted light, 22.14–22.16
  - timing and synchronization standards, 22.13–22.14
- setup for image display, 22.18–22.19
- brightness, 22.19
  - color balance and tracking, 22.19
  - contrast, 22.19
  - focus, 22.19
  - viewing environments, 22.19–22.20
- Color constancy, 10.39–10.40, 11.2
- Color constancy mechanisms, 11.71–11.72
- Color contrast, 11.2, 11.4*f*
- Color convergence, 23.1
- Color coordinate systems, 10.11–10.24
- adjusting cone spectral sensitivities, 10.17–10.18
  - colorimetric measurements, 10.23–10.24
  - color-matching functions, 10.11–10.13
  - cone fundamentals, 10.13–10.14
  - limits of color-matching data, 10.15–10.17
  - opponent and contrast spaces, 10.18–10.19
  - stimulus spaces, 10.11
  - visualizing color data, 10.19–10.23, 10.19*f*, 10.22*f*, 10.23*f*



- Color coordinates:  
 in color-deficient observers, 10.16  
 correcting for chromatic aberrations, 10.17  
 of different visual systems, 10.38–10.39  
 of surfaces, 10.36  
 transformation to CIELAB, 10.42–10.43
- Color data representations, 11.31–11.33, 11.32f
- Color direction, 11.12
- Color discrimination, 10.40–10.43, 10.41f, 14.15, 14.17
- Color LCDs, 22.37–22.40  
 colorimetry of color pixels, 22.38–22.39  
 controls and input standards, 22.39  
 geometry of color pixels, 22.37–22.38, 22.37f  
 spatial variations in output, 22.40  
 temporal variations in output, 22.39–22.40
- Color matching, 10.6–10.10  
 in color-deficient observers, 10.16  
 consistency across observers, 10.9  
 critical properties of, 10.8–10.10  
 errors in, 10.43, 10.43f, 10.44  
 Grassmann's laws, 10.8–10.9  
 maximum saturation method, 10.6–10.7, 10.6f  
 Maxwell's method, 10.8, 10.8f  
 persistence of, 10.9  
 trichromatic, 10.7–10.8  
 tristimulus values for arbitrary lights, 10.9  
 uniqueness in, 10.9
- Color opponency:  
 and color appearance, 11.62–11.66  
 and “forbidden” colors, 11.74  
 implied by test measurements, 11.12–11.15, 11.13f, 11.26–11.27, 11.28f  
 and multiplexing of color and luminance signals, 11.76–11.79  
 third-level, zero crossings of, 11.83, 11.84
- Color space transformation matrix, 10.1, 10.31
- Color spaces, 10.11, 10.18–10.19, 11.31–11.33  
 contrast, 10.19  
 defined, 11.2  
 different directions of, 11.39, 11.41–11.46, 11.44f, 11.46f  
 detection contours in L, M plane, 11.40f, 11.41–11.42  
 detection in planes other than L,M, 11.43–11.45, 11.44f  
 mechanism interactions, 11.42–11.43  
 spatial and temporal CSFs, 11.45–11.46, 11.46f
- Color spaces (*Cont.*):  
 DKL, 10.19, 11.32, 11.32f, 11.33  
 opponent, 10.19  
 specifying, 10.11  
 stimulus, 10.11  
 transformations between, 10.24t, 10.29–10.32  
 uniform, 10.40, 10.42
- Color temperature, 23.2
- Color valence, 11.2, 11.63, 11.65f, 11.66
- Color vision, age-related changes in, 14.15, 14.17
- Color vision mechanisms, 11.1–11.85  
 basic model details and limits, 11.31  
 chromatic discrimination, 11.57–11.62  
 near detection threshold, 11.58–11.59  
 pedestal experiments, 11.59, 11.60f, 11.61–11.62  
 color and contours, 11.72–11.79, 11.73f–11.75f  
 color appearance and stabilized borders, 11.74–11.75  
 contours and after-effects, 11.75–11.76, 11.75f  
 gap effect and luminance pedestals, 11.72, 11.74  
 McCollough effect, 11.76, 11.77f  
 multiplexing chromatic and achromatic signals, 11.76–11.79, 11.78f  
 color appearance and color opponency, 11.62–11.66  
 color-opponent response (valence functions), 11.63, 11.65f  
 hue scaling, 11.63, 11.64f  
 opponent-colors theory, 11.62–11.63  
 spectral properties of color-opponent mechanisms, 11.63  
 unique hues and equilibrium colors, 11.63–11.66  
 color constancy, 11.71–11.72  
 color data representations, 11.31–11.33, 11.32f  
 color-appearance mechanisms, 11.26–11.31  
 color-discrimination mechanisms vs., 11.5–11.8, 11.6f, 11.7f, 11.81–11.82  
 field measurements and first-site adaptation, 11.27, 11.29, 11.30f  
 field measurements and second-site adaptation, 11.29, 11.31  
 test measurements and opponency, 11.26–11.27, 11.28f

- Color vision mechanisms (*Cont.*):
- color-discrimination mechanisms, 11.9–11.26
  - color-appearance mechanisms *vs.*, 11.5–11.8, 11.6f, 11.7f, 11.81–11.82
  - field method, 11.11–11.12
  - first-site adaptation, 11.15–11.17, 11.16f
  - opponency implied by test measurements, 11.12–11.15, 11.13f
  - psychophysical test method, 11.9, 11.11, 11.12
  - second-site adaptation, 11.17–11.22, 11.18f, 11.19f
  - sites of limiting noise, 11.20, 11.23–11.26, 11.23f, 11.25f
- field sensitivities, 11.46–11.57
- achromatic detection and chromatic adaptation, 11.47–11.49, 11.48f
  - chromatic adaptation and the Sloan notch, 11.49, 11.51
  - detection contours and field adaptation, 11.53–11.54, 11.54f
  - field additivity, 11.51, 11.52f
  - first- and second-site adaptation, 11.51, 11.52, 11.53f
  - habituation or contrast adaptation experiments, 11.54–11.56, 11.55f
  - multiple cone inputs, 11.49, 11.50f
  - noise-masking experiments, 11.56–11.57
  - Stiles'  $\pi$ -mechanisms, 11.46, 11.47, 11.47f
- guiding principles of, 11.8–11.9
- linearity of color-opponent mechanisms, 11.66–11.70
- Bezold-Brücke effect and invariant hues, 11.67–11.68, 11.67f
  - color appearance and chromatic adaptation, 11.68
  - color appearance and chromatic detection/discrimination, 11.69
  - color appearance and habituation, 11.69–11.70
  - luminance and brightness, 11.70, 11.70f
  - tests of linearity, 11.66–11.67
- low-level and higher-order mechanisms, 11.79–11.80
- and mechanism concept, 11.9–11.11, 11.10f
- nomenclature for, 11.8
- test sensitivities, 11.34–11.46
- Color vision mechanisms, test sensitivities (*Cont.*):
- to different directions of color space, 11.39, 11.41–11.43, 11.44f, 11.45–11.46, 11.46f
  - luminance, 11.37–11.39, 11.38f, 11.40f
  - to spectral lights, 11.34, 11.35f, 11.36f, 11.37
  - three-stage zone models, 11.82–11.85, 11.85f
  - unipolar *vs.* bipolar chromatic mechanisms, 11.80–11.81
- Color-appearance mechanisms, 11.26–11.31
- color-discrimination mechanisms *vs.*, 11.5–11.8, 11.6f, 11.7f, 11.81–11.82
  - defined, 11.2
  - field measurements
    - and first-site adaptation, 11.27, 11.29, 11.30f
    - and second-site adaptation, 11.29, 11.31
  - test measurements and opponency, 11.26–11.27, 11.28f
- Color-discrimination mechanisms, 11.9–11.26, 11.10f
- color-appearance mechanisms *vs.*, 11.5–11.8, 11.6f, 11.7f, 11.81–11.82
  - defined, 11.2
  - field method, 11.11–11.12
  - first-site adaptation, 11.15–11.17, 11.16f
  - low-level and higher-order, 11.79–11.80
  - opponency implied by test measurements, 11.12–11.15, 11.13f
  - psychophysical test method, 11.9, 11.11, 11.12
  - second-site adaptation, 11.17–11.22, 11.18f, 11.19f
  - sites of limiting noise, 11.20, 11.23–11.26, 11.23f, 11.25f
- Colorimeters, 10.23
- Colorimetry, 10.1–10.45
- brightness matching and photometry, 10.43–10.45
  - color cameras, 10.38–10.40
  - color coordinate systems, 10.11–10.24
    - adjusting cone spectral sensitivities, 10.17–10.18
  - colorimetric measurements, 10.23–10.24
  - color-matching functions, 10.11–10.13
  - cone fundamentals, 10.13–10.14
  - limits of color-matching data, 10.15–10.17
  - opponent and contrast spaces, 10.18–10.19
  - stimulus spaces, 10.11
  - visualizing color data, 10.19–10.23, 10.19f, 10.22f, 10.23f

- Colorimetry (*Cont.*):  
 color discrimination, 10.40–10.43, 10.41f  
 color matching, 10.6–10.10  
 conventional terms/notation, 10.4, 10.4t, 10.5  
 errors in color-matching functions, 10.43, 10.43f, 10.44  
 image processing chain, 10.2f  
 matrix representations/calculations, 10.24–10.32  
 stimulus representation, 10.24–10.27, 10.25f, 10.26f  
 transformations between color spaces, 10.24t, 10.29–10.32  
 vector representation of data, 10.25f, 10.27–10.29, 10.27f  
 metamerism, 10.36–10.38, 10.37f  
 scope of, 10.3  
 standards for color CRTs, 22.14  
 surfaces and illuminants, 10.32–10.36, 10.33f, 10.34f  
 trichromacy, 10.4–10.6  
 univariance, 10.4  
 visual systems, 10.38–10.40
- Color-matching functions (CMFs), 10.11–10.13  
 defined, 10.1, 10.4t  
 limits of data, 10.15–10.17  
 and luminosity function, 10.10  
 and maximum saturation method, 10.7  
 online tabulation of, 10.11  
 specificity of, 10.9  
 standards for, 10.12–10.13  
 tailored to individuals, 10.15  
 transformation of, 10.10, 10.10f
- Color-opponent mechanisms:  
 linearity of, 11.66–11.70  
 Bezold-Brücke effect and invariant hues, 11.67–11.68, 11.67f  
 color appearance and chromatic adaptation, 11.68  
 color appearance and chromatic detection/discrimination, 11.69  
 color appearance and habituation, 11.69–11.70  
 luminance and brightness, 11.70, 11.70f  
 tests of linearity, 11.66–11.67  
 and opponent-colors theory, 11.3  
 spectral properties of, 11.63
- Commission Internationale de l'Éclairage (CIE), 7.9, 10.1, 10.3, 10.27
- Complex cells (cortical neurons), 2.14
- Computer Vision Syndrome (CVS), 23.1–23.12  
 disorders and eye conditions, 23.9–23.12  
 accommodation, 23.10–23.11  
 anisometropia, 23.11  
 binocular vision, 23.11  
 dry eyes, 23.9  
 presbyopia, 23.11–23.12  
 refractive error, 23.10  
 and work environment, 23.4–23.9  
 lighting, 23.4–23.5, 23.5t  
 monitor characteristics, 23.6–23.8  
 screen reflections, 23.5–23.6  
 work habits, 23.8–23.9  
 workstation arrangement, 23.8
- Computer-Automatic Virtual Environment (CAVE), 13.32
- Concomitant eye motion, 13.1
- Cone contrast spaces, 11.2, 11.32, 11.41–11.42
- Cone contrasts, 11.2, 11.32
- Cone coordinates, 10.1, 10.11 (*See also* Tristimulus values)
- Cone fundamentals, 10.13–10.14  
 adjusting cone spectral sensitivities, 10.17–10.18  
 defined, 10.1, 10.4t  
 online tabulation of, 10.11  
 primaries yielding, 10.10  
 Smith-Pokorny, 10.12  
 Stockman and Sharpe, 10.12, 10.16  
 tailored to individuals, 10.15
- Cone magno pathway, 2.9, 2.10f
- Cone mechanisms, 11.10f, 11.11  
 defined, 11.2  
 and stimulus direction, 11.33  
 and test measurements, 11.12–11.14
- Cone parvo pathway, 2.9, 2.10f
- Cone pathways, 2.9, 2.10, 2.10f
- Cone polymorphism, 10.15
- Cone-excitation spaces, 11.31–11.32
- Cone-opponent mechanisms, 11.8  
 and bichromatic test mixtures, 11.39  
 defined, 11.2  
 sensitization in, 11.69  
 unipolar vs. bipolar, 11.80–11.81

- Cones (cone photoreceptors):  
 adaptation at, 11.52  
 and age-related photopic vision changes, 14.15  
 alignment of, 8.4  
 densities and distributions of, 2.7  
 in dichromatic observers, 10.14  
 directional sensitivity of, 8.5  
 ideal “average,” 8.21, 8.21f  
 inner segments, 2.6f  
 light-collection area of, 2.8  
 linear density of, 2.6f  
 and maximum saturation color matching, 10.7  
 measuring activity of, 15.24  
 optical waveguide properties of, 14.11  
 outer segments of, 2.6f  
 pedicles of, 8.20  
 photocurrent responses of, 2.8f  
 in retinal layer of rods and cones model of biological waveguides, 8.8–8.9, 8.9f, 8.10f, 8.12–8.15  
 S-cone flicker sensitivity, 11.74–11.75  
 spatial distribution of, 2.6  
 spectral sensitivities of, 10.3, 10.13, 11.8  
 adjusting for individual differences, 10.17–10.18  
 estimates of, 10.14, 10.14f  
 loss hypothesis for, 10.14  
 in normal and dichromatic observers, 10.13  
 time constant of photopigment regeneration, 2.7  
 types and functions of, 2.4, 2.6
- Confocal microscopes, 17.1–17.10  
 clinical, 17.3, 17.6–17.9  
 clinical applications using, 17.8–17.9  
 defined, 17.1  
 development of, 17.3, 17.5  
 laser scanning, 17.3, 17.7–17.8  
 Nipkow disk, 17.4f, 17.5, 17.5f  
 scanning slit, 17.3, 17.6–17.9, 17.6f, 17.7f  
 spatial filtering with, 17.3  
 Svishchev, 17.6, 17.6f  
 theory of confocal microscopy, 17.3, 17.4f
- ConfoScan 4 microscope, 17.6, 17.7, 17.9
- Conjugate (version) eye movements/position, 1.42, 1.43, 13.1, 13.7, 13.20
- Contact lenses, 20.1–20.34  
 accommodation with, 20.26–20.30, 20.27f–20.29f, 20.29t  
 anisophoria and, 13.26  
 for aphakic patients, 12.15  
 base curve of, 12.1  
 convergence with, 20.25–20.26  
 correction with, 12.11–12.14  
 hydrogel lenses, 12.12–12.13  
 for presbyopia, 12.13–12.14, 14.27–14.28  
 rigid lenses, 12.11–12.12  
 design considerations, 20.20–20.25  
 aberrations, 20.24–20.25, 20.25f  
 aspheric lenses, 20.20–20.23, 20.21f–20.23f  
 posterior peripheral curve, 20.23–20.24, 20.24t  
 design parameters, 20.2–20.6, 20.4t–20.5t  
 base curve radius (BCR), 20.3, 20.4f, 20.5  
 center thickness, 20.6  
 edge thickness, 20.6  
 optical zone diameter (OZD), 20.5  
 overall diameter (OAD), 20.5  
 posterior peripheral curve systems, 20.5, 20.6  
 magnification, 20.31–20.33  
 relative spectacle, 20.32–20.33, 20.33f  
 spectacle, 20.31–20.32  
 materials for, 20.3  
 power of, 20.6–20.20  
 back and front vertex power, 20.7–20.8, 20.8t  
 and contact lens as thick lens, 20.6–20.7  
 effective power, 20.8–20.12, 20.9t, 20.10f, 20.11f  
 lacrimal lens consideration, 20.12–20.15, 20.12f–20.14f  
 residual astigmatism, 20.15  
 of soft lenses, 20.15–20.16  
 of toric lenses, 20.16–20.20, 20.16f–20.20f, 20.18t  
 prismatic effects with, 20.30–20.31  
 prism-ballasted contacted lenses, 20.30  
 unintentional (induced) prism, 20.31  
 (See also *specific types of lenses*)
- Contours (color vision), 11.72–11.79, 11.73f–11.75f  
 and after-effects, 11.75–11.76, 11.75f  
 color appearance and stabilized borders, 11.74–11.75

- Contours (color vision) (*Cont.*):  
 detection surface/contour, 11.12  
 defined, 11.2  
 and directions of color spaces, 11.40*f*  
 in equiluminant plane, 11.44*f*  
 and field adaptation, 11.53–11.54, 11.54*f*  
 in the L,M plane, 11.40*f*, 11.41–11.42  
 discrimination contours, 11.44*f*  
 gap effect and luminance pedestals, 11.72, 11.74  
 McCollough effect, 11.76, 11.77*f*  
 multiplexing chromatic and achromatic signals, 11.76–11.79, 11.78*f*  
 threshold surface/contour, 11.12–11.15, 11.13*f*  
 defined, 11.3  
 and loss of information, 11.20  
 and noise, 11.20, 11.23–11.26, 11.23*f*, 11.25*f*  
 and second-site adaptation to steady fields, 11.18*f*
- Contrast:  
 in color CRTs, 22.19  
 defined, 23.2  
 in monochrome CRTs, 22.7, 22.8*f*  
 in retinal imaging, 15.21–15.22  
 in vision experiments, 3.4  
 and visual acuity, 4.12, 4.12*f*
- Contrast coding:  
 first-site adaptation, 11.15–11.16  
 Weber's law and, 11.15–11.16, 11.16*f*
- Contrast (modulation) color spaces, 10.19
- Contrast constancy, 2.33
- Contrast detection, 2.19–2.31  
 adaptation and inhibition, 2.25–2.27, 2.25*f*  
 chromatic, 2.29–2.31, 2.30*f*  
 eye movements, 2.21  
 as function of spatial frequency, 2.20*f*  
 optical transfer function, 2.21–2.22, 2.22*f*  
 optical/retinal inhomogeneity, 2.24, 2.25*f*  
 receptors, 2.22–2.23  
 spatial channels, 2.23–2.24  
 temporal, 2.27–2.29, 2.27*f*, 2.29*f*
- Contrast discrimination, 2.31–2.32, 2.31*f*
- Contrast discrimination functions, 2.31–2.32, 2.31*f*
- Contrast estimation, 2.33
- Contrast masking, 2.32–2.33, 2.32*f*
- Contrast masking functions, 2.32, 2.32*f*
- Contrast sensitivity, 15.1
- Contrast sensitivity functions (CSFs), 2.20–2.23, 2.20*f*, 2.22*f*, 24.3  
 and adaptation, 2.25–2.27, 2.25*f*  
 age-related changes in, 14.17, 14.19  
 chromatic, 2.29–2.31  
 in color vision, 11.45–11.46, 11.46*f*  
 at different retinal eccentricities, 2.24, 2.25*f*  
 of ideal observer, 2.24  
 for motion detection, 2.36, 2.37  
 spatial, 11.45–11.46, 11.46*f*  
 spatio-temporal, 2.28–2.29  
 for stereopsis, 2.40, 2.41  
 temporal, 2.27–2.29, 2.29*f*, 11.45
- Contrast-transfer function, resolving capacity of the eye and, 4.5
- Convergence, 13.21  
 in color CRTs, 22.12, 22.13  
 with contact lenses, 20.25–20.26  
 defined, 13.1, 23.2  
 and egocentric direction, 13.7  
 and refractive errors, 12.15–12.16
- Convergence accommodation/convergence interaction (CA/C ratio), 13.24
- Convergence micropsia, 13.19
- Convergence responses to prism, 13.19
- Copepod, 8.1
- Copilia*, 8.4
- Cornea, 1.3*f*, 1.4–1.6, 16.3–16.4, 21.2  
 absorption of ultraviolet light at, 1.9  
 aging-related changes in, 14.5, 14.6*f*  
 asphericity in, 1.5*f*  
 and cataract (*see* Cataract)  
 confocal microscopy of (*see* Confocal microscopes)  
 as entrance pupil, 1.8  
 injury to, 7.4  
 keratometry of, 21.6  
 laser ablation, 16.11–16.19  
 ablation profiles, 16.14–16.15  
 ablation rate, 16.16–16.18, 16.18*f*  
 corneal photoablation, 16.16, 16.17*f*  
 Epi-LASIK, 16.12, 16.13, 16.13*f*  
 LASEK, 16.12, 16.13  
 LASIK, 16.13–16.14, 16.14*f*  
 photorefractive keratectomy (PRK), 16.11–16.12, 16.11*f*, 16.12*f*  
 thermal, photochemical, and photoacoustic effects, 16.18–16.19  
 optical axis of, 9.5  
 optical power of, 16.2

- Cornea (*Cont.*):  
 and refraction in the eye, 12.3  
 refractive index of, 14.5  
 refractive surgery modalities  
   corneal incisions/implants, 16.9–16.11  
   laser corneal procedures, 16.11–16.15  
 and rigid contact lenses, 12.12  
 spatially modulated excimer laser ablation of,  
   1.25  
 spectral sensitivities at, 10.18–10.19  
 stray light from, 1.20
- Corneal endothelium, 16.4  
 Corneal epithelium, 16.4  
 Corneal hydration, 16.18  
 Corneal incisions/implants, 16.9–16.11, 16.10f  
 Corneal photoablation, 16.16–16.19, 16.17f  
 ablation rate, 16.16–16.18, 16.18f  
 thermal, photochemical, and photoacoustic  
 effects, 16.18–16.19  
 Corneal relaxing incisions, 21.13  
 Corneal stroma, 16.4  
 Corneo effect, 7.5, 7.6f, 7.7  
 Corresponding retinal points, 13.1, 13.8  
 Cortical cataract, 12.14, 14.8  
 Cortical neurons, 2.14  
 Cotyledons, 8.1, 8.26, 8.27f  
 Craik-O'Brien-Cornsweet illusion, 11.78  
 Critical fusion frequency (CFF), 14.19  
 Cross-coupling, direction/distance of gaze and,  
   13.23–13.24, 13.23f  
 Crystallin proteins, 19.1  
 Crystalline lens, 1.3f, 16.3, 21.2  
 and accommodation, 1.30–1.31  
 aging-related changes in, 14.7–14.8, 14.17  
 cataract of, 12.14, 21.3 (*See also* Cataract;  
   Intraocular lenses)  
 and cone spectral sensitivities, 10.17  
 contours of refractive index in, 1.5f  
 distribution of refractive index in, 1.5–1.6,  
   1.5f  
 and fluctuations in accommodation, 1.32  
 fluorescence in, 1.21  
 gradient index structure of, 1.18, 21.2  
 growth of, 1.4  
 and refraction in the eye, 12.3  
 refractive index gradient of, 1.3, 19.12–19.13,  
   19.13f, 19.14f  
 scattered light from, 1.20  
 UV absorption in, 1.9  
 Cyclopean eye, 13.1  
 Cyclopean locus, 13.7  
 Cyclophoria, 13.35  
 Cycloplegia, 1.25, 25.11  
 Cyclovergence, 13.1, 13.22, 13.27  
 Cytochrome oxidase “blobs,” 2.14
- Da Vinci stereopsis, 13.4  
 Dark focus, 1.33, 1.34, 1.34f  
 De Valois De Valois zone model, 11.82  
 Decision tasks (in experiments), 3.2  
 Defocus:  
 and efforts to clear vision, 13.22–13.23  
 with head-mounted displays, 25.9–25.10,  
   25.9t–25.10t  
 modulation transfer with, 1.29  
 with monovision, 14.28  
 off-axis, 1.18, 1.19  
 Rayleigh units of, 1.29  
 spectacles for, 15.2  
 and visual acuity, 4.9, 4.9f, 4.10f  
 with visual instruments, 1.28  
 (*See also* Blur)
- Deformable mirrors, 15.1, 15.10, 15.10f  
 Degauss, in color CRTs, 22.9  
 Delta gun CRTs, 22.4, 22.13  
 Depth dependent sensitivity (OCT),  
   18.13–18.14, 18.13f, 18.14f  
 Depth ordering, 13.4, 13.5, 13.10–13.12  
 Depth perception, 2.39–2.40  
 age-related changes in, 14.22  
 stereodepth, 13.11–13.12  
 Depth range:  
 with OFDI, 18.16  
 in SD-OCT, 18.6–18.7  
 Depth-of-focus (DOF):  
 in human eye, 1.28–1.29  
 and pupil diameter, 1.8, 1.29, 1.30f  
 Derrington Krauskopf Lennie (DKL) space,  
   11.32, 11.32f, 11.33  
 defined, 11.2  
 equiluminant plane  
   hue scaling, 11.63  
   multiple mechanisms in, 11.57, 11.58  
 Descemet's membrane, 14.5, 16.4  
 Detection surface or contour (color vision), 11.12  
 defined, 11.2  
 and directions of color spaces, 11.40f  
 in equiluminant plane, 11.44f  
 and field adaptation, 11.53–11.54, 11.54f  
 in the L,M plane, 11.40f, 11.41–11.42

- Detection tasks, 2.15, 3.2
- Detection threshold:
- color vision
    - chromatic discrimination near, 11.58–11.59
    - two-stage model of, 11.23*f*
  - and discrimination thresholds, 4.6–4.7
- Detectors, light, 5.21, 5.22*f*
- Deuteranopes, 10.16
- DeVries-Rose law, 2.26
- Diabetic retinopathy, 14.1, 14.25–14.26
- Diaphany, 1.20
- Diattenuation, 18.1
- Dichroism, 18.1
- Dichromatic vision, 10.13, 10.14, 10.16
- Diffraction:
- in multifocal optics, 21.15–21.18
  - and retinal image quality, 1.12–1.14, 1.21
- Diffraction limit, 4.1
- Diffraction limited imaging, 17.1
- Diffraction theory, 6.1
- Diffraction contact lenses, 14.28
- Diffusers, 5.10–5.11, 5.10*f*
- Digital imaging:
- color representations in, 10.35
  - displays for vision research
    - color cathode ray tubes (color CRTs), 22.1–22.34
    - liquid crystal displays (LCDs), 22.34–22.40
- Digital Libraries, 24.7–24.8, 24.10
- Diopter (D), 12.1, 12.4, 13.21, 25.1
- Dioptric errors of focus, 1.13, 1.14, 1.28
- Diplopia (double vision), 23.2
- Dipper function, 11.59, 11.60*f*
- Direct ophthalmoscope, 12.5–12.6
- Direction, perception of, 13.7–13.10
- corresponding retinal points, 13.8
  - horopter, 13.8–13.9, 13.9*f*, 13.10*f*
  - vertical horopter, 13.9
- Direction discrimination, 2.37
- Direction selectivity, 2.14*n*
- Disability glare, 14.4, 14.12, 23.2
- Discomfort glare, 14.12, 23.2
- Disconjugate eye movements, 13.1
- Discrete actuator deformable mirrors, 15.1, 15.10, 15.10*f*
- Discrimination contours (color vision), 11.44*f*
- Discrimination experiments (*see* Pedestal experiments)
- Discrimination tasks, 2.15, 3.2
- Discrimination thresholds, 4.6–4.7
- Diseases of the eye:
- age-related, 14.3
    - age-related macular degeneration, 14.1, 14.24–14.25
    - cataract, 14.24
    - diabetic retinopathy, 14.25–14.26
    - glaucomas, 14.26–14.27
    - life-span environmental radiation damage, 14.22–14.23
    - visual impairment secondary to, 14.4 (*See also specific diseases*)
- Disjunctive eye movements (*see* Vergence eye movements)
- Displays for vision research:
- color cathode ray tubes (color CRTs), 22.1–22.34
  - colorimetric calibration of, 22.20–22.34
  - design and operation of, 22.3–22.13
  - operational characteristics of, 22.13–22.18
  - setup for image display, 22.18–22.19
  - viewing environments, 22.19–22.20
  - liquid crystal displays (LCDs), 22.34–22.40
  - color LCDs, 22.37–22.40
  - monochrome, operational principles of, 22.34–22.37
- Disruptive movements (*see* Vergence eye movements)
- Distal stimuli (human vision), 4.2
- Distance, perception of, 13.24
- Distance conflicts, in head mounted display systems, 13.31–13.32
- Distortion, with head-mounted displays, 25.7
- Divergence, 13.21
- defined, 13.1
  - in first 6 weeks of life, 13.21
- DKL color space, 10.19
- Doppler OCT, 18.18–18.19, 18.18*f*–18.20*f*
- Dot matrix, 23.2
- Dot pitch, 23.2
- Double-pass methods (retinal image quality), 1.22–1.23
- Drift, 1.44
- Droplet keratopathies, 7.6, 7.7
- Dry age-related macular degeneration, 14.1
- Dry eyes:
- with increasing age, 14.5
  - as problem with computer work, 23.9

- Dysphotopsia:  
 defined, 21.1  
 with intraocular lenses, 21.21
- Eccentricity, 2.3, 2.7, 2.24, 2.25*f*
- Eclipse burn of the retina, 7.3
- Edge clearance, 20.1
- Edge image, in diffraction-limited eye, 1.13
- Edge thickness, for contact lenses, 20.6
- Effective power:  
 defined, 20.1  
 for spectacle and contact lenses, 20.8–20.12,  
 20.9*t*, 20.10*f*, 20.11*f*
- “Effective” troland, 8.7
- Egocenter, 13.2, 13.7
- Ego-motion, 13.8, 13.9
- Electromagnetic radiation, 23.2
- Electronic imaging:  
 human vision and (*see* Human vision and  
 electronic imaging)  
 in ophthalmoscopic methods, 1.23
- Ellipse blindness, 7.7
- Ellipsoid, 8.1
- Emmetropia, 1.6, 1.32, 12.4, 16.4, 16.5*f*  
 age-related, 14.11  
 defined, 6.1, 12.1, 13.2  
 and focus of collimated light, 12.3*f*  
 and Maxwellian viewing, 5.4–5.5
- Emmetropization, 1.7
- Emotional characteristics in electronic  
 imaging (*see* Aesthetic and emotional  
 characteristics)
- Empirical horopter, 13.8
- Empty field myopia, 1.33
- Endothelium:  
 corneal, 16.4  
 defined, 16.1
- Entrance pupil:  
 and correction for SCE-1, 9.4  
 defined, 13.2  
 in Maxwellian viewing, 5.7  
 in reflectometry, 8.6–8.7  
 and retinal illuminance, 1.12  
 and retinal irradiance, 2.4  
 and stimulus specification,  
 4.2–4.3, 4.3*f*
- Epicotyl, 8.26, 8.27*f*
- Epi-LASIK (epithelial laser in situ  
 keratomileusis), 16.12, 16.13, 16.13*f*
- Epithelial layer (cornea), 14.5
- Epithelium:  
 corneal, 16.4  
 defined, 16.1
- Equatorial plane, 19.1
- Equilibrium colors:  
 additivity with, 11.66  
 and unique hues, 11.63–11.66  
 white as, 11.63
- Equilibrium level (accommodation),  
 1.33
- Equiluminance, difficulty in producing,  
 11.80
- Equiluminant plane:  
 detection and discrimination contours in,  
 11.44*f*  
 of DKL space  
 hue scaling, 11.57  
 multiple mechanisms in, 11.57, 11.58
- Equivalent reflectance, retinal, 1.11
- “Equivalent” troland, 9.2
- Equivalent veiling luminance, 1.20
- Equivalent-sphere correction, 1.6
- Ergonomics, 23.2
- Erythema, 7.1
- Estimation tasks, 2.15–2.16
- Etiolated plant tissues, 8.26–8.28
- Excimer lasers, 16.15–16.16
- Exhaustive characterization methods (ECM)  
 (color CRTs), 22.21–22.23  
 interpolation, 22.22–22.23  
 inverses, 22.23  
 out-of-gamut colors, 22.23  
 sampling, 22.21–22.22
- Exit pupil, 2.3  
 measurement of power at, 5.17  
 in reflectometry, 8.6–8.7, 8.7*f*  
 and retinal illuminance, 1.12  
 and Stiles-Crawford effects, 1.11
- Experimental conditions, 3.2
- Exposure limits (ELs) (radiation), 7.9–7.11  
 exceeding, 7.11  
 for infrared, 7.10–7.11  
 for laser light, 7.12  
 for ultraviolet, 7.9–7.10  
 for visible light, 7.10
- External limiting membrane (ELM), 8.8, 8.9,  
 8.10*f*, 8.12
- Extracapsular cataract extraction (ECCE),  
 12.14, 21.4
- Extraocular muscles, 1.42, 12.8, 12.16



- Extraretinal cues:  
 defined, 13.2  
 for eye movements, 13.7  
 in space perception, 13.3
- Exudative (wet) age-related macular degeneration, 14.1, 14.24–14.35
- Eye alignment:  
 binocular parallax, 13.22  
 cross-coupling and direction/distance of gaze, 13.23–13.24, 13.23f  
 defocus and efforts to clear vision, 13.22–13.23  
 intrinsic stimuli to vergence, 13.21–13.22  
 magnification induced errors of, 13.25–13.27  
 in Maxwellian viewing, 5.8  
 perceived distance, 13.24  
 prism induced errors of, 13.27–13.29  
 zone of clear and single binocular vision, 13.24–13.25, 13.25f
- Eye movements, 1.42–1.45, 13.19–13.20  
 analysis of, 24.7  
 characteristics of, 1.43–1.44  
 in contrast detection, 2.21  
 coordination of  
 binocular parallax, 13.22  
 cross-coupling and direction/distance of gaze, 13.23–13.24, 13.23f  
 defocus and efforts to clear vision, 13.22–13.23  
 intrinsic stimuli to vergence, 13.21–13.22  
 perceived distance, 13.24  
 zone of clear and single binocular vision, 13.24–13.25, 13.25f  
 defined, 13.2  
 extraretinal information for, 13.7  
 gaze control, 13.29–13.30  
 and optical flow, 2.39  
 stability of fixation, 1.44, 1.45f  
 types of, 13.19–13.20
- Eye protectors, for laser hazards, 7.14
- Eye relief, 1.7
- Eyes:  
 bovine lenses, 19.8–19.11, 19.11f  
 cat lenses, 19.9  
 fish lenses, 19.6, 19.6f  
 gibbon lenses, 19.14  
 guinea pig lenses, 19.8  
 of invertebrate aquatic worms, 9.14  
 model, 21.13  
 octopus lenses, 19.7, 19.7f
- Eyes (*Cont.*):  
 pig lenses, 19.11, 19.12f  
 porcine lenses, 19.11, 19.12f  
 primate eye lens, 19.13f, 19.14, 19.14f  
 rabbit lenses, 19.8  
 rat lenses, 19.7–19.8  
 (*See also* Human eye)
- Eyesight, defined, 23.2
- Eyestrain (asthenopia), 23.2
- F line, 25.1
- Far point, 12.1, 12.8, 16.5
- Farsightedness, 23.2 (*See also* Hyperopia)
- Fiber optic bundles, human receptors as, 8.3–8.5
- Fiber optic features:  
 of plant tissues, 8.26–8.28, 8.27f  
 of sponges, 8.28–8.29
- Fibroblast, 16.1
- Field additivity (color vision), 11.51, 11.52f
- Field measurements (color vision):  
 and first-site adaptation, 11.27, 11.29, 11.30f  
 and second-site adaptation, 11.29, 11.31
- Field method (color vision), 11.11–11.12  
 defined, 11.2  
 evidence for higher-order mechanisms from, 11.79–11.80
- Field of view, 8.4  
 for head-mounted displays, 25.2, 25.5, 25.7–25.8, 25.9f  
 in Maxwellian viewing, 6.6–6.7, 6.6f
- Field quality, in optical systems, 5.11
- Field sensitivities (color vision), 11.46–11.57  
 achromatic detection and chromatic adaptation, 11.47–11.49, 11.48f  
 chromatic adaptation and the Sloan notch, 11.49, 11.51  
 detection contours and field adaptation, 11.53–11.54, 11.54f  
 field additivity, 11.51, 11.52f  
 first- and second-site adaptation, 11.51, 11.52, 11.53f  
 habituation or contrast adaptation experiments, 11.54–11.56, 11.55f  
 multiple cone inputs, 11.49, 11.50f  
 noise-masking experiments, 11.56–11.57  
 Stiles'  $\pi$ -mechanisms, 11.46, 11.47, 11.47f  
 "Field" sensitivity method (color vision), 11.11–11.12
- Fields, in color CRTs, 22.9–22.10, 22.9f, 22.10f

- Filters, intensity, 5.14, 5.14*t*
- First-site adaptation (color vision), 11.11, 11.15–11.17, 11.16*f*  
 defined, 11.2  
 field measurements and, 11.27, 11.29, 11.30*f*  
 and field sensitivities, 11.51, 11.52, 11.53*f*  
 incompleteness of, 11.17  
 signals reaching second site, 11.17  
 Weber's law and contrast coding, 11.15–11.16, 11.16*f*
- Fish lenses, 19.6, 19.6*f*
- Fish-eye lenses, 19.2–19.3
- Fixation, 1.42  
 defined, 13.2  
 stability of, 1.44, 1.45*f*
- Flexure, 20.16
- Flicker photometrics, 10.34
- Flood-illuminated AO ophthalmoscope, 15.3, 15.16–15.17, 15.16*f*, 15.17*f*
- Fluence, 8.26  
 corneal photoablation, 16.16, 16.17, 16.19*f*  
 defined, 8.1
- Fluorescence:  
 lenticular, 1.21  
 stray light as result of, 1.21
- Fluoro-silicone/acrylate (F-S/A), 20.1
- Focal length, 23.2
- Focimeter, 12.9
- Focus:  
 in color CRTs, 22.19  
 in Maxwellian viewing, 5.4–5.5, 5.5*f*, 6.5–6.6  
 in monochrome CRTs, 22.7  
 and responses to distance, 13.30
- Focus of expansion, 13.2
- Font, 23.2
- Foot-candle, 23.2
- “Forbidden” colors, 11.74
- Fourier approach to optics, 4.8
- Fourier domain OCT (*see* Spectral domain OCT)
- Fourier Theory of Optics:  
 defined, 6.1  
 in Maxwellian viewing, 6.10–6.12
- Fourier transform, 2.23, 18.1
- Fovea, 1.3, 1.3*f*, 1.9, 1.15, 2.5*f*, 14.9  
 and cone spectral sensitivities, 10.17  
 defined, 4.1, 18.1  
 disparities in, 2.40  
 function of, 13.3  
 receptor size in, 2.23
- Foveal avascular zone, 14.9
- Foveal pit, 14.11
- Foves, 2.24
- Frames, in color CRTs, 22.9–22.10, 22.9*f*, 22.10*f*
- Fraunhofer diffraction pattern, 6.10, 6.12
- Free (newtonian) viewing, 5.2–5.4, 5.3*f*  
 limitations of, 5.4  
 retinal illuminance, 5.2–5.3, 5.3*f*  
 the troland, 5.3–5.4
- Fringe washout (OCT), 18.15
- Front vertex power (FVP):  
 contact lenses, 20.7–20.8, 20.8*t*  
 defined, 20.1
- Frontoparallel plane, 13.2
- Gain control, in CRTs, 22.8
- Ganglion cells, 2.10–2.11  
 contrast sensitivity functions of, 2.12*f*  
 density of, 2.11  
 and LGN, 2.12  
 linear density of, 2.6*f*  
 midget, 2.10–2.11, 2.10*n*  
 off-center, 2.10  
 on-center, 2.10  
 parasol, 2.10*n*  
 P-cells, 11.76, 11.79  
 physiology of, 2.11  
 transfer functions of, 2.11–2.12
- Ganzfeld, 1.12
- Gap effect (color vision), 11.72, 11.74
- Gas permeable (GP) contact lenses, 12.12, 20.3  
 aberrations and, 20.24  
 aspheric, 20.22  
 base curve radius for, 20.3, 20.5  
 bitoric, power of, 20.17–20.20  
 center thickness of, 20.6  
 edge thickness of, 20.6  
 lacrimal or tear lens with, 20.12–20.15  
 OAD/OZD of, 20.5  
 posterior peripheral curve systems of, 20.5, 20.6  
 and residual astigmatism, 20.15
- Gaussian noise, 2.17, 2.18
- Gaze control, 13.29–13.30
- Gaze eccentricity, 13.8
- Gaze holding, 1.42
- Gaze shifting, 1.42
- Generalized pupil function, 2.3
- Geometric optics, 8.15
- Gibbon lenses, 19.14, 19.14*f*

- Glare:  
 with computer work, 23.4–23.5, 23.5*t*  
 defined, 23.2  
 in human eye, 1.20
- Glass lenses, for spectacles, 12.9
- Glaucoma, 14.3, 18.25–18.27  
 in aging eyes, 14.26–14.27  
 cause of, 14.6  
 defined, 14.1
- Gradient index (GRIN), 19.1, 19.2
- Gradient index optics, 19.1–19.15  
 axial gradients, 19.5  
 bovine lenses, 19.8–19.11, 19.11*f*  
 cat lenses, 19.9  
 eye lens, 19.5–19.6  
 fish lenses, 19.6, 19.6*f*  
 functional considerations, 19.14–19.15  
 guinea pig lenses, 19.8  
 human/primate lenses, 19.12–19.14, 19.13*f*,  
 19.14*f*  
 nature of index gradient, 19.21–19.15  
 octopus lenses, 19.7, 19.7*f*  
 pig lenses, 19.11, 19.12*f*  
 rabbit lenses, 19.8  
 radial gradients, 19.3–19.5, 19.4*f*, 19.5*f*  
 rat lenses, 19.7–19.8  
 spherical gradients, 19.2–19.3
- Grassmann's laws, 10.8–10.9, 10.28
- Grating acuity tasks, 2.34, 2.34*f*, 2.35*f*
- “Gray world” assumption, 10.39
- Guinea pig lenses, 19.8
- Habituation (color vision), 11.54–11.56,  
 11.55*f*  
 and color appearance, 11.29, 11.69–11.70  
 defined, 11.2  
 second-site, 11.19–11.20, 11.19*f*, 11.21*f*,  
 11.22*f*
- Half-bleaching constant, 2.7
- Halftoning, low-level vision models in, 24.4
- Halos, 1.21  
 in aging eyes, 14.13  
 with glaucoma, 14.6
- Hankel functions, 8.1  
 in modal analysis, 8.22  
 for waveguides, 8.13
- Haploopia, 13.2
- Haptic, 21.1
- Heading judgments, magnification and,  
 13.14–13.15
- Head-mounted displays (HMDs), 25.1–25.12  
 accommodation with, 25.10–25.12  
 binocular vision factors in design of,  
 13.31–13.35  
 distance conflicts, 13.31–13.32  
 optical errors, 13.34–13.35  
 spatial location conflicts, 13.33, 13.34*f*  
 visual-vestibular conflicts, 13.32–13.33  
 characterizing, 25.7–25.10, 25.8*f*,  
 25.9*t*–25.10*t*  
 common design considerations for, 25.2–25.7  
 comfort, 25.2–25.5, 25.3*f*  
 functionality, 25.5–25.6, 25.6*f*, 25.7*f*  
 safety, 25.2  
 usability, 25.2
- Heat sinks, 5.15
- Heidelberg Retinal Tomograph (HRT),  
 17.7, 17.9
- Helmholtz's reciprocity theorem, 8.1, 8.25, 8.27
- Henle fibers, 8.2, 8.20
- Hering's law, 13.20, 13.26
- Hertz (HZ), 23.2
- Heterochromatic flicker photometry (HFP),  
 11.37, 11.49
- Heterochromatic modulation photometry  
 (HMP), 11.37
- Heterophoria, 13.2, 13.26
- Higher-order aberrations, 16.6, 16.7  
 correction of, 16.9  
 defined, 15.1, 20.2
- Higher-order mechanisms (color vision), 11.80
- Homogenous index, 19.1
- Horizontal cells (retina), 2.9
- Horizontal scanning, in color CRTs, 22.10–22.11,  
 22.11*f*
- Horizontal/vertical size and position, in color  
 CRTs, 22.12
- Horopter, 2.40, 13.2, 13.8–13.9, 13.9*f*, 13.10*f*
- Hue cancellation, 11.26–11.27
- Hue scaling, 11.63, 11.64*f*
- Human eye, 1.1–1.45, 1.3*f*  
 aberrations in, 1.3 (*See also* Aberrations  
 [human eye])  
 accommodation response, 1.29–1.36  
 accuracy of, 1.32–1.34  
 age-dependent changes in, 1.35–1.36  
 application to instrumentation, 1.34–1.35  
 dynamics of, 1.31–1.32  
 stability of, 1.32  
 vergence input, 1.34

Human eye (*Cont.*):

- age-related changes in, 1.5, 1.7, 14.4–14.14  
(*See also* Adaptive optics)
- accommodation, 1.35–1.36, 1.35*f*
- accommodation and presbyopia, 14.8, 14.9*f*
- anterior and posterior chambers, 14.5, 14.6
- cornea, 14.5, 14.6*f*
- eye size, 14.11
- fluorescence, 1.21
- index of diffusion for, 1.23
- lens, 14.7–14.8
- and level of aberration, 1.18
- pupil, 14.6–14.7, 14.7*f*
- pupil diameter, 1.8, 1.18
- retina, 14.9–14.11
- retinal reflectance, 1.11
- RMS wavefront error, 1.15–1.17, 1.16*t*
- scattering, 1.20
- tears, 14.4–14.5
- transmittance, 1.9
- transparency/cataract, 14.8
- ametropia, 1.6–1.7
- cornea, 16.3, 16.4, 16.4*f*
- damage from light exposure, 5.18–5.19, 5.18*f*
- depth-of-focus, 1.8, 1.28–1.29, 1.30*f*
- models of, 1.36–1.38
  - paraxial, 1.36–1.37, 1.37*f*
  - wide-angle, 1.38
- monochromatic ocular aberrations, 1.4, 1.14–1.19
  - off-axis, 1.18–1.19, 1.18*f*
  - on the visual axis, 1.15–1.18
- movements of, 1.42–1.45
  - characteristics of, 1.43–1.44
  - stability of fixation, 1.44, 1.45*f*
- ocular parameters, 1.4–1.6, 1.4*f*, 1.5*f*
- ocular radiometry, 1.11–1.12
- optical components of, 2.2
- optics of, 16.2–16.3, 16.3*f*
- pupil diameter, 1.8–1.9
- refractive elements in, 12.3
- retinal illuminance, 1.11–1.12
- retinal image quality, 1.12–1.28
  - in aberration-free eye, 1.12–1.14
  - calculation from aberration data, 1.21–1.22
- chromatic aberration, 1.19–1.20, 1.19*f*
- comparison between methods, 1.23

Human eye, retinal image quality (*Cont.*):

- effects of aberration correction, 1.25–1.26
- intraocular scattered light, 1.20–1.21
- lenticular fluorescence, 1.21
- monochromatic ocular aberrations, 1.14–1.19
- observed optical performance, 1.23–1.25
- ophthalmoscopic (double-pass) methods, 1.22–1.23
- psychophysical comparison method, 1.22
  - and pupil diameter, 1.8
  - on the visual axis, 1.21–1.28
  - with visual instruments, 1.27–1.28
- retinal reflectance, 1.11
- size of, 14.11
- stereopsis, 1.38–1.42
  - aniseikonia, 1.41–1.42
  - stereoscopic and related instruments, 1.41
  - tolerances in binocular instrumentation, 1.41–1.42
- Stiles-Crawford effect, 1.10–1.11
- structure of, 21.2–21.3
- transmittance, 1.9–1.11
- visual vs. optical axis orientation in, 1.3  
(*See also specific topics*)
- Human hair, light guide effect in, 8.24–8.26
- Human vision and electronic imaging, 24.1–24.11
  - analysis of image features, 24.6–24.9
    - attention and region of interest, 24.7
    - Digital Libraries applications, 24.7–24.8
    - user interface design, 24.8–24.9
    - visualization, 24.8
  - information sources for, 24.11
  - perception of imaging artifacts, 24.2–24.6
    - early color vision, 24.5
    - embedding digital watermarks, 24.5
    - image quality and compression, 24.3–24.4
    - limitations of early vision models, 24.5–24.6
    - rendering and halftoning, 24.4
    - target detection in medical images, 24.5
  - representation of aesthetic and emotional characteristics, 24.9–24.10
    - color, art, and emotion, 24.10
    - image quality, 24.9
    - virtual reality and presence, 24.9
- Hydrogel contact lenses (*see* Soft contact lenses)
- Hydrophilic, defined, 16.1

- Hyperacuity, 2.34, 2.35, 4.13–4.16, 4.14*f*  
 across ages, 14.22  
 defined, 4.1  
 and superresolution, 4.15
- Hypermetropia (hyperopia), 1.6, 16.5, 16.5*f*, 16.8
- Hyperopia, 1.18, 12.4  
 defined, 13.2  
 and focus of collimated light, 12.3*f*
- Hypocotyl, 8.2, 8.26, 8.27*f*
- Ideal observer, in judgment experiments, 3.6
- Ideal-observer models (color vision), 11.26
- Ideal-observer theory, 2.16–2.19, 2.19*f*, 2.24
- Identification tasks, 2.15, 2.16
- Illuminance:  
 defined, 23.2  
 measurement of, 5.17, 5.18  
 retinal, 1.11–1.12  
 SI units of, 5.3
- Illuminants:  
 metamerism for, 10.38  
 reflected from surfaces, 10.32–10.36
- Image artifacts:  
 and perceived fidelity, 24.5  
 perception of (*see* Perception, of imaging artifacts)
- Image compression, early techniques for, 24.4
- Image features in vision, 24.6–24.9  
 attention and region of interest, 24.7  
 Digital Libraries applications, 24.7–24.8  
 user interface design, 24.8–24.9  
 visualization, 24.8
- Image formation, 2.2–2.4
- Image overlap, 13.4
- Image quality:  
 for head-mounted displays, 25.2, 25.5–25.7, 25.7*f*  
 in human eye (*see* Retinal image quality)  
 subjective approach to, 24.9
- Image sampling, by photoreceptors, 2.4–2.9
- Imaging science, 24.1
- Impulse response function (IRF), 14.19–14.21, 14.20*f*
- Incoherent light sources, 6.1
- Incoherent targets, pupil size and, 6.7–6.9, 6.8*f*, 6.9*t*
- Increment contrast, 2.31, 2.32
- Increment thresholds (color vision), 11.16*f*
- Incremental cone-excitation space, 11.2, 11.31, 11.32
- Infants:  
 corneal endothelium in, 14.5  
 lenses in, 14.8  
 size of eye globe in, 14.11
- Influence function, 15.1
- Information theory, visual resolution and, 4.15–4.16
- Information-processing model, 2.15–2.16, 2.15*f*
- Infrared (IR):  
 damage from, 7.2  
 exposure limits for, 7.10–7.11
- Infrared cataract, 7.7, 7.10
- Inhibition, 2.25–2.27, 2.25*f*
- Injuries, radiation, 7.3–7.7, 7.5*f*  
 cataract, 7.5–7.6, 7.6*f*  
 droplet keratopathies, 7.6, 7.7  
 infrared cataract, 7.7  
 mechanisms of, 7.2–7.3  
 action spectra, 7.2, 7.4*f*  
 exposure duration and reciprocity, 7.2–7.3  
 photokeratitis, 7.4  
 photoretinitis, 7.7  
 pterygium, 7.6, 7.7
- In-line gun CRTs, 22.5
- Input standards, for color LCDs, 22.39
- Instrument myopia, 1.34–1.35
- Integrating spheres, 5.10, 5.11
- Intensity (color vision):  
 defined, 11.2  
 dependence of hue on, 11.67–11.68, 11.67*f*  
 of test stimulus, 11.12
- Intensity of light, controlling, 5.13–5.15, 5.14*t*
- Intensity scaling (light), 10.24, 10.25, 10.25*f*
- Interferometers, 5.24
- Interlaced monitors, 23.2
- International Commission on Illumination (*see* Commission Internationale de l'Éclairage)
- International Commission on Non-Ionizing Radiation Protection (ICNIRP), 7.9, 7.12
- International Electrotechnical Commission (IEC), 7.9, 7.12
- International Standardization Organization (ISO), 7.9
- Interocular aniso-magnification, distortion from, 13.16–13.19  
 aniseikonia, 13.17–13.18  
 interocular blur suppression with anisometropia, 13.18–13.19  
 lenses and prisms, 13.16–13.17, 13.17*f*
- Interphotoreceptor matrix (IPM), 8.8, 8.9

- Interpupillary distance (IPD), 1.39–1.41, 1.40*f*
- Interstitial matrix, 8.2
- Intracapsular cataract extraction (ICCE), 12.14
- Intracorneal ring segments, 16.10–16.11, 16.10*f*
- Intraocular lenses (IOLs), 12.15, 16.9, 21.1–21.22
- accommodating, 14.29–14.30, 21.18–21.19
  - and aging of the eye, 21.3–21.4
  - aspheric, 21.10–21.12, 21.11*f*, 21.12*t*
  - for cataract correction, 14.8
  - and cataract surgery, 21.4
  - defined, 14.1–14.2, 21.1
  - design of, 21.5–21.20
    - accommodating lenses, 21.18–21.19
    - aspheric lenses, 21.10–21.12, 21.11*f*, 21.12*t*
    - chromophores, 21.20, 21.20*f*
    - IOL aberrations, 21.9–21.10, 21.10*f*
    - IOL power, 21.5–21.9, 21.7*t*, 21.8*f*
    - multifocal lenses, 21.14–21.18, 21.15*f*, 21.16*f*, 21.16*t*, 21.18*f*
    - phakic lenses, 21.19
    - testing IOLs, 21.13–21.14
    - toric lenses, 21.13, 21.12*f*
  - effective lens position (ELP), 21.7–21.9
  - multifocal, 21.14–21.18, 21.15*f*, 21.16*f*, 21.16*t*, 21.18*f*
  - phakic, 21.19
  - for presbyopic correction, 14.28–14.29
  - side effects of, 21.20–21.22
    - dysphotopsia, 21.21
    - posterior capsule opacification, 21.21–21.22
  - and structure of the eye, 21.2–21.3
  - toric, 21.13, 21.12*f*
- Intraocular pressure (IOP), 14.2, 14.26, 14.27, 16.3
- Intraocular scatter, age-related, 14.12
- Invariant hues:
  - and Bezold-Brücke effect, 11.67–11.68, 11.67*f*
  - defined, 11.2
- Invertebrate aquatic worm, eye structures in, 9.14
- Iris, 1.3*f*, 16.3, 21.2
- Irregular astigmatism, 16.6
- Isoaccommodation circle, 13.23, 13.23*f*
- Isochromatic CFSs, 2.30, 2.30*f*
- Isodiscrimination contours, 10.40
- Isoindicial contours, 19.1
- Isolating direction (color vision), 11.2, 11.33
- Isoluminant CFSs, 2.30, 2.30*f*
- Isoluminant patterns, 2.30
- Isoplanatic patch, 15.5–15.7, 15.6*f*
- Isovergence circle, 13.23, 13.23*f*
- Jackson cross-cylinder check test, 12.7
- Javal's rule, 12.5
- Jones matrix formalism, 18.1, 18.20, 18.22–18.24
- Judd's three-stage Müller zone theory, 11.6, 11.7*f*, 11.8
- Judd-Vos modified 2° color-matching functions, 10.12, 10.45
- Judgment tasks:
  - adjustment tasks *vs.*, 3.2
  - psychophysical measurement of, 3.6–3.8
    - ideal observer, 3.6
    - rating scale, 3.6–3.8, 3.7*f*
    - response time, 3.8
    - two-alternative forced choice (2afc), 3.8
    - yes-no, 3.6
- Just noticeable distortion level (JND), 24.4
- “K,” 20.2
- Keratoconus, 9.2, 20.24
- Keratocytes, 16.1, 16.4
- Keratometer mire, 12.5
- Keratometry, 12.5, 21.6
- Kinetic cues, for perceived space, 13.4–13.7, 13.5*f*, 13.6*f*
- Knapp's law, 13.17, 13.26
- Kohler illumination, 5.6
- La Jolla colorimeters, 5.10
- Lacrimal lens:
  - and contact lens power, 20.12–20.15, 20.12*f*–20.14*f*
  - defined, 20.2
- Lag, 23.2
- Lamellas, 16.1, 16.4
- Lamp safety standards, 7.14–7.15
- Landolt C test, 4.8
- Landolt ring target, 2.34
- Large field (color matching):
  - defined, 10.9
  - standards for, 10.11, 10.12–10.13
- LASEK (laser subepithelial keratomileusis), 16.12, 16.13

- Laser ablation, **16.11–16.19**  
 ablation profiles, **16.14–16.15**  
 ablation rate, **16.16–16.18, 16.18f**  
 corneal photoablation, **16.16, 16.17f**  
 Epi-LASIK, **16.12, 16.13, 16.13f**  
 LASEK, **16.12, 16.13**  
 LASIK, **16.13–16.14, 16.14f**  
 photorefractive keratectomy (PRK),  
**16.11–16.12, 16.11f, 16.12f**  
 thermal, photochemical, and photoacoustic  
 effects, **16.18–16.19**
- Laser photocoagulation, **14.26**
- Laser ray tracing, **16.7**
- Laser scanning confocal microscopes, **17.1, 17.3, 17.7–17.8**
- Laser scanning ophthalmoscopes, **17.3**
- Lasers:  
 eye injury from, **7.3, 7.4**  
 hazards related to, **7.11–7.14, 7.12f, 7.13f**  
 accidents, **7.12–7.14**  
 and eye protectors, **7.14**  
 safety standards, **7.12**  
 and speckle fields, **5.21**  
 types of, **5.20t**
- LASIK (laser-assisted keratomileusis) surgery,  
**1.6, 12.14, 16.13–16.14, 16.14f, 23.2**
- Lateral aniseikonia, **13.17**
- Lateral chromatic aberration (*See* Transverse  
 chromatic aberration)
- Lateral geniculate nucleus (LGN), **2.12–2.14**  
 anatomy of, **2.12–2.13, 2.13f**  
 physiology of, **2.13–2.14**
- Le Grand eye model, **2.3**
- LEDs, as light source, **5.15**
- Lens capsule, **21.1, 21.2, 21.21–21.22**
- Lens pigment density:  
 and color matches, **10.9**  
 and variations in color matching, **10.15**
- Lenses:  
 bovine, **19.8–19.11, 19.11f**  
 cat, **19.9**  
 computer-assisted surfacing technologies  
 for, **12.9**  
 for correction of refractive error, **12.4**  
 distortion from interocular aniso-  
 magnification, **13.16–13.17**  
 effects on vergence and phoria, **13.25–13.27**  
 fish, **19.6, 19.6f**  
 fish-eye, **19.2–19.3**  
 gibbon, **19.14, 19.14f**
- Lenses (*Cont.*):  
 guinea pig, **19.8**  
 of human eye, **1.3f, 19.5–19.6** (*See also*  
 Cornea; Crystalline lens; Gradient index  
 optics)  
 lacrimal, **20.2, 20.12–20.15, 20.12f–20.14f**  
 for Maxwellian viewing systems, **5.4**  
 octopus, **19.7, 19.7f**  
 in optical systems, **5.11**  
 porcine, **19.11, 19.12f**  
 primate, **19.13f, 19.14, 19.14f**  
 rabbit, **19.8**  
 rat, **19.7–19.8**  
 spherocylindrical, **12.4, 12.5**  
 stimulus value of, **13.21**  
 for vision correction (*see* Contact lenses;  
 Intraocular lenses; Spectacles)  
 Wyszecki, **9.5**
- Lensometer, **12.9**
- Lenticular absorption, **1.9**
- Lenticular fluorescence, **1.21**
- Leray, P., **24.7**
- Letter acuity task, **2.34**
- Life-span environmental radiation damage,  
**14.22–14.23**
- Light, defined, **4.1, 23.3**
- Light output:  
 in color CRTs  
 spatial characteristics of, **22.14–22.15**  
 spatial uniformity of, **22.17–22.18, 22.17f, 22.18f**  
 stability of, **22.16–22.17, 22.16f, 22.17f**  
 in color LCDs  
 spatial variations in, **22.40**  
 temporal variations in, **22.39–22.40**
- Light sources:  
 coherent, **5.19, 5.21**  
 control of (*see* Optical generation of visual  
 stimulus)  
 in Maxwellian viewing, **6.2**  
 in vision laboratories, **5.19, 5.19t–5.20t**
- Lighting:  
 for aging individuals, **14.4**  
 for computer work, **23.4–23.5, 23.5t**
- Limbal relaxing incisions, **21.13**
- Limbus, **21.1**
- Limiting noise, site of, **11.24**
- Linear colorimetry models, **10.26–10.27, 10.27f**
- Linear models (matrix algebra), **10.1, 10.46–10.47**

- Linear optics, 17.1
- Linear perspective, 13.3
- Linear polarization (OCT), 18.21
- Linear visual mechanisms, 11.3
- Linearity of color-opponent mechanisms, 11.66–11.70
  - Bezold-Brücke effect and invariant hues, 11.67–11.68, 11.67*f*
  - color appearance
    - and chromatic adaptation, 11.68
    - and chromatic detection and discrimination, 11.69
    - and habituation, 11.69–11.70
    - luminance and brightness, 11.70, 11.70*f*
    - tests of linearity, 11.66–11.67
- Linearization, in optical systems, 5.15
- Line-spread function (LSF), 1.21
  - in diffraction-limited eye, 1.13
  - in ophthalmoscopic methods, 1.23
- Liou Brennan schematic eye, 15.6*f*, 15.7
- Liquid crystal displays (LCDs), 22.34–22.40
  - color, 22.37–22.40
    - colorimetry of color pixels, 22.38–22.39
    - controls and input standards, 22.39
    - geometry of color pixels, 22.37–22.38, 22.37*f*
    - spatial variations in output, 22.40
    - temporal variations in output, 22.39–22.40
  - and Computer Vision Syndrome, 23.7–23.8
  - defined, 23.2
  - monochrome, operational principles of, 22.34–22.37, 22.35*f*, 22.36*f*
  - in optical systems, 5.16
  - reflections from, 23.6
- Liquid spatial light modulators (LC-SLMs), 15.11
- Local characterization methods (LCM) (color CRTs), 22.24–22.27
  - individual colors, 22.24–22.26
  - inverses, 22.25
  - local regions of color, 22.24–22.26
  - one-dimensional color spaces, 22.25, 22.26
  - out-of-gamut colors, 22.25
- Localization, 4.14, 4.15, 8.4
- Longitudinal (axial) chromatic aberration (LCA), 1.19, 1.20, 1.28, 8.8, 15.22
- Longitudinal horopter, 13.8
- Long-range motion discrimination
  - mechanism, 2.38
- Loss hypothesis, 10.14
- Low-level mechanisms (color vision), 11.80
- Low-order aberrations, 15.1
- Luminance, 11.37–11.39
  - cone numerosity, 11.38
  - defined, 2.29*n*, 11.3, 23.3
  - luminous efficiency, 11.37
  - luminous efficiency functions, 11.37, 11.38*f*
  - multiple cone inputs, 11.49, 11.50*f*
  - multiple luminance signals, 11.38–11.39, 11.40*f*, 11.70, 11.70*f*
  - and visual acuity, 4.11, 4.11*f*, 4.12
- Luminance mechanism, 11.11
- Luminance pedestals, 11.72, 11.74
- Luminance ratios, for computer work, 23.5
- Luminosity function, 10.10, 10.16
- Luminous efficiency, 11.37
  - and brightness, 11.70, 11.70*f*
  - and chromatic adaptation, 11.48*f*
  - variations in, 11.33
- Luminous efficiency functions, 10.13, 10.44–10.45, 11.33, 11.37, 11.38*f*
- Luminous flux, 23.3
- Luminous intensity:
  - defined, 23.3
  - in stimulus specification, 4.3–4.4
- Luneberg lens, 19.3
- Lux, 23.3
- MacAdam ellipses, 10.40
- Mach-Zender interferometers, 5.24
- Macula lutea (yellow spot), 7.1, 7.11
- Macular pigment (MP), 1.9, 1.11, 14.9–14.10
- Macular pigment density:
  - adjustments for individual differences, 10.17
  - and color matches, 10.9
  - and variations in color matching, 10.15
- Magnetic resonance imaging (MRI), 19.1
- Magnification, with contact lenses, 20.31–20.33
  - relative spectacle, 20.32–20.33, 20.33*f*
  - spectacle, 20.31–20.32
- Magnification and translation of images:
  - distortion by monocular magnification, 13.13–13.16
    - bifocal jump, 13.15–13.16, 13.15*f*
    - from convergence responses to prism, 13.19
    - discrepant views of objects/images, 13.16
    - motion parallax, 13.14–13.15
    - perspective distortion, 13.13, 13.14*f*
    - stereopsis, 13.13, 13.14



- Magnification and translation of images (*Cont.*):  
 distortion from interocular aniso-  
 magnification, 13.16–13.19  
 aniseikonia, 13.17–13.18  
 interocular blur suppression with  
 anisometropia, 13.18–13.19  
 lenses and prisms, 13.16–13.17, 13.17*f*  
 errors of eye alignment induced by,  
 13.25–13.27
- Magnitude (optical aberrations), 15.4
- Magnitude estimation, in psychophysical  
 measurement, 3.8
- Magnitude production, in adjustment  
 experiments, 3.5
- Magnocellular laminae, 2.12, 2.13, 2.13*f*
- Mandelbaum's phenomenon, 25.11–25.12
- Mandell-Moore bitoric lens guide, 20.17–20.19,  
 20.18*t*, 20.19*f*
- Maréchal criterion, 1.15
- Masks (in experiments), 3.2
- Matching:  
 in adjustment experiments, 3.4, 3.5  
 color (*see* Color matching)
- Matrix algebra, 10.45–10.48  
 addition and multiplication, 10.45–10.46  
 colorimetry, 10.24–10.32  
 stimulus representation, 10.24–10.27,  
 10.25*f*, 10.26*f*  
 transformations between color spaces,  
 10.29–10.32  
 vector representation of data, 10.25*f*,  
 10.27–10.29, 10.27*f*  
 glossary of notation in, 10.5*t*  
 linear models, 10.46–10.47  
 matrix transposition, 10.46  
 simultaneous linear equations,  
 10.47–10.48  
 singular value decomposition, 10.48  
 special matrices and vectors, 10.46  
 vectors and matrices, 10.45
- Maximum saturation method (color  
 matching), 10.6–10.7, 10.6*f*
- Maxwellian view (viewing), 5.4–5.8, 5.5*f*,  
 6.1–6.14  
 advantages of, 5.7  
 control of focus, 5.4–5.5, 5.5*f*  
 defined, 7.1  
 field of view, 6.6–6.7, 6.6*f*  
 focus, 6.5–6.6  
 interferometers, 5.24
- Maxwellian view (viewing) (*Cont.*):  
 partial coherence, 6.12–6.14  
 positioning of pupil in, 5.9  
 pupil size, 6.7–6.12  
   coherent illumination, 6.9–6.12, 6.11*t*  
   incoherent target, 6.7–6.9, 6.8*f*, 6.9*t*  
 retinal conjugate plane, 5.5*f*  
 retinal illuminance, 5.6–5.7, 6.3–6.5  
 size, 5.5–5.6  
 spatial frequency content of stimuli, 5.8  
 two-channel, 5.21, 5.23–5.24, 5.23*f*
- Maxwell's method, for color matching,  
 10.8, 10.8*f*
- McCullough effect, 11.76, 11.77*f*
- Mean-sphere correction, 1.6
- Measured horopter, 13.8
- Mechanism direction (color vision), 11.3, 11.33
- Mechanist approach (color vision) (*see* Color  
 vision mechanisms)
- Melanin, 14.9–14.10
- Membrane mirrors (AO), 15.1, 15.10–15.12,  
 15.10*f*
- Mesocotyl, 8.2
- Metamerism, 10.7, 10.36–10.38, 10.37*f*
- Metamers, 10.1, 10.7
- Method of constant stimuli, 3.9
- MHz (MegaHertz), 23.3
- Michaelis-Menton function, 2.14
- Michelson contrast, in vision experiments, 3.4
- Michelson formula, 4.8
- Michelson interferometer, 18.1
- Microelectromechanical systems (MEMS),  
 15.11
- Microkeratomes, 16.13, 16.14
- Micro lens Nipkow disk confocal microscope,  
 17.1–17.2
- Micromachined membrane MEMS mirrors, 15.11
- Microscacades, 1.44
- Midget ganglion cells, 2.10–2.11, 2.10*n*
- Midsagittal plane, 13.2
- Miniature eye movements, 1.44
- Minimally distinct border (MDB), 11.37
- Minimum motion (MM), 11.37
- Mode:  
 defined, 8.2  
 and energy density distribution, 8.15  
 in hair cells, 8.25–8.26  
 in human photoreceptors, 8.16*f*, 8.18*f*  
 in monkey/human retinal receptors,  
 8.19–8.24, 8.22*f*

- Model eyes, 21.13
- Model-dependent characterization methods (MDCM) (color CRTs), 22.27–22.33  
 conditions for use, 22.30  
 gun independence, 22.27–22.28  
 inverse transformations, 22.32  
 measurement of parameters, 22.31  
 normalization coefficients, 22.31–22.32  
 out-of-gamut colors, 22.32–22.33  
 partial models, 22.30  
 phosphor constancy, 22.28–22.29, 22.32  
 phosphor output models, 22.29–22.30
- Modulation (contrast) color spaces, 10.19
- Modulation thresholds, 1.22
- Modulation transfer:  
 with defocus, 1.29  
 TCA effect on, 1.20
- Modulation transfer function (MTF):  
 aberration-derived, 1.22  
 in aging eyes, 14.13  
 in diffraction-limited eye, 1.13, 1.13f  
 double-pass, 1.23  
 and observed optical performance, 1.24–1.25  
 and off-axis image quality, 1.26–1.27  
 in ophthalmoscopic methods, 1.23  
 and optical quality of IOLs, 21.14  
 and scattered light, 1.21  
 of visual instruments, 1.28  
 in young adult eyes, 1.24f
- Modulators, intensity, 5.14
- Monitors:  
 and Computer Vision Syndrome, 23.6–23.8  
 interlaced, 23.2  
 noninterlaced, 23.3  
 in optical systems, 5.16
- Monkey photoreceptors, 8.19–8.24, 8.22f
- Monochromatic ocular aberrations, 1.4, 1.14–1.19  
 age-related, 14.12–14.14, 14.13f  
 correction of, 1.25, 1.26, 1.26f  
 off-axis, 1.18–1.19, 1.18f  
 on the visual axis, 1.15–1.18
- Monochromatic vision, 10.16
- Monochrome CRTs, 22.3–22.4, 22.3f  
 controls for, 22.7, 22.8f  
 design and operation of, 22.3–22.4, 22.3f  
 standards for, 22.14
- Monochrome LCDs, operational principles of, 22.34–22.37
- Monocular cues, for perceived space, 13.3–13.7
- Monocular field:  
 horizontal angular extent of, 1.38f  
 and stereopsis, 1.38–1.42
- Monocular magnification, distortion by, 13.13–13.16  
 bifocal jump, 13.15–13.16, 13.15f  
 from convergence responses to prism, 13.19  
 discrepant views of objects/images, 13.16  
 motion parallax, 13.14–13.15  
 perspective distortion, 13.13, 13.14f  
 stereopsis, 13.13, 13.14
- Monovision:  
 with contact lenses, 14.28  
 defined, 12.13  
 with head mounted visual displays, 13.32
- Motion artifacts (OCT), 18.15
- Motion detection/discrimination, 2.36–2.40  
 optic flow fields, 2.39f  
 thresholds, 2.37f
- Motion parallax, 13.4, 13.5  
 defined, 13.2  
 with monocular magnification, 13.14–13.15
- Motion perception, 13.6–13.7
- Moving entry and exit pupils method (photo-receptor directionality), 8.6
- Mueller matrix formalism, 18.1, 18.20, 18.22
- Müller cells, 8.20–8.21, 14.11
- Müller zone model, 11.65f, 11.84, 11.85f
- Müller zone theories, 11.6, 11.7f
- Multifocal lenses:  
 contact lenses, 14.28, 20.2  
 intraocular lenses, 21.1, 21.14–21.18, 21.15f, 21.16f, 21.16t, 21.18f
- Multiplicative adaptation, 2.26
- Multizone intraocular lenses, 14.28–14.29
- Munnerlyn's equations, 16.15
- Musculoskeletal, 23.3
- Myoid, 8.2
- Myopia (nearsightedness), 1.6, 1.18, 12.4, 16.5  
 correction of, 13.17–13.18, 16.7–16.8  
 defined, 13.2, 19.1, 23.3  
 and early presbyopia, 12.8  
 empty field, 1.33  
 and focus of collimated light, 12.3f  
 instrument, 1.34–1.35  
 night (twilight), 1.34  
 refractive surgery for, 12.14
- Naperian absorption coefficient, 8.9
- Near triad, 1.30

- Nearpoint, 23.3
- Nearsightedness (*see* Myopia)
- Neuronal receptive fields, resolving capacity of the eye and, 4.6
- Newtonian viewing (*see* Optical generation of visual stimulus, free (newtonian) viewing)
- Night (twilight) myopia, 1.34
- Nipkow disk tandem-scanning confocal microscope, 17.2, 17.4f, 17.5, 17.5f
- Noise (color vision), threshold contours for, 11.20, 11.23–11.26, 11.23f, 11.25f
- Noise masking (color vision):  
defined, 11.3  
experiments in, 11.56–11.57
- Nomogram, 16.1
- Nonconcomitant movement of eyes, 13.2
- Nonexudative (dry) age-related macular degeneration, 14.1, 14.24–14.35
- Noninterlaced monitors, 23.3
- Nonlinear optics, 17.2, 17.3
- Nonlinear visual mechanisms, 11.3
- Nuclear cataract, 7.5, 12.14, 14.8
- Nulling, in adjustment experiments, 3.4
- Numerical aperture, 17.2
- Nyquist limit, 2.6, 2.7, 2.11, 2.22–2.23
- Obedience to Abney's law, 10.44, 11.37 (*See also* Additivity)
- Objective amplitude of accommodation, 1.32
- Objective tasks, 2.15n
- Oblique effect (pattern discrimination), 2.35
- Octopus lenses, 19.7, 19.7f
- Ocular motility, 23.3
- Ocular parameters, 1.4–1.6
- Ocular radiation hazards, 7.1–7.15  
examples of, 7.8–7.9  
exposure limits, 7.9–7.11  
exceeding, 7.11  
guidelines for visible light, 7.10  
IR, 7.10–7.11  
UV, 7.9–7.10  
injury mechanisms, 7.2–7.3  
action spectra, 7.2, 7.4f  
exposure duration and reciprocity, 7.2–7.3  
lamp safety standards, 7.14–7.15  
laser hazards, 7.11–7.14, 7.12f, 7.13f  
accidents, 7.12–7.14  
and eye protectors, 7.14  
safety standards, 7.12  
retinal irradiance calculations, 7.7–7.8, 7.8f
- Ocular radiation hazards (*Cont.*):  
types of injury, 7.3–7.7, 7.5f  
cataract, 7.5–7.6, 7.6f  
droplet keratopathies, 7.6, 7.7  
infrared cataract, 7.7  
photokeratitis, 7.4  
photoretinitis, 7.7  
pterygium, 7.6, 7.7
- Ocular radiometry, 1.11–1.12
- Ocular wavefronts, 16.6–16.7, 16.7t
- Off-axis image quality, 1.18–1.19, 1.18f, 1.26–1.27, 1.27f
- Olympus water immersion microscope, 17.10
- One mode components analysis, 10.34
- Ophthalmoheliosis, 7.1, 7.5
- Ophthalmoscopes:  
defined, 15.1  
flood-illuminated AO, 15.3, 15.16–15.17  
scanning laser, 15.2, 15.3, 15.17–15.19, 15.18f, 15.19f
- Ophthalmoscopic (double-pass) methods (retinal image quality), 1.22–1.23
- Opponent color spaces, 10.18–10.19
- Opponent-colors theory, 11.3, 11.5, 11.6f, 11.62–11.63
- Optic disc, 1.3f
- Optic fiber, defined, 19.1
- Optic flow, 13.4  
and bifocal jump, 13.15, 13.16  
defined, 13.2
- Optic flow fields, 2.38–2.39, 2.39f
- Optic nerve head, 18.2
- Optical aberrations:  
categories of, 17.2  
in human eyes (*see* Aberrations [human eye])
- Optical axis, 1.3f
- Optical circulator, 18.2
- Optical coherence tomography (OCT), 15.3, 15.19, 18.1–18.30  
at 1050 nm, 18.15–18.17, 18.15f–18.17f  
autocorrelation noise, 18.11–18.12  
combined with adaptive optics, 15.19–15.21, 15.20f, 15.21f  
defined, 15.2, 18.2  
depth dependent sensitivity, 18.13–18.14, 18.13f, 18.14f  
Doppler OCT, 18.18–18.19, 18.18f–18.20f  
fringe washout, 18.15  
motion artifacts, 18.15

- Optical coherence tomography (OCT) (*Cont.*):  
 optical frequency domain imaging, 18.7–18.9, 18.7f, 18.8f  
 polarization sensitive OCT, 18.18, 18.20–18.27, 18.21f, 18.23f, 18.25f–18.27f  
 shot-noise-limited detection, 18.12–18.13  
 signal to noise ratio, 18.11  
 spectral domain OCT, 18.5–18.7, 18.5f, 18.6f, 18.9  
   noise analysis of, 18.9–18.10, 18.12f  
   retinal imaging with, 18.27–18.29, 18.28f, 18.29f  
   sensitivity advantage of, 18.9  
 Stratus OCT 3, 15.20–15.21, 15.21f  
 time domain OCT, 18.3–18.5, 18.4f
- Optical design:  
 binocular vision factors in, 13.1–13.35  
 and refractive errors, 12.1–12.17  
   assessment of, 12.5–12.8  
   binocular factors, 12.15–12.17  
   correction of, 12.8–12.15  
   types of, 12.4–12.5
- Optical errors, in head mounted display systems, 13.34–13.35
- Optical fiber perform, 19.1
- Optical fibers, 5.10, 5.11, 19.4
- Optical frequency domain imaging (OFDI), 18.3, 18.7–18.9, 18.7f, 18.8f  
 at 1050 nm, 18.15–18.17  
 defined, 18.2  
 SD-OCT *vs.*, 18.9
- Optical generation of visual stimulus, 5.1–5.25  
 building an optical system, 5.8–5.18  
   alternating of source and retinal planes, 5.8–5.9  
   calibration, 5.17–5.19, 5.17f  
   combining lights, 5.9–5.11, 5.9f, 5.10f  
   controlling intensity, 5.13–5.15  
   controlling wavelength, 5.11–5.13, 5.12t  
   field quality, 5.11  
   generating complex patterns, 5.16, 5.16f  
   lenses, 5.11  
   turning field on/off, 5.13, 5.13t  
 coherent radiation, 5.19, 5.21  
 detectors, 5.21, 5.22t  
 free (newtonian) viewing, 5.2–5.4, 5.3f  
   limitations of, 5.4  
   retinal illuminance, 5.2–5.3, 5.3f  
   the troland, 5.3–5.4
- Optical generation of visual stimulus (*Cont.*):  
 light exposure and optical safety, 5.18–5.19, 5.18f  
 light sources, 5.19, 5.19t–5.20t  
 Maxwellian viewing, 5.4–5.8, 5.5f  
   advantages of, 5.7  
   control of focus, 5.4–5.5, 5.5f  
   interferometers, 5.24  
   positioning of pupil in, 5.9  
   retinal conjugate plane, 5.5f  
   retinal illuminance, 5.6–5.7  
   size, 5.5–5.6  
   spatial frequency content of stimuli, 5.8  
   two-channel, 5.21, 5.23–5.24, 5.23f  
   size of stimulus, 5.2
- Optical power (of cornea), 16.2
- Optical safety, 5.18–5.19, 5.18f
- Optical Stiles-Crawford effect, 9.2
- Optical transfer function (OTF), 2.21–2.22, 2.22f  
 defined, 4.1  
 for human eye, 1.21–1.22  
 in ophthalmoscopic methods, 1.23  
 and optical quality of IOLs, 21.14  
 with visual instruments, 1.27
- Optical zone diameter (OZD):  
 for contact lenses, 20.5  
 defined, 20.2
- Opticaldiagnostics.com, 12.17
- Optics:  
 of aging eyes, 14.4–14.14  
   accommodation and presbyopia, 14.8, 14.9f  
   anterior and posterior chambers, 14.5, 14.6  
   cornea, 14.5, 14.6f  
   eye size, 14.11  
   lens, 14.7–14.8  
   pupil, 14.6–14.7, 14.7f  
   retina, 14.9–14.11  
   tears, 14.4–14.5  
   transparency/cataract, 14.8  
 of photoreceptors, 8.3, 8.10f  
 (*See also* Adaptive optics)
- Optokinetic nystagmus, 1.44
- Optokinetic reflex, 13.20
- Organ of Corti, 8.2, 8.24–8.26
- Orthokeratology, 12.12
- Overall diameter (OAD):  
 for contact lenses, 20.5  
 defined, 20.2

- Overscan, in color CRTs, 22.12  
 Oxygen permeability (Dk), 20.2
- Panum's fusional area (PFA), 12.15, 13.12, 25.5  
 Parabolic louver, 23.3  
 Parasol ganglion cells, 2.10*n*  
 Paraxial models (of human eye), 1.36–1.37, 1.37*f*  
 Partial coherence, in Maxwellian viewing, 6.12–6.14  
 Partial coherence interferometry, 21.7  
 Parvocellular laminae, 2.12, 2.13, 2.13*f*  
 Pattern discrimination, 2.35–2.36  
 Pedestal contrast, 2.31  
 Pedestal control, in CRTs, 22.8  
 Pedestal effects (color vision), 11.60*f*  
   and chromatic discrimination, 11.69  
   in crossed conditions, 11.59  
   defined, 11.3  
   and gap effect, 11.72, 11.74  
   in uncrossed conditions, 11.59, 11.61  
 Pedestal experiments, 3.2, 11.59–11.62, 11.60*f*  
 Pedicles (cones), 8.2, 8.20  
 Percept rivalry suppression, 13.14  
 Perception:  
   of color, 14.17  
   defined, 13.2, 23.3  
   of direction, 13.7–13.10  
     corresponding retinal points, 13.8  
     horopter, 13.8–13.9, 13.9*f*, 13.10*f*  
     vertical horopter, 13.9  
   of distance, 13.24  
   Gestalt principles in, 24.8  
   of imaging artifacts in vision, 24.2–24.6  
     early color vision, 24.5  
     embedding digital watermarks, 24.5  
     image quality and compression, 24.3–24.4  
     limitations of early vision models, 24.5–24.6  
     rendering and halftoning, 24.4  
     target detection in medical images, 24.5  
   measuring (*see* Psychophysical measurement)  
   of size, 13.19  
   of space, 13.3–13.7  
     binocular cues, 13.7  
     distortion of, 13.16–13.19  
     extraretinal information for eye movements, 13.7  
   of space (*Cont.*):  
     kinetic cues, 13.4–13.7, 13.5*f*, 13.6*f*  
     monocular cues, 13.3–13.7  
     and visual effects created by artists, 24.10  
 Percepts:  
   3-D, 13.3  
   defined, 13.2  
 Perceptual Subband Image Coder (PIC), 24.4  
 Perceptually based image compression, 24.4  
 Performance, measuring (*see* Psychophysical measurement)  
 Perimetry, 14.21  
 Peripheral field, 1.3  
   and retinal illuminance, 1.12  
   TCA in, 1.20  
 Peripheral retina, 13.3, 14.10  
 Perspective:  
   defined, 13.2  
   distorted, with monocular magnification, 13.13, 13.14*f*  
 Phacoemulsification, 12.14, 21.2, 21.4, 21.5  
 Phakia, 14.2  
 Phakic lenses:  
   defined, 21.2  
   intraocular, 16.9, 21.19  
 Phase discrimination, 2.36  
 Phase retardation, 18.2  
 Phase stability, Doppler OCT and, 18.18–18.19, 18.18*f*–18.20*f*  
 Phase transfer function (PTF):  
   in diffraction-limited eye, 1.13  
   and off-axis image quality, 1.26, 1.27  
   in ophthalmoscopic methods, 1.23  
 Phoria, 13.21–13.22  
   defined, 13.2  
   effect of lenses and prisms on, 13.25–13.27  
 Phoropter, 12.6  
 Phosphor, 23.3  
 Photocoagulation, 7.4  
 Photographic recording, in ophthalmoscopic methods, 1.23  
 Photokeratitis, 1.9, 7.4*f*  
   defined, 7.1  
   from radiation, 7.4  
   UV, 7.3  
 Photometric efficiency (PE) factor, 9.3  
 Photometry, 9.1–9.15, 10.10, 10.43–10.45, 11.37 (*See also* Chapter 37 in Volume II)  
 Photon noise, 2.4  
 Photon-flux irradiance (retina), 2.4, 2.7

- Photophobia, 23.3
- Photopic luminosity function, 10.1, 10.4t
- Photopic luminous efficiency function, 10.44
- Photopic retinal illuminance, 2.4
- Photopic troland, 9.2
- Photopic vision:
  - accommodation response, 1.32–1.34, 1.33f, 1.34f
  - age-related changes in, 14.15, 14.16f
  - color, 10.3
  - Stiles-Crawford effect, 1.10, 1.10f
- Photopigment opsin genes, 10.14
  - in color-deficient observers, 10.16
  - and variations in color matching, 10.15
- Photopigment optical density:
  - and adjustment of cone spectral sensitivities, 10.17
  - adjustments for individual differences, 10.18
  - and color matches, 10.9
- Photopigments:
  - absorption spectra of, 2.7
  - adjustments for individual differences, 10.18
  - and color matches, 10.9
  - in color-deficient observers, 10.16
  - time constant of regeneration, 2.7
  - variability in, 10.15
- Photoreceptors, 10.3, 10.4
  - biological waveguide models of, 8.8–8.9, 8.9f, 8.10f, 8.12–8.15
  - and color appearance, 11.62
  - and contrast detection, 2.22–2.23
  - defined, 15.2
  - directional sensitivity of, 8.5
  - dynamic range of, 2.9
  - image sampling by, 2.4–2.9
  - inhomogeneous, 10.16
  - length of outer segments, 2.7–2.8
  - modal patterns in, 8.16f, 8.18f, 8.19–8.24, 8.22f
    - for neighboring waveguides, 8.22
    - radial distribution of transmitted energy, 8.22–8.23, 8.23f
    - Snyder and Pask cone model, 8.21f
    - transfer function, 8.24f
  - optical waveguide properties of, 14.11
  - optics of, 8.3, 8.10f
  - orientation and alignment of, 8.5–8.8, 8.6f, 8.7f
  - quantitative observations of, 8.15, 8.16f, 8.17, 8.18f, 8.19f
- Photoreceptors (*Cont.*):
  - and resolving capacity of the eye, 4.5, 4.6f
  - schematic diagram of, 8.9f
  - temporal response properties of, 2.8–2.9
  - types and functions of, 2.4, 2.6
  - waveguiding in, 8.3–8.5
  - (*See also* Cones; Rods)
- Photorefractive keratectomy (PRK), 12.14, 16.11–16.12, 16.11f, 16.12f
- Photoretinitis, 7.1, 7.7
- Phototropism, 8.2
- Physiological optics, 2.2
- Pig lenses, 19.11, 19.12f
- Pigments:
  - macular, 1.9, 1.11, 14.9–14.10
  - of the retina, 14.9–14.11
- Pixel, 23.3
- Plant tissues, fiber-optic, 8.26–8.28
- Plastic lenses, for spectacles, 12.9
- Poincaré sphere, 18.2
- Point-spread function (PSF), in human eye, 1.21
  - defined, 4.1, 15.2
  - in diffraction-limited eye, 1.12–1.14
  - direct measurements of, 2.3–2.4
  - and image formation, 2.3
  - and resolving capacity of the eye, 4.4, 4.5
- Polarity, 23.3
- Polarization, of light entering eye, 1.10
- Polarization sensitive OCT (PS-OCT), 18.18, 18.20–18.27, 18.21f, 18.23f, 18.25f–18.27f
- Polycarbonate lenses, for spectacles, 12.9
- Polymethylmethacrylate (PMMA) contact lenses, 12.11–12.12, 20.3
- Porcine lenses, 19.11, 19.12f
- Posterior capsule opacification (PCO):
  - defined, 21.2
  - with intraocular lenses, 21.21–21.22
- Posterior chamber, 14.5, 14.6, 16.3
- Posterior limiting lamina (*see* Descemet's membrane)
- Posterior peripheral curve (contact lenses), 20.5, 20.6, 20.23–20.24, 20.24t
- Posterior subcapsular cataract, 14.8
- Prentice's rule, 12.16, 13.15, 13.18, 13.28, 20.31
- Presbyopia (old eye/old sight), 1.7, 12.3, 16.5
  - accommodation restoration for, 14.29–14.30
  - as age-related, 14.8, 14.9f
  - assessment of, 12.7–12.8
  - clinical onset of, 12.3, 12.7–12.8

- Presbyopia (old eye/old sight) (*Cont.*):  
 correction of, 14.27–14.30  
   accommodation restoration, 14.29–14.30  
   bifocals for, 12.8  
   contact lenses for, 12.13–12.14,  
     14.27–14.28  
   hyperchromatic lens design for, 14.14  
   intraocular lenses for, 14.28–14.29  
   noncataract related refractive surgeries for,  
     14.29  
   spectacles for, 12.10–12.11, 14.27  
 defined, 12.1, 14.2, 21.2, 23.3  
 as problem with computer work, 23.11–23.12  
 and UV radiation exposure, 14.23
- Primary lights (primaries), 10.4  
 additive mixing of, 10.4, 10.5*f*  
 CMFs for, 10.10, 10.11  
 in color matching, 10.6, 10.7  
 defined, 10.1, 10.4*t*  
 for DKL color space, 10.19  
 imaginary, 10.10  
 in Maxwell's matching method, 10.8, 10.8*f*  
 perceptual vs. physical matches of, 10.7  
 for stimulus spaces, 10.11  
 in trichromatic color matching, 10.7, 10.8  
 and vector representations, 10.27–10.29
- Primary position, 1.42, 13.2
- Primary visual cortex, 2.12, 2.14
- Primate eye lens, 19.13*f*, 19.14, 19.14*f*
- Prism(s):  
 in contact lenses, 20.2  
 distortion from interocular aniso-magnifica-  
   tion, 13.16–13.17, 13.17*f*  
 effects on vergence and phoria, 13.25–13.27  
 errors of alignment with, 13.27–13.29  
 nonuniform magnification of, 13.10*f*  
 stimulus value of, 13.21
- Prism diopter, 12.2, 25.1
- Prismatic effects, 12.8  
 and bifocal jump, 13.15  
 with contact lenses, 20.30–20.31  
   prism-ballasted contacted lenses, 20.30  
   unintentional (induced) prism, 20.31  
 with corrected ametropia, 12.16  
 defined, 12.2
- Prism-ballasted contacted lenses, 20.30
- Progressive addition lenses (PALs), 12.10–  
 12.11, 12.11*f*
- Proportionality (color matching), 10.8
- Protanopes, 10.16
- Proximal (psychic) convergence, 1.35
- Proximal stimuli (human vision), 4.2
- Pseudoaccommodation, 14.27, 14.29, 14.30
- Pseudophakia, 12.15, 14.28  
 correction of, 12.14–12.15  
 defined, 14.2, 21.2
- Psychic convergence, 1.35
- Psychophysical comparison method (retinal  
 image quality), 1.22
- Psychophysical measurement, 3.1–3.10  
 of adjustment tasks, 3.4–3.5  
   magnitude production, 3.5  
   matching, 3.4–3.5  
   nulling, 3.4  
   threshold, 3.4, 3.5*f*  
 definitions related to, 3.2–3.3  
 of judgment tasks, 3.6–3.8  
   ideal observer, 3.6  
   rating scale, 3.6–3.8, 3.7*f*  
   response time for, 3.8  
   two-alternative forced choice (2afc), 3.8  
   yes-no, 3.6  
 magnitude estimation in, 3.8  
 professional tips for, 3.10  
 stimulus sequencing in, 3.9  
   method of constant stimuli, 3.9  
   sequential estimation methods, 3.9  
 of visual acuity, 4.6–4.7  
 visual stimuli, 3.3–3.4
- Psychophysical test method (color vision), 11.9,  
 11.11, 11.12
- Psychophysics, 4.1, 15.2
- Pterygium, 7.1, 7.6, 7.7
- Punctum remotum*, 12.8
- Pupil:  
 center of, 1.20  
 defined, 6.1  
 diameter of (*see* Pupil diameter/size)  
 interpupillary distance, 1.39–1.41, 1.40*f*  
 in Maxwellian viewing, 5.9  
 in ophthalmoscopic methods, 1.23  
 and retinal illuminance, 1.12  
 true area of, 1.10  
 and visual acuity, 4.9  
 and visual instruments, 1.27–1.28  
 (*See also* Entrance pupil; Exit pupil)
- Pupil conjugate plane:  
 in Maxwellian viewing, 5.8–5.9  
 rotating a mirror in, 5.16, 5.16*f*  
 shutters at, 5.13, 5.13*t*

- Pupil diameter/size, **1.8–1.9, 1.8f**  
 age-related changes in, **1.18, 14.6–14.7, 14.7f, 14.14**  
 change in, **16.3**  
 and correction for refractive error, **1.23, 1.24**  
 and correction for SCE-1, **9.4, 9.13**  
 and depth-of-focus, **1.29, 1.30f**  
 in Maxwellian viewing, **6.3–6.5, 6.7–6.12**  
 coherent illumination, **6.9–6.12, 6.11t**  
 incoherent target, **6.7–6.9, 6.8f, 6.9t**  
 optimal, **15.22**  
 and RMS wavefront error, **1.15–1.16**  
 and Stiles-Crawford effect, **1.10**  
 variation in parameters as function of, **1.13**
- Pupil function, OTF and, **1.21**
- Purity, in color CRTs, **22.9**
- Pursuit movements, **1.43**
- QUEST, **3.9**
- Rabbit lenses, **19.8**
- Radial edge lift, **20.2**
- Radial gradients, **19.3–19.5, 19.4f, 19.5f**
- Radial keratotomy, **12.14, 16.9–16.10, 16.10f**
- Radiation damage, life-span, **14.22–14.23**
- Radiation hazards (*see* Ocular radiation hazards)
- Radiometry (*See also* Chapter 37 in Volume II):  
 for color matching, **10.12**  
 correction for SCE-1 in, **9.1–9.15**  
 ocular, **1.11–1.12**
- Raman microspectroscopy, **19.1**
- Random-dot stereograms, **2.40**
- Raster CRTs, **22.9–22.13, 22.13f**
- Raster scanning, to despeckle light sources, **5.21**
- Rat lenses, **19.7–19.8**
- Rating scale judgments, **3.6–3.8, 3.7f**
- Ray optics, **8.15**
- Ray tracing, **19.1**
- Rayleigh units, **1.13, 1.29**
- Receiver operating characteristic (ROC) curves, **3.7–3.8, 3.7f**
- Reciprocity theorem of optics (Helmholtz), **8.25, 8.27**
- Red to far-red (R/FR) ratio, **8.28**
- Reduced eye model, **1.36, 1.37, 2.3, 2.4**
- Reflectance, retinal, **1.11**
- Reflection:  
 from computer screens, **23.5–23.6**  
 in fiber-optic plant tissues, **8.26, 8.28**  
 upon interfaces separating different media, **8.14–8.15, 8.14f**
- Reflection model, **10.32**
- Reflective apertures, **5.10**
- Reflectometric sensors, **19.1**
- Refraction:  
 in human eye, **12.3, 16.1**  
 and retinal image quality, **1.21**  
 subjective, **12.6–12.7**  
 upon interfaces separating different media, **8.14–8.15, 8.14f**
- Refractive, defined, **23.3**
- Refractive ametropia, **20.2**
- Refractive error correction, **12.8–12.15, 16.7–16.19**  
 ametropias, **16.8f**  
 aphakia and pseudophakia, **12.14–12.15**  
 with contact lenses, **12.11–12.14**  
 hydrogel, **12.12–12.13**  
 for presbyopia, **12.13–12.14**  
 rigid, **12.11–12.12**  
 with laser ablation, **16.11–16.19**  
 ablation rate, **16.16–16.18, 16.18f**  
 corneal photoablation, **16.16, 16.17f**  
 refractive surgery modalities, **16.11–16.15, 16.11f–16.14f**  
 thermal, photochemical, and photoacoustic effects, **16.18–16.19**
- prescribing, **12.8**  
 with refractive surgery, **12.14, 16.9–16.15**  
 corneal incisions/implants, **16.9–16.11, 16.10f**  
 intraocular lenses, **16.9**  
 laser ablation modalities, **16.11f–16.14f**  
 modalities for, **16.11–16.15**  
 with spectacles, **12.9–12.11, 12.10f, 12.11f**
- Refractive errors, **12.1–12.17, 12.3f, 12.4f, 16.4–16.19, 16.5f**  
 assessment of, **12.5–12.8**  
 objective tests, **12.5–12.6**  
 presbyopia, **12.7–12.8**  
 subjective techniques, **12.6–12.7**
- astigmatism, **16.5–16.6**
- binocular factors, **12.15–12.17**  
 aniseikonia, **12.16–12.17**  
 anisometropia, **12.16–12.17**  
 convergence and accommodation, **12.15–12.16**



- Refractive errors (*Cont.*):  
 consequences for optical design, 12.18  
 correction of (*see* Refractive error correction)  
 ocular wavefronts, 16.6–16.7, 16.7*t*  
 as problem with computer work, 23.10  
 spherical ametropias, 16.5  
 types of, 12.4–12.5
- Refractive index, 8.12  
 of cornea, 14.5  
 defined, 19.1  
 in lens of human eye, 1.5–1.6, 1.5*f*  
 (*See also* Gradient index optics)
- Refractive lens exchange (RLE), 21.18
- Refractive power, defined, 19.1
- Refractive surgery, 1.15, 1.25, 12.14, 16.9–16.15  
 change in cornea with, 16.3  
 corneal incisions/implants, 16.9–16.11,  
 16.10*f*  
 intraocular lenses, 16.9  
 laser corneal procedures, 16.11–16.15  
 ablation profiles, 16.14–16.15  
 Epi-LASIK, 16.12, 16.13, 16.13*f*  
 LASEK, 16.12, 16.13  
 LASIK, 16.13–16.14, 16.14*f*  
 photorefractive keratectomy (PRK),  
 16.11–16.12, 16.11*f*, 16.12*f*  
 for presbyopic correction, 14.29
- Refresh rate, 23.3
- Region of interest, in human vision, 24.7
- Regular astigmatism, 1.6
- Relative directional sensitivity, 9.6
- Relative intensity noise (RIN):  
 defined, 18.2  
 with SD-OCT, 18.10
- Relative spectacle magnification (RSM), 20.2,  
 20.32–20.33
- Rendering, low-level vision models in, 24.4
- Residual astigmatism:  
 and contact lens power, 20.15  
 defined, 20.2
- Resolution (visual acuity):  
 defined, 23.3  
 and information theory, 4.15–4.16  
 optics of resolving capacity, 4.4–4.5, 4.4*f*  
 contrast-transfer function, 4.5  
 point-spread function, 4.4, 4.5  
 spatial-frequency coordinates, 4.5  
 in retinal imaging, 15.21  
 and superresolution, 4.15  
 thresholds of, 4.6–4.7, 4.15
- Response time, as measure of performance, 3.8
- Response to direction (gaze control), 13.29–13.30
- Resting point of accommodation (RPA), 23.3
- Resting positions (eyes), 13.21
- Resting state (accommodation), 1.33
- Retina, 1.3*f*, 21.2  
 aging-related changes in, 14.9–14.11  
 anatomy of, 2.9–2.11  
 AO-controlled light delivery to, 15.22–15.24  
 alignment, 15.23  
 in an AO SLO, 15.23  
 conventional AO vision systems, 15.23  
 to generate aberrations, 15.24  
 longitudinal chromatic aberration, 15.22  
 measuring activity of individual cones,  
 15.24  
 transverse chromatic aberration, 15.22–15.23  
 uses of, 15.23–15.24  
 control of alignment in, 8.7  
 direction-corresponding points, 13.8  
 fovea, 2.24  
 injury to, 7.4  
 neural pathways in, 2.9, 2.10*f*  
 nonfoveal areas of, 2.24  
 OCT image of, 18.4*f*  
 optic flow fields, 2.38–2.39, 2.39*f*  
 physiology of, 2.11–2.12  
 pigments of, 14.9–14.11  
 and refraction in the eye, 12.3  
 and resolving capacity of the eye, 4.5–4.6,  
 4.6*f*  
 schematic diagram of, 8.9*f*, 8.10*f*  
 UV light damage to, 14.23
- Retina cameras, AO, 15.3, 15.12 (*See also*  
 Ophthalmoscopes)
- Retina pigment epithelium (RPE), 8.12
- Retinal burn, 7.7
- Retinal conjugate plane, in Maxwellian  
 viewing, 5.5*f*, 5.8–5.9
- Retinal disorders, photoreceptor orientation/  
 realignment after, 8.7
- Retinal disparity, 13.2
- Retinal eccentricity, visual acuity and,  
 4.10–4.11, 4.11*f*
- Retinal illuminance, 1.11–1.12  
 in free (newtonian) viewing, 5.2–5.4, 5.3*f*  
 maximum permissible, 5.18*f*  
 in Maxwellian viewing, 5.6–5.7, 6.3–6.5  
 in normal viewing, 6.3  
 and pupil diameter, 1.8, 1.9

- Retinal image:
- AO ophthalmic applications, **15.16–15.22**
  - contrast and resolution, **15.21–15.22**
  - flood-illuminated AO ophthalmoscope, **15.16–15.17, 15.16f, 15.17f**
  - optical coherence tomography, **15.19–15.21, 15.20f, 15.21f**
  - scanning laser ophthalmoscope, **15.17–15.19, 15.18f, 15.19f**
  - light spread in, **4.4–4.5**
  - relating actual object and, **4.2**
- Retinal image disparity, **13.22**
- Retinal image quality, **1.12–1.28**
- in aberration-free eye, **1.12–1.14, 1.13f, 1.14f**
  - aging-related changes in, **14.11–14.14**
  - chromatic aberration, **14.14**
  - intraocular scatter, **14.12**
  - monochromatic aberrations, **14.12–14.14, 14.13f**
  - chromatic aberration, **1.19–1.20**
  - computing, **2.3**
  - intraocular scattered light, **1.20–1.21**
  - lenticular fluorescence, **1.21**
  - monochromatic ocular aberrations, **1.14–1.19**
  - off-axis, **1.18–1.19, 1.18f, 1.26–1.27, 1.27f**
  - on the visual axis, **1.15–1.18**
  - in peripheral field, **1.3**
  - and pupil diameter, **1.8**
  - variation with field location, **15.5**
  - on the visual axis, **1.21–1.28**
  - calculation from aberration data, **1.21–1.22**
  - comparison between methods, **1.23**
  - effects of aberration correction, **1.25–1.26**
  - observed optical performance, **1.23–1.25**
  - ophthalmoscopic (double-pass) methods, **1.22–1.23**
  - psychophysical comparison method, **1.22**
- Retinal imaging, with spectral domain OCT, **18.27–18.29, 18.28f, 18.29f**
- Retinal irradiance, **2.4, 7.7–7.8, 7.8f**
- Retinal layer of rods and cones model of
- biological waveguides, **8.8–8.9, 8.9f, 8.10f, 8.12–8.15**
  - assumptions and approximations for, **8.9, 8.12–8.13**
  - electromagnetic validity of, **8.13**
- Retinal microscopy:
- adaptive optics in, **15.1–15.24**
  - AO-controlled light delivery to the retina, **15.22–15.24**
  - control system, **15.12–15.15**
  - history of, **15.2–15.3**
  - imaging of the retina, **15.16–15.22**
  - implementation of, **15.7–15.15**
  - in ophthalmic applications, **15.15–15.24**
  - properties of ocular aberrations, **15.4–15.7, 15.5f, 15.6f**
  - wavefront corrector, **15.9–15.12, 15.10f, 15.11f**
  - wavefront sensor, **15.8–15.9, 15.8f**
  - defined, **15.2**
- Retinal nerve fiber layer (RNFL), **18.25–18.27**
- Retinal neurons, **2.5f**
- ganglion cells, **2.10–2.11**
  - information transmission by, **2.11**
- Retinal pigmented epithelium (RPE), **14.25, 18.2**
- Retinal processing, **2.9–2.12**
- Retinal reflectance, **1.11**
- layers of occurrence for, **1.23**
  - and ophthalmoscopic methods, **1.23**
  - and stray light, **1.20**
- Retinex theory (Land), **11.71, 11.72**
- Retinopathy:
- diabetic, **14.1, 14.25–14.26**
  - solar, **7.1, 7.3**
- Retinoscopy, **12.5–12.6**
- Rhabdomeric photoreceptors, **8.3**
- Rigid contact lenses:
- correction of refractive errors with, **12.11–12.12**
  - early types of, **20.2, 20.3**
  - (*See also* Gas permeable contact lenses)
- Rimless mounting, **12.2, 12.9**
- RMS (root mean square) contrast, **3.4**
- Rod amacrine cells, **2.5f**
- Rod bipolar cells, **2.5f**
- Rod pathway, **2.9, 2.10, 2.10f**
- Rods, **2.5f, 10.3**
- and age-related scotopic vision changes, **14.15**
  - alignment of, **8.4**
  - amacrine cells, **2.10**
  - and color matching, **10.17**
  - in color-deficient observers, **10.16**
  - diameter of, **2.6f**

- Rods (*Cont.*):  
 directional sensitivity of, 8.5  
 function of, 2.4  
 linear density of, 2.6f  
 and maximum saturation color matching, 10.7  
 optical standing waves in, 8.17, 8.19f  
 optical waveguide properties of, 14.11  
 photocurrent responses of, 2.8f  
 in retinal layer of rods and cones model of biological waveguides, 8.8–8.9, 8.9f, 8.10f, 8.12–8.15  
 spatial distribution of, 2.6  
 spectral sensitivities of, 10.18  
 time constant of photopigment regeneration, 2.7
- Root-mean-square (RMS) wavefront error, 1.15–1.17, 1.16f
- Rostock Cornea Module (RCM) microscope, 17.7, 17.9
- Runs (trial sequences), 3.9
- Saccades, 1.42, 1.43, 1.43f, 13.20
- Saccadic suppression, 1.43
- Safety, of head-mounted displays, 25.2
- Safety standards, for laser hazards, 7.12
- Sagittal plane, 19.1
- Scanning laser ophthalmoscope (SLO), 15.3, 15.17, 17.9  
 applying adaptive optics to, 15.17–15.19, 15.18f, 15.19f, 15.23  
 AO-controlled stimulus delivery, 15.17–15.19  
 visual acuity with, 15.23–15.24  
 defined, 15.2
- Scanning slit confocal microscopes, 17.2, 17.3, 17.6–17.9, 17.6f, 17.7f
- Scattered light:  
 with cataracts, 14.24  
 intraocular, 1.20–1.21  
 and OTF calculation, 1.21  
 and psychophysical comparison method, 1.22
- Schematic eye model, 1.36, 1.37
- Schlemm's canal, 14.5
- Sclera, 16.3, 21.2
- Scotoma, 7.7
- Scotopic troland, 9.2
- Scotopic vision, 10.3  
 age-related changes in, 14.15, 14.16f  
 in color-deficient observers, 10.16  
 Stiles-Crawford effect, 1.10, 1.10f
- Screen reflections, with computer work, 23.5–23.6
- Second-harmonic generation (SHG), 17.2, 17.9–17.10
- Second-site adaptation (color vision), 11.17–11.22, 11.18f, 11.19f  
 desensitization by steady fields, 11.17, 11.18, 11.18f  
 field measurements and, 11.29, 11.31  
 and field sensitivities, 11.51, 11.52, 11.53f  
 habituation, 11.19–11.20, 11.19f, 11.21f, 11.22f
- Second-site desensitization (color vision):  
 defined, 11.3  
 by steady fields, 11.17, 11.18, 11.18f, 11.54f
- Segmented corrector, 15.2
- Segmented piston and piston/tip/tilt wavefront correctors, 15.10, 15.10f
- Seidel aberrations, 12.9
- Self-motion, 13.8, 13.9
- Semiconductor optical amplifiers, 18.2
- Senescent changes in vision (*see* Age-related changes in vision)
- Sequential stimulus sequencing, 3.9
- Shack-Hartmann wavefront sensing (SHWS), 15.3, 15.8–15.9, 15.8f
- Shack-Hartmann wavefront sensors, 15.2, 16.7
- Shadowmask color CRTs, 22.4–22.6  
 common problems in, 22.5, 22.6  
 geometry of, 22.4–22.5, 22.5f, 22.6f
- Shear, 13.2
- Short-range motion discrimination  
 mechanism, 2.38
- Shot noise, 18.9  
 defined, 18.2  
 with SD-OCT, 18.10
- Shot-noise-limited detection, 18.12–18.13
- Signal detection theory, 2.17, 11.20, 11.23f, 11.24
- Signal to noise ratio (SNR), 18.11  
 defined, 18.2  
 for in-vivo high speed human retinal imaging, 18.15  
 and preferred frequency, 2.24  
 with SD-OCT, 18.7, 18.10

- Silicone hydrogel contact lenses, 20.2, 20.3
- Simple cells (cortical neurons), 2.14
- Simplified schematic eye model, 1.36, 1.37
- Simultaneous vision contact lenses,  
12.13–12.14, 14.28, 20.22
- Site of limiting noise, 11.24
- Size, perceived, 13.19
- Skew movement, 13.2
- Slit-lamp microscopes, 17.2
- Slit-scanning arrangements, in  
ophthalmoscopic methods, 1.23
- Sloan notch, 11.34, 11.35f, 11.37  
and chromatic adaptation, 11.49, 11.51  
and chromatic discrimination, 11.57, 11.58
- Small field (color matching):  
defined, 10.9  
standards for, 10.11, 10.12
- Snakes, effect of laser exposures in, 9.13
- Snellen letters, 4.1, 4.8
- Snell's law, biological waveguides and,  
8.14, 8.15
- Snow blindness, 1.9, 7.3, 7.4
- Soft (hydrogel) contact lenses:  
aberrations and, 20.24  
base curve radius for, 20.3, 20.5  
center thickness of, 20.6  
correction of refractive errors with,  
12.12–12.13  
edge thickness of, 20.6  
OAD/OZD of, 20.5  
posterior peripheral curve systems of, 20.6  
power of, 20.15–20.16  
toric lenses  
defined, 20.2  
power of, 20.16–20.17
- Solar keratitis, 14.23
- Solar retinitis, 7.7
- Solar retinopathy, 7.1, 7.3
- Space, perception of, 13.3–13.7  
binocular cues, 13.7  
distortion of, 13.16–13.19  
extraretinal information for eye movements,  
13.7  
kinetic cues, 13.4–13.7, 13.5f, 13.6f  
monocular cues, 13.3–13.7
- Spatial channels, in contrast detection,  
2.23–2.24
- Spatial equichromatic CSFs, 11.45–11.46,  
11.46f
- Spatial equiluminant CSFs, 11.45–11.46, 11.46f
- Spatial fidelity (optical aberrations), 15.4
- Spatial filtering:  
with confocal microscopes, 17.3  
to despeckle light sources, 5.21
- Spatial frequency, 4.1
- Spatial frequency channels (in vision), 3.2
- Spatial impulse response, of photoreceptors, 8.23
- Spatial location conflicts, in head mounted  
display systems, 13.33, 13.34f
- Spatial sinusoids, 2.20
- Spatial vision, age-related changes in, 14.7,  
14.17–14.19, 14.18f
- Spatial-frequency coordinates, resolving  
capacity of the eye and, 4.5
- Spatially-modulated excimer laser ablation of  
cornea, 1.25
- Spatio-temporal CSFs, 2.29
- Speckle fields, 5.19, 5.21
- Spectacle blur, 12.12
- Spectacle lenses:  
anisophoria and, 13.26  
base curve of, 12.1, 12.9  
effective power of, 20.8, 20.9t  
and gaze control, 13.29–13.30  
materials for, 12.9–12.10  
optical power of, 12.4  
power of, 20.6
- Spectacle magnification (SM), 20.2, 20.31–20.32
- Spectacles, 15.2  
correction by (*See also* Spectacle lenses)  
for anisometropia, 12.16, 13.17–13.18  
aphakia, 12.15  
for presbyopia, 14.27  
for refractive errors, 12.9–12.11, 12.10f, 12.11f  
defined, 15.2  
and response to distance, 13.30
- Spectral color sensitivities:  
test method for, 11.34, 11.35f, 11.36f, 11.37  
in three-stage Müller zone model, 11.7f  
two-color measurements, 11.34
- Spectral (Fourier) domain OCT, 18.5–18.7,  
18.5f, 18.6f, 18.9  
defined, 18.2  
noise analysis of, 18.9–18.10, 18.12f  
OFDI vs., 18.9  
retinal imaging with, 18.27–18.29, 18.28f,  
18.29f  
sensitivity advantage of, 18.9
- Spectral power density, vector representation of,  
10.24, 10.25f

- Spectral power distribution, **10.3**  
 in colorimetric measurements, **10.23**  
 and tristimulus values, **10.11, 10.36–10.37**
- Specular microscopes, **17.2**
- Specular reflection, **23.3**
- Speed discrimination, **2.37**
- Spherical ametropias, **1.6–1.7, 16.5**  
 hypermetropia, **16.5**  
 myopia, **16.5**
- Spherical gradients, **19.2–19.3**
- Spherocylindrical lenses, **12.4, 12.5**
- Spherule, **8.2**
- Sponges, **8.28–8.29**
- Square root law, **2.26**
- Stability of fixation, **1.44, 1.45f**
- Standard Notation (lenses), **12.5**
- Standard observer:  
 CIE 1931, **10.12**  
 CIE 1964, **10.13**  
 defined, **10.1, 10.4t**  
 limitations of, **10.15**  
 standardized CMFs as, **10.9**
- “Star” patterns (human vision), **1.21**
- Steradian, **6.3**
- Stereoacuity, **2.41**
- Stereopsis, **1.38–1.42, 2.40**  
 acuity, **1.39–1.40**  
 age-related changes in, **14.22**  
 aniseikonia, **1.41–1.42**  
 Da Vinci, **13.4**  
 defined, **13.2**  
 factors affecting, **2.40, 2.41**  
 with head mounted visual displays, **13.31, 13.32**  
 with monocular magnification, **13.13, 13.14**  
 with monovision, **14.28**  
 in perception of depth, **13.11–13.12, 13.11f**  
 stereoscopic and related instruments, **1.41**  
 tolerances in binocular instrumentation, **1.41–1.42**
- Stereoscopic instruments, **1.41**
- Stiles and Burch (1955) 2° color-matching functions, **10.12**
- Stiles and Burch (1959) 10° color-matching functions, **10.12, 10.13**
- Stiles’  $\pi$ -mechanisms, **11.16, 11.16f, 11.46, 11.47, 11.47f**  
 and contrast coding, **11.16, 11.16f**  
 and field additivity, **11.51**  
 limitation of, **11.34**
- Stiles-Crawford effect, **1.29, 2.8**  
 of the first kind, **1.21, 8.5** (*See also* Stiles-Crawford effect of the first kind)  
 correction for, **9.1–9.15**  
 just prior to phototransduction, **6.14**  
 and Maxwellian viewing, **6.2**  
 optical, **9.2**  
 and pupil size, **6.7**  
 of the second kind, **8.3, 8.5, 9.2**  
 and human eye, **1.11**  
 and waveguide models, **8.21**
- Stiles-Crawford Effect of the first kind (SCE I/SCE-1), **1.21, 8.5**  
 and biological waveguides, **8.3, 8.6f**  
 correction for, **9.1–9.15**  
 adaptive optics techniques for, **9.5**  
 application of approach, **9.13–9.14**  
 confounds in, **9.4–9.5**  
 history of, **9.5**  
 with nonmonochromatic stimulus to vision, **9.5**  
 Photometric efficiency factor, **9.3**  
 sample point-by-point SCE-1 estimates, **9.6–9.13, 9.7f, 9.9f–9.13f**  
 teleological and developmental factors in, **9.14**  
 and trolands, **9.2, 9.3**  
 without adaptive optics, **9.5–9.6**  
 defined, **9.2**  
 and human eye, **1.10, 1.10f, 1.11**  
 and photoreceptor directionality, **8.6–8.8**  
 and waveguide analysis, **8.21**
- Stimulus color spaces (colorimetry), **10.11**
- Stimulus sequencing, in psychophysical measurement, **3.9**
- Stimulus specification (visual acuity), **4.2–4.4, 4.3f**
- Stokes parameters, **18.20**
- Stokes vector, **18.2**
- Strabismus, **13.2**
- Strehl ratio, **1.23–1.24, 1.24f, 1.28, 15.2**
- Stroma (cornea), **14.5**
- Subadditivity, **11.51, 11.52f**
- Subjective amplitude of accommodation, **1.32**
- Subjective refraction, **12.6–12.7**
- Subjective tasks, **2.15n**
- Subtractive adaptation, **2.26**
- Superadditivity, **11.51, 11.52f**
- Superposition (colorimetry), **10.25, 10.26f**
- Superresolution (visual acuity), **4.1, 4.15**

- Suppression (binocular vision), 13.12–13.13  
of blur, 13.18–13.19  
defined, 23.3  
with head mounted visual displays, 13.33
- Suprathreshold models (color vision), 11.41
- Surface micromachined MEMS devices, 15.11
- Surfaces:  
illuminants reflected from, 10.32–10.36  
metamerism for, 10.38
- Surround, visual acuity and, 4.12, 4.13f
- SVGA (super video graphics array), 23.3
- Svishchev confocal microscope, 17.6, 17.6f
- Swept Source OCT (SS-OCT), 18.2 (*See also*  
Optical frequency domain imaging)
- Symmetry (color matching), 10.8
- Synchronization standards, for color CRTs,  
22.13–22.14
- TABO, 12.5, 12.7
- Taflove, A., 8.17
- “Target” sensitivity method (color vision)  
(*see* Test sensitivity method (color vision))
- Tear lens, 12.12
- Tears, aging-related changes in, 14.4–14.5
- Telecentric principle, 6.6–6.7
- Temporal coherence, 18.2
- Temporal contrast detection, 2.27–2.29, 2.27f,  
2.29f
- Temporal despeckling, 5.21
- Temporal vision, age-related changes in,  
14.19–14.21, 14.20f
- Test (“target”) sensitivity method (color  
vision), 11.11, 11.12, 11.34–11.46  
defined, 11.3  
and different directions of color space, 11.39,  
11.41–11.43, 11.44f, 11.45–11.46, 11.46f  
luminance, 11.37–11.39, 11.38f, 11.40f  
and spectral lights, 11.34, 11.35f, 11.36f, 11.37
- Tests of visual acuity, 4.7–4.8, 4.7f  
for infants, 4.8  
traditional visual acuity chart, 4.6
- Theoretical horopter, 13.8, 13.9f
- Third-harmonic generation (THG), 17.2,  
17.9–17.10
- Three-dimensional color data, 10.19–10.20, 10.20f
- Three-segment model of biological waveguides,  
8.9, 8.11–8.15, 8.11f  
assumptions and approximations for, 8.9,  
8.12–8.13  
electromagnetic validity of, 8.13
- Three-stage zone models (color vision), 11.6,  
11.82–11.85, 11.85f  
De Valois and De Valois model, 11.82  
Müller model, 11.6, 11.7f, 11.65f, 11.84,  
11.85f
- Threshold experiments, 11.17
- Threshold surface or contour (color vision),  
11.12–11.15, 11.13f  
defined, 11.3  
and loss of information, 11.20  
and noise, 11.20, 11.23–11.26, 11.23f, 11.25f  
and second-site adaptation to steady fields,  
11.18f
- Threshold-limit values (TLVs), 7.9
- Thresholds, 3.3f  
in adjustment experiments, 3.4  
defined, 3, 3.2  
detection (color vision)  
chromatic discrimination near,  
11.58–11.59  
two-stage model of, 11.23f  
discrimination, 4.6–4.7  
estimation of, 3.9  
increment (color vision), 11.16f  
modulation, 1.22  
for motion detection/discrimination, 2.37f  
in psychophysics, 3.2–3.3  
suprathreshold models (color vision), 11.41  
of visual resolution, 4.6–4.7, 4.15
- Tilt, 13.2
- Time, visual acuity and, 4.12, 4.13f
- Time domain OCT, 18.2–18.5, 18.4f
- Timing standards, for color CRTs, 22.13–22.14
- Tonic level of accommodation, 1.33–1.34
- Toric lenses:  
for astigmatism, 12.13  
bitoric, 20.17–20.20, 20.18f–20.20f, 20.18t  
intraocular, 21.13, 21.12f  
power of, 20.16–20.20, 20.16f, 20.17f  
prism with, 20.30
- Torsion, 13.2, 13.8, 13.22
- Torsional vergence, 13.22, 13.27
- Transfer function (human vision), 2.3  
defined, 6.1  
of photoreceptors, 8.23–8.24, 8.24f  
temporal and spatial components of, 2.11,  
2.12f
- Transitivity (color matching), 10.8
- Transmittance, in the human eye, 1.9–1.11, 1.9f

- Transparency, aging-related changes in, **14.8**  
(*See also* Cataract)
- Transparent HMDs, **25.4–25.5**
- Transverse (lateral) chromatic aberration  
(TCA), **1.19, 1.20, 15.22–15.23**  
measurement of, **8.8**  
with visual instruments, **1.28**
- Tremor, **1.44**
- Trials, experimental, **3.2**
- Trichromacy (trichromatic vision), **10.4–10.6**  
color appearance vs., **11.5**  
in color matching, **10.7–10.9**  
in color-deficient observers, **10.16**
- Trifocal lenses, **12.10, 14.27**
- Trinitron CRTs, **22.5**
- Tristimulus values (cone coordinates), **10.4**  
for arbitrary lights, **10.9**  
in colorimetric measurements, **10.23**  
and consistency across observers, **10.9**  
defined, **10.1, 10.4f**  
in maximum stimulus method, **10.7**  
negative, **10.7**  
and spectral power distribution, **10.11**  
spectral power distributions from,  
**10.36–10.37**  
tailored to individuals, **10.15**  
uniqueness of, **10.9**  
in vector representations, **10.27–10.29**
- Tritanopes, **10.16**
- Trivex lenses, **12.9**
- The troland, **5.3–5.4, 6.3**  
defined, **2.7n**  
“effective,” **8.7**  
“equivalent,” **9.2**  
limitations of, **9.3**  
as unit of retinal illuminance, **9.2**
- Twilight myopia, **1.34**
- Two-alternative forced choice (2afc), **3.8**
- Two-channel Maxwellian viewing, **5.21,**  
**5.23–5.24, 5.23f**
- Two-color threshold experiments, **11.34, 11.37**
- Ultraviolet (UV) radiation, **14.23**  
cataract from, **7.5–7.6**  
damage from, **7.2**  
exposure limits for, **7.9–7.10**  
pterygium and droplet keratopathies from,  
**7.6, 7.7**
- Unaccommodated eye, **16.5**
- Underscan, in color CRTs, **22.12, 22.13f**
- Uniform color spaces, **10.40, 10.42**
- Unintentional (induced) prism, with contact  
lenses, **20.31**
- Unipolar chromatic mechanisms,  
**11.80–11.81**
- Unipolar mechanism (color vision), **11.3**
- Unique hues, **11.27**  
in color discrimination vs. color appearance  
tests, **11.81–11.82**  
defined, **11.3**  
and equilibrium colors, **11.63–11.66**  
zero crossings, **11.62–11.63**
- Unitary matrix, **18.2**
- Univariance, **10.4**
- Univariant mechanism (color vision):  
defined, **11.3**  
flicker, **11.45**  
suprathreshold, **11.58**
- User interfaces, human image features and,  
**24.8–24.9**
- Van Cittert-Zernike theorem, **6.12**
- Vanishing point, **13.4, 13.4f**
- Vantage point, **13.2**
- Vergence (disjunctive) eye movements, **1.42,**  
**1.44, 13.20, 13.21**
- Vergence system, **1.29, 1.30, 1.32, 1.34,**  
**1.43–1.44**  
binocular parallax, **13.22**  
defocus and effort to clear vision,  
**13.22–13.23**  
effect of lenses and prisms on, **13.25–13.27**  
with head mounted visual displays,  
**13.31–13.32**  
intrinsic stimuli to, **13.21–13.22**  
and variations of cross-coupling,  
**13.23–13.24**  
zone of clear and single binocular vision,  
**13.24–13.25, 13.25f**
- Vernier acuity, **4.1**
- Vernier acuity task, **2.34, 2.35f**
- Version eye movements (*see* Conjugate eye  
movements/position)
- Vertex distance, **12.9, 16.8**
- Vertical horopter, **13.9**
- Vertical scanning, in color CRTs, **22.11–22.12,**  
**22.12f**
- Vestibulo-ocular reflex (VOR), **13.20**
- Vestibulo-ocular responses, **1.42**
- Video display terminal (VDT), **23.3**

- Video head sets (*see* Head-mounted displays)
- Video monitors, in optical systems, 5.16
- Video Quality Experts Group (VQEG), 24.4
- Vieth-Müller circle, 13.8, 13.9, 13.9f
- Vieth-Müller horopter, 2.40
- Viewing environments:  
   for color CRTs, 22.19–22.20  
   for computer work, 23.4–23.9  
     lighting, 23.4–23.5, 23.5t  
     monitor characteristics, 23.6–23.8  
     screen reflections, 23.5–23.6  
     work habits, 23.8–23.9  
     workstation arrangement, 23.8
- Virtual reality:  
   head-mounted displays for, 25.1–25.12  
   characterizing, 25.7–25.10, 25.8f, 25.9t–25.10t  
   design considerations, 25.2–25.7  
   research in development of, 24.9
- Visible light, hazards from, 7.10
- Vision, defined, 23.4
- Vision experiments/research:  
   displays for, 22.1–22.40  
     color cathode ray tubes (color CRTs), 22.1–22.34  
     liquid crystal displays (LCDs), 22.34–22.40  
   using Maxwellian view in (*see* Maxwellian view)  
   (*See also* Psychophysical measurement)
- Vision laboratories, control of light sources in (*see* Optical generation of visual stimulus)
- Vision therapy, 23.4
- Visual acuity, 2.33–2.35, 2.34f, 2.35f, 4.1–4.13, 12.6, 12.7  
   age-related changes in, 14.17, 14.19  
   and color, 4.10  
   and contrast, 4.12, 4.12f  
   defined, 4.1, 12.2, 15.2, 23.1  
   and defocus, 4.9, 4.9f, 4.10f  
   factors affecting, 4.9–4.13  
   hyperacuity, 4.13–4.16, 4.14f  
     defined, 4.1  
     and superresolution, 4.15  
   and luminance, 4.11, 4.11f, 4.12  
   practice effect on, 4.12  
   and pupil, 4.9  
   resolution  
     and information, 4.15–4.16  
     and superresolution, 4.15
- Visual acuity (*Cont.*):  
   resolving capacity of the eye, 4.4–4.5, 4.4f  
     contrast-transfer function, 4.5  
     point-spread function, 4.4, 4.5  
     spatial-frequency coordinates, 4.5  
   and retinal eccentricity, 4.10–4.11, 4.11f  
   retinal limitations, 4.5–4.6, 4.6f  
   and stage of development/aging, 4.13  
   stimulus specification, 4.2–4.4, 4.3f  
   and surround, 4.12, 4.13f  
   tests of visual acuity, 4.7–4.8, 4.7f  
   and time, 4.12, 4.13f  
   visual resolution threshold determination, 4.6–4.7
- Visual angles:  
   angular extent, 5.2  
   defined, 2.3, 4.1, 10.1, 10.4t  
   degrees of, 10.9  
   and image formation, 2.3  
   steradian of, 6.3
- Visual axis, 2.3
- Visual differences predictor, 24.3
- Visual displays, measuring performance of (*see* Psychophysical measurement)
- Visual fields:  
   age-related changes in, 14.21, 14.21f  
   binocular overlap of, 13.7  
   defined, 13.2  
   perception of, 13.3
- Visual instruments:  
   and accommodation response in eye, 1.34–1.35  
   and chromostereopsis, 1.20  
   exit pupil of, 1.27–1.28  
   retinal image quality with, 1.27–1.28  
   stereoscopic, 1.41
- Visual performance, 2.1–2.41  
   aging-related changes in, 14.14–14.22  
   color vision, 14.15, 14.17  
   depth and stereovision, 14.22  
   minimal, 14.22  
   sensitivity under scotopic and photopic conditions, 14.15, 14.16f  
   spatial vision, 14.17–14.19, 14.18f  
   temporal vision, 14.19–14.21, 14.20f  
   visual field, 14.21, 14.21f  
   binocular stereoscopic discrimination, 2.40–2.41, 2.41f



- Visual performance (*Cont.*):
- central visual processing, 2.12–2.14
    - anatomy of LGN, 2.12–2.13, 2.13f
    - physiology of LGN, 2.13–2.14
  - contrast detection, 2.19–2.31
    - adaptation and inhibition, 2.25–2.27
    - chromatic, 2.29–2.31, 2.30f
    - eye movements, 2.21
    - as function of spatial frequency, 2.20f
    - optical transfer function, 2.21–2.22, 2.22f
    - optical/retinal inhomogeneity, 2.24, 2.25f
    - receptors, 2.22–2.23
    - spatial channels, 2.23–2.24
    - temporal, 2.27–2.29, 2.27f, 2.29f
  - contrast discrimination, 2.31–2.32, 2.31f
  - contrast estimation, 2.33
  - contrast masking, 2.32–2.33, 2.32f
  - ideal-observer theory, 2.16–2.19, 2.19f
  - image formation, 2.2–2.4
  - image sampling by photoreceptors, 2.4–2.9
  - information-processing model for, 2.15–2.16, 2.15f
  - motion detection/discrimination, 2.36–2.40
    - optic flow fields, 2.39f
    - thresholds, 2.37f
  - pattern discrimination, 2.35–2.36
  - retinal processing, 2.9–2.12
    - anatomy of retina, 2.9–2.11
    - physiology of retina, 2.11–2.12
  - visual acuity, 2.33–2.35, 2.34f, 2.35f
- Visual plane, 13.2
- Visual resolution threshold, determination of, 4.6–4.7
- Visual stimuli:
- in decision tasks, 3.2
  - optical generation of (*see* Optical generation of visual stimulus)
  - in psychophysical experiments, 3.3–3.4
  - spatial frequency content of, 5.8
- Visual stress, 23.4
- Visual systems (colorimetry), 10.38–10.40
- Visual tasks, 2.15
- Visualization, in electronic imaging, 24.8
- Visual-vestibular conflicts, in head mounted display systems, 13.32–13.33
- Vitreous humor/gel, 1.3f, 14.9, 16.3, 18.2
- Von Bezold spreading, 11.3, 11.4f
- Von Kries adaptation, 11.3, 11.49, 11.51, 11.68
- Watercolor illusion/effect, 11.72, 11.73f
- Wave aberration, 15.2
- Wavefront aberration, 1.15–1.19, 1.16f, 1.16t, 1.17f
  - and defocus of visual instruments, 1.28
  - and final retinal image quality, 1.21–1.22
- Wavefront correctors (AO systems for the eye), 15.3, 15.7f, 15.9–15.12, 15.10f, 15.11f
- Wavefront sensor (AO systems for the eye), 15.3, 15.7f, 15.8–15.9, 15.8f
- Waveguide parameter (*V*-parameter), 8.2, 8.17, 8.20–8.21
- Waveguides, biological (*see* Biological waveguides)
- Wavelength:
- color data as functions of, 10.21, 10.23, 10.23f
  - in optical systems, 5.11–5.13, 5.12t
- Weber contrast, in vision experiments, 3.4
- Weber's law, 11.16f
  - and cone contrast spaces, 11.32
  - and contrast coding, 11.15–11.16, 11.16f
  - defined, 11.3
  - first-site adaptation, 11.15, 11.16f
  - second-site adaptation, 11.18
- Webvision, 11.5
- Welder's flash, 1.9, 7.4
- Wet age-related macular degeneration, 14.1, 14.24
- White balance, in color CRTs, 22.9
- White noise, 18.2
- Wide-angle models (of human eye), 1.38
- With-the-rule astigmatism, 1.6
- Wood lens, 19.3–19.4, 19.4f
- Working distance (retinoscopy), 12.2, 12.6
- Workstation arrangement, for computer work, 23.8
- World Health Organization (WHO), 7.9
- Wyszecki lens, 9.5
- Yes-no judgments, 3.6
- Yoked eye movements, 13.2, 13.20
- Young-Helmholtz trichromatic theory, 11.5
- Young's fringes, 1.22

- Z-dependent oscillary term, **8.22, 8.23**
- Zernike coefficient, for RMS wavefront error,  
    **1.15–1.19, 1.16t, 1.17f**
- Zernike polynomials:
  - defined, **15.2**
  - for representing ocular aberrations, **15.4,**  
    **15.5f, 16.6–16.7, 16.7t**
- ZEST, **3.9**
- Zonal refractive lenses (IOLs), **21.15**
- Zone models (color vision), **11.5**
  - three-stage, **11.6, 11.7f, 11.8, 11.82–11.85,**  
    **11.85f**
  - two-stage, **11.5, 11.6f**
- Zonules, **1.3f, 21.2**