SPEAKinVR: validation of a virtual audience

ETIENNE Elodie
HEC Liège
University of Liège
Liège, Belgium
Elodie.Etienne@uliege.be

LECLERCQ Anne-Lise
Département de Logopédie
University of Liège
Liège, Belgium
AL.Leclercq@uliege.be

REMACLE Angélique
Département de Logopédie
University of Liège
Liège, Belgium
angelique.remacle@uliege.be

PETERS Florian
HEC Liège
University of Liège
Liège, Belgium
fpeters@uliege.be

SCHYNS Michaël
HEC Liège
University of Liège
Liège, Belgium
M.Schyns@uliege.be

Keywords— Virtual reality, public speaking, soft skills, training

I. Introduction and context

Nowadays public speaking is one of the most important skills a person should develop. Such a skill is vital in many circumstances and in very different fields: the sales representative who presents a product to customers, the tourist guide visiting a city with a group, the manager who defends his project in front of stakeholders, the candidate during a job interview, the professor in front of students… Unfortunately, many firms complain about the too low level of this skill within their staff. The global theme of our work is to look for innovative solutions leading to an improvement of public speaking performances.

Social anxiety, i.e. the fear of social situations that involve interaction with other people (Wiederhold & Bouchard, 2014) is one of the main reasons for these poor performances. In most individuals, public-speaking anxiety can be controlled. Repeated training in front of an audience can help to better control the speaker's emotions and skills (Wallach, Safir & Bar-Zvi, 2009; Wiederhold & Bouchard, 2014). Even in the case of non-pathological anxiety, many techniques practiced with speech therapists and/or coaches can greatly improve the speaking performances. Although training in front of a real audience can be logistically difficult to organize, virtual reality (VR) can be the solution and is the framework we will consider. Learning by doing has many benefits and it is well known that VR has a huge potential in this area. A virtual environment with an interactive virtual audience allows to practice in the same situation you will be confronted with in your real life. You can practice where, when and as often as you need. Moreover, the training can be progressive and fully controlled; which leads to much more efficient learning processes and to a faster progression.

The overall research project SPEAKinVR is divided in two parts. The first part, developed in this paper, focuses on investigating how people perceive an audience delivered by a VR technology and on selecting representative attitudes. These results will be used in the second part of the project, which is dedicated to public speaking training.

We are not the first to consider VR in this context; e.g. Chollet et al. showed that VR can indeed improve speaking performances, in particular when the audience is interactive (Chollet, Wortwein, Morency, Shapiro & Scherer, 2015; Goberman et al., 2011). A virtual environment is highly configurable: number of people, the sex ratio, the overall behaviour (positive, neutral or negative), the level of interactivity with the public and the quality of the graphics. The question is: "what are the right parameters to use for the audience?". In this paper the focus is on these parameters. The quality and the validity of such environments are key elements which require an in depth analysis before being able to deal with the more general question on public speaking training. The validation of a virtual audience is the main goal of this paper and our expected contribution.

II. Research questions

The main goal of this paper is to validate a virtual reality environment for public speaking training. Following Slater's terminology (2003), there are two important concepts in VR: "*immersion which stands for what the technology delivers from an objective point of view"* and *presence* which is *"the human reaction to immersion", i.e.,* the participant's subjective sense of being in the virtual place.

The reactions of the audience can have a significant impact on the speaker's emotions and performance. At a first level, our hypothesis is that interactivity has a positive impact on the presence feeling. At a higher level, as already shown by Chollet et al. (2015), interactivity in VR is also a major ingredient in the training process. It is therefore essential to know if the users perceive the interactions in the virtual environment as representative of the reality and how each one is interpreted. There are two main dimensions in the context of emotion and affect: arousal and valence. As defined by Chollet and Scherer (2017), *"arousal can be understood as an audience member's level of alertness, and valence corresponds to how positively or negatively the person feels toward the speaker or the presentation"*. In their paper, they tried to understand how users perceive virtual audience based on the nonverbal behavior of audience members. Our first question is to investigate which attitudes the characters must display and how people perceive the individual members of the audience in terms of their states of arousal and valence.

A second related question is linked to the level of reality used to represent the public. The characters in virtual environments, i.e. avatars, are most often synthetic images. In some cases, photorealistic representations are used but the level of animation is then generally extremely limited. In this context, our second research question investigates whether the use of fully rigged 3D photogrammetric models, i.e. with a skeleton we can animate, can significantly improve the user's *presence*.

III. ENVIRONMENTS AND METHODOLOGY

We developed several environments in our lab for this research. The development platform is Unity 3D. The virtual audience can be represented by 3D animated drawings or by 3D photogrammetric models. The photogrammetric models are built from photos of real people by our 3D artists. They are then rigged and animated by our 3D animators. The same animations are available for both drawings and photogrammetric models. Each avatar is able to change its posture (forward, backward, neutral), head orientation, headshake, facial expression (lips and smile, eyebrows and squinting eyes…), and eye-gaze. The combination of all these features allow characters to display many attitudes. We defined different poses, facial postures, and head movements representing more than 180 possibilities. Furthermore, the virtual characters are able to do some activities such as taking notes (handwriting notes or computer notes) and talking to each other.

In order to validate the audience, we are creating a sample of animated clips corresponding to a certain amount of combinations of the behaviour parameters for a male or female character. We are using either drawings or 3D photogrammetric models. Among the 180 possible combinations, only one part will be tested in the clips since some combinations may not reflect attitudes that real people might display.

While the full environment is built for high-end full VR headsets, all the clips can also be experienced in VR through Google Cardboard basic devices or anaglyph glasses. A cardboard is a simple box made of paper with two cheap plastic lenses. Such a device is able to transform most smartphones into basic VR headsets. Anaglyph glasses are paper glasses with improved red-cyan filters. Nowadays smartphones are powerful enough to watch 3D stereoscopic clips as the ones we have created.

The use of cardboards or anaglyph glasses has huge advantages. First, they both cost nearly nothing with respect to the use of high-end VR headsets. We can therefore conduct surveys with hundreds of people at a time. It allows us to consider larger samples and increase the reliability of our results. Second, their format allows for easy delivery to participants in case of online survey. Second, by contrast to Chollet et al. (2017) who performed their survey using traditional 2D screen, the responders are in our case immersed in VR, i.e. in conditions close to the ones they will encounter in the final VR environment. Our results will be directly transferable to the final context.

About four hundred people are contacted to participate in the study. This number is determined on the basis of a power analysis for statistical tests. A first online survey with about 200 participants will be conducted in the coming months and a second survey with about 200 participants will be organised in September 2020. Based on the clips they are watching through a cardboard or through anaglyph glasses, they are asked to identify individual member states. They should assess the degree of arousal (seven-point Likert scale from very low to very high) and valence (seven-point Likert scale from very negative to very positive). We are also asking the participants their level of certainty for each of their answers.

Our final goal is to create a full VR environment where a speaker will train his/her speaking skills in front of a realistic and challenging audience. A statistical analysis will therefore be conducted on the survey results. This will allow us to measure how participants feel towards the model and whether that feeling depends on the gender of the model or the method used (drawings or photogrammetric). Furthermore, it will allow us to select the most pertinent attitudes and 3D models, and to define which animations should be used to represent faithfully a set of audience reactions for different degrees of arousal and valence.

IV. CONCLUSIONS

The objective of this first step is to validate the virtual audience. Unity 3D engine was used to create a first basic version of the tool. 3D avatars have been designed and animated to represent some common audience postures, corresponding to different degrees of arousal and valence, and some classical situations as characters typing on laptops or playing with smartphones. 3D photogrammetric versions of these avatars have been completed for testing increased realism. A methodology based on a survey was developed. This is still work in progress. Pre-tests on the VR environment were made and the preliminary results are encouraging. The first results of the survey are expected in June 2020 and the full results in November 2020. By the end of the year we should have developed our paper for publication.

V. FUTURE WORK

Based on the results of the present study related to the first part of the project, we will be able to address the question of public speaking training. For the second part of the project, we have already created different virtual rooms where participants will be invited to hold a presentation in front of a virtual audience. We have also started to work on automatic methods based on statistical, machine learning and natural language processing methods to implement real-time biofeedback of the audience to the speaker's presentation.

ACKNOWLEDGEMENTS

We are grateful to three anonymous referees for helpful suggestions and comments.

REFERENCES

[1] Chollet, M., & Scherer, S. (2017). Perception of virtual audiences. *IEEE computer graphics and applications*, *37*(4), 50-59.

[2] Chollet, M., Wörtwein, T., Morency, L. P., Shapiro, A., & Scherer, S. (2015, September). 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* (pp. 1143-1154).

[3] Fukuda, M., Huang, H. H., Ohta, N., & Kuwabara, K. (2017, October). Proposal of a parameterized atmosphere generation model in a virtual classroom. In *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 11-16).

[4] Glémarec, Y., Bosser, A. G., Buche, C., Lugrin, J. L., Landeck, M., Latoschik, M. E., & Chollet, M. (2019, November). A Scalability Benchmark for a Virtual Audience Perception Model in Virtual Reality. In *25th ACM Symposium on Virtual Reality Software and Technology* (pp. 1-1).

[5] Latoschik, M. E., Lugrin, J. L., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016, November). Breaking bad behavior: Immersive training of class room management. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (pp. 317-318).

[6] Mostajeran, F., Balci, M. B., Steinicke, F., Kühn, S., & Gallinat, J. (2020). The Effects of Virtual Audience Size on Social Anxiety during Public Speaking. In *Proceedings of IEEE Virtual Reality (VR)*, (pp. 303-312).

[7] Kang, N., Brinkman, W. P., van Riemsdijk, M. B., & Neerincx, M. (2016). The design of virtual audiences: noticeable and recognizable behavioral styles. *Computers in Human Behavior*, *55*, 680-694.

[8] Pertaub, D. P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators & Virtual Environments*, *11*(1), 68-78.

[9] Slater, M. (2003). A note on presence terminology. *Presence connect*, *3*(3), 1-5.

[10] Wallach, H. S., Safir, M. P., & Bar-Zvi, M. (2009). Virtual reality cognitive behavior therapy for public speaking anxiety: a randomized clinical trial. *Behavior modification*, *33*(3), 314-338.

[11] Wiederhold, B. K., & Bouchard, S. (2014). *Advances in virtual reality and anxiety disorders*. Springer.