An Interactive Virtual Audience Platform for Public Speaking Training

(Demonstration)

Mathieu Chollet, Giota Sratou, Ari Shapiro, Louis-Philippe Morency and Stefan Scherer
Institute for Creative Technologies, Los Angeles, CA, United States
mchollet@enst.fr, {stratou, shapiro, morency, scherer}@ict.usc.edu

ABSTRACT

We have developed an interactive virtual audience platform for public speaking training. Users' public speaking behavior is automatically analyzed using audiovisual sensors. The virtual characters display indirect feedback depending on user's behavior descriptors correlated with public speaking performance. We used the system to collect a dataset of public speaking performances in different training conditions.

Categories and Subject Descriptors

H.1.2 [**User/Machine systems**]: [Human information processing];
I.5.4 [**Pattern Recognition Applications**]: Computer Vision, Signal Processing;
K.3.1 [**Computers and Education**]: Computers Uses in Education

Keywords

Virtual audience, public speaking training, automatic behavior recognition

1. INTRODUCTION

Interpersonal skills such as public speaking are essential assets for a large variety of professions and in everyday life. Nonverbal communication (affect, demeanor, posture, eye contact, speech tone and fluency) is a key aspect of successful public speaking and interpersonal communication [5]. Audiences provide indirect feedback during presentations by signaling nonverbally, as they continuously rate and sense the presenter's speaking style, such as nodding and leaning forward in presentations they enjoy, or averting their gaze when they are not interested [3]. Paying attention to these feedback behaviors allows speakers to improve their performance.

However, an actual human audience is not always available or sometimes too intimidating for an anxious speaker. Virtual audiences have already been used successfully in Virtual Reality Therapy (VRT) to mitigate public speaking anxiety [4]. Pushing further, we want to investigate if public speaking performance can be improved using virtual training.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

Figure 1: Screenshot of the running system

In a previous study [1], we extracted automatic behavior descriptors correlated with experts' ratings of public speaking performances. Building on those results, we present an interactive virtual audience prototype capable of providing feedback to the speaker using these descriptors. We then describe a study we conducted to assess how an interactive virtual audience is perceived as a public speaking training tool, compared to a passive audience, and a passive audience enriched with direct feedback on the speaker's public speaking performance.

2. SYSTEM ARCHITECTURE

The architecture of the system is as follows. One computer acts as a *Performance Handler* and hosts a central message broker. The computers in the system connected to sensors run an instance of the *Behavior Recognizer*, a perception framework built to automatically extract behavior descriptors of public speaking performances [1]. The computers connected to a screen run a *Feedback Producer* module.

Each *Behavior Recognizer* instance sends detected behaviors to the *Performance Handler*. Alternatively, a Wizard-of-Oz interface can be used to act as another *Behavior Recognizer*. The *Performance Handler* aggregates all these behavior descriptors, and forwards the ones that are relevant to the training condition to the *Feedback Producer* instances. It can be configured to compute additional features from raw behavior values (e.g. fraction of spoken words over a time window).

The *Feedback Producers* instances run a Unity player dis-

playing a virtual audience, and optional colored gauges (See figure 1 for an example). The gauges are configured to display the value of a behavioral descriptor, giving direct and objective feedback to the speaker about their performance. Each *Feedback Producers* instance is configured with an audience layout, as well as a feedback profile for each audience member. These profiles define behaviors the virtual character will play when specific conditions are met (e.g. smile when the speaker looks in the character's direction).

3. STUDY DESIGN

Lane *et al.* have shown that the feedback strategy used while practicing intercultural communication skills has an effect on learning experience and outcomes [2]. As a first study on the use of our virtual audience for training, we decided to explore different feedback strategies:
(1) Hints given before training. Direct feedback during training: display an objective measure of performance (e.g. a gauge).
(2) Hints given before training. Indirect feedback during training: the audience behaves positively when the speaker is performing well (e.g. nodding), negatively when not (e.g. looking away).
(3) Hints given before training. No feedback during training.

One of our goals is to assess whether virtual audiences can improve overall public speaking performance. However, it would be impossible to conclude if a speaker had improved overall if, for instance, he had improved vocal qualities (e.g. flow of speech, enunciation) while worsening other behaviors (e.g. not looking at the audience). Therefore, we decided to focus on the improvement of specific behaviors independently of the others. We chose behaviors that we had found to be correlated with experts' ratings of public performances in our previous study [1]: gaze (i.e. it is judged good public speaking behavior to look at the whole audience) and speech, specifically filler words (i.e. hesitation words such as *"err"*, *"um"* or *"uh"*, are evaluated as poor public speaking performance).

A few days before their participation in the study, subjects were instructed they would be asked to present two topics in 5 minute presentations. They were sent some material (i.e. abstract and slides) to use, though they were instructed they would be free not to use them if preferred. The first and fourth presentations consisted of a pre-test and a post-test, where the participants were asked to present the first subject in front of a passive virtual audience. Between these two tests, the participants would train gaze and speech in two separate presentations, using the second topic. Every participant was given an information sheet with quotes from public speaking experts[1] about how gaze and filler words impact a public speaking performance. They were separated into three conditions, a control condition training with a passive audience and no gauges (*Passive VA*), a condition training with a passive audience and gauges giving direct feedback (*Passive VA + Gauges*), and a condition training with an interactive virtual audience giving indirect feedback (*Interactive VA*).

The study was run using two large screens for making a larger audience (see Figure 2), forcing the participants 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, and with a camera and a Microsoft Kinect placed in the middle of the two screens. 51 participants were recorded (17 per condition). Our next step consists of having public speaking experts evaluate whether the participants' performance improved from the pre-test condition to the post-test condition, overall and on gaze and speech behaviors.

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

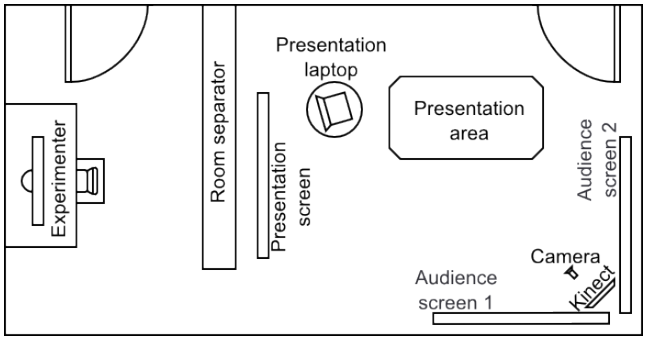

Figure 2: Study room setup

4. CONCLUSION

We presented a public speaking training system using an interactive virtual audience. It is designed around a distributed modular architecture, and relies on configuration files allowing to easily modify training conditions. To assess which type of feedback during training has the best impact, we collected a dataset of 51 participants doing a public speaking presentation in 3 different conditions: no feedback, direct feedback, indirect feedback. This dataset will be used to evaluate which feedback strategy is most efficient. In future work, we want to compare the effects of exaggerated audience behaviors (e.g. yawning, falling asleep), which could be used to create salience in the speaker, against more believable audience behaviors.

5. REFERENCES

[1] L. M. Batrinca, G. Stratou, A. Shapiro, L.-P. Morency, and S. Scherer. Cicero - towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents*, page 116–128, 2013.

[2] H. C. Lane, M. J. Hays, M. G. Core, and . Auerbach. Learning intercultural communication skills with virtual humans: Feedback and fidelity. 105:1026–1035, 2013.

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

[4] 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*, 11(1):68–78, Feb. 2002.

[5] E. Strangert and J. Gustafson. What makes a good speaker? subject ratings, acoustic measurements and perceptual evaluations. In *Interspeech*, page 1688–1691. ISCA, 2008.