7th International XR Conference: extended abstract Validation of a virtual framework for public speaking training

Elodie Etienne[a]
Anne-Lise Leclercq[bcd]
Angélique Remacle[b]
Michaël Schyns[a]

[a]QuantOM , HEC Liège, University of Liège
[b]Département de Logopédie, Université de Liège
[c]Unité de Recherche Enfances
[d]Clinique Psychologique et Logopédique de l'Université de Liège

Keywords— Virtual reality, public speaking, soft skills, training, speech performance, avatars, photo-realism

To cite this work: Etienne, E. *et al.* (2022, April 27-29). Validation of a virtual framework for public speaking training *[Paper presentation]. 7th International XR Conference, Lisbon, Portugal.*

I. Introduction

Speaking out in public or during meetings is a challenge for many of us. It is nonetheless a competency required in many activities: 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. To fight public speaking anxiety, which may impede oral presentation performances, we need to prepare and train the speakers. It can be done efficiently by immersing the speaker in totally controlled Virtual Reality (VR) environments. This paper first aims to validate the emotional valence and level of arousal of the attitudes of a virtual public through an experiment with 125 participants. Based on these results, a library of public attitudes corresponding to different levels of arousal and valence has been created. Our experiment also investigates the benefits of using low-end and high-end VR headsets, as well as the use of photo-realistic public compared to sketched avatars. In future work, we will use our library during the training of the speakers to confront them in various situations in response to their speech.

II. Literature review

Public speaking anxiety, the fear of giving a speech or presentation in public because of the expectation of being negatively evaluated or humiliated by others (Vandenbos, 2007), is one of the most common fear (Pollard and Henderson, 1988) and the main reasons for people's poor performance in public speaking (Rothwell, 2010 and Menzel, 1994). 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 speaking skills (Rust et al. 2020, Tsang 2020, Howe and Cionea 2021).

Learning by doing has many benefits and it is well known that virtual reality and augmented reality has a huge potential in this area as shown by Franceschi et al. (2009), Howard et al. (2021), and Eckert and Mower (2020) for VR, by Abraham and Annuziata (2017) for AR, and by Stephen et al. (2019), for both VR and AR.

In the context of public speaking, Harris et al. (2002), Kahlon et al. (2019), and Farell et al. (2018), among others, have shown the benefits of VR training.

A virtual environment with an interactive virtual audience, i.e. an audience providing non-verbal behaviour as feedback (Chollet and Scherer, 2017 and Kang et al. 2016), allows speakers to practice in a situation very similar to that they will be confronted to in real life. People can practice where, when, and as often as they need to increase their sense of competence and be ready to hold a speech (Scheveneels et al. 2019, and Craske et al., 2014). Moreover, the training can be progressive and controlled in terms of the number of audience members, their attitudes, etc. which leads to much more efficient learning processes and a faster progression (Eckert and Mower, 2020).

In this context, Chollet and Scherer (2017) showed that VR training can improve speaking performances, when the audience is interactive (Wörtwein et al. 2015 and Goberman et al. 2011). The reactions of the audience can have a significant impact on the speaker's emotions and performance. As already shown by Chollet et al. (2015), interactivity in VR is 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 reality and how each one is interpreted. Two dimensions in the context of emotion and affect are especially important: the level of arousal and the emotional valence. As defined by Chollet et al. (2015), *arousal* can be understood as an audience member's level of alertness, and *valence* corresponds to how positively or negatively the audience member feels toward the speaker or the presentation. The authors tried to understand how users perceive virtual audiences based on the nonverbal behaviour of their members, also named avatars in this digital context. They explored which behaviours are relevant to be perceived by the speaker as expressing high or low arousal, and positive or negative attitude in terms of arousal and valence.

The main goal of this paper is to validate a VR environment for public speaking training. Following Slater's terminology (2003), we refer to two important concepts in VR: immersion and presence. *Immersion* stands for what the technology delivers from an objective point of view or the objective level of sensory fidelity a VR system provides. *Presence* is defined as the human reaction to immersion, i.e., the participant's subjective sense of being in the virtual place. It is the observer's sense of psychologically leaving their real location and feeling as if transported to a virtual environment (Weech et al. 2019). Presence can affect the user's performance (Nash et al. 2000).

Our first question is to investigate which attitudes the avatars must display and how people perceive the individual members of the audience in terms of their states of arousal and valence.

Our 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, photo-realistic 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 photo-realistic models, i.e., with a skeleton we can animate, can significantly improve participant's perception of avatar's arousal and valence or their confidence levels interpretation of the audience. The desire of this comparison lies in the ease of making photo-realistic avatars. In our context, we can even think about a virtual environment with virtual avatars who really look like the people who are supposed to be present in the real context.

Our last question is related to a more in-depth analysis of the quality of immersion and the feeling of presence. We want to investigate whether the use of high-end or low-end headsets has an impact on the quality of immersion and the feeling of presence. Due to its hardware limitations, the low-end headset is not expected to perform better in any way than the high-end headset and the latest is often the preferred choice in terms of quality of immersion. One might ask why to compare the two types of headsets. As explained in Amin et al. (2016), who compared the immersion between cardboard and Oculus Rift headset, the answer lies in cost, ease of operations, and portability. A high-end headset is expensive, requires professional technical operations, and is less mobile compared to cardboard (with its affordability (only a smartphone is needed in addition to the cardboard) and its ease of use). This can make VR inaccessible for everyone or for mass diffusion (a hundred people at the same time). If we can show that our results are similar in terms of the evaluation of valence and arousal with both the low-end and high-end headsets, we can overcome the affordability and the mass diffusion problems, and this, for a lower cost.

III. 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 sketched avatars or by 3D photo-realistic models. The photo-realistic 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 sketched and photo-realistic models. Each avatar can change its posture (forward, backward, neutral), facial expression (smiling, frowning, eyebrows raised, neutral), and head movements (nod, shake, question, neutral). The combination of all these features allows characters to display many attitudes. We defined different poses, facial postures, and head movements representing more than 180 possibilities.

About 125 participants were contacted to take part in the study. This number was determined based on a power analysis for statistical tests. After they had been given the definitions of valence and arousal, the participants were asked to wear the headset (written at the beginning of the online questionnaire). The VR application contains a list of animations to be watched. The 20 animated sequences were presented in an arbitrary order. Each participant had to watch each animation for at least 15 seconds (the sequence repeated endlessly at the end of the 15 seconds until the participant moved on to the next one). This ensured that each animation was played fully before switching to the next one. Moreover, the sequence restarted as often as they needed. For each of the 20 animations, they answered (with their computer or smartphone) an online questionnaire set up for this experiment. The participants had to remove the headset after each sequence in order to answer the questionnaire. They were asked to assess the degree of arousal (7-point Likert scale from very low to very high) and valence (seven-point Likert scale from very negative to very positive) as well as their confidence level for each of their answers. They could also write down open-ended comments about the video they watched. After watching all the sequences (post-immersion), participants answered (in the online questionnaire) some questions about the feeling experienced in the VR environment.

The questionnaires was the Gatineau Presence Questionnaire (GPQ) (Laforest et al. 2016). We used the French-Canadian version of the GPQ, which is a four-item questionnaire rated in percentage scale. As explained by Laforest et al. (2016), the GPQ consists of four items: 1) the presence of being there, 2) appraising the experience as being real, 3) the awareness of the virtual environment as being artificial, and 4) the feeling of being in the physical office instead of the virtual environment". We added a last question about the negative side effects of the immersion (i.e., cybersickness). At the end of the experiment, participants completed the online questionnaire on socio-demographic data and reported any recent events that could alter their perception

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

IV. FINDINGS

Regarding our first question about the attitudes the avatars must display and how participants perceive them, our study shows for each parameter (posture and hands, facial expression, and head movements) the associated level of valence and arousal. Our results are coherent with Chollet's work (2017) and we confirm their 2D findings in a 3D VR setting. Furthermore, among the combinations of behaviour selected, we now have a library of avatar attitudes associated with some levels of valence and arousal. For example, for the emotional valence, an avatar with an upright posture, the hands on top of each other, a neutral facial expression, and who is nodding his head expresses the highest level of valence. At the opposite, an avatar with a backward posture, the hands behind the head, frowning eyes, and who is shaking his head expresses the lowest level of valence. For the level of arousal, an avatar with a backward posture, with elbows on the table, a smiling face, and who is shaking his head expresses the highest level of arousal. On the contrary, an avatar with a forward posture, the hands together, a neutral face and who is nodding the head, expresses the lowest level of arousal. Additional results were obtained and will be discussed in the full paper. We observe that some gestures dominate other ones. For example, each time an avatar shakes his head, it is perceived as a negative emotional valence and high level of arousal, regardless of the posture or facial expression associated. Similarly, when the avatar nods his head, it is mostly perceived as a positive emotional valence. Sometimes, it is perceived as neutral depending on the other parameters (posture and facial expression), but never negative as expected as well. We also observe some links between valence and arousal. For example, if the emotional valence of the avatar is perceived as positive or negative, the associated level of arousal is never low. It is not surprising because if the avatar seems to agree or disagree with the participant, he should appear awake. Thanks to these results, we know which sequences to choose to show a specific level of arousal and emotional valence.

Considering our second related question linked to the level of reality used to represent the public, there does not seem to be any significant difference between the assessment of the emotional valence (respectively level of arousal) for sketched avatars and photo-realistic avatars. It means that the attitudes of the avatars are perceived in the same way whatever the quality of graphics used to represent them. However, it seems that the confidence level is improved when assessing the level of valence and arousal for photo-realistic avatars. Indeed, the assessment of the confidence level when assessing the emotional valence or the level of arousal shows fewer negative answers (corresponding to 1^{st}, 2^{nd}, and 3^{rd} levels of the Likert scale) and more totally positive answers (corresponding to the 7^{th} level of the Likert scale). In conclusion, there is a positive impact of using photo-realistic avatars. Indeed, while keeping the participants' judgment of valence and arousal unchanged, the use of photo-realistic avatars improves the confidence level in their judgment. Based on the participant's comments, we hypothesise that the use of photo-realistic avatars makes it easier for the participant to enter the avatar's attitudes. This seems coherent with Seymour et al.'s results (2021) about trustworthiness and affinity with human-realistic avatars.

Regarding our last question about the quality of immersion and the feeling of presence, there does not seem to be any significant difference between the assessment of the emotional valence using both headsets. It means that the emotional valence is perceived in the same way whatever the quality of the headset used. However, for the level of arousal, it seems that arousal is evaluated higher using the high-end headset. Indeed, we can see that the percentages of negative answers (corresponding to 1^{st}, 2^{nd}, and 3^{rd} levels of the Likert scale) are higher for the low-end headset than for the high-end headset. It means that the level of arousal is perceived higher when using the high-end headset. We hypothesise that it is due to the more precise details (especially for the avatar's eyes) that can be seen through the high-end headset. As expected, the major finding for this question is about the quality of immersion. Except for the level of artificiality, the quality of immersion statistically increases when we compare these parameters (higher feeling of presence, higher level of realism and lower spatial awareness (meaning that participants forgot more that they are in the actual room)) for participants who wore the low-end headset compared to participants who wore the high-end headset.

V. CONCLUSIONS

This research is the first step of a long-term project. The final goal is to create a full VR environment where an user will train his/her speaking skills in front of a realistic, challenging, and interactive audience. In this project, the focus is on speaking skills, but the VR environment created can have many other applications to develop different business skills. The first part, developed in this paper, aims to validate the emotional valence and level of arousal of the attitudes of a virtual public through an experiment with 125 participants. A library of public attitudes corresponding to different levels of arousal and valence has been created. Our experiment also shows the benefits of using low-end and high-end VR headsets, as well as the use of photo-realistic public compared to sketched avatars

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.

VI. References

Abraham, M., & Annunziata, M. (2017). Augmented reality is already improving worker performance. *Harvard Business Review, 13*, 1–5.

Amin, A., Gromala, D., Tong, X., & Shaw, C. (2016). Immersion in cardboard VR compared to a traditional head-mounted display. *International Conference on Virtual, Augmented and Mixed Reality*, (pp. 269–276).

Benbasat, I., & Wang, W. (2005). Trust in and adoption of online recommendation agents. *Journal of the association for information systems, 6*, 4.

Blöte, A. W., Kint, M. J., Miers, A. C., & Westenberg, P. M. (2009). The relation between public speaking anxiety and social anxiety: A review. *Journal of anxiety disorders, 23*, 305–313.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., . . . Roucaut, F.-X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: a three-arm randomised controlled trial. *The British Journal of Psychiatry, 210*, 276–283.

Cafaro, A., Vilhjálmsson, H. H., & Bickmore, T. (2016). First impressions in human–agent virtual encounters. *ACM Transactions on Computer-Human Interaction (TOCHI), 23*, 1–40.

Chollet, M., & Scherer, S. (2017). Perception of virtual audiences. *IEEE computer graphics and applications, 37*, 50–59.

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S. (2015). Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (pp. 1143–1154).

Collange, J., & Guegan, J. (2020). Using virtual reality to induce gratitude through virtual social interaction. *Computers in Human Behavior, 113*, 106473.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., & Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour research and therapy, 58*, 10–23.

Eckert, D., & Mower, A. (2020). The effectiveness of virtual reality soft skills training in the enterprise: a study.

Etemad-Sajadi, R. (2016). The impact of online real-time interactivity on patronage intention: The use of avatars. *Computers in human behavior, 61*, 227–232.

Farrell, W. A. (2018). Learning becomes doing: Applying augmented and virtual reality to improve performance. *Performance Improvement, 57*, 19–28.

Franceschi, K., Lee, R. M., Zanakis, S. H., & Hinds, D. (2009). Engaging group e-learning in virtual worlds. *Journal of Management Information Systems, 26*, 73–100.

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

Goberman, A. M., Hughes, S., & Haydock, T. (2011). Acoustic characteristics of public speaking: Anxiety and practice effects. *Speech communication, 53*, 867–876.

Harris, S. R., Kemmerling, R. L., & North, M. M. (2002). Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & behavior, 5*, 543–550.

Hibbeln, M. T., Jenkins, J. L., Schneider, C., Valacich, J., & Weinmann, M. (2017). How is your user feeling? Inferring emotion through human-computer interaction devices. *Mis Quarterly, 41*, 1–21.

Howard, M. C., Gutworth, M. B., & Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior, 121*, 106808.

Howe, W. T., & Cionea, I. A. (2021). Exploring the associations between debate participation, communication competence, communication apprehension, and argumentativeness with a global sample. *Argumentation and Advocacy, 57*, 103–122.

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality.* Morgan & Claypool.

Kahlon, S., Lindner, P., & Nordgreen, T. (2019). Virtual reality exposure therapy for adolescents with fear of public speaking: a non-randomized feasibility and pilot study. *Child and adolescent psychiatry and mental health, 13*, 47.

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.

Kleinsmith, A., Rivera-Gutierrez, D., Finney, G., Cendan, J., & Lok, B. (2015). Understanding empathy training with virtual patients. *Computers in human behavior, 52*, 151–158.

Krokos, E., Plaisant, C., & Varshney, A. (2019). Virtual memory palaces: immersion aids recall. *Virtual reality, 23*, 1–15.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., & Rosin, P. L. (2009). Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior, 33*, 1–15.

Laboratoire de Cyberpsychologie de l'UQO. (2006). Questionnaire de présence de l'UQO (QP-UQO).

Laforest, M., Bouchard, S., Crétu, A.-M., & Mesly, O. (2016). Inducing an anxiety response using a contaminated Virtual environment: validation of a therapeutic tool for obsessive–compulsive disorder. *Frontiers in ICT, 3*, 18.

Lindner, P., Miloff, A., Fagernäs, S., Andersen, J., Sigeman, M., Andersson, G., . . . Carlbring, P. (2019). Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: A randomized controlled trial. *Journal of anxiety disorders, 61*, 45–54.

Lugrin, J.-L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., & Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT, 3*, 26.

Menzel, K. E., & Carrell, L. J. (1994). The relationship between preparation and performance in public speaking. *Communication Education, 43*, 17–26.

Mostajeran, F., Balci, M. B., Steinicke, F., Kühn, S., & Gallinat, J. (2020). The effects of virtual audience size on social anxiety during public speaking. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, (pp. 303–312).

Nash, E. B., Edwards, G. W., Thompson, J. A., & Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of human-computer Interaction, 12*, 1–41.

Perkis, A., Timmerer, C., Baraković, S., Husić, J. B., Bech, S., Bosse, S., . . . others. (2020). QUALINET white paper on definitions of immersive media experience (IMEx). *arXiv preprint arXiv:2007.07032*.

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

Pollard, C. A., & Henderson, J. G. (1988). Four types of social phobia in a community sample. *Journal of Nervous and Mental Disease*.

Rothwell, J. D. (2010). *In the company of others: An introduction to communication.* Oxford University Press New York.

Rust, C., Gentry, W. M., & Ford, H. (2020). Assessment of the effect of communication skills training on communication apprehension in first year pharmacy students–A two-year study. *Currents in Pharmacy Teaching and Learning, 12*, 142–146.

Scheveneels, S., Boddez, Y., Van Daele, T., & Hermans, D. (2019). Virtually unexpected: No role for expectancy violation in virtual reality exposure for public speaking anxiety. *Frontiers in psychology, 10*, 2849.

Schwind, V., Knierim, P., Haas, N., & Henze, N. (2019). Using presence questionnaires in virtual reality. *Proceedings of the 2019 CHI conference on human factors in computing systems*, (pp. 1–12).

Seymour, M., Riemer, K., & Kay, J. (2018). Actors, avatars and agents: Potentials and implications of natural face technology for the creation of realistic visual presence. *Journal of the association for Information Systems, 19*, 4.

Seymour, M., Yuan, L. I., Dennis, A., Riemer, K., & others. (2021). Have We Crossed the Uncanny Valley? Understanding Affinity, Trustworthiness, and Preference for Realistic Digital Humans in Immersive Environments. *Journal of the Association for Information Systems, 22*, 9.

Slater, M. (2003). A note on presence terminology. *Presence connect, 3*, 1–5.

Slater, M., & Steed, A. (2000). A virtual presence counter. *Presence, 9*, 413–434.

Steffen, J. H., Gaskin, J. E., Meservy, T. O., Jenkins, J. L., & Wolman, I. (2019). Framework of affordances for virtual reality and augmented reality. *Journal of Management Information Systems, 36*, 683–729.

Tsang, A. (2020). The relationship between tertiary-level students' self-perceived presentation delivery and public speaking anxiety: A mixed-methods study. *Assessment & Evaluation in Higher Education, 45*, 1060–1072.

VandenBos, G. R. (2007). *APA dictionary of psychology.* American Psychological Association.

Weech, S., Kenny, S., & Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: a review. *Frontiers in psychology, 10*, 158.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*, 225–240.

Wörtwein, T., Chollet, M., Schauerte, B., Morency, L.-P., Stiefelhagen, R., & Scherer, S. (2015). Multimodal public speaking performance assessment. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, (pp. 43–50).