

Development of a Responsive Virtual Audience for Public Speaking Training in Virtual Reality

1st Sarah Saufnay
HEC Liège
University of Liège
Liège, Belgium
sarah.saufnay@uliege.be

2nd Michaël Schyns
HEC Liège
University of Liège
Liège, Belgium
m.schyns@uliege.be

Abstract—Public speaking is a common activity in many areas of daily and professional life. However, communication skills are often difficult to master, and opportunities for practice remain limited, as the presence of an audience is difficult to replicate. To address this need, Virtual Reality training systems have been developed. Once immersed in the environment, users can practice in realistic conditions and in front of an audience, albeit virtually. However, to further improve fidelity and training outcomes, audience responsiveness is essential. A framework for integrating Artificial Intelligence (AI) is therefore proposed. The underlying objective is to enrich the experience through AI-driven virtual agents that, over time, respond appropriately to the speaker's performance and emotional state. This framework is presented along with the main design choices of the application.

Index Terms—public speaking, virtual audience, artificial intelligence

I. INTRODUCTION

Thanks to Virtual Reality (VR) capabilities, business processes can be accurately recreated in secure and tailored environments to develop competencies. Recently, the relevance of VR for improving public speaking skills has been highlighted [10], as they are essential in management situations. However, such soft skills are often overlooked and acquired directly through on-the-job experience. Training opportunities also remain limited due to the difficulty of reproducing realistic presentation conditions. Thanks to VR, public speaking tasks can be simulated in controlled environments using virtual agents [11]. Users can repeatedly practice while immersed in front of virtual audiences, thereby approximating real-life settings. Integrating Artificial Intelligence (AI) into these applications could further enhance effectiveness by enriching interactions and increasing realism. The design of AI-driven virtual agents would then expose speakers not only to the presence of an audience but also to contextually appropriate reactions, thereby better preparing users for real-world speaking situations [2].

II. VIRTUAL AUDIENCE DESIGN

As the objective is to create intelligent virtual audiences displaying appropriate reactions, speaker performance must be tracked in real time during immersion. One of the most influential factors shaping performance is vocal delivery (*e.g.*, intensity, speech rate, fundamental frequency), which is also a strong predictor of anxiety [3]. Although nonverbal behavior

(*e.g.*, posture, gaze) and presentation content are also relevant, acoustic parameters alone enable a useful approximation of performance [3] and can additionally be used to detect the speaker's emotional state. Emotion expression also influences perceived speaker performance, as it may be valued in some contexts but considered inappropriate in others.

Designing plausible and appropriate virtual audience reactions is both a technical and conceptual challenge. The virtual agents' behaviors must indeed be carefully designed as well, given the variety of factors influencing how attitudes are perceived [6]. Several studies have examined how characteristics, such as agents' realism or gender, affect the perception of audience nonverbal behavior. Perceptual studies often rely on the dimensions of valence (*i.e.*, sympathy toward the speaker) and arousal (*i.e.*, level of activation) to model different audience attitudes and assess their perception [6]. Such cues directly influence the perceived appreciation of the presentation by the audience and, consequently, the stress elicited during the task. The proposed pipeline therefore relies on a valence–arousal coding system previously validated through perceptual studies to ensure coherence [4].

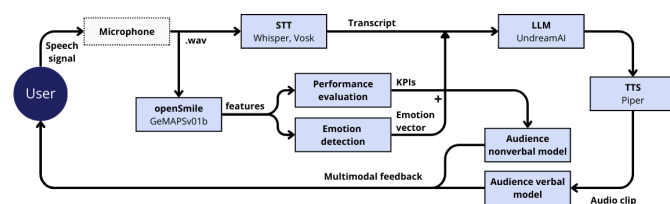
III. TECHNICAL IMPLEMENTATION

The proposed pipeline focuses exclusively on users' speech in order to address VR technical constraints. In fact, although nonverbal cues constitute valuable performance indicators, they are difficult to track without using external devices. Conversely, verbal behavior remains the most accessible modality in VR settings as it can be captured directly through headset microphones, thereby supporting simple and easy-to-deploy solutions. Another key objective is to ensure that all processing is performed locally within the headset to preserve user data privacy. The development of dedicated machine learning algorithms and the local integration of AI models are therefore essential, while still representing a technical challenge. An overview of the pipeline is presented in Figure 1.

First, an emotion detection module analyzes users' vocal delivery. Acoustic features are extracted from the microphone signal using the GeMAPSv01b openSMILE configuration, selected for its compact 68-feature set yet relevant for emotion and performance assessment [7]. The models are deployed in

In parallel, the user’s audio signal is processed by a speech-to-text module, for which two solutions embedded directly into the VR headset are considered and will be compared (*i.e.*, Vosk [13] and Whisper [8]). This transcript is then used as input to a Large Language Model, integrated directly into the headset using UndreamAI and the LLaMA 3.2 1B model [12], to implement a conversational agent. The LLM also considers the detected emotional state and the computed performance indicators to generate appropriate verbal responses.

It is worth noting that the system is designed to be multilingual, with the user preselecting the language before the session starts, after which the modules adapt accordingly. The whole system operates in a loop, with continuous processing until the end of the presentation.



- [1] Akçay, M. B., & Öğüz, K. (2019). Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication, 116*, 56–76. <https://doi.org/10.1016/j.specom.2019.12.001>
- [2] Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., & Scherer, S. (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 (UbiComp '15)*. <https://doi.org/10.1145/2750858.2806060>
- [3] Chollet, M., Wörtwein, T., Morency, L.-P., & Scherer, S. (2016). A Multimodal Corpus for the Assessment of Public Speaking Ability and Anxiety. 10th International Conference on Language Resources and Evaluation, LREC 2016, 488–495.
- [4] Etienne, E., Leclercq, A.-L., Remacle, A., Dessart, L., & Schyns, M. (2023). Perception of avatars nonverbal behaviors in virtual reality. *Psychology and Marketing, 40* (11), 2464–2481. <https://doi.org/10.1002/mar.21871>
- [5] Etienne, E., Leclercq, A.-L., Remacle, A., & Schyns, M. (2025). EVE: Emotional validated expressions, an acted audiovisual corpus. *Proceedings of the 25th ACM Int. Conf. Intelligent Virtual Agents (IVA '25)*, Berlin, Germany, <https://doi.org/10.1145/3717511.3749303>.
- [6] Etienne, E., Ristorcelli, M., Saufnay, S., Quilez, A., Casanova, R., Schyns, M., & Ochs, M. (2024). A Systematic Review on the Socio-affective Perception of IVAs' Multi-modal behaviour. *Proceedings of the 24th ACM International Conference on Intelligent Virtual Agents*, Glasgow, UK, ACM, <https://doi.org/10.1145/3652988.3673943>
- [7] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, and K. P. Truong, “The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing,” *IEEE Transactions on Affective Computing*, vol. 7, no. 2, pp. 190–202, Jul. 2015, doi: 10.1109/TAFFC.2015.2457417.
- [8] OpenAI: Whisper GitHub repository. <https://github.com/openai/whisper>
- [9] Piper: Piper unity GitHub repository. <https://github.com/skykim/piper-unity>.
- [10] PwC. (2020). The effectiveness of Virtual Reality Soft Skills Training in the Enterprise: A Study. <https://www.pwc.co.uk/services/technology/immersive-technologies/study-into-vr-training-effectiveness.html>
- [11] Saufnay, S., Etienne, E., & Schyns, M. (2024). Improvement of Public Speaking Skills Using Virtual Reality: Development of a Training System. In 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). <https://doi.org/10.1109/ACIIW63320.2024.00025>
- [12] UndreamAI: LLM unity GitHub repository. <https://github.com/undreamai/LLMUnity>.
- [13] Vosk: Vosk API GitHub repository. <https://github.com/alphacep/vosk-api>.