Automatic Assessment and Analysis of Public Speaking Anxiety: A Virtual Audience Case Study

Torsten Wörtwein
Institute of Anthropomatics and Robotics
Karlsruhe Institute of Technology
Karlsruhe, Germany
torsten.woertwein@student.kit.edu

Louis-Philippe Morency
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA, USA
morency@cs.cmu.edu

Stefan Scherer
Institute for Creative Technologies
University of Southern California
Los Angeles, CA, USA
scherer@ict.usc.edu

***Abstract*—Public speaking has become an integral part of many professions and is central to career building opportunities. Yet, public speaking anxiety is often referred to as the most common fear in everyday life and can hinder one's ability to speak in public severely. While virtual and real audiences have been successfully utilized to treat public speaking anxiety in the past, little work has been done on identifying behavioral characteristics of speakers suffering from anxiety. In this work, we focus on the characterization of behavioral indicators and the automatic assessment of public speaking anxiety. We identify several indicators for public speaking anxiety, among them are less eye contact with the audience, reduced variability in the voice, and more pauses. We automatically assess the public speaking anxiety as reported by the speakers through a self-assessment questionnaire using a speaker independent paradigm. Our approach using ensemble trees achieves a high correlation between *ground truth* and our estimation (r=0.825). Complementary to automatic measures of anxiety, we are also interested in speakers' perceptual differences when interacting with a virtual audience based on their level of anxiety in order to improve and further the development of virtual audiences for the training of public speaking and the reduction of anxiety.**

I. Introduction

Public speaking anxiety is often referred to as the greatest and most common fear even though it does not result in a life threatening situation. Good public speaking skills are essential for everyday situations and career opportunities. Public speaking anxiety affects our public speaking performance. People anxious about public speaking tend to have a higher heart rate [1], are more self-focused [2] while presenting, and cannot visualize themselves being successful in public speaking [3].

Three different methods are commonly used to treat public speaking anxiety, namely systematic desensitization, cognitive modification, and skill training [3]. The first one assumes that a tendency of becoming hyper-aroused is the cause for public speaking anxiety [4]. By relaxing and imagining to give a speech over and over again with increased difficulty, systematic desensitization tries to dissociate public speaking from anxiety. Cognitive modification sees the reason for public speaking anxiety in problematic cognition [5]. Therefore, it tries to replace problematic cognition about public speaking with positive cognition. This is usually achieved by three steps [6]; firstly, fears of public speaking are discussed; secondly, one makes negative self statements, so that lastly, a trained therapist can disagree with these beliefs and encourage the person. The last method to treat public speaking anxiety is skill training [7]. Thereby, a lack of skills is assumed. Similar to *normal* public speaking classes skills including organization, voice, and non-verbal behavior are taught and through improved performance confidence in one's abilities can foster.

In the recent past, virtual audiences and virtual humans have become more prevalent for such training applications and researchers investigated virtual audiences specifically for the reduction of anxiety in public speaking [8]. Other virtual human training applications focus on the training of specific interpersonal skills rather than the reduction of anxiety. Notable examples include job interview training [9], [10], social skills [11], [12], and public speaking [13], [14].

Here, we extend this work and aim to understand what behaviors are characteristic for a speaker suffering from anxiety and improve automatic measures of anxiety severity. In related work, self-assessment questionnaires and intrusive systems, e.g. heart rate monitors, are used to assess and evaluate changes in public speaking anxiety severity [3]. We investigate less intrusive methods to approximate public speaking anxiety using machine learning. This approach could find applications in the assessment of public speaking anxiety as well as the capability to adapt the behavior of the virtual audience dependent on the public speaking anxiety to give a more appropriate and non-threatening experience for an anxious user.

In this work, we investigate acoustic and visual behavioral indicators, such as pause time and eye contact, which correlate with a self-assessed public speaking anxiety score, i.e. the Personal Report of Confidence as a Public Speaker (PRCS) [15]. Additionally, we attempt to automatically evaluate public speaking anxiety based on acoustic and visual features.

In particular, we identify three main research questions for this work:

Q1: What nonverbal acoustic and visual behaviors are correlated with public speaking anxiety?

Q2: Is it possible to automatically approximate anxiety scores using multimodal machine learning?

Q3: What effect does training with a virtual audience have on the performance of anxious presenters? Do anxious

speakers engage differently with a virtual audience and how is the virtual audience perceived by anxious speakers?

II. Related Work

Traditional methods mainly rely on intrusive measures of the autonomous nervous system and physiology to objectively reason about public speaking anxiety [3]. Signals such as blood pressure [16], heart rate [1], and electrodermal activity [17] are commonly used. Other research focused more on the neural activity of the brain, e.g. [18]. All these methods have in common that they are invasive and might interfere with a person giving a speech. A less intrusive way is to assess a speaker's behavior from third party perspective or the use of self-assessment questionnaires. For example, people who are anxious about public speaking tend to be not only more self-focused while presenting but they also think more negative about their public speaking performance [2]. The resulting lack of focus on an audience and their surroundings is one explanation why anxious speakers tend to underperform in public speaking tasks [2].

Less intrusive automatic measures of public speaking anxiety involve audiovisual behavior assessments. For example, in [19] researchers have investigated acoustic features related to self-assessed public speaking anxiety as well as to audience-assessed public speaking anxiety. Earlier work [20] indicated that the fundamental frequency (f_0), its variations, and the speech rate are important features that indicate public speaking anxiety. Based on the self-assessed public speaking anxiety and acoustic features extracted with PRAAT [21], the researchers in [19] found that anxious people tend to have a higher percentage of pause time. Despite a higher pause time, they found a correlation that apprehensive participants tend to have a higher speaking rate based on the syllables per breath group. Based on the audience-assessed anxiety they only found a link between anxiety and the variation of f_0.

Visual behaviors were manually assessed to identify foreign language anxiety in an oral foreign language examination [22]. Eight female students from a Midwestern university had a video taped examination in French. The first four minutes were evaluated by three rates by assessing non-verbal behavior such as facial expressions, gaze behavior, body movement, gesture, and posture. Anxious students expressed less facial activity compared to non-anxious students, especially less smiling, less brow behaviors, but a higher blinking rate. In addition, anxious students had less eye contact and had more often a rigid and closed posture. Lastly the gesture behavior varied between these two groups, anxious students had less speech related gestures. However, visual behaviors were manually annotated and not automatically detect.

Virtual audiences have been further investigated to treat public speaking anxiety [23], [24], [25]. Early works on virtual reality used to treat public speaking anxiety suggest that virtual reality could indeed be useful in treating public speaking anxiety and self-reported levels of anxiety could be reduced [23]. Further, a study involving university students with prominent public speaking anxiety underlined prior findings and suggests that virtual reality treatment sessions are indeed effective in reducing public speaking anxiety [24]. Researchers investigated the effect of three different types of virtual audiences, namely a neutral, a positive, and a negative audience, consisting of eight virtual characters [8]. They showed that the three settings had an influence on participants, generating anxiety in participants who scored high on the Personal Report of Confidence as a Public Speaker (PRCS) [15], underlining the immersive characteristic of such virtual audiences.

In addition, findings suggest that virtual humans can reduce the stress and fear associated with the perception of being judged [26], and thereby, lower emotional barriers to seeking help or increase willingness to train [27]. Overall, this puts virtual humans in a unique position capable of aiding socially anxious individuals to improve their social skills and potentially reduce their anxiety over time with frequent exposure. Virtual humans are excellent in captivating individuals' attention, in creating rapport and engaging the learner [28], which are essential prerequisites for successful learning outcomes.

In contrast to previous work, we focus on non-intrusive automatic computable features to automatically assess public speaking anxiety severity. In order to provide a training tool for anxious public speakers, we evaluate their perception of a virtual audience to adjust it for anxious speakers.

III. Methods

A. Experimental Design

For this study we make use of a virtual audience for public speaking training. In the present work we focus on the automatic characterization and assessment of public speaking anxiety. In addition to behavioral changes in public speaking anxiety, we assess users in a pre- to post-training test paradigm, i.e. we compare nonverbal behavior between a pre-training performance and a post-training performance. By following this paradigm, we can assess speakers' relative behavioral change while compensating for their initial public speaking anxiety. We inquire speakers' public speaking anxiety through the use of a self-assessment questionnaire. In the following we describe the study protocol in detail.

1) Study Protocol: A few days before their participation in the study, participants were instructed they would be asked to present two topics during 5-minute presentations. They were sent material about those presentations (i.e. abstract and slides) to prepare the day of the study. Before recording the first presentation, participants completed questionnaires on demographics, self-assessment, and public-speaking anxiety. Each participant gave four presentations. The first and fourth consisted of the pre-training and post-training presentations, where the participants were asked to present the same topic in front of a passive virtual audience. Note that all analyses presented here in this work are focusing on behavior in either the pre- or post-training presentations.

Between these two tests, the participants trained for *eye contact* and *avoiding pause fillers* in two separate presentations,

using the second topic. We chose these two basic behavioral aspects of good public speaking performances following discussions with Toastmasters[1] experts. In the second and third presentations, the audience was configured according to one of three training conditions:

1) **Control condition (Non-interactive virtual audience):** Hints given before training. No feedback during presentation.
2) **Direct feedback condition:** Hints given before training. Direct feedback during training: displayed as an objective measure of performance, i.e. a color-coded gauge at the top of the audience display.
3) **Interactive virtual audience condition:** Hints given before training. Indirect feedback during training: the audience behaves positively when the speaker is performing well (e.g. nodding, leaning forward), negatively when not (e.g. looking away, shaking head).

The condition was randomly assigned to participants when they came in; differences based on training conditions are not investigated in the present work.

In the study the virtual audience was displayed using two projections to render the audience in life-size. The projections were positioned such that the participants would be forced to move their head slightly to look at the whole audience, thereby making it easier to evaluate gaze performance. The participants were recorded with a head mounted microphone, with a Logitech web camera capturing facial expressions, and a Microsoft Kinect placed in the middle of the two screens capturing the body of the presenter.

After the post-training presentation, the participants were asked to complete a self-assessment questionnaire including questions about the learning experience and felt rapport with the audience, which took between 10 and 20 minutes. Participants were then debriefed and paid.

B. Participants and Dataset

Participants were recruited from Craigslist[2] and paid USD 25. In total, 47 people participated (29 male and 18 female) with an average age of 37 years ($SD = 12.05$). Out of the 47 participants 30 have some college education. Two recordings had technical problems leaving a total of 45 participants. On average the pre-training presentations lasted for 3:57 minutes (SD=1:56 minutes) and the post-training presentation 3:54 minutes (SD=2:17 minutes) respectively. Overall, there is no significant difference in presentation length between pre- and post-training presentations.

C. Measures

1) Self-Assessment Questionnaires: All participants completed questionnaires before the pre-training presentation, namely a demographics questionnaire and the 30-item 'Personal Report of Confidence as a Speaker (PRCS)' questionnaire [15]. The PRCS questionnaire is commonly used to estimate the public speaking anxiety. We used the original scoring method for the 30-item PRCS questionnaire proposed in [15], i.e. we inverted negative questions and summed all questions up (no: 0, yes:1). Afterwards, we linearly transformed the score to the interval from 0 (not anxious about public speaking, 0 on original scale) to 1 (anxious about public speaking, 30 on original scale). No significant difference ($t(43) = 0.25$, $p = 0.173$, $g = 0.414$) in the PRCS anxiety score between female ($M = 0.39$, $SD = 0.30$) and male participants ($M = 0.29$, $SD = 0.22$) was observed. Directly after the post-training the participants completed a 32-item self assessment questionnaire (SA)[3] adapted from the immersive experience questionnaire [29]. This questionnaire will be used to evaluate the participants' perception of the interactive virtual audience.

2) Objective measures: To assess behavioral characteristics of anxious participants, we evaluated public speaking performance using two objective measures, namely *eye contact* and the *avoidance of pause fillers*. The presenters were specifically informed about these two aspects in the training presentations. In order to create objective individual baselines, we annotated both measures for all pre-training and post-training test presentations. Two annotators manually marked periods of *eye contact* with the virtual audience and the occurrence of *pause fillers* using the annotation tool ELAN [30]. For both aspects we observed high inter-rater agreement for a randomly selected subset of four videos that both annotators assessed. The Krippendorff α for eye contact is $\alpha = 0.751$ and pause fillers $\alpha = 0.957$ respectively. Krippendorff's α is computed on a frame-wise basis at 30 Hz.

For eye contact we computed a ratio for looking at the audience $\in [0, 1]$, with 0 = never looks at the audience and 1 = always looks at the audience, over the full length of the presentation based on the manual annotations. The number of pause filler words were normalized by the duration of the presentation in seconds.

In order to compare learning differences, we used the normalized difference index ndi between pre-training and post-training test presentations for both objectively assessed behaviors and was calculated by

$$ndi = \frac{post - pre}{post + pre}. \quad (1)$$

3) Automatic Behavior Assessment: To approximate the PRCS anxiety score we utilize a regression approach. For this, we use Matlab's implementation of a least squared boosted regression ensemble tree. We evaluate our predictions with a leave one speaker out cross-evaluation. We use the forward feature selection to find a subset of features for each speaker independent fold. This kind of feature selection starts with an empty set of features and iteratively adds the feature that decreases together with the chosen features a criterion function the most. As a criterion function we use $(1 - \text{corr}(\hat{y}, y))^2$, where $\hat{y}$ are the predictions of the leave one speaker out cross-evaluation and y the ground truth. The following features of the pre-training presentations are used.

[1] http://www.toastmasters.org/

[2] http://www.craigslist.org/

[3] Self-assessment questionnaire: http://tinyurl.com/psonwly

a) Acoustic Behavior Assessment: For the processing of the audio signals, we use the freely available COVAREP toolbox (v1.2.0), a collaborative speech analysis repository available for Matlab and Octave [31][4]. COVAREP provides an extensive selection of open-source robust and tested speech processing algorithms enabling comparative and cooperative research within the speech community.

All following acoustic features are masked with voiced-unvoiced (VUV) [32], which determines whether the participant is voicing or not. After masking, we use the average and the standard deviation of the temporal information of our features.

Using COVAREP, we extract the following acoustic features: the parabolic spectral parameter (PSP) [33], the difference in amplitude of the first two harmonics of the differentiated glottal source spectrum (H1H2) [34], and the estimation of the R_d shape parameter of the Liljencrants-Fant glottal model [35]. Beside these features we also use the fundamental frequency (f_0) [32] and the first three formants ($F_{1/2/3}$) with two different filtering methods, ARMA and KARMA [36]. Additionally, we use the first four Mel-frequency cepstral coefficients ($\text{MFCC}_{0/1/2/3}$) and extract the voice intensity in dB. We do not use additional MFCCs nor further formants, since they may contain too much noise and are not easily interpretable. In addition to the acoustic features extracted by COVAREP we used a robust algorithm to detect syllable nuclei to estimate the speaking rate [37].

b) Visual Behavior Assessment: Gestures are measured by the change of upper body joints' angles captured using the Microsoft Kinect. In detail, we take the sum of differences in angles (from the following left and right body joints: shoulder, elbow, hand, and wrist). To eliminate noise, we set the difference to zero when both hands are below the hips. To avoid over weighting voluminous gestures, we truncate the differences when the difference is higher than a threshold, which we calculate from manual gesture annotations of 20 presentations. We use the mean of the absolute differences as an indicator for gesturing during the presentation. Additionally, we use the orientation of the torso to judge whether the participant is orientated to the audience. After binarizing the angle with a threshold determined by annotations, we use the average of being oriented and not being oriented towards the audience as a feature.

We evaluate eye contact with the audience based on two eye gaze estimations; eye gaze estimation from Omron's OKAO [38] and head orientation from constrained local neural field (CLNF) [39] are used separately to automatically assess whether a participant is looking at the audience or not. On a side note, using the CLNF algorithm for face gaze assessment [39], we observe a high correlation between the manually annotated and automatically assessed eye contact behavior. In particular, we observe a Pearson's $r = 0.71$ which is a highly significant correlation with $p < 0.01$. We use the ratio of looking at the audience relative to the length of the presentation as a feature.

[4]http://covarep.github.io/covarep/

Emotions, such as anger, sadness, and contempt, are extracted with FACET [40]. After filtering out frames that have a low confidence using the provided threshold by FACET, we take the mean of the emotions' intensity as features.

IV. Results

Since we conduct a Pearson correlation in several places (i.e. PRCS anxiety score with features), we test whether our data is normally distributed. Almost all features are not rejected by the Kolmogorov-Smirnov test at $alpha = 0.05$. This indicates that these features are normally distributed. However, features from the commercial FACET tool, namely *fear*, *sadness*, and *negative* facial expressions are not normally distributed.

A. Q1 - Behavioral Indicators of Public Speaking Anxiety

In the following section, we report the linear Pearson correlation coefficient along with the degrees of freedom and the p-value. Below we report the detailed statistical findings with the pre-training presentations.

The most correlated feature from the acoustic features is the vocal expressivity measured by the standard deviation of the first formant: ARMA-filtered ($r(43) = -0.30$, $p < 0.05$), KARMA-filtered ($r(43) = -0.41$, $p < 0.01$). Additionally, the standard deviation of MFCC_0 negatively correlates with the PRCS anxiety score ($r(43) = -0.36$, $p < 0.05$). Lastly, the pause time estimated by the ratio of unvoiced phonemes and voicing correlates positively with the anxiety score ($r(43) = 0.35$, $p < 0.05$). FACET's average facial *fear* expression intensity significantly correlates with the PRCS anxiety score ($r(43) = 0.41$, $p < 0.01$). Furthermore, both automatically extracted eye contact scores and the annotated eye contact score negatively correlate with the PRCS anxiety score: eye contact score based on CLNF ($r(43) = -0.41$, $p < 0.01$), based on OKAO ($r(43) = -0.54$, $p < 0.001$), and the annotated eye contact score ($r(43) = -0.32$, $p < 0.05$). We did not observe a correlation between the PRCS anxiety score and the variation of the pitch (standard variation of f_0; $r(43) = 0.04$, $p = 0.787$).

B. Q2 - Automatic Assessment of Public Speaking Anxiety Score

For the automatic assessment of the PRCS anxiety score we report Pearson's linear correlation coefficient and the mean absolute error (MAE). To test whether the three feature groups (visual, acoustic, and both) are significantly different, we conducted a one-way analysis of variance. We present the p-values of two-tailed t-tests as well as Hedges' g values as a measure of the effect size. The g value denotes the estimated difference between the two population means in magnitudes of standard deviations [41].

Tables I and II summarize the achieved performance and the selected features by the greedy feature selection of the leave-one out evaluation.

With a paired t-test we observe a significant improvement in the absolute error between the unimodal approaches and the multimodal approach ($t_{acoustic}(44) = 0.12$, $p_{acoustic} <$

Modality	Pearson's r	Mean Absolute Error
Acoustic	0.653	0.148
Visual	0.640	0.154
Acoustic+Visual	**0.825**	**0.118**
Baseline	n/a	0.215

TABLE I
PERFORMANCE (PEARSON'S r AND MEAN ABSOLUTE ERROR) FOR ALL MODALITIES WHEN PREDICTING THE PRCS ANXIETY SCORE ON THE PRE-TRAINING PRESENTATION.

Modality	Selected Features	Pearson's r
Acoustic	pause time	0.35*
	PSP mean	0.28
	$MFCC_2$ mean	0.28
	$MFCC_3$ std	–0.03
	speech rate (PRAAT)	–0.14
Visual	facial sadness expressions	0.21
	gesture	0.05
	orientation towards audience	–0.22
	eye contact (OKAO)	–0.54**
Acoustic+Visual	facial fear expressions	0.41**
	pause time	0.35*
	F_3 std	0.24
	facial negative expressions	0.23
	PSP mean	0.28
	H1H2 mean	0.16
	RD mean	0.01
	speech intensity std	–0.04
	speech intensity mean	–0.17
	F_1 std	–0.30*
	eye contact (OKAO)	–0.54**

TABLE II
SELECTED FEATURES FOR ALL MODALITIES WHEN PREDICTING THE PRCS ANXIETY SCORE ON THE PRE-TRAINING PRESENTATION WITH THEIR RESPECTIVE PEARSON'S r. * INDICATES $p < 0.05$ AND ** INDICATES $p < 0.01$.

0.05, $g_{acoustic} = -0.369$, $t_{visual}(44) = 0.12$, $p_{visual} < 0.05$, $g_{visual} = -0.336$): only acoustic features ($M = 0.16$, $SD = 0.12$), only visual features ($M = 0.15$, $SD = 0.12$), and both feature sets combined ($M = 0.12$, $SD = 0.09$).

In addition to between modalities, we compared our automatic assessment errors with the error of a constant mean prediction. Our multimodal pre-training prediction ($M = 0.12$, $SD = 0.09$) is significantly more accurate than the constant prediction ($M = 0.21$, $SD = 0.14$; $t(88) = 0.11$, $p < 0.001$, $g = -0.840$). Figure 1 shows a scatter plot of the ground truth against the predicted anxiety with both modalities.

C. Q3 - Effect of Virtual Audience on Anxious Public Speakers

To measure behavioral changes of anxious and non-anxious participants between the presentations, we correlate the PRCS anxiety score with the ndi (Eq. 1) of both targeted training behaviors, namely eye contact and avoiding pause fillers (see Section III-A1). We found that subjects with higher anxiety scores improved in their eye gaze patterns as observed with high correlations in manual and automatic measures. In particular, the change in the annotated eye contact ($r(43) = 0.33$, $p < 0.05$) as well as the change of the eye contact measured

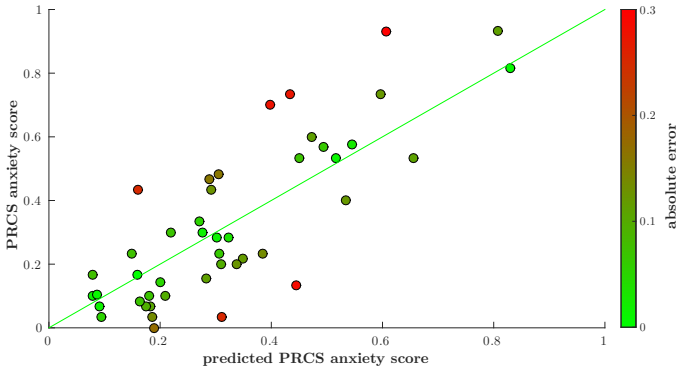

Fig. 1. Scatter plot of the predicted PRCS anxiety score against its ground truth.

by OKAO ($r(43) = 0.44$, $p < 0.01$) are positively correlated with the public speaking anxiety score. This indicates that anxious participants improved their eye contact with the virtual audience relative to non-anxious participants more. We could not observe a correlation between the PRCS anxiety score and the change of used pause fillers ($r(43) = 0.01$, $p = 0.930$). The change of the speech rate approximated by the number of syllables also correlates with the PRCS anxiety score ($r(43) = 0.32$, $p < 0.05$), meaning that anxious people spoke more after training.

In addition to behavioral changes between pre- and post-training performances, we evaluate the perception of the audience of participants by correlating surveyed questions with the public speaking anxiety score. Since our questionnaire is based on [42], we use the proposed joint-evaluation for this questionnaire. Here, we use Spearman's rank correlation coefficient since the questionnaire score is not continuous. The *real world dissociation* is positively correlated with the PRCS anxiety score ($r_s(43) = 0.36$, $p < 0.05$). This indicates that more anxious participants lose track of their surroundings. They do not focus on the virtual audience and are not emotionally attached to it. This finding is inline with [2], who associate an increased (negative) self-focus with public speaking anxiety. Hence, anxious participants do neither focus on their surroundings nor the audience.

Other joint evaluation scores (e.g. *cognitive involvement* and *challenge*) and single questions are highly uncorrelated with PRCS anxiety score. This indicates that anxious speakers experience learning and usefulness of the virtual audience similarly compared to non-anxious participants.

V. DISCUSSION

A. Q1 - Behavioral Indicators of Public Speaking Anxiety

In the first research question we investigate acoustic and visual non-intrusive behavioral indicators which correlate with the self-assessed public speaking anxiety score using the PRCS scale.Several of our findings are in accordance with findings in prior work [19]. For example, anxious participants exhibit more pauses in both studies. In line with Goberman et al., we could not find a relation between the pitch variation

(standard deviation of f_0) and self-assessed public speaking anxiety severity. However, we found a lower variation of the first formant and MFCC_0 which is closely related to speech intensity. These findings can be interpreted as a trend toward anxious people engaging in more monotonic speech.

Aside from exploring a higher negative self-focus of anxious people, the human audience in [2] attested to high anxiety participants exhibiting significantly less eye contact compared to low anxiety participants. Our experiment strengthens this finding. However, instead of subjectively validating eye contact by an audience, we not only automatically estimated eye contact but also objectively validated eye contact through manual annotations.

B. Q2 - Automatic Assessment of Public Speaking Anxiety

The second research question aimed to test whether it is possible to automatically assess public speaking anxiety with acoustic and visual features. We found that using both acoustic and visual features increased performance compared to a single modality both with respect to mean absolute error and Pearson's correlation (cf. Table I). The automatically selected features are summarized in Table II. The features selected from both modalities are comprised by correlating features (cf. **Q1**) or closely related features, such as being sad, being more quiet, speaking at a slower pace, gesturing less, and a lack of orientation towards to the audience in the visual modality. Our multimodal prediction significantly outperforms a static baseline approach in terms of mean absolute error and highly correlates with the ground truth labels ($r = 0.825$; cf. Figure 1). In the future, we plan to integrate an estimation of a speaker's anxiety in our interactive virtual audience to deliver a more engaging experience that is non-threatening.

C. Q3 - Effect of Virtual Audience on Anxious Public Speakers

In the last research question we investigate whether anxious participants engage differently with a virtual audience and how participants' behavior changes after the interaction with a virtual audience.

Based on the assessed questionnaires, anxious participants mostly experience the virtual audience experience similar to non-anxious participants. Most notably, there is no significant difference in the experienced usefulness, joy, and motivation when interacting with the virtual training platform. In addition, highly anxious speakers feel that they made similar progress as non-anxious speakers. However, we found evidence that anxious people need a more immersive and personalized experience to be more emotionally attached to the virtual audience and less self-focused. These findings confirm the discussed increased self-focus observed for anxious speaker in prior work [2]. In future experiments, we hope to overcome this lack of immersion in the interaction through the use of more engaging virtual audience behavior and appearance.

In addition to self-assessment questionnaires, we investigated whether automatically assessed behavior changes are stronger for anxious participants relative to non-anxious participants with the normalized difference index (cf. Eq. 1). This measure examines behavioral changes between pre-training and post-training performances. Hence, this assessment provides us with an evaluation of speakers' learning progress and perceived threat through the virtual audience training. In particular, we investigate eye contact and a high pause time as indicators for public speaking anxiety [2], [19]. Both measures change significantly more for anxious participants between the pre- and post-training presentation, i.e. anxious participants engage in more eye contact and have fewer pauses after training. This indicates that they reduced their public speaking anxiety to some degree and internalized qualities for successful public speaking.

VI. Conclusion

Our three research question focused on the behavior of participants with public speaking anxiety, the automatic assessment of public speaking anxiety scores, and finally the interaction of anxious participants with a virtual audience and their propensity to engage in learning. The first research question (**Q1**) aimed at identifying acoustic and visual non-invasive behavioral indicators related to public speaking anxiety. We found several indicators for public speaking anxiety: a higher pause time, a lower variation in speech parameters, less eye contact with the audience, and more fearful facial expressions. Based on these behavioral differences we attempt to automatically estimate the self-assessed PRCS score in research question **Q2**. The multimodal estimation yielded a high correlation with the ground truth ($r = 0.825$) and a small mean absolute error ($mae = 0.12$). The last research question **Q3** focused on the interaction of anxious participants with a virtual audience and their behavioral changes. Highly anxious participants were more self-focused and less emotionally attached to the virtual audience. However, behaviors relevant to public speaking anxiety and performance, especially eye contact and pause time, improved more for anxious participants after training.

We are confident that an automatic assessment of public speaking anxiety is an important step to provide anxious speakers a more immersive user experience when engaging with a virtual audience by appropriately varying its behavior and challenging them without posing a threat.

Acknowledgment

We would like to thank the Toastmasters group Funny Bones of Culver City, CA for their engagement and their helpful comments and discussions. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1421330. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] R. R. Behnke and L. W. Carlile, "Heart rate as an index of speech anxiety," *Speech Monographs*, vol. 38, no. 1, pp. 65–69, 1971.

[2] J. A. Daly, A. L. Vangelisti, and S. G. Lawrence, "Self-focused attention and public speaking anxiety," *Personality and Individual Differences*, vol. 10, no. 8, pp. 903–913, 1989.

[3] G. D. Bodie, "A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety," *Communication Education*, vol. 59, no. 1, pp. 70–105, 2010.

[4] G. Friedrich, B. Goss, T. Cunconan, D. Lane, J. Daly, J. McCroskey, J. Ayres, T. Hopf, and D. Ayres, "Systematic desensitization," *Avoiding communication: Shyness, reticence, and communication apprehension*, pp. 305–330, 1997.

[5] M. Allen, J. E. Hunter, and W. A. Donohue, "Metaanalysis of self-report data on the effectiveness of public speaking anxiety treatment techniques," *Communication Education*, vol. 38, no. 1, pp. 54–76, 1989.

[6] F. Glogower, W. Fremouw, and J. McCroskey, "A component analysis of cognitive restructuring," *Cognitive Therapy and Research*, vol. 2, no. 3, pp. 209–223, 1978.

[7] L. Kelly, "Skills training as a treatment for communication problems," *Avoiding communication: Shyness, reticence, and communication apprehension*, vol. 2, pp. 331–365, 1997.

[8] D.-P. Pertaub, M. Slater, and C. Barker, "An experiment on public speaking anxiety in response to three different types of virtual audience," *Presence: Teleoperators and Virtual Environments*, vol. 11, no. 1, pp. 68–78, 2002.

[9] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "MACH: My automated conversation coach," in *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing*, 2013, pp. 697–706.

[10] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard, H. Jones, M. Ochs, C. Pelachaud, K. Porayska-Pomsta, P. Rizzo, and N. Sabouret, "The TARDIS framework: Intelligent virtual agents for social coaching in job interviews," in *Proceedings of the International Conference on Advances in Computer Entertainment*, 2013, pp. 476–491.

[11] H. C. Lane, M. J. Hays, M. G. Core, and D. Auerbach, "Learning intercultural communication skills with virtual humans: Feedback and fidelity," *Journal of Educational Psychology Special Issue on Advanced Learning Technologies*, vol. 105, no. 4, pp. 1026–1035, 2013.

[12] H. Tanaka, S. Sakti, G. Neubig, T. Toda, H. Negoro, H. Iwasaka, and S. Nakamura, "Automated social skills trainer," in *Proceedings of the International Conference on Intelligent User Interfaces*, 2015.

[13] L. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer, "Cicero - towards a multimodal virtual audience platform for public speaking training," in *Proceedings of the International Conference on Intelligent Virtual Agents*, 2013, pp. 116–128.

[14] M. Tanveer, E. Lin, and M. E. Hoque, "Rhema: A real-time in-situ intelligent interface to help people with public speaking," in *Proceedings of the International Conference on Intelligent User Interfaces*, 2015.

[15] G. L. Paul, *Insight vs. Desensitization in Psychotherapy: An Experiment in Anxiety Reduction*. Stanford University Press, 1966.

[16] M. Dickens and W. R. Parker, "An experimental study of certain physiological, introspective and ratingscale techniques for the measurement of stage fright," *Speech Monographs*, vol. 18, no. 4, pp. 251–259, 1951.

[17] K. Clements and G. Turpin, "Physiological effects of public speaking assessed using a measure of palmar sweating," *Journal of Psychophysiology*, vol. 10, no. 4, pp. 283–290, 1996.

[18] J. B. Roberts, C. R. Sawyer, and R. R. Behnke, "A neurological representation of speech state anxiety: Mapping salivary cortisol levels of public speakers," *Western Journal of Communication*, vol. 68, no. 2, pp. 219–231, 2004.

[19] A. M. Goberman, S. Hughes, and T. Haydock, "Acoustic characteristics of public speaking: Anxiety and practice effects," *Speech Communication*, vol. 53, no. 6, pp. 867–876, 2011.

[20] M. A. Hagenaars and A. van Minnen, "The effect of fear on paralinguistic aspects of speech in patients with panic disorder with agoraphobia," *Journal of Anxiety Disorders*, vol. 19, no. 5, pp. 521–537, 2005.

[21] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[22] T. S. Gregersen, "Nonverbal cues: Clues to the detection of foreign language anxiety," *Foreign Language Annals*, vol. 38, no. 3, pp. 388–400, 2008.

[23] M. M. North, S. M. North, and J. R. Coble, "Virtual reality therapy: An effective treatment for the fear of public speaking," *International Journal of Virtual Reality*, vol. 3, pp. 2–6, 1998.

[24] S. R. Harris, R. L. Kemmerling, and M. M. North, "Brief virtual reality therapy for public speaking anxiety," *Cyberpsychology and Behavior*, vol. 5, pp. 543–550, 2002.

[25] L. Aymerich-Franch and J. Bailenson, "The use of doppelgangers in virtual reality to treat public speaking anxiety: a gender comparison," in *Proceedings of the International Society for Presence Research Annual Conference*, 2014, pp. 173–186.

[26] G. Lucas, J. Gratch, A. King, and L.-P. Morency, "It's only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior*, vol. 37, pp. 94–100, 2014.

[27] J. Hart, J. Gratch, and S. Marsella, *How Virtual Reality Training Can Win Friends and Influence People*, ser. Human Factors in Defence. Ashgate, 2013, ch. 21, pp. 235–249.

[28] N. Wang and J. Gratch, "Don't Just Stare at Me!" in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Chicago, IL, 2010, pp. 1241–1250.

[29] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, pp. 641–661, 2008.

[30] H. Sloetjes and P. Wittenburg, "Annotation by category: ELAN and ISO DCR," in *Proceedings of the International Conference on Language Resources and Evaluation*, 2008.

[31] G. Degottex, J. Kane, T. Drugman, T. Raitio, and S. Scherer, "Covarep - a collaborative voice analysis repository for speech technologies," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, 2014, pp. 960–964.

[32] T. Drugman and A. Abeer, "Joint robust voicing detection and pitch estimation based on residual harmonics," in *Proceedings of Interspeech 2011*. ISCA, 2011, pp. 1973–1976.

[33] P. Alku, H. Strik, and E. Vilkman, "Parabolic spectral parameter - a new method for quantification of the glottal flow," *Speech Communication*, vol. 22, no. 1, pp. 67–79, 1997.

[34] I. Titze and J. Sundberg, "Vocal intensity in speakers and singers," *Journal of the Acoustical Society of America*, vol. 91, no. 5, pp. 2936–2946, 1992.

[35] G. Fant, J. Liljencrants, and Q. Lin, "The LF-model revisited. transformations and frequency domain analysis," *Speech Transmission Laboratory, Quarterly Report, Royal Institute of Technology*, vol. 2, no. 1, pp. 119–156, 1995.

[36] D. D. Mehta, D. Rudoy, and P. J. Wolfe, "KARMA: Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking," *Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1732–1746, 2011.

[37] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior research methods*, vol. 41, no. 2, pp. 385–390, 2009.

[38] S. Lao and M. Kawade, "Vision-based face understanding technologies and their applications," in *Proceedings of the Conference on Advances in Biometric Person Authentication*, 2004, pp. 339–348.

[39] T. Baltrusaitis, P. Robinson, and L.-P. Morency, "Constrained local neural fields for robust facial landmark detection in the wild," in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, ser. IEEE, 2013, pp. 354–361.

[40] Emotient, "FACET SDK," 2014, http://www.emotient.com/products.

[41] L. V. Hedges, "Distribution theory for glass's estimator of effect size and related estimators," *Journal of Educational Statistics*, vol. 6, no. 2, pp. 107–128, 1981.

[42] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton, "Measuring and defining the experience of immersion in games," *International Journal of Human-Computer Studies*, vol. 66, no. 9, pp. 641–661, 2008.