Exploring Feedback Strategies to Improve Public Speaking: An Interactive Virtual Audience Framework

Mathieu Chollet[*1], **Torsten Wörtwein**[†1], **Louis-Philippe Morency**[‡], **Ari Shapiro**[*], **Stefan Scherer**[*]
[*]Institute for Creative Technologies, University of Southern California, Los Angeles, CA, USA
[†]Institute of Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany
[‡]Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA
{chollet, shapiro, scherer}@ict.usc.edu, torsten.woertwein@student.kit.edu, morency@cs.cmu.edu

ABSTRACT
Good public speaking skills convey strong and effective communication, which is critical in many professions and used in everyday life. The ability to speak publicly requires a lot of training and practice. Recent technological developments enable new approaches for public speaking training that allow users to practice in a safe and engaging environment. We explore feedback strategies for public speaking training that are based on an interactive virtual audience paradigm. We investigate three study conditions: (1) a non-interactive virtual audience (control condition), (2) direct visual feedback, and (3) nonverbal feedback from an interactive virtual audience. We perform a threefold evaluation based on self-assessment questionnaires, expert assessments, and two objectively annotated measures of eye-contact and avoidance of pause fillers. Our experiments show that the interactive virtual audience brings together the best of both worlds: increased engagement and challenge as well as improved public speaking skills as judged by experts.

Author Keywords
Multimodal Interfaces; Virtual Reality; Public Speaking; Training

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation: Miscellaneous

INTRODUCTION
Interpersonal skills such as public speaking are essential assets for a large variety of professions and in everyday life. The ability to communicate in social and public environments can greatly influence a person's career development, help build relationships, resolve conflict, or even gain the upper hand in negotiations. Nonverbal communication expressed through behaviors, such as gestures, facial expressions, and prosody, is a key aspect of successful public speaking and interpersonal communication. This was shown in many domains including healthcare, education, and negotiations where nonverbal communication was shown to be predictive of patient and user satisfaction [7], negotiation performance [25], and proficiency in public speaking [38, 33, 31, 3, 4]. However, public speaking with good nonverbal communication is not a skill that is innate to everyone, but can be mastered through extensive training [11]. In addition, even mild forms of public speaking anxiety can hinder one's ability to speak in public. Frequent exposure to presentation scenarios (even virtual ones), can help control public speaking anxiety [27]. The best form of training often is to present in familiar and forgiving environments and by receiving the audience's feedback during and after the presentation. Audiences may also provide indirect feedback during presentations by signaling nonverbally [23]. For instance, while an audience in a lecture may show signs of high attention (e.g. mutual gaze or forward leaning posture) and cues of rapport (e.g. nodding or smiling) in presentations they are engaged in, they may also show no interest (e.g. averted gaze or lack of backchannel behavior) or disagreement otherwise. Although these examples show the potential of live audiences to help interpersonal skill training, live audience based skill training is a process that is difficult to formalize as every audience behaves differently and the feedback might be unspecific to the speaker's performance. Other current practices involve a combination of text-book learning, practice with expert human role-players, and critiquing videos of the trainees' performances [11, 37, 28].

Recent developments in nonverbal behavior tracking and virtual human technologies enable new approaches for public speaking training that allow users to practice in a safe and interactive environment [6, 2, 39]. In particular, virtual human based training has shown considerable potential in the recent past, as they proved to be effective and engaging [17, 30, 1, 11]. Further, the virtual humans' replicability and consistency enable the development of new formalized training strategies for interpersonal skill training.

In this work we explore learning strategies and investigate the efficacy of a virtual audience in public speaking training. We present and evaluate a virtual audience prototype providing realtime feedback to the speaker. The audience can provide

[1]First and second author contributed equally to this work.

UbiComp '15 September 7-11, 2015, Osaka, Japan.

ACM 978-1-4503-3574-4/15/09...$15.00.
http://dx.doi.org/10.1145/2750858.2806060

Figure 1. We propose and explore feedback strategies for public speaking training with an interactive virtual audience framework. (A) Our learning architecture automatically identifies feedback strategies and enables multimodal realtime feedback mechanisms based on the speaker's audiovisual behavior, assessed in a wizard of Oz setting. (B) We evaluate three different and exclusive feedback strategies, namely the use of a (1) non-interactive virtual audience as the control condition of our evaluation study, (2) direct visual feedback, and (3) an interactive virtual audience. (C) To evaluate our framework, we conducted a study where participants trained with the virtual audience system and we assessed their performance improvement in a pre- vs. post-training evaluation paradigm. Each participant received feedback during training produced with one of the three strategies. Our evaluation addresses three research questions using the three assessment perspectives: (Q1) the presenters themselves, (Q2) public speaking experts , and (Q3) objectively quantified behavioral measures.

either explicit directly visualized performance measures or implicit nonverbal feedback from the virtual audience members (cf. Figure 1 (B)). Here, we investigate if these forms of feedback have a varying effect on learning outcomes. To approach this, we compared three feedback strategies, namely (1) a passive non-interactive audience as our control condition, (2) a passive audience enriched with direct visual feedback on the speaker's public speaking performance, and (3) an interactive virtual audience providing nonverbal behavior as feedback. We present a threefold evaluation of a pre- vs. post-training study to assess how a virtual audience is perceived as a public speaking training tool, and how it improves public speaking performance both from an expert's point of view as well as with objectively annotated behavioral measures.

RELATED WORK

Virtual humans are used in a wide range of healthcare applications, including psychological disorder assessment and treatment [6, 22] as well as social skills training applications, such as job interview training [15, 1], public speaking training [3, 5], conversational skills for autism spectrum disorder [40], and intercultural communicative skills training [19]. Virtual human social skills training holds several advantages over traditional role-play based or non-experiential training (e.g. classroom lectures) [11, 39]. Unlike traditional role-play based social skill training, virtual human based training is less dependent on trainer availability, scheduling, and costs. Further, human role-playing actors might introduce unnecessary and distracting variables to the training (e.g. culture, gender, language, age, etc.). Conversely, virtual humans' appearance and behavioral patterns can be precisely programmed, controlled and systematically presented to pace the exposure or interaction. It is also possible to start at a level that the user is most capable of successfully interacting with and to gradually ramp up in difficulty. In addition, once developed and polished a virtual human's availability and reach is only limited by access to technology.

On the other hand, integrating virtual human technologies into a working training application can be a complex endeavor, requiring expertise in different areas such as multimodal perception, speech synthesis or computer animation as well as in the considered application domain. Also, even realistic virtual humans in immersive settings still cannot compete with the level of realism of their actual human counterparts [18].

Further, findings suggest that virtual humans can reduce the stress and fear associated with the perception of being judged [22], and thereby, lower emotional barriers to seeking help or increase willingness to train [11]. 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. In addition, virtual humans are excellent in captivating individuals' attention, in creating rapport and engaging the learner [42], which are essential prerequisites for successful learning outcomes.

In the recent past, virtual humans were employed in social skills training applications, such as giving job interviews [15]. The My Automated Conversation Coach (MACH) job interview training system, for example, was tested with 90 undergraduate students (53 female and 37 males) from the MIT campus [15]. The experiment design consisted of three phases where the participants first interacted with a human counselor (considered the baseline assessment), then depending on condition they interacted with a specific version of MACH (with feedback or not) or simply watched a 30 minute educational video. Finally they interacted with the same human counselor for post-intervention assessment. The human counselors were blind to the conditions (educational video, MACH no feedback and MACH with feedback). The re-

sults showed a significant improvement in job interview skills between the MACH no-feedback and MACH with-feedback conditions and MACH with-feedback and control condition.

A Japanese research group developed a dialogue system called "Automated Social Skills Trainer", which primarily focuses to improve social skill training for people on the autism spectrum [40]. Their system closely follows traditional social skill training approaches: their training starts with an example video of a *good* dialog. After watching the example dialog video, the users themselves talk to a virtual human character and receive non-verbal feedback, such as head nods to encourage the user to continue speaking. The virtual human's feedback is not tied to the user's behavior [40]. Specific performance related feedback is only provided at the end of a session. The system generates a comparison report between the users' performances and the example dialog. At present only acoustic features are investigated. After receiving this direct feedback on how similar the performance was to the example dialog, the participants additionally received some positive comments about their performances to encourage them. In this work, no feedback based on the participant's performance was given during the interaction.

Virtual audiences have been further investigated to treat public speaking anxiety. 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 [24]. Further, a study involving university students with public speaking anxiety underlined prior findings and suggests that virtual reality treatment sessions are indeed effective in reducing public speaking anxiety [10]. 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 [27]. 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) [26], underlining the immersive characteristic of such virtual audiences.

The researchers, who developed MACH, recently investigated alternative direct feedback mechanisms with Google Glass [41]. The system, named Rhema, provides the speaker with feedback on speaking volume, i.e. speech intensity, and speaking rate. In a study with 30 students from the University of Rochester, the researchers evaluated continuous as well as sparse feedback systems (such as line plots and words, e.g. *LOUDER*). All participants gave three presentations (average duration of 3 minutes) with a continuous, a sparse, and no feedback system. The participants preferred the sparse feedback system which only provided brief periodical feedback. A post-hoc mechanical Turk survey was conducted but differences in performances between the feedback strategies were not found.

While previous work has investigated the use of virtual humans to train social skills, none of these works investigated the impact of multimodal feedback, *i.e.* feedback in the form of verbal, vocal or non-verbal signals, on public speaking skills training outcomes. On the other hand, Rhema investigated whether direct feedback can enhance user's training experience, however it did not involve a virtual audience. To the best of our knowledge, the present work is the first to specifically explore and evaluate ad hoc multimodal feedback strategies to improve public speaking skills using a virtual audience paradigm. The use of a virtual audience enables us to explore a wide range of feedback strategies both involving direct feedback, as utilized in [41], as well as less intrusive, potentially less distracting, and natural feedback from a virtual audience. In addition, the presented work complements prior work that found that public speaking anxiety was reduced when practicing in less threatening virtual environments, with the findings that in fact public speaking skills also improve when presenting in front of a virtual audience. Lastly, the present study is the first to investigate public speaking performance improvement using a thorough three-fold evaluation (1) using self-assessment questionnaires, (2) public speaking expert opinions, and (3) annotated behavioral measures.

RESEARCH QUESTIONS

In this paper, we set out to answer three research questions and identify effects of different feedback strategies for public speaking training with virtual audiences. We did this from the perspective of the learners themselves, third-party public speaking experts, and objectively quantified behaviors. We aim to develop a captivating learning experience that motivates users to enhance their public speaking skills through repeated use of our training framework. In particular, we are interested in measuring learners' engagement and perceived challenge in the task since these factors can effectively gauge improved performance. Furthermore, we utilize the expertise of experienced speakers to assess the complex characteristics of public speaking performances. Concepts such as *confidence* in presenting in front of an audience are essential for a good performance, however, they are multifaceted and difficult to formalize. Public speaking experts, such as graduates of the Toastmasters program, have a unique disposition to professionally assess a speaker's improvement on a wide range of complex performance aspects. However, this expert assessment has its challenges as well, such as potentially subjectively biased ratings. Therefore, we complemented expert criteria with objective and quantifiable measures to assess improvement. We chose two basic behavioral aspects, *eye contact* (*i.e.* how much does the participant looks at the audience) and *number of pause fillers* (hesitation vocalisations, *e.g. err* or *hmm*), as indicators of good public speaking performances following discussions with Toastmasters experts. In addition, these two aspects are clearly defined and can be objectively quantified using manual annotation enabling our threefold evaluation. We address the following research questions within this work:

Q1: Which feedback strategy provides the most engaging and challenging learning experience from the study *participants'* (i.e. the learners) point-of-view?

Q2: Which feedback strategy leads to the most improvement over the participants' public speaking skills, as assessed by *experts*?

Q3: Based on two *objectively* quantified behaviors of participants - eye contact and number of pause fillers - which

feedback strategy improves the participants' performance the most?

INTERACTIVE LEARNING FRAMEWORK

We developed an interactive learning framework based on audiovisual behavior sensing and feedback strategies (cf. Figure 1 (A)). In particular, the speaker's audiovisual nonverbal behavior was registered in the architecture and feedback was provided to the speaker according to the defined feedback strategies. We investigated three such strategies: (1) no feedback, i.e. the control condition, (2) direct visual feedback using a color-coded bar directly reflecting the speaker's performance, and (3) an interactive virtual audience producing nonverbal feedback (cf. Figure 1 (B)).

An interesting paradigm for a control condition would have been to use other human participants to act as a real audience for the training of a participant. However, such a solution involves logistics that were beyond our means. Additionally, an actual audience cannot be completely controlled and we would not have been able to standardize the amount of feedback of the audience to every participant and their reactions to every participant. Instead, we chose to use a passive virtual audience as a *control condition*, which is not subject to these two problems. Moreover, virtual humans applications for social training have been shown to trigger stress comparable to the stress induced by social training with a human partner [27, 18].

The color-coded *direct visual feedback* elements were configured to display the internal value of a behavioral descriptor directly, giving immediate feedback to the speaker about his or her performance. For our study, we used colored gauges (cf. Figure 1 (B), top) to indicate his/her performance to the participant. For instance, when training gaze behavior (*e.g.* look at the audience and not elsewhere), a fully green bar would indicate to the participant that his/her performance is perfect, *i.e.* he/she has been continuously looking towards the audience. Conversely, as the participant's performance worsens, the colored bar turns red, reflecting the amount of time when he/she did not look at the audience.

For the *interactive virtual audience*, the virtual characters were configured with a feedback profile. These profiles define behaviors the virtual characters will enact when specific conditions were met (e.g. smile when the speaker looks in the character's direction). Thus, the virtual characters can be used to provide natural, nonverbal feedback to the users according to their performance [5]. For the purpose of this study, the characters could nod or lean forward in an engaged manner when the participant's performance was good, and they could lean backwards or shake their head in disagreement when the participant's performance was poor.

To obtain perceptual information on the speaker's performance to send to the learning framework, we made use of a wizard of Oz interface. In order to ensure correct detection of the target behaviors (i.e. eye contact and pause fillers), this interface provides a frontal view of the participant and the output of a headset microphone worn by the participant to the wizard of Oz. In the future, we will utilize automatic audiovisual behavior tracking and machine learned models to automatically assess the speaker's performance, as suggested in [3][2]. Within our architecture, the perceptual information was aggregated into public speaking performance descriptors that directly influenced either the virtual audience's nonverbal behavior as feedback or the direct visual overlays. Behavioral information was coded in realtime as Perception Markup Language (PML) messages [32].

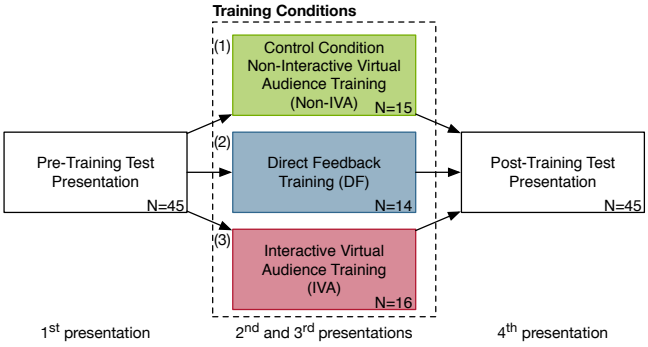

Figure 2. Study design and participant assignment to training conditions. In each training condition the participants gave two trial presentations; one focussing on their gaze behavior and one on avoiding pause fillers.

Our learning framework's output component was designed as a virtual audience scene built in the Unity 3D engine. The virtual characters were controlled with the Smartbody character animation platform [34]. The scene consisted of a simple room with tiered levels, allowing us to arrange the audience in different rows (cf. Figure 1 (B)). Visual overlays (e.g. colored gauges) can be displayed and used as direct feedback to the user as in the direct feedback study condition (cf. Figure 1 (B)).

METHODS AND MATERIALS

Experimental Design

As an initial study on the use of virtual audiences for public speaking training, we explore different feedback strategies. To this effect, we had users train with our virtual audience prototype with a pre- to post-training test paradigm (cf. Figure 2), i.e. we compare learning outcomes between a pre-training performance and a post-training performance. By following this paradigm, we can assess speakers' relative performance improvement while adjusting for their initial public speaking expertise. The three investigated and compared conditions reflect three feedback strategies.

Study Protocol

A few days before their participation in the study, participants were instructed they would be asked to present two

[2]As a first step, we investigated automatic assessment of gaze behavior with manual annotations of eye contact. Using the constrained local neural field algorithm for face gaze assessment [2], 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 $p < 0.01$.

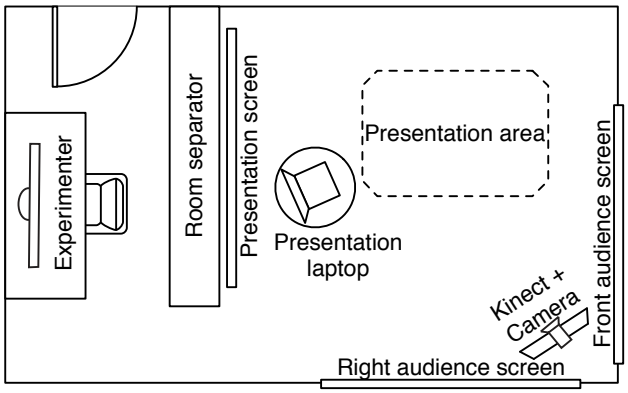

Figure 3. Study room setup.

topics during 5-minute presentations. They were sent material about those presentations (i.e. abstract and slides) to prepare before 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 (cf. Figure 2). 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 non-interactive virtual audience. Between these two tests, the participants trained for *eye contact* and *avoiding pause fillers* in two separate presentations, using the second topic. Every participant was given an information sheet with quotes from public speaking experts, namely Toastmasters, about how gaze and pause fillers impact a public speaking performance[3]. The Toastmaster hints were provided to the participants before each of the two training presentations (i.e. second and third presentation; cf. Figure 2).In the second and third presentations, the audience was configured according to one of the following three training conditions (cf. Figure 2 and Figure 1 (B)).

1. **Control condition (Non-interactive virtual audience; Non-IVA):** Hints given before training. No feedback during training.
2. **Direct feedback condition (DF):** 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 (IVA):** Hints given before training. Non-verbal 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. For conditions (2) and (3), a wizard provided the virtual audience with information about the speaker's performance in one of the two training behaviors, i.e. eye contact behavior and number of pause fillers. Note, that no wizard intervention was present during pre- and post-training test presentations.

[3]Written hints provided before training: **http://tinyurl.com/m4t6l62**

In the study the virtual audience was displayed using two projections to render the audience in life-size (cf. Figure 3). 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, paid, and escorted out.

Participants and Dataset

Participants were recruited from Craigslist[4] and paid 25USD. 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 (i.e. two or four year college).
Two recordings had technical problems leaving a total of 45 participants, with 15 participants assigned to the control condition (i.e. non-interactive virtual audience), 14 to the direct feedback condition, and 16 to the interactive virtual audience condition. 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.

Experts

To compare the pre- with the post-training presentations, three experts of the Toastmasters worldwide organization[5] were invited and paid 125USD. Their average age is 43.3 year ($SD = 11.5$), one was female and two were male. All experts had given more than eleven presentations in front of an audience during the past two years. The experts rated their public speaking experience and comfort on 7-point Likert scales. On average they felt very comfortable presenting in front of a public audience ($M = 6.3$, with 1 - not comfortable, 7 - totally comfortable). They have extensive training in speaking in front of an audience ($M = 6$, with 1 - no experience, 7 - a lot of experience).

Measures

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 [26], which is used to estimate public speaking anxiety [14]. Directly after the post-training the participants completed a 32-item self assessment questionnaire (SA)[6] (31 Likert-scale questions and one free form question) adapted from the immersive experience questionnaire [16].

[4]**http://www.craigslist.org/**

[5]**http://www.toastmasters.org/**

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

Expert Assessment

Three Toastmasters experts, who were blind to the conditions, evaluated whether participants of the three training conditions improved their public speaking skills. Experts viewed videos of the pre- and post-training presentations given in front of the virtual audiences. The videos were presented pairwise for a direct comparison in a random order blind to the experts. The experts could watch both videos independently as many times as they deemed necessary. Each video showed both the participant's upper body as well as facial expressions (cf. Figure 1 (C)).

The position of the pre- and post-training video, i.e. left or right, was randomized for each pair, as well as the order of participants. Additionally, experts were unaware of the participant's training condition.

The experts evaluated the performances on 7-point Likert scales. In particular, they assessed whether performance aspects - derived from prior work on public speaking assessment [33, 3, 31, 29] and targeted discussions with experts - apply more to the pre- or post-training presentation[7]:

1. Eye Contact
2. Body Posture
3. Flow of Speech
4. Gesture Usage
5. Intonation
6. Confidence Level
7. Stage Usage
8. Avoids pause fillers
9. Presentation Structure
10. Overall Performance

The pairwise agreement between the three experts is measured by the absolute distance between the experts' Likert scale ratings. The percentage of agreement with a maximal distance of 1 ranges between 63.70% and 81.48% for all 10 aspects, indicating high overall agreement between raters.

All the Likert scales of the survey originally ranged from 1 (applies more to the left video) over 4 (applies equally) to 7 (applies more to the right video). These scales were derandomized and linearly transformed so that -3 indicates that the aspect applies more to the pre-training presentation, 0 it applies equally, and 3 meaning it applies more to post-training presentation. Thus, improvement occurred if the values are positive.

Objective measures

To complement the self-assessment and expert ratings, we evaluated public speaking performance improvement using two objective measures, namely *eye contact* and the *occurence of pause fillers*. The presenters were specifically informed about these two aspects with the hints given before the training presentations for all three conditions. 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 [36]. For both measures 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 (in seconds) based on the manual annotations. The number of pause filler words were normalized by the duration of the presentation in seconds.

The behavioral change is measured by the normalized difference index ndi between the pre-training and post-training test presentations for both objectively assessed behaviors. This allows us to capture the difference of performance from one presentation to the other while compensating for individual differences; ndi was calculated by

$$ndi = \frac{post - pre}{post + pre} \in [-1, 1], \quad (1)$$

with pre ($post$) the eye contact ratio or pause fillers per second values for the pre-training and post-training presentations respectively.

RESULTS

We report statistical evaluation results below with M denoting the arithmetic mean and SD the standard deviation. In addition, we present the p-values of two-tailed t-tests and 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 [12]. Hedges' g is a commonly used standardized mean difference measure that can be transferred into other measures like Cohen's d [8]. Following the three research questions **Q1** - **Q3**, we report findings based on self-assessment questionnaires (**Q1**), expert assessment (**Q2**), and objective evaluation (**Q3**).

Q1 - Self-Assessment Questionnaires

Condition Dependent:

We first consider differences in self-assessment questionnaire measures by training feedback condition (cf. Figure 2). A one-way analysis of variance and two-tailed t-tests between the three conditions are applied. Out of the total 31 Likert-scale self-assessment questions we report the most relevant for the present work. The questionnaire's identifiers (e.g. SA_{Q1}) and exact questions are available online (**`http://tinyurl.com/psonwly`**).

We observe a significant difference among conditions whether the virtual audience is successful in holding the participants' attention (SA_{Q1}; $F(2, 44) = 4.628$, $p = 0.015$). Participants in the interactive virtual audience condition felt that the virtual audience held their attention significantly more ($M = 4.50$, $SD = 0.52$), when compared to the control condition ($M = 3.44$, $SD = 1.09$; $t(30) = 0.86$, $p = 0.001$, $g = 1.211$), and the direct feedback condition ($M = 3.60$, $SD = 1.40$; $t(29) = 1.04$, $p = 0.023$, $g = 0.840$; cf. Figure 4). No significant difference is found between control and the direct feedback conditions.

No significant difference among the three conditions is found regarding, whether the participants felt consciously aware of

[7]Aspect definitions and an online version of the questionnaire are available: **`http://tinyurl.com/ovtp67x`**

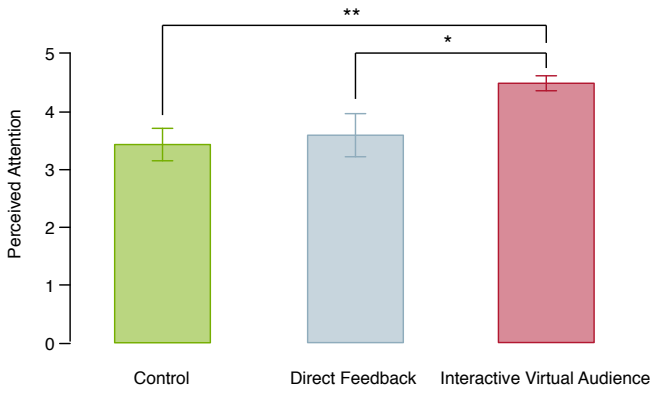

Figure 4. Perceived Attention (Q1). Visualization of perceived attention (SA_{Q1}) as assessed by using self-assessment questionnaires, with mean and standard error. Interactive virtual audience condition is perceived to significantly hold the speakers' attention more than the control and direct feedback conditions, with $p < 0.01$ as marked with ** and $p < 0.05$ with * respectively.

presenting in front of a virtual audience (SA_{Q6}; $F(2, 44) = 2.229$, $p = 0.120$). However, the presenters in the interactive virtual audience condition ($M = 3.56$, $SD = 1.63$) felt significantly more consciously aware of presenting in front of a virtual audience than the control condition ($M = 2.62$, $SD = 1.20$; $t(30) = 1.43$, $p = 0.049$, $g = 0.696$).

We found a marginally significant effect between training conditions, whether it was more challenging to present in front of a virtual audience (SA_{Q17}; $F(2, 44) = 2.745$, $p = 0.075$). The interactive virtual audience condition ($M = 2.94$, $SD = 1.12$) was rated as significantly more challenging than the control condition ($M = 2.00$, $SD = 1.03$; $t(28) = 1.02$, $p = 0.025$, $g = 0.801$). There is no statistical difference for perceived challenge between the interactive virtual audience and the direct feedback conditions.

Condition Independent:

Independent of the three training conditions, we investigate if participants in general express interest in a virtual audience as a training platform for public speaking. We measure this with two-tailed t-tests with the null-hypothesis being that the participants' average opinion coincides with the scale's middle point value (i.e. $M = 3.00$).

On average, participants scored significantly above the mean of the scale for the question whether they were focused on the virtual audience (SA_{Q2}; $M = 3.87$, $SD = 1.19$; $t(46) = 1.19$, $p < 0.001$, $g = 1.027$), and whether presenting in front of the virtual audience was easy (SA_{Q19}; $M = 4.02$, $SD = 0.99$; $t(46) = 0.99$, $p < 0.001$, $g = 1.449$).

Further, participants scored significantly above the scale's middle point when asked whether they want to repeat this experience (SA_{Q28}; $M = 4.64$, $SD = 0.79$; $t(46) = 0.79$, $p < 0.001$, $g = 2.902$). In addition, participants rated the virtual audience as a very useful tool (SA_{Q29}; $M = 4.81$, $SD = 0.50$); $t(46) = 0.50$, $p < 0.001$, $g = 5.123$) and that they would like to train with this tool to improve their public speaking skills (SA_{Q30}; $M = 4.77$, $SD = 0.56$; $t(46) = 0.56$, $p < 0.001$, $g = 4.423$).

Aspect	Non-IVA	DF	IVA
Eye Contact	0.40 (1.37)	0.02 (1.32)	0.27 (1.27)
Body Posture	0.29 (1.12)	0.00 (1.13)	0.19 (1.12)
Flow of Speech	0.16 (1.33)	0.17 (1.25)	0.40 (1.30)
Gesture Usage	0.42 (1.39)	0.26 (1.15)	0.33 (1.24)
Intonation	0.29 (1.38)	−0.02 (1.09)	0.50 (1.35)
Confidence Level	0.33 (1.49)	0.05 (1.45)	0.44 (1.58)
Stage Usage	0.42 (1.25)	−0.12 (0.99)	0.40 (0.89)
Avoids pause fillers	0.47 (1.01)	−0.07 (0.84)	0.35 (0.76)
Presentation Structure	0.22 (1.35)	0.17 (1.38)	0.42 (1.15)
Overall Performance	0.49 (1.42)	0.05 (1.45)	0.60 (1.32)

Table 1. Expert Assessment (Q2). Mean values and standard deviation (in parentheses) for all aspects for all three conditions, namely non-interactive virtual audience (Non-IVA), direct feedback (DF), and interactive virtual audience (IVA).

Q2 - Expert Assessment

Here, we report differences in expert assessment measures by training feedback condition. For each of the ten investigated aspects a one-way analysis of variance and two-tailed t-tests between the three conditions are conducted. Note, that the expert assessments are pairwise and hence positive values indicate that the participant improved in the post-training presentation and negative values vice versa; all values are $\in [-3, 3]$. To improve readability and due to space restrictions we reduce our report to aspects with significant differences only. Table 1 summarizes the mean M and standard deviations SD for all aspects.

For the *stage usage* aspect, i.e. the speaker's ability to use the space and stage to their advantage, a significant difference is observed between conditions ($F(2, 132) = 3.627$, $p = 0.029$). Stage usage improves significantly more for the interactive virtual audience condition ($M = 0.40$; $t(88) = 0.94$, $p = 0.011$, $g = 0.543$) and the control condition ($M = 0.42$; $t(85) = 1.13$, $p = 0.029$, $g = 0.473$) respectively, when compared to the direct feedback condition ($M = -0.12$).

For the *avoids pause fillers* aspect a significant difference is observed between conditions ($F(2, 132) = 4.550$, $p = 0.012$). Participants improve significantly more on average in the interactive virtual audience condition ($M = 0.35$; $t(88) = 0.80$, $p = 0.013$, $g = 0.530$) and control condition ($M = 0.47$; $t(85) = 0.93$, $p = 0.009$, $g = 0.572$) respectively as assessed by experts, when compared to the improvement in the direct feedback condition ($M = -0.07$).

For the *overall performance* aspect no significant difference is observed for the three conditions ($F(2, 132) = 1.945$, $p = 0.147$). However, between the interactive virtual audience condition ($M = 0.60$) and the direct feedback condition ($M = 0.05$) the overall performance improvement is approaching significance ($t(88) = 1.38$, $p = 0.059$, $g = 0.400$).

When comparing the mean over all experts for all aspects jointly, a significant difference between the three conditions is observed ($F(2, 447) = 5.814$, $p = 0.003$; cf. Figure 5). Overall the participants in the interactive virtual audience condition ($M = 0.39$, $SD = 0.83$; $t(298) = 0.86$, $p = 0.001$, $g = 0.395$) and control condition ($M = 0.35$,

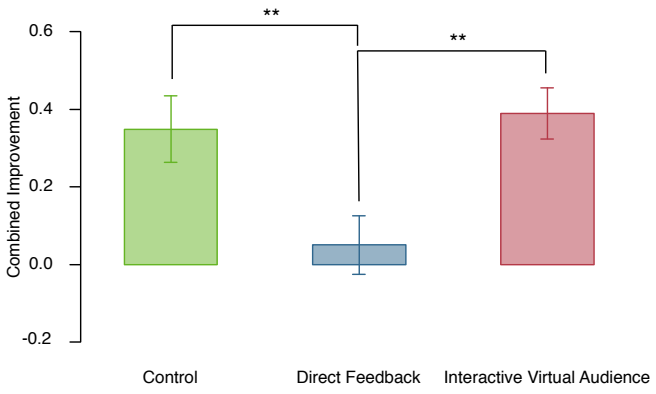

Figure 5. Combined overall expert assessment (Q2). Visualization of joint aspect improvement as assessed by experts, with mean and standard error. Both control and interactive virtual audience conditions outperform direct feedback condition significantly, with $p < 0.01$ as marked with **.

$SD = 1.05$; $t(288) = 0.98$, $p = 0.010$, $g = 0.305$) respectively improve significantly, when compared to the direct feedback condition ($M = 0.05$, $SD = 0.89$). No significant difference between the control condition and the interactive virtual audience condition is observed.

Q3 - Objective Evaluation

Here, we report differences between training conditions using manually assessed measures. For each of the two investigated aspects, namely eye contact and pause fillers, a one-way analysis of variance and two-tailed t-tests between the three conditions are conducted.

Based on the manually assessed eye contact, we observe no significant difference between the three conditions ($F(2, 42) = 0.923$, $p = 0.405$). However, presenters in all three conditions improved their eye contact behavior: control condition $M = 0.21$, $SD = 0.29$; direct feedback condition $M = 0.10$, $SD = 0.19$; and interactive virtual audience condition $M = 0.12$, $SD = 0.18$.

The usage of pause fillers also improved for all conditions as the number of pause fillers that is observed is reduced in the post-training presentation: control condition $M = -0.39$, $SD = 0.35$; direct feedback condition $M = -0.36$, $SD = 0.35$; and interactive virtual audience condition $M = -0.37$, $SD = 0.39$. Again, no significant difference between the three conditions is observed ($F(2, 42) = 0.018$, $p = 0.982$).

Finally, we explored eye contact improvement differences between groups of participants with different initial skills. We grouped participants based on the manual annotation of eye contact in the pre-training presentation, i.e. we divide the population of speakers into three equally sized tertiles (i.e. weak, moderate, and good speakers; cf. Figure 6). Weak speakers are defined as those that hold less eye contact with the audience in the pre-training presentation. Overall, we observe a significant difference for the three tertiles ($F(2, 42) = 36.762$, $p < 0.001$). All subsequent t-tests between tertiles are highly significant with $p < 0.01$.

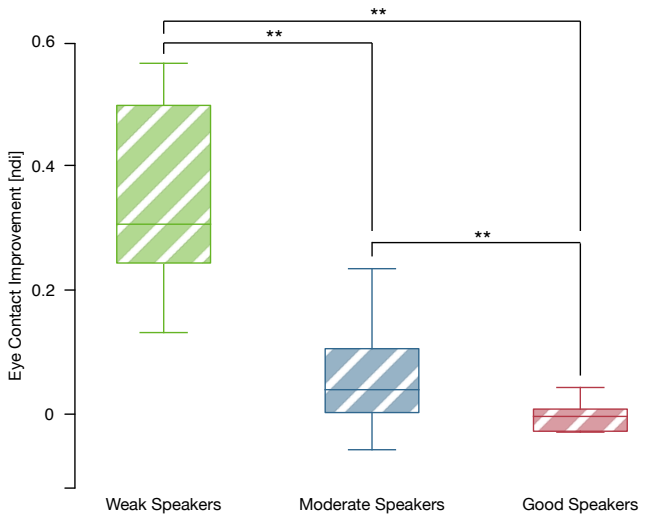

Figure 6. Objective Evaluation (Q3). Visualization of the ceiling effect of learning for weak, moderate, and good speakers with respect to their pre-training performance of eye contact. Weak speakers, i.e. those that did not look at the audience in the pre-training presentation, improve the most. Significant differences with $p < 0.01$ are marked with **.

DISCUSSION

In this section, we discuss our results with respect to the research questions previously introduced:

Q1 - Learners' Experience

Our first research questions aims at assessing the learning experience from the participants' point-of-view. The analysis of the participants' questionnaires revealed some interesting trends among and across feedback conditions:

The first result is that the participants' found the learning experience to hold much potential in terms of public speaking skills improvement across all conditions (SA_{Q29}). Participants were overall very eager to use the system in the future to hone their skills (SA_{Q30}). Although the very high scores obtained by the system on these questions should partly be put in perspective (i.e. it is likely that there is positive bias from the novelty of the experience and the highly immersive experimental setup with life-size virtual human characters), these results are very encouraging and suggest that the system could show good adherence when used repeatedly over longer time spans. As mentioned earlier, virtual audiences in general have the great advantage that their appearance and behavioral patterns can be consistently and systematically manipulated in order to customize and pace the training.

Second, interacting with the system was found to be rather easy and non-threatening (SA_{Q19}). This finding is further in accordance with findings in prior work where frequent exposure to presentation scenarios, can help control public speaking anxiety [27]. Similarly, it was found that participants reported experiencing lower fear of negative evaluation and engaged in less impression management when a virtual character was framed as autonomous than when it was framed as human-controlled [22]. The control condition was found to be the easiest to interact with, perhaps unsurprisingly as

there were no visual stimuli in this condition that gave direct feedback about the speaker's performance, hence there was no interaction required by the participants with the relatively static characters.

Lastly, the participants found the system to be very engaging (SA_{Q2} and SA_{Q10}). Perhaps unsurprisingly, the interactive virtual audience condition was the most engaging (SA_{Q1}; cf. Figure 4). In addition, the interactive virtual audience condition was found to be significantly more challenging than the control condition (SA_{Q17}). Both these findings are very promising: indeed, prior work in intelligent tutoring systems have shown a high correlation between the learners' engagement and the actual learning outcomes and performance in post-training tests [30]. While, high engagement and challenge in a learning task are not a guarantee for improved learning outcomes, it favorably influences the learners' attitude towards the training paradigm and could improve performance in the long run.

Q2 - Experts' Assessment

With our second research question, we investigated whether participants improved between the pre- and post-training test presentation based on expert assessments. The experts assess the speakers' improvement with ten selected categories referring to audiovisual nonverbal behavior, the structure of the presentation, as well as the overall performance of the presenters. We selected three experienced public speakers as independent experts from the Toastmasters organization. The experts assessed performances based on videos of the pre- and post-training presentations side-by-side for a direct comparison. By comparing the performances of pre- and post-training presentations we can compensate for both the presenters' level of expertise and the experts' critical opinion.

In general, we observe that overall, all the considered performance aspects improved across all training conditions, although the effect is only moderate. The overall performance improvement was the strongest for the interactive virtual audience conditions. The effect is approaching significance with $p = 0.059$ when compared to the direct feedback condition.

When comparing all the assessed aspects together, the interactive virtual audience and control conditions both lead to statistically significantly better expert ratings than the direct feedback condition (cf. Figure 5).

In addition, we found significant differences on some particular aspects across conditions: namely, speech intonation, stage usage and pause fillers improved more in the interactive virtual audience condition than in the direct feedback condition, while stage usage and pause fillers also improved more in the control condition when compared to the direct feedback condition.

In conclusion, the system generally shows promise for improving presenters' public speaking skills across all investigated aspects. It seems however that direct visual feedback performed poorly compared to the other conditions. This effect can be explained in a way that the additional visual stimuli (i.e. color coded gauges) proved to be more of a distraction than a benefit for the participants. This finding is in line with prior findings in related work where researchers found that users' preferred sparse direct visual feedback that is only available at some instances during a presentation rather than continuously [41].

Aspect	Expert 1	Expert 2	Expert 3
Eye Contact	0.58	0.76	0.68
Body Posture	0.63	0.72	0.68
Flow of Speech	0.74	0.86	0.71
Gesture Usage	0.70	0.78	0.71
Intonation	0.66	0.92	0.70
Confidence Level	0.83	0.89	0.81
Stage Usage	0.63	0.74	0.69
Avoids pause fillers	0.50	0.64	0.77
Presentation Structure	0.73	0.50	0.85

Table 2. Expert Preferences (Q2). Pearson's r values denoting linear correlation coefficients between all rated aspects with the overall performance for each reviewer (all observed correlations are significant with $p < 0.01$).

In the present work, the interactive virtual audience condition producing nonverbal feedback was not significantly better than the control condition after the investigated minimal training of only two short presentations. However, the fact that the interactive virtual audience has a positive learning effect along with the findings regarding perceived engagement (SA_{Q1}) and challenge (SA_{Q17}) discussed for research question Q1 could prove pivotal in the long run and keep the learner engaged and present a more challenging task (cf. figures 4 and 5). We see this result as an encouragement for us to investigate strategies of audience nonverbal feedback that could prove significantly more efficient than a still virtual audience. In particular, we plan to investigate and validate different types of feedback behavior produced by the virtual audience.

While the expert agreement is reasonably high with respect to the investigated aspects, we observe different expert preferences, i.e. the importance of aspects varies from expert to expert. In order to investigate this further, we correlate the aspect assessments for each expert with the expert's corresponding overall performance assessment (cf. Table 2). For example, it can be seen that Expert 2 highly values *flow of speech* with $r = 0.86$ and *intonation* with $r = 0.92$. In general, *confidence* highly correlates for all experts with the overall performance. We plan to further investigate expert preferences and plan to identify elements of improvement for each aspect in targeted follow-up investigations and discussions with the experts.

Q3 - Objective Behavioral Assessment

Last, we investigate whether participants improve between the pre- and post-training test presentations based on objective assessments. In particular, we investigate if presenters improve their *eye contact* behavior, i.e. increase eye contact with the virtual audience, and *avoid pause fillers*. For this evaluation we rely on manual annotations of these behaviors.

We observe that participants improve both behaviors in all conditions consistently. However, no significant effect for feedback condition is observed. This suggests that increased awareness of certain behaviors in fact improves public speaking skills and nonverbal behaviors regardless of feedback.

Based on our observations we could identify that speakers that already perform well with respect to the investigated behaviors do not benefit as much from the additional training. This ceiling effect becomes very clear for the improvement of *eye contact* when we divide the population of speakers into three equally sized tertiles (i.e. weak, moderate, and good speakers; cf. Figure 6). While these learning effects are encouraging, they might only be short-term improvements and possibly fade away in the long run. We plan to investigate the speakers' ability to retain their improved abilities in a more longitudinal assessment. The next section details some of our more concrete plans for the future.

Future Work

In the future we plan to compare different types of nonverbal feedback. For example, we would like to investigate if extreme or exaggerated audience behaviors, such as excessive yawning or even falling asleep, can help to create salience and ultimately improve the student's learning outcome as compared to a virtual audience that is showing less exaggerated behaviors but more believable and naturalistic ones. Even Toastmasters public speaking experts provide explicit (e.g. raising hands and signaling mistakes) as well as implicit feedback (e.g. nodding or smiling) to speakers during a presentation alongside post-hoc verbal feedback. While we expect that the exaggerating virtual audience might be perceived as less natural or realistic, it remains to be seen if their more stereotypical behavior can improve learning outcomes. This notion has been previously investigated under the term of *pedagogical experience manipulation* for cultural and social skill training with virtual humans in [20], where virtual characters would react with heated expressions of anger if the human interactant made a *cultural error*.

We also plan to evaluate our system in a longitudinal study. Participants will train on a regular basis over a longer course of time, and we plan to use presentations in front of real audiences for pre-training and post-training assessments. With such a training paradigm, we will be able to assess whether our prototype can improve public speaking skills consistently over time, and whether the improved public speaking skills of the participants actually transfer to real public speaking opportunities. We will also investigate more thoroughly differences in learning outcomes between speakers with varying initial public speaking skill levels. We further plan to train an automatic machine learning algorithm capable of modeling aspects of public speaking performances to automatically drive the virtual audience feedback behavior and assess public speaking performances.

Further, we plan to investigate the use of naturalistic characters and behaviors to enhance the appearance and versatility of the audience. As it can be seen from our preliminary version of the virtual audience (cf. Figure 1 (B)) we lack character model variability. In particular, the audience being only composed of male characters could have resulted in gender effects depending on the gender of participants. With the recent availability of inexpensive 3D scanning technology, an increasing number of researchers are investigating the possibility of scanning humans and creating personalized virtual characters. While most of the related literature is investigating static scans [43], we plan to utilize a virtual character development pipeline [9] to automatically acquire and utilize characters generated from such scans as virtual audience members [35]. This will enable the capture of virtual audience members with realistic proportions and appearances of any size, gender, ethnicity, and clothing as could be found among a real audience. Such scans only require a few minutes to capture and process, and are suitable for body gestures, weight shifts, head movements, and other body-centric nonverbal feedback. In addition, facial scanning [13] and animation techniques [21] will allow the virtual audience to perform emotive facial expressions that would render the audience more versatile (cf. Figure 7).

Figure 7. Example automatically generated facial expressions based on human face scans.

CONCLUSION

In this paper, we proposed and explored learning feedback conditions for an interactive virtual audience framework for public speaking training. We evaluated three different and exclusive feedback strategies, namely the use of an (1) interactive virtual audience, (2) direct visual feedback, and (3) non-interactive virtual audience as the control condition. We conducted a study with participants training with the virtual audience system and assessed their performance improvement in a pre- vs. post-training evaluation paradigm. We analyzed three research questions using three assessment perspectives: **Q1** the presenters themselves, **Q2** public speaking experts, and **Q3** objectively annotated behavioral data. Based on these we could identify the following three major findings: **Q1** Presenters enjoyed interacting and training their public speaking skills with the virtual audience in general. In addition, the interactive virtual audience was more engaging, captivating, and challenging overall when compared to the other conditions, which could prove beneficial for training outcomes in the long run. **Q2** Experts identified consistent improvement of public speaking skills from pre- to post-training for both the control and the interactive virtual audience conditions, with no significant differences between the two conditions. **Q3** Objective assessments of two basic behaviors, namely *eye contact* and *avoid pause fillers*, show consistent improvement regardless of feedback condition. Overall, we believe that a virtual audience can act as an effective platform to both improve public speaking skills as well as regulate or reduce public speaking anxiety, as public speaking with good nonverbal

communication is a skill that can be mastered through extensive training.

ACKNOWLEDGMENTS

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 Grants No. IIS-1421330 and No. IIS-1118018 and U.S. Army Research Laboratory (ARL) under contract number W911NF-14-D-0005. 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 or the Government, and no official endorsement should be inferred.

REFERENCES

1. Anderson, K., and et al. The TARDIS framework: Intelligent virtual agents for social coaching in job interviews. In *Proceedings of International Conference on Advances in Computer Entertainment* (2013), 476–491.
2. Baltrusaitis, T., Robinson, P., and Morency, L.-P. Constrained local neural fields for robust facial landmark detection in the wild. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, IEEE (2013), 354–361.
3. Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. Cicero - towards a multimodal virtual audience platform for public speaking training. In *Proceedings of Intelligent Virtual Agents (IVA) 2013*, Springer (2013), 116–128.
4. Chen, L., Feng, G., Joe, J., Leong, C. W., Kitchen, C., and Lee, C. M. Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16th International Conference on Multimodal Interaction*, 200–203.
5. Chollet, M., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. An interactive virtual audience platform for public speaking training. In *Proceedings of International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (2014), 1657–1658.
6. DeVault, D., and et al. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS)* (2014), 1061–1068.
7. DiMatteo, M. R., Hays, R. D., and Prince, L. M. Relationship of physicians' nonverbal communication skill to patient satisfaction, appointment noncompliance, and physician workload. *Health Psychology 5*, 6 (1986), 581.
8. Durlak, J. A. How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology 34*, 9 (2009), 917–928.
9. Feng, A., Huang, Y., Xu, Y., and Shapiro, A. Fast, automatic character animation pipelines. *Computer Animation and Virtual Worlds* (2013).
10. Harris, S. R., Kemmerling, R. L., and North, M. M. Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology and Behavior 5* (2002), 543–550.
11. Hart, J., Gratch, J., and Marsella, S. *How Virtual Reality Training Can Win Friends and Influence People*. Human Factors in Defence. Ashgate, 2013, ch. 21, 235–249.
12. Hedges, L. V. Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics 6*, 2 (1981), 107–128.
13. Hernandez, M., Choi, J., and Medioni, G. Laser scan quality 3-d face modeling using a low-cost depth camera. In *Proceedings of the 20th European Signal Processing Conference* (2012), 1995–1999.
14. Hook, J. N., Smith, C. A., and Valentiner, D. P. A short-form of the personal report of confidence as a speaker. *Personality and Individual Differences 44*, 6 (2008), 1306–1313.
15. Hoque, M., Courgeon, M., Martin, J.-C., Bilge, M., and Picard, R. Mach: My automated conversation coach. In *Proceedings of International Joint Conference on Pervasive and Ubiquitous Computing* (2013).
16. Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies 66* (2008), 641–661.
17. Johnson, W. L., Rickel, J. W., and Lester, J. C. Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education 11*, 1 (2000), 47–78.
18. Kwon, J. H., Powell, J., and Chalmers, A. How level of realism influences anxiety in virtual reality environments for a job interview. *International Journal of Human-Computer Studies 71*, 10 (2013), 978–987.
19. Lane, H. C., Hays, M. J., Core, M. G., and Auerbach, D. Learning intercultural communication skills with virtual humans: Feedback and fidelity. *Journal of Educational Psychology Special Issue on Advanced Learning Technologies 105*, 4 (2013), 1026–1035.
20. Lane, H. C., and Wray, R. E. *Individualized Cultural and Social Skills Learning with Virtual Humans*. Adaptive Technologies for Training and Education. Cambridge University Press, 2012, ch. 10.
21. Li, H., Weise, T., and Pauly, M. Example-based facial rigging. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2010) 29*, 3 (July 2010).
22. Lucas, G., Gratch, J., King, A., and Morency, L.-P. It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior 37* (2014), 94–100.

23. MacIntyre, P. D., Thivierge, K. A., and MacDonald, J. R. The effects of audience interest, responsiveness, and evaluation on public speaking anxiety and related variables. *Communication Research Reports 14*, 2 (1997), 157–168.

24. North, M. M., North, S. M., and Coble, J. R. Virtual reality therapy: An effective treatment for the fear of public speaking. *International Journal of Virtual Reality 3* (1998), 2–6.

25. Park, S., Shoemark, P., and Morency, L.-P. Toward crowdsourcing micro-level behavior annotations: the challenges of interface, training, and generalization. In *Proceedings of the 18th International Conference on Intelligent User Interfaces (IUI '14)*, ACM (2014), 37–46.

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

27. Pertaub, D. P., Slater, M., and Barker, C. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and virtual environments 11* (2002), 68–78.

28. Peterson, R. T. An examination of the relative effectiveness of training in nonverbal communication: Personal selling implications. *Journal of Marketing Education 27*, 2 (2005), 143–150.

29. Rosenberg, A., and Hirschberg, J. Acoustic/prosodic and lexical correlates of charismatic speech. In *Proceedings of Interspeech 2005*, ISCA (2005), 513–516.

30. Rowe, J., Shores, L., Mott, B., and Lester, J. C. Integrating learning and engagement in narrative-centered learning environments. In *Proceedings of the Tenth International Conference on Intelligent Tutoring Systems* (2010).

31. Scherer, S., Layher, G., Kane, J., Neumann, H., and Campbell, N. An audiovisual political speech analysis incorporating eye-tracking and perception data. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, ELRA (2012), 1114–1120.

32. Scherer, S., Marsella, S., Stratou, G., Xu, Y., Morbini, F., Egan, A., Rizzo, A., and Morency, L.-P. Perception markup language: Towards a standardized representation of perceived nonverbal behaviors. In *Proceedings of Intelligent Virtual Agents (IVA'12)*, LNAI 7502, Springer (2012), 455–463.

33. Schreiber, L. M., Gregory, D. P., and Shibley, L. R. The development and test of the public speaking competence rubric. *Communication Education 61*, 3 (2012), 205–233.

34. Shapiro, A. Building a character animation system. In *Motion in Games*, J. Allbeck and P. Faloutsos, Eds., vol. 7060 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2011, 98–109.

35. Shapiro, A., Feng, A., Wang, R., Li, H., Bolas, M., Medioni, G., and Suma, E. Rapid avatar capture and simulation using commodity depth sensors. *Computer Animation and Virtual Worlds 25*, 3-4 (2014), 201–211.

36. Sloetjes, H., and Wittenburg, P. Annotation by category: Elan and iso dcr. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, European Language Resources Association (ELRA) (2008).

37. Spence, S. H. Social skills training with children and young people: Theory, evidence, and practice. *Child and Adolescent Mental Health 8*, 2 (2003), 84–96.

38. Strangert, E., and Gustafson, J. What makes a good speaker? subject ratings, acoustic measurements and perceptual evaluations. In *Proceedings of Interspeech 2008*, ISCA (2008), 1688–1691.

39. Swartout, W., Artstein, R., Forbell, E., Foutz, S., Lane, H. C., Lange, B., Morie, J., Rizzo, A., and Traum, D. Virtual humans for learning. *AI Magazine 34*, 4 (2013), 13–30.

40. Tanaka, H., Sakti, S., Neubig, G., Toda, T., Negoro, H., Iwasaka, H., and Nakamura, S. Automated social skills trainer. In *ACM International Conference on Intelligent User Interfaces (IUI)* (2015).

41. Tanveer, M., Lin, E., and Hoque, M. E. Rhema: A real-time in-situ intelligent interface to help people with public speaking,. In *ACM International Conference on Intelligent User Interfaces (IUI)* (2015).

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

43. Wang, R., Choi, J., and Medioni, G. Accurate full body scanning from a single fixed 3d camera. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, IEEE (2012), 432–439.