21 Emotionally resonant media

Advances in sensing, understanding and influencing human emotion through interactive media[1]

Jonathan Gratch

Introduction

In *the Diamond Age*, novelist Neal Stephenson envisions the early twenty-first century as a time when mass media is radically transformed (Stephenson 1995). Rather than watching *passives* – traditional linear media such as TV and movies – the public engages in *ractives*, a form of interactive entertainment that effortless senses and stimulates human emotion. This creates something akin to a virtual theater-in-the-round, where everyone can be the star of their own personal drama.

Traditional mass media can certainly evoke emotion, but it is ultimately limited by the medium's static nature. In contrast, human emotion is inherently relational and dynamic (Parkinson 2001). As social animals, we evolved to survive through close relationships and emotions are a fundamental building block of effective social interchange (Keltner and Haidt 1999). Emotions arise from our moment-to-moment understanding of the social environment but, further, are broadcast to other social actors through words and expressions, changing the behavior of other social actors. This creates cycles of emotional resonance such as rapport (Tickle-Degnen and Rosenthal 1990) or emotional contagion (Hatfield et al. 1994). Although performers or advertisers sometimes anticipate and simulate such feedback loops, mass media is inherently impoverished, 'one-sided, non-dialectical, controlled by the performer, and not susceptible of mutual development' (Horton and Wohl 1956).

Technological advances are slowly breaking down the rigidity of traditional mass media but the work is proceeding in fits and starts. Virtual worlds such as *Second Life*™ seek to create a portal whereby people can establish real emotional relationships through media, although these technological systems strip out much of the subtly of human interpersonal communication. Computer games allow rich opportunities for interactivity but typically reduce social intercourse to stylized archetypes such as war, as in *Call of Duty*; sex, as in *The Sims* or *Grand Theft Auto*; or parenting, as in *Petz*. More recently, intelligent entities called 'virtual humans' are beginning to simulate some of the basic aspects of resonant emotional behavior. For example, Figure 21.1 illustrates (clockwise from the upper left) Mr. Bubb, an emotionally intelligent agent that simulates the emotions a small child might express when playing a game of catch (Loyall et al. 2004); Façade, an interactive game that puts the player in the middle of a failing marriage (Mateas and Stern 2003); Justina, an educational tool that allows psychiatry students to practice their skills at interviewing a synthetic rape victim (Kenny et al. 2008) and SASO, a system that allows people to practice negotiation and conflict resolution in a virtual war-torn Iraq (Swartout et al. 2006).

Figure 21.1 Some examples of emotionally resonant media. Clockwise from the upper left: Mr. Bubb, Façade, Justina, and SASO.

In this chapter, I review recent technological advances that are creating the building blocks for a revolution in mass communication. Already, computational techniques are beginning to recognize, understand, synthesize, and respond to human emotions. Ultimately, these tools will enable the creation of personalized and emotionally resonant experiences that will allow society to re-envision how we teach, entertain, and communicate in the twenty-first century.

An architecture for emotionally resonant media

The plot of *The Diamond Age* revolves around a girl's relationship with new media. A street urchin, Nell, stumbles across an experimental device designed to guide a child's cognitive and emotional development. Cast in the form of an incredibly sophisticated interactive book, this device acts both as an expert tutor and an improvisational artist, stimulating learning through a series of interactive games and stories. In an abstract sense, this tool can be seen as an 'intelligent agent.' This is a term used in artificial intelligence to refer to a software artifact that observes and acts upon an environment and directs its activity toward achieving goals (Russell and Norvig 2002). The technology to construct such a device exists today, albeit in a more limited form. Imagining how such

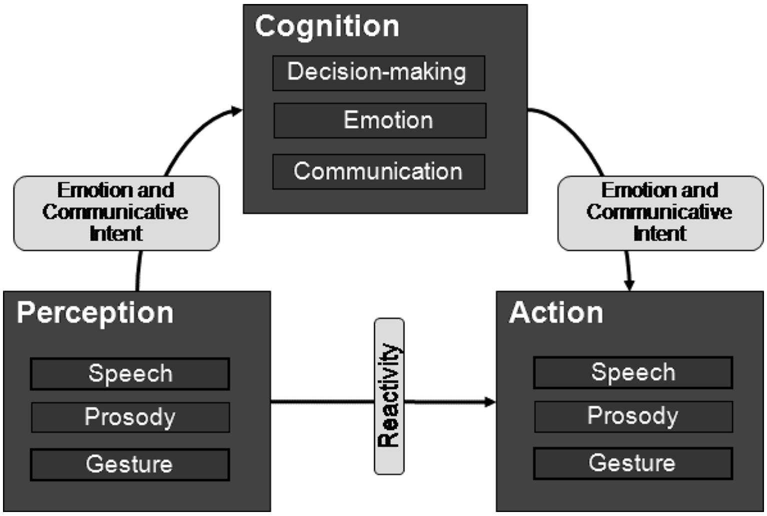

Figure 21.2 A typical software architecture for an emotionally resonant system. Such techniques must recognize and interpret human emotional signals, reason about the social context from which they arise, and generate appropriate responses.

a tool could be constructed will serve to illustrate the state-of-the-art in computational methods to sense and respond to human emotion and suggest one possible approach to building emotional mass media.

Figure 21.2 illustrates a likely architectural diagram of Nell's ractive tutor. In fact, this basic design is shared by several existing applications that will be reviewed below. To create a personalized media experience, such a system must be able to sense and respond to human intentions and also emotions. This starts with *perception*: when Nell speaks, she broadcasts a myriad of signals that can be diagnostic, not only of explicit communicative acts (e.g., 'I command my dinosaur to enter the castle'), but also underlying emotional undertones (e.g., Nell is frustrated and may give up on today's lesson). These signals include not only words but also the rhythm, stress, and intonation of speech and the accompanying nonverbal cues including facial expressions, gestures, and body posture. Then it moves to *cognition*: the device must relate these percepts to its goals. This allows it to craft an engaging story that teaches Nell to cope with bullies, select an appropriate high-level response to Nell's preceding behavior to advance the story, conveys characters' emotions and mental state, as well as stimulate social effects, such as inducing empathy and shifting Nell's focus from her own frustration to the emotions of other social actors. Finally the device must *act*, translating this high-level decision into rich and emotional behaviors of story characters.

Each step in this cycle – perception, cognition, action – involves significant challenges, in terms of both our limited scientific understanding of human emotion and the practical difficulties of translating existing knowledge into a working media system. In the following sections, I will discuss the obstacles that arise at each step and illustrate how computational methods are beginning to overcome them. I will then illustrate some working media systems that exploit these techniques in creating a personalized and emotional media experience.

Progress in perceiving human emotional cues

Computer scientists and engineers have made significant strides in perceiving human emotional cues. Computers can now reliably recognize facial movements and gestures associated with emotion. They can extract affective cues from word choice or from the subtle aspects of the human voice. They can even, in some circumstances, use these cues to predict and influence human behavior. Here, I give a high-level overview of the state-of-the-art in emotion perception techniques.[2]

In human-to-human interaction, participants can never truly know what is transpiring in the mind and heart of their conversation partner, and the same holds for human–computer interaction. Just as occurs between people, the computer must intuit a person's mental state from a variety of physical cues and contextual information. Giving the computer the capability to perceive surfaces cues is the first link in creating emotionally responsive media. For example, before a media system can respond appropriately when Nell is happy, it must infer she is happy from behavioral cues (e.g., she is smiling). But this in itself is a complex process. The computer must first find Nell's face in a video image, distinguishing it from distracters such as the doll she's holding. Next, it must find specific features in her face such as the eyes and the corners of her mouth. Even subtle features such as the wrinkles of the eye can be crucial as this may distinguish, for example, if Nell is genuinely happy or merely being polite (Ekman et al.1990). Further, the system must also understand how expressions shift over time. For example, the dynamics of a smile – how quickly the corners of the mouth turn up, how long the expression is held and how long it takes to fade away – can be diagnostic of true feeling (Krumhuber et al. 2007).

Computer scientists have made rapid progress in automatically identifying and tracking a wealth of low-level behavioral features (see also Ahn et al. this volume). Figure 21.3 illustrates some contemporary methods for automatically detecting gaze, facial expressions, gestures, and speech quality. To achieve this capability, researchers must hypothesize the appropriate level of abstraction (i.e., language) with which to describe emotionally diagnostic behaviors, and then develop techniques that accurately identify behaviors described at this level. For example, facial activity can be characterized at a low level by tracking the movement of individual feature points on the face (e.g., the corners of the mouth). Under this approach, a facial expression is characterized by a *pose* (i.e., the position and orientation of each of these feature points) and techniques must faithfully follow the movements of these individual features. Examples of such an approach include Ashish and Rosalind's (2002). Alternatively, some approaches try to describe facial activity in a more abstract way by identifying *facial action units* (Bartlett et al. 2006; Lucey et al. 2006), which is a common coding system for describing facial expressions. For example, CERT uses this approach to detect – many times per second – the raising and lowing of inner and outer brows, the pulling of the lip corners, the wrinkling of the nose, etc. (Bartlett et al. 2006; also Ahn et al., this volume). Even more abstractly, one could attempt to directly recognize conceptually meaningful features (e.g., a happy or sad face), with *FaceReader*™ by Noldus being an example of such an approach. Each representational choice has its own advantages and disadvantages:

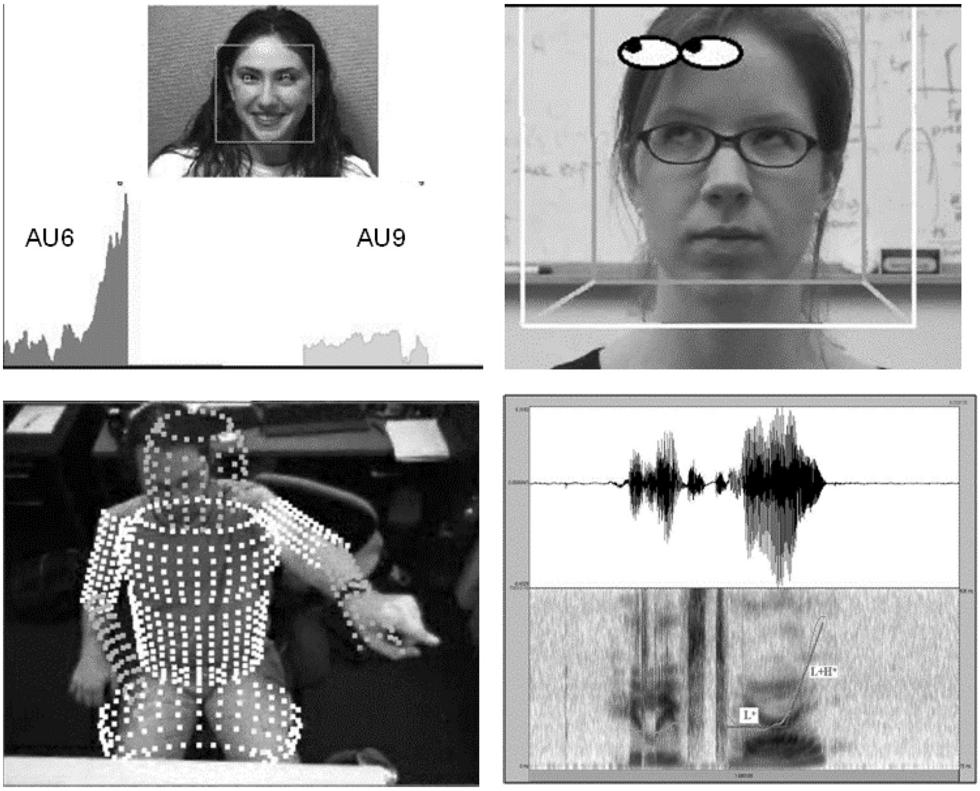

Figure 21.3 This figure illustrates some of the modalities emotionally resonant media must recognize and some of the current techniques for achieving this, including recognizing facial expressions (Bartlett et al. 2006), gaze (Morency et al. 2008), gestures (Demirdjian et al. 2005) and emotional speech (Busso et al. 2009).

low-level representations require further processing to extract semantically meaningful information – however, abstract representations can overlook important distinctions such as the fact that expressions may reflect a mixture of different emotions.

The face, of course, is merely one channel of emotional expressivity. The Watson system uses computer vision techniques to track the position and orientation of the head and eyes, from which one could recognize a bowed head or furtive gaze (Morency et al. 2008). Speech-processing techniques can extract a variety of features – such as tremors in the voice, as well as its pitch and loudness – that can be diagnostic of a variety of affective states.

Emotionally diagnostic cues are usually too ambiguous to infer a person's emotional state – people smile for many reasons including pleasure, embarrassment, and adherence to social norms. Distinguishing between these states generally involves a deeper understanding of emotional processes and the social context, as will be discussed in the next section – however, it is often possible to build meaningful media systems that respond to these purely surface behaviors. One approach is to respond to surface behaviors without any understanding of what they might mean. For example, simply mirroring emotional behaviors can produce useful social effects such as increased persuasion (Bailenson and Yee, 2005) and more fluent interaction (Gratch et al. 2007). Alternatively, some systems narrow the context of a social interaction to the extent that they can make reliable inferences about a user's mental state purely from surface behavior. For example, in the restricted domain of telephone calls to customer support centers, Narayanan et al. have shown good reliability in identifying angry customers based on features of the audio signal (Lee and Narayanan 2005); in the domain of intelligent tutoring systems, Whitehill et al. showed that recognized facial expressions predicted the difficulty a student had with understanding a lecture (Whitehill et al. 2008); and in the context of automobile driver behavior, Vural et al. demonstrated that certain facial movements could predict, with 90 percent accuracy, driver crash and sleep episodes within a driving simulator (Vural et al. 2007). In some cases, such automated techniques have even showed an ability to outperform human observers. For example, Littlewort et al. showed that computer techniques could discriminate true versus fake expressions of pain with an accuracy of 72 percent, whereas human observers were far less accurate (Littlewort et al. 2007). Together, these findings illustrate that media systems can, with increasing reliability, recognize aspects of human emotional behavior and mental life.

Progress in reasoning about human emotion

Perceiving emotional cues is only the first step in creating emotionally resonant media. Simply recognizing surface behaviors is usually insufficient to inform meaningful social exchange. A media system must understand what this cue reveals about the user's mental state and leverage this information to enhance the quality of interaction. For example, just because Nell is smiling does not mean she is happy, and if she is happy, the smile does not reveal the task or social factors that caused her happiness, nor does it suggest which system behaviors will maintain or alter this state. All of us have some common-sense notion of how people 'work' that we use when faced with similar challenges. When we see an ice cream scoop roll off a child's cone, we can easily imagine what she feels and her next likely actions. Referred to as 'theory of mind' (Whiten 1991), such a model of another's thoughts and feelings is a powerful construct that helps people understand, anticipate, and sometimes control the outcome of a social interaction. Here we discuss recent progress on developing a computational *theory of emotional mind*, with which interactive media can better recognize, understand and shape emotional experiences (Gratch et al. 2009).[3]

A theory of emotional mind must explain the *antecedents of emotion* (i.e., the task, situational and bodily factors that contribute to the genesis of an emotion) and the *consequences of emotion* (i.e., the way emotion shapes our focus of attention, our perceptions, the nature of decision processes, and ultimately our decisions and observable behavior). The computational sciences have long studied how to simulate and reason about human thought, but the central focus of this research comes from a research tradition that disparages or simply ignores the role of emotion in decision-making. Methods such as logic, decision theory, and game theory emphasize the optimal method for arriving at solutions and are often held in opposition to heuristic or emotional thought. As such, they neither represent nor consider the role of emotion and are of no immediate use in predicting the antecedents of an emotional experience. Nor do these methods capture the pervasive and systematic consequences of emotion on human perceptions and decisions. For example, angry people are more likely to blame others (Keltner et al. 1993), punish them more severely (Goldberg et al. 1999) and take on greater risks in the process (Lerner and Keltner 2000), whereas happy people tend to think more shallowly and are more influenced by their pre-conceived notions and stereotypes (Schwarz et al. 1991). Consequently, a theory of emotional mind must either augment or replace classical computational models of intelligence and, indeed, researchers draw on a variety of nonclassical theories in the social sciences in an attempt to model human emotional processes.

A diversity of computational theories of emotional mind now exist, reflecting the diversity of theoretical perspectives on emotion, though most contemporary computational models of emotion grow out of a single theoretical tradition. Appraisal theory, since its emergence in the 1960s (Arnold 1960; Frijda 1987; Lazarus 1991), has been one of the dominant theoretical perspectives on emotion and is especially congenial to a computational treatment, given its emphasis on symbolic reasoning processes that have already been extensively studied in nonemotional models of human cognitive processes. Appraisal theory argues that emotion arises from some cognitive *appraisal* of how events (external or mental) impact the individual's beliefs, desires, and intentions (cf. Barlett and Gentile this volume). These events are characterized along a number of specific dimensions referred to as appraisal dimensions. These include the desirability of the event (e.g., did the child want to eat the ice cream?), its expectedness (e.g., did the child know the ice cream would fall?), and perceptions of control (e.g., could the child have been more careful?). The specific pattern of appraisals elicited by an event will tend to determine the emotional response (e.g., an uncertain future undesirable event will provoke fear). The specific emotion elicited will then bias certain action and coping tendencies such as approach and avoidance.

Computer programs can simulate many of these reasoning processes that, according to appraisal theory, underlie human emotional actions. For example, EMA (Gratch and Marsella 2004a) is a computer program that forecasts how people may respond in emotional situations. The system must be given an initial *domain theory* that describes a person's goals (e.g., ice cream is good) and a set of actions they (or others) can use to act on the world (e.g., mommies can buy ice cream, they do this at most once a day, you can't eat ice cream that falls on the ground, etc.). With this domain model, the model can project the consequences of actions, calculate how these consequences should be appraised and suggest plausible emotional responses to specific actions within this domain. The model has even shown good results at predicting human emotional responses in a variety of emotional situations – for example, predicting both the type and the intensity of response, how these responses change as a situation evolves, and how these emotions shift an individual's beliefs and preferences (Gratch and Marsella 2004b; Mao and Gratch 2006).

Such computational models of emotional mind can be used in a variety of ways to improve human–computer interaction. One possibility is to use the model to help resolve the ambiguity inherent in many expressions of emotion. The fact that Nell is smiling gives us some indication that she is happy, but the added knowledge that the system just revealed a hidden treasure that Nell has long been seeking should considerably improve the reliability of this inference. One practical application of this idea is in the area of intelligent tutoring systems. Emotion plays an important role in educational settings and expert tutors both make use of emotional cues to assess a student's motivational state and use emotional signals to influence student engagement and learning (Lepper 1988; Lester et al. 2000). Although it is quite difficult to automatically infer a student's emotional state from their surface behaviors, some researchers have had reasonable success in using a model of emotion mind to improve such inferences. For example, Conati et al. used a computational appraisal model to infer how events in a learning game might impact their emotions and thereby provide more informed tutorial feedback (Conati and MacLaren 2004).

Computational models of emotional mind can also significantly enhance the believability and engagingness of characters in interactive story worlds. Although traditional linear narratives (e.g., movies or books) allow the author to carefully script the emotional reactions of characters, authors of interactive narratives face the challenge that they cannot always anticipate the user's actions and the sequence of events that will confront a specific character. Ideally, the storytelling system would have some emotional calculus for calculating the emotional significance of user actions for individual story characters. A theory of emotional mind provides exactly this calculus and a number of research projects have successfully employed such models to create realistic expressions of emotion contingent on the user's actions (El Nasr et al. 2000; Elliott 1992; Swartout et al. 2006).

Progress in synthesizing emotional cues

A theory of emotional mind can determine how a system should emotionally respond, but this high-level guidance must be translated into a rich and emotionally evocative presentation. Much of effective media is about performance. Human actors and computer animators are extensively trained in creating carefully crafted performances designed both to convey a sense of emotional authenticity and to provoke empathy, sympathy or antipathy in the audience. How can interactive media, which must carefully craft and tailor a performance based on the user's moment-to-moment input, create this same sense of authenticity and evocativeness?

Emotion is conveyed in several ways through several modalities. Some behaviors are implicit and obtain their emotional meaning only through context. For example, if we see a person avoid an object, the emotional meaning is quite different depending on if the object is a rattlesnake or an enticing piece of chocolate cake and such interpretations can only be generated by a system that reasons about this context through some theory of emotional mind. On the other hand, some behaviors (e.g., facial expressions) seem unambiguously emotional and in this section I describe recent advances in synthetically conveying emotion through overt displays such as facial expressions, body posture, voice quality, and word choice.

Facial expressions are an obvious way to convey emotional meaning and technology has long been available to create graphical characters that change their expressions. For example, the simplest approach is to define a set of 'morph targets' (pre-defined posed facial expressions) and to interpolate between these poses based on the character's current calculated emotional state. Such simple approaches can reliably communicate the intended emotion when the character is quite simple (e.g., a smiley face) but becomes increasingly problematic as the visual fidelity of the character increases. Through a phenomenon known as the 'uncanny valley' (Mori 2005), people become much more sensitive to, and disturbed by, stylized behavioral movements when they are applied to photorealistic representations of living organisms. Thus, using linear interpolation to move a face into a smile, rather than conveying happiness, is more likely to convey insanity. More recent work in synthesizing facial expressions has delved much more deeply into the psychology and physiology of natural facial expressions.

Computer animation has explored different approaches for capturing the subtly of facial displays. *Data-driven approaches* seek to simply capture natural human facial movement and use it to control the movements of an animated face. An example of the approach is facial performance capture in which an actor is recorded, typically with the help of a large number of reference points physically affixed to the face. Work in this vein by Ma et al. is illustrated in Figure 21.4 (Ma et al. 2008). These local facial movements can then drive the motion of a computer-generated face, creating a realistic though inflexible display of emotion that faithfully replicate the nuances of the original actor. Current research attempts to expand the flexibility of this approach by splitting this natural motion into small segments that can be rearranged and reassembled in real time (Yong et al. 2005).

Theory-driven approaches build on psychological and physiological studies of how the human face moves and how these movements convey meaning. One aspect that has received a lot of attention is the fact that people rarely exhibit 'pure' expressions of emotion. Natural facial expressions are a mixture of movements and different mixtures of facial components that create quite different interpretations in the mind of the observer. For example, simply raising the corners of the mouth may be insufficient to convey true happiness. The so-called Duchenne smile also involves movement around the eyes and is more likely to be perceived as an authentic display of joy (Ekman et al. 1990). The way in which a face moves into a pose can also dramatically impact interpretation. For example, a facial expression can be dynamically described

Figure 21.4 This illustrates the technique of performance capture, using the motion of a human actor to animate computer-generated media. In Stephenson's *Diamond Age*, this was achieved by surgically embedding sensors in actors' faces but contemporary methods use computer vision.

in terms of its onset time (the time it takes to move from a resting position to the apex of the expression), its apex time (the time it is held), and its offset time (the time it takes to return to rest position). Smiles with a short onset/offset time relative to the apex time are perceived as less authentic and these interpretations can even impact people's decisions, such as their willingness to cooperate (Krumhuber et al. 2007). Theory-driven animation approaches attempt to divide facial expressions into smaller components such as the facial actions units described above. For example, the Artificial Actor project at the Filmakademie Baden-Wuerttemberg provides a complete FACS-based animation system that allows rich and naturalistic emotional displays (see Figure 21.5). A difficulty faced by theory-driven methods is how to control a large set of individual facial components in realistic ways and researchers have begun to explore methods for composing these individual elements into naturalistic facial expressions that can convey a mixture of emotions (Martin et al. 2006).

Of course, emotion can be conveyed through other modalities as well. How one turns one's gaze toward a stimulus can convey one's feeling about the object – someone looking at you from the corner of their eyes might seem shy or flirtatious whereas a face-on stare suggests anger or aggression. To simulate such behaviors, Lance developed a data-driven approach that controls relative velocities of a character's eye, neck, and spine to reliably convey different emotional impressions (Lance and Marsella 2007). Other researchers are exploring a variety of other modalities including how to convey emotion by tuning the intensity and prosody of a speech signal (Bulut et al. 2008), selecting specific words to convey emotional nuances – for example, 'my opponent' vs. 'that one' (Fleischman and Hovy 2002), and even physiological displays such as flushing, perspiration and tears (de Melo and Gratch 2009).

The behaviors we have described up to this point are individualistic and de-contextualized (they obtain their meaning solely through the movements of the individual); however, many displays obtain their meaning in relation to another individual. The example of emotional gaze hints at this social nature of some emotional displays and others more fundamentally depend on the coupling of behaviors between individuals to extract their meaning. For example, synchrony and coupling of movements often conveys a sense of liking and rapport (Tickle-Degnen

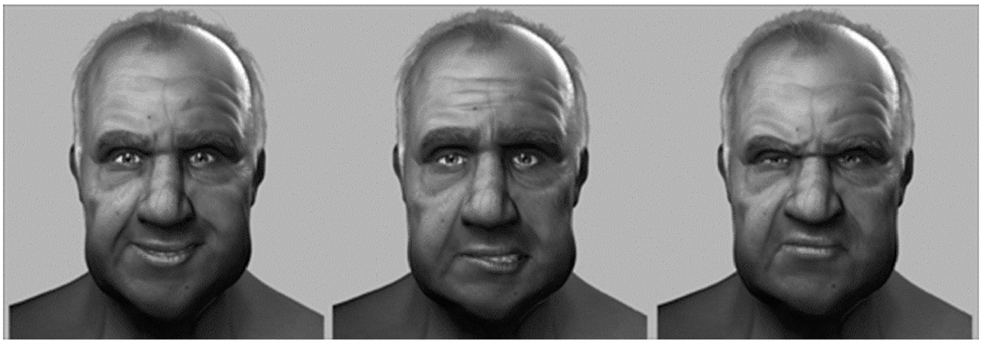

Figure 21.5 The 'Artificial Actors' project (Helzle 2003–6) has created an animation system that allows dynamic facial expressions to be created using the facial action coding system (Ekman and Friesen 1978).

and Rosenthal 1990). Complementary movements can convey discord or power asymmetries between individuals (Tiedens and Fragale 2003). Increasingly, computational researchers are exploring ways to create the social sense of emotion that emerges through the coordination of movements between virtual characters and, more ambitiously, attempts to evoke them in interactions between the machine and the human.

The Rapport Agent (see Figure 21.6) is an example of one such system that tries to evoke a sense of emergent social emotion through synchronizing agent behaviors to the emotional movement and expressions of the human user (Gratch et al. 2007). The rapport agent uses many of the behavior recognition techniques described above to perceive characteristics of the person's verbal and nonverbal behavior. For example, it uses a stereo camera system to recognize the position and orientation of a person's gaze and basic gestures such as head nods and shakes, as well as posture shifts. It also uses audio processing techniques to extract a wide variety of acoustic features, such as pitch and energy shifts. These features then automatically trigger reciprocal body movements in a virtual character the user is speaking with. These subtle movements induce feelings of rapport between the person and the character, but also have behavioral effects such as enhancing engagement (people speak longer when these contingent behaviors are present) and enhancing speech fluency (people stutter less and use fewer 'filler words' such as 'um').

Whether individualistic or relational, synthetic emotional behaviors are becoming increasingly compelling and a growing body of research has documented the social impact they can have on people. We next turn to a variety of systems that attempt to translate these potential effects into practical working applications.

Applications

Techniques for sensing, reasoning about, and responding to human emotion have been used in a number of applications. Of course, emotionally resonant media have received a great deal of interest from the entertainment industry, but the techniques are also beginning to shape how we communicate, how we care for our bodies and

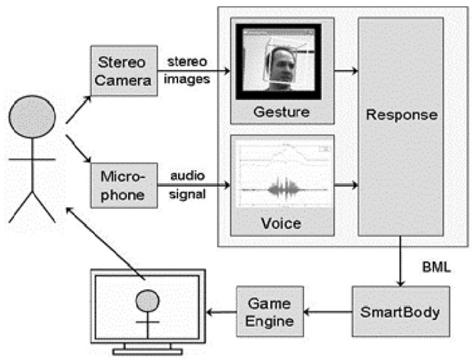

Figure 21.6 The Rapport Agent is an emotionally resonant system that detects subtle face and voice signals and reflects them back on a user in order to promote a sense of rapport (Gratch et al. 2007).

minds, and, more fundamentally, how we learn about ourselves. Educational systems incorporate emotional techniques to teach social competencies. Other systems are used for medical assessment and treatment and can infer certain socio-emotional disorders by examining how an emotionally resonant interaction unfolds. Some systems apply emotion methods to persuade, for instance, using emotional communication in health communications or advertising. Finally, this technology has become an increasingly important methodological tool for communications research.

Emotion is used in a variety of different ways to support learning. Some systems use emotional signals to motivate or alter student behavior. For example, Lester's COSMO system employed a virtual character that praised students when they gave correct answers and gave sympathetic emotional displays when they failed so as to reinforce task feedback and increase student motivation (Lester et al. 2000). Other systems are designed to allow people to practice high-stakes emotional encounters through the relative safety of virtual reality. For example, the SASO-EN system (Figure 21.1) allows people to practice conflict resolution by speaking to life-sized virtual people with opposing viewpoints (Traum et al. 2008). In this system, the student's actions influence the emotional state of simulated characters, which is reflected in their physical behaviors (e.g., gestures and facial expressions) and also in their choice of negotiation tactics. Another example in this vein is the Virtual Patient Project, which allows psychiatry patients to practice diagnostic skills by interviewing a virtual rape victim (Kenny et al. 2008). Finally, other approaches use sensing technology to infer student emotions and adjust system behavior accordingly. For example, some experimental versions of the AutoTutor tutoring system detect student boredom or frustration through their posture and facial expressions (D'Mello et al. 2007; Graesser et al. 2008).

Emotion plays an important role in many psychosocial disorders, and systems that can automatically detect and respond to human emotional displays can play a factor in assessing or even treating these ailments. For instance, people suffering anxiety disorders experience emotional arousal in certain situations or even media portrayals of those situations – for example, a soldier suffering post-traumatic stress disorder (PTSD), an anxiety disorder that can develop after exposure to a terrifying event or ordeal, might experience physiological arousal when watching combat footage on the news. Emotionally resonant media can facilitate diagnosis of such conditions by simulating the anxiety-invoking situations associated with specific conditions and measuring any emotional reactions in the patient. An example of this approach is the virtual reality cognitive performance assessment test (VRCPAT) which makes use of emotionally evocative virtual environments and a battery of neuropsychological measures to help diagnose socio-emotional disorders such as PTSD (Parsons and Rizzo 2008). Similar techniques can further be used to treat such disorders. In what is known as virtual reality exposure therapy (VRET), therapists use computer-generated media to evoke anxiety in a patient and then use techniques such as cognitive behavioral therapy to help patients help regulate their emotions (Jarrell et al. 2006; Rothbaum et al. 2006). Other applications have used a theory of emotional mind to simulate how stress impacts people's reasoning to allow patients to better reflect on their own emotional reactions to situations. For example, Carmen's Bright Ideas is a system designed to teach emotion regulation techniques to mothers of pediatric cancer patients (Marsella et al. 2003). The system allows patients to develop meta-cognitive skills by observing and also shaping a simulated therapy session (see Figure 21.1).

Several applications have explored the use of emotional techniques to create persuasive communication. For example, Timothy Bickmore's Relational Agents Group has created a series of healthcare applications that employ empathetic dialogues and emotional nonverbal cues to promote adherence to medical recommendations (e.g., see Bickmore et al. 2009). Other research suggests a variety of ways to use emotionally resonant media to improve the persuasiveness of messages. For instance, Bailenson demonstrated that Stanford undergraduates were more convinced to accept new security policy when the agent that presented it mimicked their body movements (Bailenson and Yee 2005) and Moon showed that an agent using empathy and reciprocity could induce more self-disclosure during an interview (Moon 2000). Negative emotions can also be persuasive. For example, van Kleef et al. showed that displays of anger would elicit greater concessions in a multi-issue bargaining task (van Kleef et al. 2004).

An interesting and growing application of emotionally resonant technology is as a methodological tool for communication research. Past research on how emotional cues impact communication has been hampered by the difficulty in obtaining reliable measures of human emotional behavior. Typically this information is obtained at great cost by manual annotation. Techniques for automatically recognizing emotional behaviors provide one method for rapidly coding large quantities of behavioral data and providing a more objective and reproducible measure than these more traditional methods. A separate issue is how to systematically manipulate emotional cues within the context of a controlled experiment. Confederates are often used to manipulate these factors in order explore the impact of low-level behaviors on communication processes – for example, a confederate might be told to copy the posture and facial expressions of a subject during their communication. A persistent concern with the use of confederates is that the act of performing these unnatural behaviors might alter communication in ways that makes study results difficult to interpret. Techniques for automatically synthesizing emotional behaviors can address these methodological concerns. Emotionally expressive characters can be programmed to systematically manipulate one channel of information (e.g., gesture, facial expression, race or gender) while leaving other channels identical. Indeed, a large body of work has replicated, and in some cases extended, prior research findings through use of virtual interactions (Bailenson et al. 2004; Baylor and Kim 2008; Bente et al. 2001; Hoyt et al. 2003; Kang et al. 2008; Kraemer 2008; Slater and Steed 2002; Van Vugt 2008; Van Vugt et al. 2006).

Techniques for recognizing, understanding, and synthesizing emotion have only recently begun to transition from the laboratory and we can expect rapid transformation and innovation in how these tools impact applications. Methods are still brittle, their effectiveness inconsistent, and the appropriateness of their use still a matter of debate. What is clear, however, is that future computers will become increasingly competent in navigating human emotional life.

Conclusion

I began this chapter with a science fiction story about emotionally resonant media. For good or ill, such media will soon be science fact. Already, there are cameras that will only take our picture when we smile (Thangham 2007), games that 'read our emotions' (Radd 2007) and digital nurses that persuade us to take our medication

(Bickmore and Pfeifer 2008). Interactive media in the twenty-first century will certainly be more emotionally sophisticated, more evocative and more persuasive.

Many research challenges remain in realizing this potential. The last fifteen years have seen an explosion in the range and scope of computational models of emotional processes but the field is far from mature, and competing methods contain a host of incompatible and sometimes poorly articulated processing assumptions. Underlying this is a more fundamental lack of consensus in theoretical perspectives on human emotion that inform these computational methods: emotion theories differ in which components are intrinsic to an emotion (e.g., cognitions, somatic processes, behavioral tendencies and responses), the relationships between components (e.g., do cognitions precede or follow somatic processes), and representational distinctions (e.g., is anger a prototype or a natural kind). This lack of theoretical clarity remains a key challenge for model development.

Interestingly, emotionally resonant media may hold the seeds of a solution to this conundrum. The exercise of translating an emotion theory into a concrete working model can highlight inconsistencies and unarticulated assumptions of a given theoretical perspective on emotion. Further, the behavior of such working systems, as they interact with dynamic environments and human users can highlight implications of a theory that were difficult to foresee with pen and paper. This can lead to a continual cycle of theory inspiring model and model inspiring theory. Already, this has begun to spark an interdisciplinary partnership between theoreticians and modelers that may transform the science of emotion and, ultimately, our understanding of human emotions (see Scherer et al. in press).

Whether or not they change the way scientists theorize about emotion, emotional resonance will certainly change how we conceive ourselves. Humans have long defined their identity in contrast to 'less capable' others. Throughout history we have valued rational intellect as something that elevates us from emotional, brutish animals. Current fashion has us valuing emotion as something that elevates us from soulless, rational machines – perhaps in response to recent successes in machines defeating our intellectual champions in chess, backgammon, and the like. How will we chose to bolster our egos when our machines become our emotional equals?

Notes

1 This work was sponsored by the U.S. Army Research, Development and Engineering Command, and the National Science Foundation, under grants 0713603 and 0729287. The content does not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred.

2 Those interested in a more detailed discussion of current methods can consult these excellent recent survey articles on the topic: Cowie et al. 2001; Zeng et al. 2009.

3 Research on theory of mind often emphasizes emotions, although this is less true for computational systems that represent and reason about the mental state of other agents (human or synthetic). I use the term 'theory of emotional mind' to emphasize the specific attention to modeling emotional processes.

References

Arnold, M. (1960) *Emotion and Personality*. NY: Columbia University Press.

Ashish, K. and Rosalind, W.P. (2002) 'Real-time, fully automatic upper facial feature tracking', paper presented at the Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition.

Bailenson, J. and Yee, N. (2005). 'Digital chameleons: automatic assimilation of nonverbal gestures in immersive virtual environments', *Psychological Science*, 16: 814–19.
Bailenson, J., Beall, A., Loomis, J., Blascovich, J. and Turk, M. (2004) 'Transformed social interaction: decoupling representation from behavior and form in collaborative virtual environments', *PRESENCE: Teleoperators and Virtual Environments*, 13: 428–41.
Bartlett, M.S., Littlewort, B.C., Frank, M.G., Lainscsek, C., Fasel, I.R. and Movellan, J.R. (2006) 'Automatic recognition of facial actions in spontaneous expressions', *Journal of Multimedia*, 1: 22–35.
Baylor, A.L. and Kim, S. (2008) 'The effects of agent nonverbal communication on procedural and attitudinal learning outcomes', paper presented at the International Conference on Intelligent Virtual Agents.
Bente, G., Kraemer, N.C., Petersen, A. and de Ruiter, J.P. (2001) 'Computer animated movement and person perception: methodological advances in nonverbal behavior research', *Journal of Nonverbal Behavior*, 25: 151–66.
Bickmore, T. and Pfeifer, L. (2008) 'Relational agents for antipsychotic medication adherence', paper presented at the CHI'08 workshop on Technology in Mental Health.
Bickmore, T., Pfeifer, L. and Jack, B. (2009) 'Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents', paper presented at the SIGCHI Conference on Human Factors in Computing Systems.
Bulut, M., Sungbok, L. and Narayanan, S. (2008). 'Recognition for synthesis: automatic parameter selection for resynthesis of emotional speech from neutral speech', paper presented at the Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on.
Busso, C., Lee, S. and Narayanan, S. (2009) 'Analysis of emotionally salient aspects of fundamental frequency for emotion detection' *IEEE Transactions on Speech, Audio and Language Processing*, 17: 528–96.
Conati, C. and MacLaren, H. (2004) 'Evaluating a probabilistic model of student affect', paper presented at the 7th International Conference on Intelligent Tutoring Systems, Maceio, Brazil.
Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. et al. (2001) 'Emotion recognition in human-computer interaction'. *IEEE Signal Processing Magazine*, 32–80.
D'Mello, S.K., Picard, R.W. and Graesser, A.C. (2007). 'Toward an affect-sensitive autotutor'. *IEEE Intelligent Systems*, 22: 53–61.
de Melo, C. and Gratch, J. (2009) 'Expressing emotions in virtual humans through wrinkles, blushing and tears', paper presented at the 22nd Annual Conference on Computer Animation and Social Agents.
Demirdjian, D., Ko, T. and Darrell, T. (2005) 'Untethered gesture acquisition and recognition for virtual world manipulation'. *Virtual Reality*, 8: 222–30.
Ekman, P. and Friesen, W. (1978) *Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Palo Alto: Consulting Psychologists Press.
Ekman, P., Davidson, R.J. and Friesen, W.V. (1990) 'The duchenne smile: emotional expression and brain psysiology II'. *Journal of Personality and Social Psychology*, 58: 342–53.
Elliott, C. (1992) 'The affective reasoner: a process model of emotions in a multi-agent system' Ph.D dissertation No. 32, Northwestern, IL: Northwestern University Institute for the Learning Scienceso.
El Nasr, M.S., Yen, J. and Ioerger, T. (2000) 'FLAME: fuzzy logic adaptive model of emotions', *Autonomous Agents and Multi-Agent Systems*, 3: 219–57.
Fleischman, M. and Hovy, E. (2002) 'Emotional variation in speech-based natural language generation', paper presented at the International Natural Language Generation Conference, Arden House, NY.

Frijda, N. (1987) 'Emotion, cognitive structure, and action tendency', *Cognition and Emotion*, 1: 115–43.

Goldberg, J.H., Lerner, J.S. and Tetlock, P.E. (1999) 'Rage and reason: the psychology of the intuitive prosecutor', *European Journal of Social Psychology*, 29: 781–95.

Graesser, A.C., D'Mello, S.K., Craig, S.D., Witherspoon, A., Sullins, J., McDaniel, B. et al. (2008) 'The relationship between affect states and dialogue patterns during interactions with AutoTutor', *Journal of Interactive Learning Research*, 19: 293–312.

Gratch, J. and Marsella, S. (2004a) 'A domain independent framework for modeling emotion', *Journal of Cognitive Systems Research*, 5: 269–306.

Gratch, J. and Marsella, S. (2004b) 'Evaluating the modeling and use of emotion in virtual humans', paper presented at the 3rd International Joint Conference on Autonomous Agents and Multiagent Systems, New York.

Gratch, J., Marsella, S. and Petta, P. (2009) 'Modeling the antecedents and consequences of emotion', *Journal of Cognitive Systems Research*, 10: 1–5.

Gratch, J., Wang, N., Gerten, J. and Fast, E. (2007) 'Creating rapport with virtual agents', paper presented at the 7th International Conference on Intelligent Virtual Agents.

Hatfield, E., Cacioppo, J.T. and Rapson, R.L. (eds) (1994) *Emotional Contagion*, Cambridge: Cambridge University Press.

Helzle, V. (2003–6). ' "Artificial actors" Project filmakademie Baden-Wuerttemberg / Institute of Animation'.

Horton, D. and Wohl, R.R. (1956) 'Mass communication and para-social interaction: Observations on intimacy at a distance', *Psychiatry* 19: 215–29.

Hoyt, C., Blascovich, J. and Swinth, K. (2003) 'Social inhibition in immersive virtual environments', *Presence*, 12: 183–95.

Jarrell, P., Brian, A., Matthieu, D., Anton, T., Matt, L., Ken, G. et al. (2006) 'A virtual reality exposure therapy application for iraq war post traumatic stress disorder', paper presented at the Proceedings of the IEEE conference on Virtual Reality.

Kang, S.-H., Gratch, J., Wang, N. and Watt, J. (2008) 'Does contingency of agents' nonverbal feedback affect users' social anxiety?' paper presented at the 7th International Conference on Autonomous Agents and Multiagent Systems, Estoril, Portugal.

Keltner, D. and Haidt, J. (1999) 'Social functions of emotions at four levels of analysis', *Cognition and Emotion*, 13: 505–21.

Keltner, D., Ellsworth, P. and Edwards, K. (1993) 'Beyond simple pessimism: effects of sadness and anger on social perception', *Journal of Personality and Social Psychology*, 64: 740–52.

Kenny, P., Parsons, T., Gratch, J. and Rizzo, A. (2008) 'Evaluation of justina: a virtual patient with PTSD', paper presented at the 8th International Conference on Intelligent Virtual Agents.

Kraemer, N. (2008) 'Social effects of virtual assistants. A review of empirical results with regard to communication', paper presented at the Proceedings of the 8th International Conference on Intelligent Virtual Agents.

Krumhuber, E., Manstead, A., Cosker, D., Marshall, D., Rosin, P.L. and Kappas, A. (2007) 'Facial dynamics as indicators of trustworthiness and cooperative behavior', *Emotion*, 7: 730–5.

Lance, B. and Marsella, S. (2007) 'Emotionally expressive head and body movement during gaze shifts', paper presented at the International Conference on Intelligent Virtual Agents.

Lazarus, R. (1991) *Emotion and Adaptation*. NY: Oxford University Press.

Lee, C.M. and Narayanan, S. (2005) 'Towards detecting emotions in spoken dialogs', *IEEE Transactions on Speech and Audio Processing*, 13: 293–302.

Lepper, M.R. (1988) 'Motivational considerations in the study of instruction', *Cognition and Instruction*, 5: 289–309.

Lerner, J.S. and Keltner, D. (2000) 'Beyond valence: toward a model of emotion-specific influences on judgement and choice', *Cognition and Emotion*, 14: 473–93.

Lester, J.C., Towns, S.G., Callaway, C.B., Voerman, J.L. and FitzGerald, P.J. (2000) 'Deictic and emotive communication in animated pedagogical agents', in J. Cassell, S. Prevost, J. Sullivan and E. Churchill (eds), *Embodied Conversational Agents*. Cambridge: MIT Press.

Littlewort, G., Bartlett, M.S. and Lee, K. (2007) 'Automated measurement of spontaneous facial expressions of genuine and posed pain', paper presented at the International Conference on Multimodal Interfaces.

Loyall, A.B., Reilly, W.S.N., Joseph, B. and Peter, W. (2004) 'System for authoring highly interactive, personality-rich interactive characters', paper presented at the Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation.

Lucey, S., Matthews, I., Changbo, H., Ambadar, Z., de la Torre, F. and Cohn, J. (2006) 'AAM derived face representations for robust facial action recognition', paper presented at the Automatic Face and Gesture Recognition, 2006, 7th International Conference.

Ma, W.-C., Chiang, A. J.J.-Y., Hawkins, T., Frederiksen, S., Peers, P., Vukovic, M. et al. (2008). 'Facial performance synthesis using deformation-driven polynomial displacement maps', *ACM Transactions on Graphics*, 27: 121:1–121:10.

Mao, W. and Gratch, J. (2006) 'Evaluating a computational model of social causality and responsibility', paper presented at the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, Japan.

Marsella, S., Johnson, W.L. and LaBore, C. (2003) 'Interactive pedagogical drama for health interventions', paper presented at the Conference on Artificial Intelligence in Education, Sydney, Australia.

Martin, J.-C., Niewiadomski, R., Devillers, L., Buisine, S. and Pelachaud, C. (2006) 'Multimodal complex emotions: gesture expressivity and blended facial expressions', *International Journal of Humanoid Robotics*, 3: 269–91.

Mateas, M. and Stern, A. (2003) 'Integrating plot, character and natural language processing in the interactive drama façade', paper presented at the Technologies for Interactive Digital Storytelling and Entertainment (TIDSE).

Moon, Y.M. (2000) 'Intimate exchanges: using computers to elicit self-disclosure from consumers', *Journal of Consumer Research*, 26: 323–39.

Morency, L.-P., Whitehill, J. and Movellan, J. (2008). 'Generalized adaptive view-based appearance model: integrated framework for monocular head pose estimation', paper presented at the 8th International Conference on Automatic Face and Gesture Recognition.

Mori, M. (2005) 'On the uncanny valley', paper presented at the Proceedings of the Humanoids-2005 workshop: Views of the Uncanny Valley.

Parkinson, B. (2001) 'Putting appraisal in context', in K. Scherer, A. Schorr and T. Johnstone (eds), *Appraisal Processes in Emotion: Theory, Methods, Research* (pp. 173–86), London: Oxford University Press.

Parsons, T.D and Rizzo, A.A. (2008) 'Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test', *CyberPsychology and Behavior*, 11: 17–25.

Radd, D. (2007) 'EmSense: Tell us how you *really* feel', *Game Daily*, October 1.

Rothbaum, B.O., Anderson, P., Zimand, E., Hodges, L., Lang, D. and Wilson, J. (2006) 'Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying', *Behavior Therapy*, 37: 80–90.

Russell, S. and Norvig, P. (2002) *Artificial Intelligence: A Modern Approach*, 2nd edn, New York: Prentice Hall.

Scherer, K.R., Bänziger, T. and Roesch, E. (eds) (in press), *A Blueprint for an Affectively Competent Agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing*: Oxford University Press.

Schwarz, N., Bless, H. and Bohner, G. (1991) 'Mood and persuasion: affective states influence the processing of persuasive communications', *Advances in Experimental Social Psychology*, 24: 161–99.

Slater, M., and Steed, A. (2002) 'Meeting people virtually: experiments in shared virtual environments', in R. Schroeder (ed.), *The Social Life of Avatars*, London: Springer.

Stephenson, N. (1995), *The Diamond Age*: Spectra.

Swartout, W., Gratch, J., Hill, R., Hovy, E., Marsella, S., Rickel, J. et al. (2006) 'Toward virtual humans', *AI Magazine*, 27.

Thangham, C.V. (2007) 'New sony camer technology: no smile no picture', *Digital Journal*, September 14.

Tickle-Degnen, L. and Rosenthal, R. (1990) 'The nature of rapport and its nonverbal correlates', *Psychological Inquiry*, 1: 285–93.

Tiedens, L.Z. and Fragale, A.R. (2003) 'Power moves: complementarity in dominant and submissive nonverbal behavior', *Journal of Personality and Social Psychology*, 84: 558–68.

Traum, D., Gratch, J., Marsella, S., Lee, J. and Hartholt, A. (2008) 'Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents', paper presented at the 8th International Conference on Intelligent Virtual Agents.

van Kleef, G.A., De Dreu, C.K.W. and Manstead, A.S.R. (2004) 'The interpersonal effects of anger and happiness in negotiations', *Journal of Personality and Social Psychology*, 86: 57–76.

Van Vugt, H. C. (2008) *Embodied Agents from a User's Perspective*. Amsterdam: VU University Amsterdam.

Van Vugt, H.C., Hoorn, J.F., Konijn, E.A. and de Bie Dimitriadou, A. (2006) 'Affective affordances: improving interface character engagement through interaction', *International Journal of Human-Computer Studies*, 64(9): 874–88.

Vural, E., Cetin, M., Ercil, A.G.L., Bartlett, M.S. and Movellan, J.R. (2007) 'Drowsy driver detection through facial movement analysis', paper presented at the International Conference on Computer Vision.

Whitehill, J., Bartlett, M.S. and Movellan, J.R. (2008) 'Automatic facial expression recognition for intelligent tutoring systems', paper presented at the CVPR 2008 Workshop on Human Communicative Behavior Analysis.

Whiten, A. (ed.) (1991) *Natural Theories of Mind*, Oxford: Basil Blackwell.

Yong, C., Wen, C.T., Petros, F. and Pighin, Fr. (2005) 'Expressive speech-driven facial animation', *ACM Transactions on Graphics*, 24(4): 1283–302.

Zeng, Z., Pantic, M., Roisman, G.I. and Huang, T.S. (2009) 'A survey of affect recognition methods: audio, visual, and spontaneous expressions', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31.