Public Speaking Simulator with Speech and Audience Feedback

Bao Truong*
Department of Computer Science
University of Dayton

Trung-Nghia Le †
University of Science
Vietnam National University
Ho Chi Minh City

Khanh-Duy Le ‡
University of Science
Vietnam National University
Ho Chi Minh City

Minh-Triet Tran§
University of Science
John von Neumann Institute
Vietnam National University
Ho Chi Minh City

Tam V. Nguyen¶
Department of Computer Science
University of Dayton

ABSTRACT

Public speaking is one of the most important ways to share ideas with many people in different domains such as education, training, marketing, or healthcare. Being able to master this skill allows the speaker to clearly advocate for their subject and greatly influence others. However, most of the population reported having public speaking anxiety or glossophobia, which prevents them from effectively conveying their messages to others. One of the best solutions is to have a safe and private space to practice speaking in front of others. As a result, this research work is proposed with the overarching goal of providing people with virtual environments to practice in front of simulated audiences. In addition, the proposed work will aim to have live audience feedback and speech analysis details which could be useful for the users. The experiments via a user study provide insights into the proposed public speaking simulator.

Keywords: Virtual reality, public speaking, simulator, speech analysis, audience feedback.

Index Terms: Human-centered computing—Virtual reality

Figure 1. The user interface of our proposed public speaking simulator from the viewpoint of the speaker inside the virtual environment.

1 INTRODUCTION

Communication helps people connect with others in both personal and professional life. Public speaking is a form of communication that has the potential to reach hundreds to millions of people across the world. Nevertheless, many people struggle with public speaking, which results in poor performance due to anxiety. According to a research article, about 77% of the general population fears public speaking [1]. Public speaking is also important in the viewpoint of health science. A good speech can make an emotional or intellectual connection between the speaker and the audience. To overcome the fear and anxiety of public speaking, it is important to regularly practice topics and get familiar with a large audience of the environment.

Several prior works have been proposed to address the public speaking simulator [2-9]. However, most of these works lack the real time feedback to users. Therefore, in this paper, we aim to develop a public speaking simulation in virtual reality (VR) to help people with trouble or practice public speaking and provide them with real time feedback. In reality, people often do not have a place or a big crowd to practice public speaking except for themselves and a mirror. This motivates us to develop a simulated stage and interactive audiences through a virtual environment. When the speaker is giving a speech, the simulation should be able to pick up and analyze the voice into data. Some possible collected data are the rate of speaking, pauses, filler words, grammar, to name a few. The simulation tracks human gaze to help the speaker with eye contact during the presentation. These data would then be fed to virtual audiences in terms of giving back real-time reactions, depending on how the speaker performs. Following the conclusion

* email: truongb1@udayton.edu
† email: ltnghia@fit.hcmus.edu.vn
‡ email: lkduy@fit.hcmus.edu.vn
§ email: tmtriet@fit.hcmus.edu.vn
¶ email: tamnguyen@udayton.edu

Figure 2. Example scenario of virtual interaction for students in a health science class.

of the practice session, more in-depth details are shared with the speaker for further self-improvement.

Figure 1 shows the viewpoint of the speaker inside the virtual environment. Furthermore, in a virtual environment, speakers can also practice their interaction with virtual objects to attract audience in their presentations. For example, in Figure 2, we illustrate a scenario of our system when a presenter can easily interact with virtual models of human body and heart to give lectures to young kids and students in a health science class.

The remainder of this paper is organized as follows. Section 2 summarizes the related works. In Section 3, we introduce the proposed framework. Section 4 presents the evaluation. Finally, Section 5 concludes the paper and paves the way for future work.

2 Related Works

VR is a trending and growing research field whereas public speaking has a long history. There have been different effective ways people have come up with traditional methods to help with public speaking, but combining with VR it allows for new and informative ways to practice.

Pellett et al. [2] presented a simple VR training simulator, in which users present in a virtual auditorium with their speech appearing on a virtual cue card. Poeschl et al. [3] used a Quest-VR framework to evaluate public speaking of students in a VR environment. However, these systems do not support virtual audiences or feedback to the users. Ovation [4], a paid application for professional public speaking, has an implementation of a training tool that gives real-time feedback. It also helps optimize workflow by allowing the import of speech notes and slides. Another aspect is that it has a realistic environment and audience. Nevertheless, audience reaction to the speech and body language is still lacking and very basic. Another widely known virtual reality communication application is VRChat [5]. It has a large active audience and is able to talk with other real people. It has a lot of different environments and diverse tools. Though, the program is geared toward more chatting one-on-one or in small groups instead of public speaking. It also requires real-life participation from others which is not always available.

Research on emotion responsive audiences for VR public speaking simulators based on speakers' voices have been conducted by El-Yamri et al. [6]. They employed a game environment, in which a virtual individual reacted to the user's voice, body, and gaze. It features a speech audio processor that analyzes the speech, tone, and emotions of the speaker. On the other hand, Piyush [7] proposed to analyze audio signals in virtual talking. In particular, reference audio from the user is recorded and compared with the audio input obtained from the user in the virtual talking using a speaker verification algorithm. In another effort, El-Yamri et al. [8] proposed a VR videogame for training the ability of public speaking. The speaker talks in front of a virtual audience who reacts in real-time to the speaker's features, such as voice and gestures. Thus, the speaker can adapt his/her speech according to the reaction of the virtual audience. However, these works only focus on audio instead of gaze and body tracking. Recently, Hangyu et al. [9] discussed and analyzed the effectiveness of VR playback in public speaking training. They conducted a pilot experiment to test hypotheses about the relations between the public speaking anxiety level and the attention-concentrating effect of the virtual avatar, and the immersion of VR.

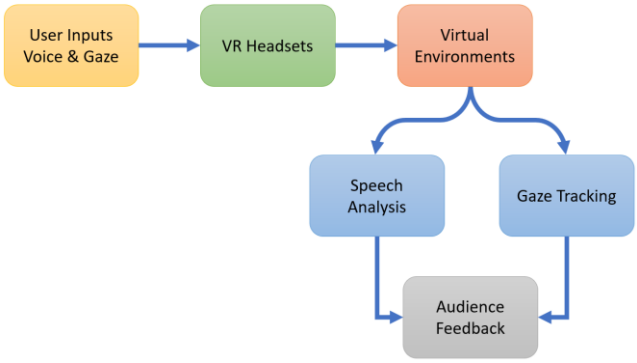

Figure 3. The overview of our proposed framework for public speaking simulator.

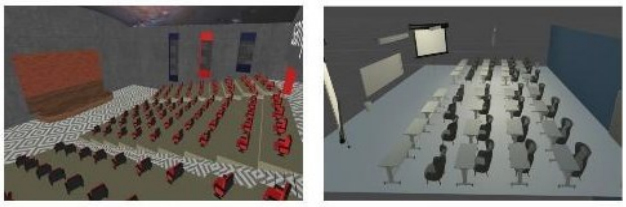

Figure 4. Our virtual environments include auditorium and classroom.

3 Proposed Framework

As mentioned earlier, we propose a new framework for public speaking simulation. Figure 3 shows the flowchart of our proposed framework. There are major components that would make up the main VR application: 1) We develop multiple environments for different types of public speaking situations. 2) We integrate the speech analysis to provide details on faults and improvements. 3) We introduce audience feedback based on user inputs in the virtual environment.

3.1 Virtual Environment

We design and develop different virtual environments for the public speaking simulator. Figure 4 shows some existing locations to practice in our VR system including auditorium and classroom. Using a 3D modeling tool, we model and apply texture to the auditorium and classroom. Note that we can use the 3D

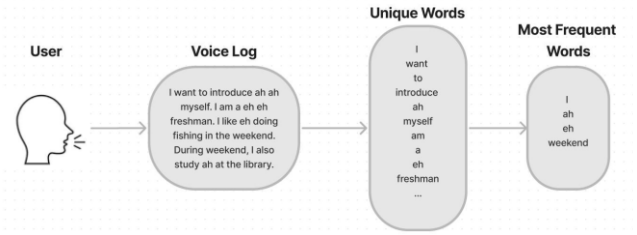

Figure 5. The flowchart of the speech analysis in our simulator.

reconstruction techniques such as 3D Scanner App [10] to capture 3D scenes of the real environment. Then, we can manually refine the scanned models to speed up the 3D modeling process. For the audience, we use avatars instead of real photos.

3.2 Speech Analysis

We integrate the speech analysis into the simulator. In particular, we use the built-in Window Speech library [11] to transcribe voice into text format. This allows us to log a text file of the speech after it has ended. Figure 5 shows the flowchart of the speech analysis in our simulator. We extract the unique words and then the most frequent words. We match the recognized words with a predefined list of stop words. Stop words are any word in a stop list (or stoplist or negative dictionary) which are filtered out (i.e. stopped) before or after processing of natural language data (text). This helps the system detect if the speaker tends to repeat any stop words very often. In order to help the speaker avoid stammering, we also detect the frequencies of stuttering words. Stuttering words are whole words or parts of the word more than once, or pause awkwardly between words. This is known as stuttering. For example, the speaker may speak fast and jam words together, or say "ah" often. All of this analyzed information is provided to the speakers after every practice session.

3.3 Audience Feedback

Taking into account the speech and gaze tracker and depending how the speaker performs, each virtual individual audience is assigned with unique attentional and emotional scores. Each stage of emotions is simplified using faces from our list of predefined avatars. Figure 6 illustrates the human gaze in the public speaking simulation. Whenever the audience receives the attention, i.e., 4-second timespan (as recommended in [12]), his/her emotion and attention scores, ranging from 0 to 100, are updated positively. On the contrary, the audience feedback scores reduce if the virtual individual is neglected by the speaker. Figure 6 shows some examples of our audience avatars with emotions. The emotion score also considers the user performance such as pause words, filler words, speaking rate and more.

3.4 Implementation

Regarding implementation, we develop our public speaking simulator with Unity 3D [13] and C# programming language. All models such as auditorium, classroom, facilities, and audience are designed and modeled in Blender [14]. Then, we imported the 3D models within the virtual environment in Unity3D. As mentioned, we use Window Speech [11] to transcribe voice into the text. We also implement both speech and audience feedback via the update functions in Unity3D. Finally, we deploy our simulator onto Oculus Rift VR headsets. The headsets render 2 stereoscopic views, one for each eye, to help viewers/participants experience the full immersion.

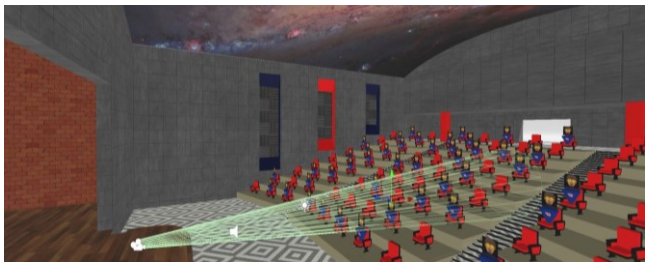

Figure 6. The visualization of the human gaze in the public speaking simulator. The virtual audience receives the attention from the speaker.

Figure 7. The examples of our avatars with emotions. Note that the emotions depend on the attention of the speaker in the virtual environment.

4 Evaluation

To evaluate the quality of this study, we built 3 different versions of the simulator, namely, basic, lite and full version.

- Basic Version: Users can only observe the virtual environment and practice public speaking without speech and audience feedback.
- Lite Version: Users can only observe the virtual environment and practice public speaking with speech feedback.
- Full Version: Users can observe the virtual environment and practice public speaking with both speech and audience feedback.

We invited 10 participants who are university students from freshman to senior. All users trying the simulator are those who have different levels of glossophobia. We let users experience and collect their feedback on all 3 versions. Each participant took part in a 30-minute session, namely, a 10-minute trial for each version. For each trial, each participant was tasked to present about him/herself, and pick one of the following topics such as global warming, university experience, technology, favorite sports, freedom of speech, and prospective careers. After the participant finished all three trials with three versions of the system, he/she went through a post-study interview. In the interview, we collected his/her qualitative feedback on how he/she felt about the effectiveness of three versions of the system on public speaking training and what can be improved.

Generally, we received positive comments after the user study. Some insightful discussions of our participants are as follows. The participants reported that they felt confident to talk in the simulator

in all three versions. They practiced their public speaking without any pressure by looking at the human avatars.

The lite version supports the speech feedback which is useful to avoid repeating certain words or stop words. Most participants prefer the full version. The full version provides the attentional feedback from the audience so that the speaker knows where to focus in the virtual auditorium. The speaker can pay attention to the entire room instead of looking straight only to the audience at front.

In addition, we also received constructive comments to improve our simulator. For example, the body pose tracking of the speaker is also useful. Indeed body language helps people have a better understanding of the overall situation and thus adjust the content of the talk and voice intonation. To address this, we plan to integrate some sensors such as Kinect to detect the body pose and reflect the body pose in the virtual environment.

Another key additional feature should be the ability to record the practice sessions in the virtual environment. The users can replay the recordings to better understand their upcoming public speaking in the future. Another possible improvement is to have more unique audience emotions feedback. Currently there is only implementation of audience facial feedback. Some participants suggested adding real faces to the audience. This is very helpful since the real faces may give different impressions to the user than the avatars. The participants also recommended adding more audience expressions and gestures such as sleepy, tired, bored, or looking at the phone.

5 Conclusion and Future Works

In this paper, we introduce a new public speaking simulator. In particular, we present two features, namely, speech analysis and audience attention feedback. We conduct a user study on this simulator with different versions. The participants highly praised the safe space that the simulator offers. They also gave positive feedback for the improvement of the simulator.

In the future, we plan to extend the simulator by adding more features such as body pose analysis, more natural audience gestures, and the ability to record and replay the public speaking practice sessions. For a more realistic approach, we plan to use sensors such as LIDAR and photogrammetry to scan real world environments and human faces. We will import the scanned models into the simulator for a better immersive experience. Last but not least, we aim to evaluate our simulator in various domains such as education, training, marketing or healthcare.

Acknowledgments

This research is funded by Vietnam National University HoChiMinh City (VNU-HCM) under grant number DS2020-42-01. This work is partially funded by National Science Foundation (NSF) under Grant 2025234. Bao Truong is supported by the UD Dean's Summer Fellowship.

References

[1] Heeren A, Ceschi G, Valentiner DP, Dethier V, Philippot P. Assessing public speaking fear with the short form of the Personal Report of Confidence as a Speaker scale: confirmatory factor analyses among a French-speaking community sample. Neuropsychiatr Dis Treat. 2013;9:609-618.

[2] Kerry Pellett and Syed Fawad Mustafa Zaidi. 2019. A Framework for Virtual Reality Training to Improve Public Speaking. In the 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19). Association for Computing Machinery, New York, NY, USA, Article 44, 1–2.

[3] Poeschl S (2017) Virtual Reality Training for Public Speaking—A QUEST-VR Framework Validation. Front. ICT 4:13.

[4] OvationVR, https://www.ovationvr.com/, last retrieved on June 15, 2022.

[5] VRChat, https://hello.vrchat.com /, last retrieved on June 15, 2022.

[6] M. El-Yamri, A. Romero-Hernandez, M. Gonzalez-Riojo and B. Manero, "Emotions-Responsive Audiences for VR Public Speaking Simulators Based on the Speakers' Voice," 2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT), 2019, pp. 349-353.

[7] Piyush, "Public Speaking Enhancement via Virtual Reality," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-4.

[8] M. El-Yamri, A. Romero-Hernandez, M. Gonzalez-Riojo, B. Manero (2019) ComunicArte: A Virtual Reality Game To Iimprove Public Speaking Skills, EDULEARN19 Proceedings, pp. 9061-9066.

[9] Hangyu Zhou, Yuichiro Fujimoto, Masayuki Kanbara, and Hirokazu Kato. 2020. Effectiveness of Virtual Reality Playback in Public Speaking Training. In the Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion). Association for Computing Machinery, New York, NY, USA, 462–466.

[10] 3D Scanner App, https://apps.apple.com/us/app/3d-scanner-app/id1419913995, last retrieved on June 15, 2022.

[11] Windows Speech library, https://docs.unity3d.com/ScriptReference/Windows.Speech.DictationRecognizer.html, last retrieved on June 15, 2022.

[12] Eye Contact in Public Speaking, https://mannerofspeaking.org/2009/09/04/eye-contact/, , last retrieved on June 15, 2022.

[13] Unity3D, https://docs.unity3d.com/Manual/index.html, last retrieved on June 15, 2022.

[14] Blender software, https://blender.org, last retrieved on June 15, 2022.