Effectiveness of Virtual Reality Playback in Public Speaking Training

Hangyu Zhou
Nara Institute of Science and Technology
Nara, Japan
zhou.hangyu.zh1@is.naist.jp

Yuichiro Fujimoto
Nara Institute of Science and Technology
Nara, Japan
yfujimoto@is.naist.jp

Masayuki Kanbara
Nara Institute of Science and Technology
Nara, Japan
kanbara@is.naist.jp

Hirokazu Kato
Nara Institute of Science and Technology
Nara, Japan
kato@is.naist.jp

ABSTRACT

In this paper, factors with positive effects in the playback of virtual reality (VR) presentation in training are discussed. To date, the effectiveness of VR public speaking training in both anxiety reduction and skills improvement has been reported. Though the playback using videotape is an effective way in original public speaking training, very few researchers focused on the effectiveness and possibility of VR playback. In this research, A VR playback system for public speaking training is proposed, and a pilot experiment is carried out, so as to figure out the effects of the virtual agent, immersion and public speaking anxiety level in VR playback.

KEYWORDS

public speaking training, public speaking anxiety, virtual reality playback, videotape feedback, virtual agent

ACM Reference Format:
Hangyu Zhou, Yuichiro Fujimoto, Masayuki Kanbara, and Hirokazu Kato. 2020. Effectiveness of Virtual Reality Playback in Public Speaking Training. In *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3395035.3425220

1 INTRODUCTION

For a majority of people, public speaking may be a rare but most valued part of social communication, and good public speaking skills could be important for acquiring professional and academic recognition. However, public speaking anxiety, which is one of the most common subtypes of social phobia among the general population, could present a particular challenge to it. Given the high level of fear associated with public speaking, many universities have offered public speaking courses, which effectively relieve anxiety and improve public speaking skills [3].

ICMI '20 Companion, October 25–29, 2020, Virtual event, Netherlands

ACM ISBN 978-1-4503-8002-7/20/10…$15.00
https://doi.org/10.1145/3395035.3425220

Accompanied by the excellent performance of VR in exposure therapy for anxiety disorders, a considerable amount of research was conducted on its application in public speaking training in the past few years. However, few works investigated the potentially effective factors.

Feedback is considered as an indispensable element in both communication research and communication systems since it keeps us informed of progress necessary for subsequent adjustments [17]. The feedback usually provides or stimulates the evaluation to improve the outcome of future training. Feedback in public speaking training can be classified into immediate and delayed feedback based on timing. The existing research indicates that immediate feedback helps train automatic and nonconscious behavior, such as hand gestures and filled pauses. While delayed feedback aids in the executive processes requiring conscious and effortful deliberation and preplanning, such as the connection between slides [11].

Since various kinds of virtual information are very likely to be added to the VR system, a number of publications have been devoted to studying the methods and effect of real-time feedback in VR public speaking training systems. However, the playback of the whole presentation record in VR systems seems to be ignored though videotape feedback has been recognized as a classic and effective way to evoke self-reflection and widely used in public speaking courses.

Besides, individual differences could also influence VR public speaking training results. Individuals with public speaking anxiety are generally not confident and rate themselves as having performed worse than others. They think more negatively and focus more on themselves during public speaking. As a result, their true performances would become worse and they may facing the task more difficult and less enjoyable. However, it's them who could benefit more from training [5].

In this work, the following research questions will be addressed:

Effectiveness of VR playback: *Could the avatar and immersion of VR in the record guide more appropriate self-reflection and help with global improvement of public speaking?*

Individual differences: *Will these changes caused by the avatar and immersion of VR be different among the people with different public speaking anxiety levels?*

To tackle these two questions, we first develop a public speaking system with VR playback. A VR agent is used as the avatar of the presenter, and some performances related to speeches such as

types of speech delivery and body movements are caught during the presentation. After presentation, the presenter can watch the demonstration given by his/her avatar as an audience using a head mounted display (HMD). Section 3 describes the system we proposed in detail. A pilot experiment will be conducted to check this system in section 4. In section 5, we discuss the planned experiment and other future works. The paper is summarized in section 6.

2 RELATED WORK

2.1 Public Speaking Training System

Automated assessment and anxiety reduction are two research hotspot as well as advantage on public speaking training in VR. They both improve the performance while the former is more about external manifestation and the latter more about inner cognition.

Being evaluated by system could know the shortcomings without feeling judged [20]. Researchers tend to use multimodal cues in automated assessment. Batrinca Ligia et al. [1] identified voice, skeleton, and gaze as factors contributing to the assessment of a presentation, while the interrupted speech and excessive gestures showed no correlation. Chen Lei et al. [4] developed a scoring model using data collected by audio, video, and 3D motion capturing devices, and found that this model could effectively predict the presentation scores based on the voice (subdivided into 3 aspects: fluency, pronunciation, and prosody) and the head, body, and hand motions. Pfister Tomas and Peter Robinson [15] presented an algorithm for real-time inference of emotional states according to such non-verbal features of speech as pitch, loudness and voice quality.

A lot of efforts have been made to relieve anxiety and improve public speaking skills. Public speaking anxiety is commonly treated with cognitive behavioral therapy and exposure therapy. Recently, the VR technology has been confirmed effective in alleviating anxiety by creating a controllable environment containing stimuli [9]. The virtual audience has a great impact on the effect of the VR exposure therapy of public speaking anxiety. The more interested the audience is perceived, the higher the self-rating score and the lower the public speaking anxiety. Moreover, this trend is reinforced by a higher sense of co-presence [18]. In addition, Chollet Mathieu et al. [6] established that the virtual audience could better improve the overall public speaking performance than direct visual feedback.

Furthermore, the task difficulty also affects the presenter's reaction to the virtual audience. Poeschl Sandra [16] concluded that there were less cognitive resources left to notice in difficult tasks. However, the ability to concentrate was independent of the task difficulty. In addition, an increase of task difficulty is found to relate positively to the anxiety, but anxiety decreased more over time during the exposure.

These studies all discussed the effectiveness of the whole VR public speaking training system or the real-time feedback, but ignored the delayed feedback, especially the VR presentation playback.

2.2 Videotape Feedback

The videotaped feedback has proven to be a useful pedagogical tool in basic public speaking courses. The nonverbal and verbal elements of speaking performances can be recorded on videotapes for subsequent review and analysis. Videotape feedback improves skill acquisition, speech content, objective test performance and the recall of the actual speech [2], reduces anxiety and perceived social costs, and increases positive appraisals of performance [12]. What's more, the videotaped feedback can inspire self-reflection and adjustment which do not have accurate rules.

However, people with public speaking anxiety who could have improved a lot through the videotaped feedback usually have lower self-esteem and self-efficacy, and are more likely to avoid the videotaped feedback because they do not want to confirm their negative performance [10]. A similar result was reported in the real-time video image feedback [8]. Many presenters were more aware of aspects of their personal appearance than their performance or evaluation, which was very distracting. To solve this problem, we are thinking of showing the avatar in side of presenters themselves in the record, which may cover some unimportant or even annoying information to fasten people's attention on the playback and guide more appropriate self-reflection.

On the other hand, we speculate that the third-person point of view offered by videos improves self-efficacy although there is no direct evidence. As VR has a stronger immersion sense than screen images on the third-person point of view [13], we expect the playback in a virtual environment can improve self-efficacy. Therefore, the effect of immersion of VR on self-reflection needs to be clarified.

2.3 Public Speaking Anxiety

A series of differences were reported during and after training according to different level of anxiety. Actually, a more significant decrease in the anxiety is observed in individuals with a high level of public speaking anxiety after VR exposure therapy [5, 19], which may be because even the virtual situation can refresh their memories of fear of public speaking.

Individuals with a high level of public speaking anxiety are more likely to have a distorted picture of how they appear to other people, and global performance is more inclined to be misrepresented than local information over time. No such changes are found in individuals with a lower level of public speaking anxiety, and they consistently rate global items more positively than local items instead. Cody Meghan W. and Bethany A. [7] attributed it to the difference between positive and negative self-schemas. We also assume different aspects that people pay attention to during the playback as one possible reason. As individuals with a high level of public speaking anxiety expect to make a good impression on others, they may lay more stress on details in presentation, which lead to a gradually blurred memory of global information over time.

Therefore, we expect that high public speaking anxiety can enlarge the effectiveness of the virtual avatar, but does not change it to immersive effectiveness, as individual with a high level of public speaking anxiety would pay more attention to their global performance when watching the virtual avatar in the record.

3 METHOD

Based on above works and thoughts, following hypotheses are derived:

H1: *Using the virtual avatar in the VR presentation record could hide some unimportant or even disturbing local performances, thus*

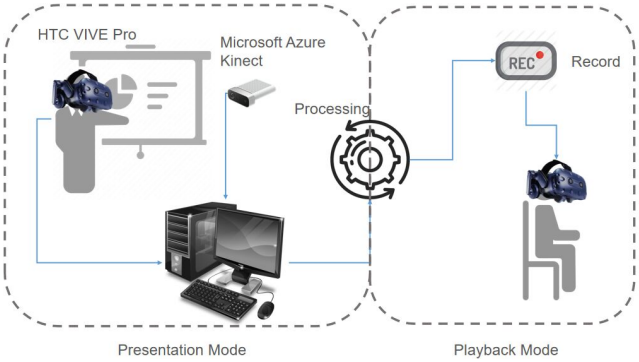


Figure 1: General architecture of our VR public speaking training system.

focusing the user's attention to global information and aspects that can enhance performance.

H2: *Users could better improve their public speaking skills by watching VR presentation records using the avatar.*

H3: *Watching records in VR enables users to immerse in it and substitute themselves into the audience's perspective, thereby reducing the gap between self-rating and other ratings.*

H4: *Users could get more confidence through VR immersion records.*

H5: *A stronger trend could be found among individuals with a high level of public speaking anxiety in H1 to H4.*

To test these hypotheses, we develop a public speaking training system and conduct an experimental study with it.

3.1 Overview of the System

The general architecture of our system is shown in Figure 1. It has two modes at use: the presentation mode and the playback mode. In the presentation mode, users could do simulation presentations in a virtual environment, and the system will detect user behavior, including speeches, body movement and eye contact.

After processing user's behavior, the system could create the presentation record with the virtual avatar. Then, users could watch the VR presentation record using HMD in the playback mode.

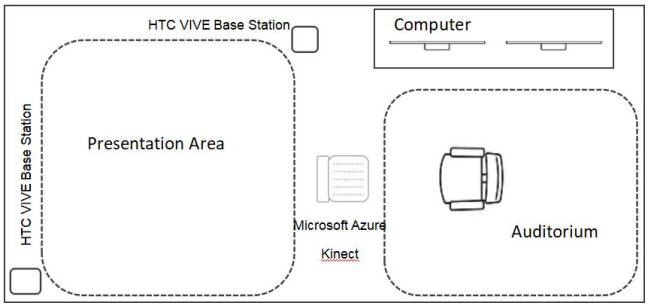
3.2 Presentation Mode

Before the presentation, we set up the system environment (Figure 2a) and build the presentation room in the virtual environment (Figure 2b).

The HTC Vive Pro VR headset is used to show the VR environment and record the head position and eye gaze. The base station tracks the position of HMD. One controller of Vive controls the slides. We use the body tracking SDK developed by Microsoft Azure Kinect to record 3D position information of presenter's main joints. An external microphone records sound. During presentation, the presenter should wear the headset and the microphone.

The presenter could practice in this virtual meeting room by using VR HMD, or they could watch the VR presentation record with HMD and practice in real environments without HMD.

There are three virtual audiences placed in the virtual meeting room since this population can cause more social anxiety [14].



(a) real room

(b) Virtual meeting room

Figure 2: The setup of the system environment.

The presenter can take place of one audience to attend his/her presentation during the playback.

3.3 Playback Mode

Considering both public speaking skills and equipment, we mainly capture following features:

- Stance: According to the height of real presenter, the virtual avatar was scaled. Then the main joins of avatar was placed by the position of presenter's joints so the avatar could have the same stance as real presenter. Though the movable range during presentation is limited, the stance is a significant sign showing the general mental state. For example, standing straight could show the confidence and respect, but too straight would become stiffness and tension and too relaxing may be neglected.
- Hand Gestures: Hand gestures can be a supplement of language, and guide audiences' gaze. However, due to the limit of the tracking accuracy of Kinect Azure, only two status, namely, opening the palm and making a fist, can be recorded. We hope the presenter pays more attention to global performance for self-reflection at least in the playback mode. Besides, more precise gestures related to behavior require additional equipment.
- Eye Contact: Positive eye contact is conducive to building rapport with audiences and keeping them engaged with the presentation. The eye gaze is tracked by HTC Vive Pro. The eye movement can hardly be seen in VR because of the low resolution, so we use a laser to show it. Users can choose to

(a) video (b) HMD-avatar

Figure 3: Video and VR recording.

display or remove the eye gaze laser since it is a bit distracting even though the light is thin.

- Vocal Performance: The vocal performance includes strain, roughness, pitch, loudness and pronunciation, which can impact a listener's perception of the speech. It is an obscure way to convey emotions as well. The vocal performance is recorded by the external microphone without other processing.

The following information is not included in the VR presentation record:

- Personal Face Mode: The personal face should not be taken as a factor of evaluating the presentation since it sometimes even interferes with the training. The use of the VR avatar can hide the true face of presenter in the record.
- Facial Expression: Facial expression is also a way to convey emotions of the speaker, but it is more related to nonconscious behavior, and same facial expression on the virtual avatar might also be distracting.
- Dress: Formal dressing itself can create a serious atmosphere, but paying attention to dress at every time training is unnecessary.

4 PILOT EXPERIMENT

We invited four participants of different native languages and different gender into a pilot experiment to check and get feedback of the system. One watched the video image record (Figure 3a). One watched the VR presentation record with the avatar in it on the monitor and two experienced the VR presentation record using HMD (Figure 3b).

Each participant did a presentation at first, and only the one who watched the video image record did not wear HMD. After the first presentation, they watched the record in their condition, then filled NASA-TLX about the playback task and did self-assessment.

The trend could not be revealed by a test with such a small sample size, but we could still get some information from the questionnaire. A consistency was observed in NASA-TLX, showing that our system required neither additional mental nor physical demand. Two VR playback participants had a stronger motivation to replay the record than the videotape participant, which might be attributed to the novelty of our system. However, the participant watching the VR playback using HMD was slightly more frustrated than others, which was possibly ascribed to a high public speaking anxiety level of this participant whose self-assessment score was lower.

According to the self-assessment results, the participant watching the monitor VR pointed out a lot of deficiencies during the playback, such as "too many uuuuhs", "slides move before me" and "contradicted myself". He was more concerned with his voice, and he mentioned he gesticulated a lot anyway. The participant with the immersive VR record condition also reported similar strong attention to his voice, and he commented that "I considered a lot about would my vocal performance be accepted by others, but neglected the presentation contents". The participant watching the video image did not mention about her voice in self-assessment. It seems that when we hide the appearance, people may focus more on their vocal performance.

After the pilot experiment, we got some comments from the participants. A participant reported the jump of hand tracking with the Kinect, and recommended us to improve that. Another participant said he concerned about hand movement, especially the unconscious use and sketchy gestures. He held that the detail of unnatural gestures could be ignored easily.

The participant who watched the video image record with the avatar was more concerned with the posture. She commented that she preferred video images to the virtual avatar, since the latter might hide many posture details that she wanted to change. However, other participants did not show the willingness to watch the video image after experiment, so we cannot have the conclusion of their preference.

5 FUTURE WORK

The pilot experiment suggests that the system can be ameliorated by improving hand gestures. We may try the basic motion capture jacket or leap motion in the future. The system shows an optimistic performance anyway.

Moreover, we plan to do an experiment to test our hypotheses, and there would be three groups of participants to experience each kind of the playback in our user experiment. We use the original videotape feedback as the control group, and the VR playback as the experimental group. To discuss the effectiveness of the avatar and immersion separately, the VR playback shown on monitor is added as another experimental group.

The experiment we are planning to do also has many limitations, and we may need a further research to clarify our hypotheses on the VR playback. Considering that the VR playback is a way of delayed feedback, we intend to research on public speaking training using real-time feedback subsequently.

6 CONCUSSION

In this paper, we focus on the playback of VR presentation in training public speaking skills and the possible factors leading to success. The attention-concentrating effect of the virtual avatar, immersion of VR and the public speaking anxiety level are discussed, and hypotheses are put forward. In addition, we develop a VR public speaking training system and do a pilot experiment. Despite of the comment about poor hand gesture tracking effect, the pilot experiment shows the availability of the proposed system. In the next step, we will do the user experiment to test our hypotheses.

7 ACKNOWLEDGEMENTS

This work was supported by JST CREST Grant Number JPMJCR19A, Japan.

REFERENCES

[1] Ligia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2013. Cicero-towards a multimodal virtual audience platform for public speaking training. In *International workshop on intelligent virtual agents*. Springer, Berlin, 116–128.

[2] John Bourhis and Mike Allen. 1998. The role of videotaped feedback in the instruction of public speaking: A quantitative synthesis of published empirical research. *Communication Research Reports* 15, 3 (Nov. 1998), 256–261.

[3] Robert E Carlson and Deborah Smith-Howell. 1995. Classroom public speaking assessment: Reliability and validity of selected evaluation instruments. *Communication education* 44, 2 (Apr. 1995), 87–97.

[4] Lei Chen, Gary Feng, Jilliam Joe, Chee Wee Leong, Christopher Kitchen, and Chong Min Lee. 2014. Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16th International Conference on Multimodal Interaction*. Turkey, 200–203.

[5] Mathieu Chollet, Pranav Ghate, Catherine Neubauer, and Stefan Scherer. 2018. Influence of individual differences when training public speaking with virtual audiences. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 1–7.

[6] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, Ari Shapiro, and Stefan Scherer. 2015. Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1143–1154.

[7] Meghan W Cody and Bethany A Teachman. 2011. Global and local evaluations of public speaking performance in social anxiety. *Behavior Therapy* 42, 4 (Apr. 2011), 601–611.

[8] Fiona Dermody and Alistair Sutherland. 2018. Recommendations from a study of a multimodal positive computing system for public speaking. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Springer, 81–92.

[9] Sandra R Harris, Robert L Kemmerling, and Max M North. 2002. Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & behavior* 5, 6 (Jul. 2002), 543–550.

[10] Katherine A Karl and Jerry M Kopf. 1994. Will individuals who need to improve their performance the most, volunteer to receive videotaped feedback? *The Journal of Business Communication (1973)* 31, 3 (Jul. 1994), 213–223.

[11] Paul E King, Melissa J Young, and Ralph R Behnke. 2000. Public speaking performance improvement as a function of information processing in immediate and delayed feedback interventions. *Communication Education* 49, 4 (May. 2000), 365–374.

[12] Judith M Laposa and Neil A Rector. 2014. Effects of videotaped feedback in group cognitive behavioral therapy for social anxiety disorder. *International Journal of Cognitive Therapy* 7, 4 (Dec. 2014), 360–372.

[13] Diego Monteiro, Hai-Ning Liang, Wenge Xu, Marvin Brucker, Vijayakumar Nanjappan, and Yong Yue. 2018. Evaluating enjoyment, presence, and emulator sickness in VR games based on first-and third-person viewing perspectives. *Computer Animation and Virtual Worlds* 29, 3-4 (May. 2018), e1830.

[14] Fariba Mostajeran, Melik Berk Balci, Frank Steinicke, Simone Kühn, and Jürgen Gallinat. 2020. The effects of virtual audience size on social anxiety during public speaking. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 303–312.

[15] Tomas Pfister and Peter Robinson. 2010. Speech emotion classification and public speaking skill assessment. In *International Workshop on Human Behavior Understanding*. Springer, 151–162.

[16] Sandra Poeschl. 2017. Virtual reality training for public speaking—a QUEST-VR framework validation. *Frontiers in ICT* 4 (Jun. 2017), 13.

[17] Arkalgud Ramaprasad. 1983. On the definition of feedback. *Behavioral science* 28, 1 (Jan. 1983), 4–13.

[18] Mel Slater, D-P Pertaub, and Anthony Steed. 1999. Public speaking in virtual reality: Facing an audience of avatars. *IEEE Computer Graphics and Applications* 19, 2 (Mar/Apr. 1999), 6–9.

[19] Snežana Stupar-Rutenfrans, Loes EH Ketelaars, and Marnix S van Gisbergen. 2017. Beat the fear of public speaking: Mobile 360 video virtual reality exposure training in home environment reduces public speaking anxiety. *Cyberpsychology, Behavior, and Social Networking* 20, 10 (Oct. 2017), 624–633.

[20] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S Shyam Sundar. 2020. Alexa as coach: Leveraging smart speakers to build social agents that reduce public speaking anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.