Enhancing Public Speaking Skills through Advanced VR Training Systems

Sai Kiran Reddy Beeram
Human Computer Interaction, Software Engineering
Saint Louis University

Sri Harika Vaddineni
Human Computer Interaction, Computer Science
Saint Louis University

Anjali Putta
Human Computer Interaction, Software Engineering
Saint Louis University

Abstract—The introduction of Virtual Reality (VR) technology has brought about a revolution in a number of fields, such as rehabilitation, training, and education. In particular, it has become a viable tool in the field of public speaking training, helping to improve presentation skills and address glossophobia, or the fear of public speaking. This paper reviews the current literature on VR applications in public speaking training, identifying key features and gaps. It then proposes a comprehensive VR training system that integrates customizable environments, advanced speech analysis, body language evaluation, and interactive elements for an immersive and effective learning experience.

I. INTRODUCTION

Public speaking is a critical skill in both professional and personal contexts. However, it is also a common source of anxiety, which can impede an individual's ability to effectively communicate. Traditional training methods, while beneficial, may not fully address the immersive experience needed to simulate real-life speaking scenarios. The emergence of Virtual Reality (VR) offers a novel approach, allowing individuals to practice in safe, controlled, yet realistic environments. This paper explores the integration of VR into public speaking training, examining existing solutions, and proposing an advanced VR training system.

II. LITERATURE SURVEY

One significant area of focus in recent research has been the design and deployment of virtual reality (VR) environments for public speaking. Study [5] investigated the acceptance and effectiveness of a VR public speaking training program, focusing on user experience and learning outcomes. Their findings suggest high acceptance rates among participants and significant improvements in public speaking skills, reinforcing the potential of VR as a highly engaging and effective training tool for public speaking, with positive implications for user motivation and learning efficacy.

However, despite these advancements, existing VR systems exhibit notable limitations that impede their effectiveness and broader application. For instance, studies such as [1] and [3] detailed the development of VR systems tailored for university seminars and training programs, which enable speakers to practice in front of virtual audiences. These systems, while innovative, often lack in simulating realistic audience dynamics comprehensively, offering limited customization options for scenarios such as room size and audience personality factors, which are crucial for the effectiveness of training systems.

Moreover, investigations into VR training effectiveness, including user perceptions of virtual audiences ([2], [3]), highlight the significance of presence and the immersive nature of VR. Yet, these studies also reveal a gap in integrating comprehensive feedback mechanisms that address both verbal and non-verbal communication skills effectively. The study [6], for example, explores the animation and perception of virtual audiences in VR but underscores the absence of advanced feedback mechanisms like advanced speech analysis and body language evaluation, critical for a holistic training experience.

The adaptive nature of VR training systems, as discussed in [4] and [5], allows for customization of audience reactions and feedback based on the speaker's performance. Nonetheless, the literature reveals a fragmentation in the approach to VR training for public speaking, with each system addressing specific facets while overlooking others. This fragmentation suggests a pressing need for a VR training system that not only offers immersive and interactive environments but also integrates a comprehensive evaluation of the speaker's performance, including advanced speech and body language analysis.

Our proposed VR training system aims to address these identified gaps by offering customizable environments, immersive audience interaction powered by advanced AI algorithms, interactive presentation management, and a comprehensive evaluation of both verbal and non-verbal communication skills.

III. PROPOSED METHODOLOGY

Our proposed VR training system aims to address the gaps identified in the literature by offering the following features:

1) Customizable Environments: Our system allows users to create a realistic practice space by selecting the room size, audience type, and other environmental settings. This customization extends to the selection of auditory background noise levels, lighting conditions, and even the attire of the audience, enabling users to simulate

a wide range of speaking scenarios, from an intimate boardroom presentation to a large-scale conference speech.

2) Immersive Audience Interaction: The virtual audience in our system is powered by sophisticated AI algorithms capable of generating a spectrum of emotional reactions and interactions, mirroring real-world audience behaviors. This includes dynamic changes in audience attention and engagement—audience members may lean forward in interest, display various facial expressions, or even exhibit signs of disengagement, such as looking away or whispering to neighbors, contingent upon the speaker's performance.

3) Interactive Presentation Management: To further simulate the real-world experience, our system integrates gesture-based slide navigation, allowing speakers to transition through their presentation slides using intuitive hand gestures. This feature is designed to recognize a range of gestures, from simple swipe motions for moving to the next slide to more complex gestures for activating embedded multimedia within the presentation. This capability not only adds to the realism of the practice session but also aids speakers in mastering the art of seamless presentation flow.

4) Advanced Speech Analysis: Leveraging state-of-the-art speech recognition and natural language processing technologies, our system provides in-depth analysis of speech patterns, including pace, volume, and filler word usage. It goes beyond these fundamentals to assess the speech's relevance and logical coherence, offering constructive feedback on how effectively the speaker is conveying their message and engaging with the audience.

5) Body Language and Non-Verbal Communication Evaluation: Utilizing advanced pose estimation and motion capture technologies, our system evaluates the speaker's body language, providing feedback on posture, hand movements, and overall non-verbal communication effectiveness. This includes analysis of gesture variety and appropriateness, eye contact with the audience, and the use of space, helping speakers to develop a commanding presence and connect more profoundly with their virtual audience.

This methodology integrates the benefits of VR technology with specific training needs in public speaking, offering a holistic tool for individuals looking to improve their presentation skills in a safe, controlled, yet challenging environment.

IV. Study Design

Through the user study, we would like to embark on a journey to explore the impact of a web-based public speaking platform on individuals, both anxious and confident in their public speaking abilities. With a participant group of 60, equally divided between those who experience public speaking anxiety and those who do not, our study delves into how the use of our system can influence public speaking skills and the overall user experience.

Upon beginning the study, each participant is introduced to the platform—not through traditional training sessions but via a brief orientation. Here, they are acquainted with the project's objectives and how to navigate the system. This includes utilizing hand gestures to advance slides during their presentations, mimicking the action of presenting in a real-life scenario. This preparation is pivotal, ensuring participants can fully engage with the system as they would in a genuine public speaking event.

The core of the study unfolds over three consecutive sessions in a single day, designed to closely simulate the experience of public speaking in various settings. During these sessions, participants are not merely users of a system but active presenters, tasked with discussing a predetermined topic. This setup aims to closely mirror the pressures and dynamics of public speaking, allowing for an authentic response to the virtual environment.

For those with public speaking anxiety, an additional layer of observation is introduced through wearable heart rate monitors. These devices silently track stress levels, providing a window into the physiological effects of public speaking in a controlled yet immersive setup.

Following these sessions, every participant is invited to delve into a detailed survey, crafted to grasp the diverse impacts of their experience with our system. This survey aims to unearth a broad range of outcomes: it queries participants on perceived enhancements in speech quality, confidence, and overall presentation skills, and delves into the nuances of body posture, eye contact, and ease of articulation. A significant addition to this evaluation is the focus on user interaction, highlighting the critical aspect of how participants navigate and utilize the system's features and controls, which is instrumental to their overall experience.

Incorporated into this comprehensive feedback mechanism is the NASA Task Load Index (NASA-TLX), a proven tool designed to assess the perceived workload experienced by participants while engaging with the system. By integrating the NASA-TLX, the survey extends beyond subjective assessments of public speaking improvements to include an objective measure of the cognitive, emotional, and physical demands placed on users during their interactions with the platform. This inclusion is pivotal, offering insights into the system's usability and the efficiency with which users can employ its features under the pressure of simulating a public speaking scenario.

Expected outcomes of this study are multifaceted. For those grappling with anxiety, a significant finding would be a marked reduction in stress levels, as evidenced by heart rate data, signaling a decrease in public speaking anxiety through repeated exposure to the simulated environment and constructive feedback provided by our system. Participants are anticipated to report enhancements in their public speaking skills, including more confident delivery, better engagement with their audience through eye contact and gestures, and

an improved ability to navigate their presentations smoothly. Furthermore, feedback on system usability and the quality of user interaction will provide invaluable insights, helping refine the platform to better serve its users' needs.

By weaving these elements together, our study aims to shed light on the potential of web-based platforms as tools for enhancing public speaking skills, offering both a solution for those seeking to overcome anxiety and a practice arena for seasoned speakers aiming to polish their skills. Through the lens of this research, we aspire to contribute to the broader dialogue on technology's role in personal and professional development, particularly in the realm of communication and public speaking.

V. SYSTEM IMPLEMENTATION

This section outlines the implementation of our web application, designed to enhance public speaking skills through an interactive and dynamic interface that simulates the immersive aspects of Virtual Reality environments. While the system does not utilize VR hardware, it incorporates responsive design principles and interactive elements that closely mimic the engaging experience VR offers.

Our web application is designed with a clean, intuitive interface using aesthetic colors and thoughtful design elements to ensure a pleasant and productive user experience. Here's a detailed walkthrough of the system features and their implementation:

1) Recording a Session:
   Environment Selection: Users can select their desired speaking environment from options such as a classroom, conference hall, or auditorium. This feature simulates different acoustic and audience settings, helping users practice in varied contexts. This dynamic environment selection is inspired by the flexibility found in traditional VR applications for public speaking, as discussed in generic studies on VR training environments.
   Audience Selection: After selecting the environment, users choose their audience type—students, professionals, or a custom demographic. This customization allows the speaker to tailor their content and delivery style to the audience, enhancing the realism of the practice session.
   Document Handling: Users have the option to upload supporting documents (PDF or DOCX) for their presentation or proceed without any visual aids. This flexibility is crucial for accommodating different presentation styles and requirements.
   User Image Display: During the session, a live feed of the user's image is displayed on the left-hand side of the screen in a small frame. This feature allows speakers to see themselves as they present, enabling real-time self-assessment of their posture, facial expressions, and overall delivery style.
2) Session Interaction:
   Emotionally Reactive Audience: The audience on the screen reacts emotionally to the speaker's delivery, start-

ing with neutral expressions and adjusting based on performance cues. This dynamic interaction is designed to provide real-time feedback, similar to the responsive audience dynamics as implemented in [6]
   Start/Stop Capture: Users can initiate the recording of their session by clicking on the "Start Capture" button, which records both audio and video. The "Stop Capture" button ceases recording once the user completes their presentation.
   Video Upload: After the session, users have the option to upload their video for storage and later review. This feature supports continuous improvement by allowing users to revisit their performance and assess their progress.
   Document Interaction: If a document is uploaded, it appears on the right side of the screen, with the audience and environmental simulation on the left. Users can interact with the document using natural hand gestures, such as swiping up or down, to scroll through pages, enhancing the realism of the simulation.
3) Feedback and Analysis:
   Post-session, the application directs the user to a feedback page that offers comprehensive insights:
   Pace and Confidence: The system analyzes the pacing of the speech and the speaker's confidence, providing pointers for improvement.
   Eye Contact: It evaluates the speaker's ability to maintain eye contact with the virtual audience, which is critical for engaging and effective communication.
   Verbal and Non-Verbal Communication: Detailed feedback on verbal clarity and non-verbal cues like body language and gestures is provided, helping users refine both aspects of their presentation skills.
4) Review and Continual Improvement:
   The "My Recordings" page catalogs all user sessions, complete with playback controls. This repository allows for self-assessment and detailed review of past performances, facilitating ongoing improvement and skill development.

This system is built with a focus on user-friendly interfaces, customizable settings, and detailed feedback mechanisms, ensuring that users not only enhance their speaking skills but also enjoy a visually appealing and engaging training experience.

VI. RESULTS

The survey data reveals distinct patterns in user ratings and their correlation with overall satisfaction. In the average feature ratings (Fig. 1), all major features of the public speaking practice app—Realism of Virtual Environment, Audience Engagement and Immersion, Effectiveness of Gesture-Based Slide Navigation, App Performance, and Ease of Navigation—received high average scores, all above 4 out of a possible 5. This indicates strong user satisfaction with these key aspects of the app.

The improvement in public speaking skills by anxiety level (Fig. 2) displays a clear trend: individuals with high initial anxiety benefit most from using the app, showing the greatest
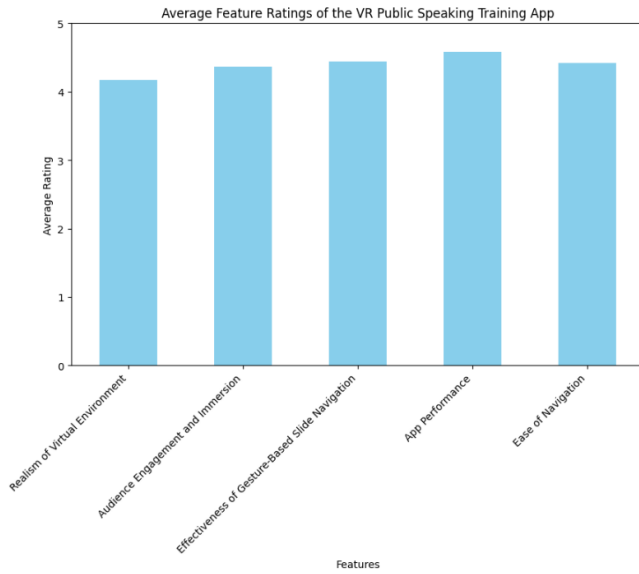
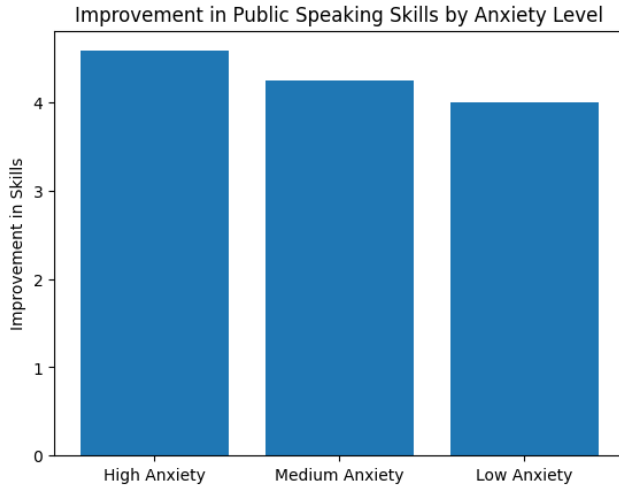Fig. 1. App features Vs Average Ratings

Fig. 2. User's Anxiety Vs Improvement in public speaking skills
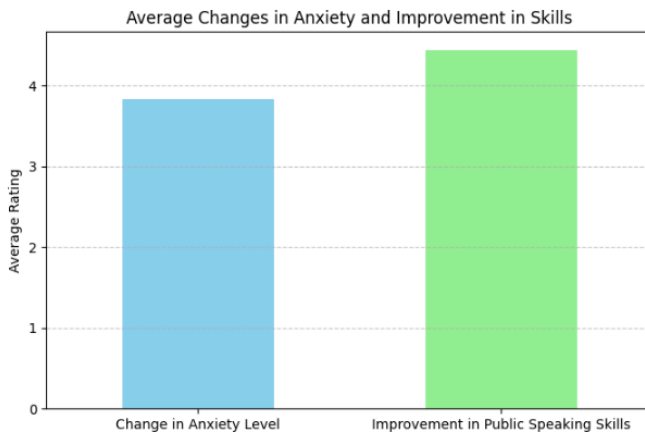

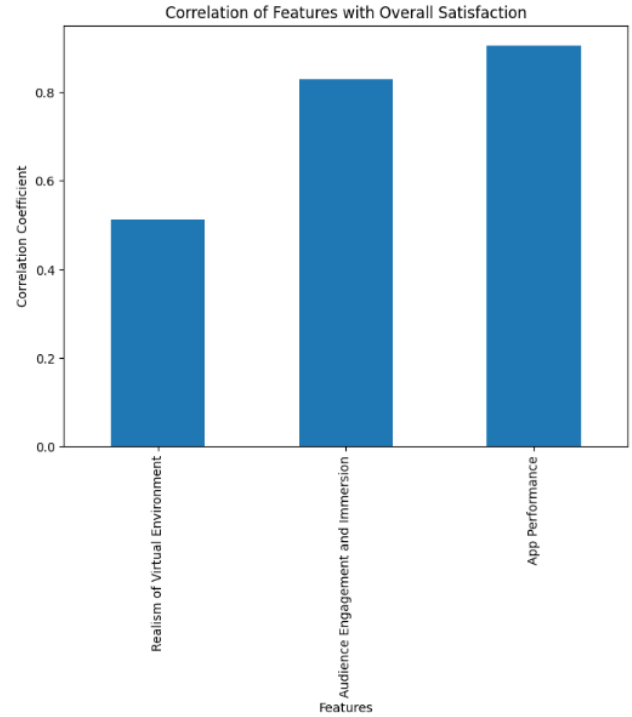Fig. 3. Improvement in Anxiety level and Improvement in skills

Fig. 4. Correlation of features with overall satisfaction

improvement in their speaking skills. In terms of changes in anxiety and improvements in public speaking skills (Fig. 3), the graph shows that there is a substantial improvement in skills, with a smaller yet significant reduction in anxiety levels.

The correlation of features with overall satisfaction (Fig. 4) further highlights which features are most valued by users. Here, 'App Performance' exhibited the strongest positive correlation with overall satisfaction, suggesting that users place high importance on the smooth functioning of the app. 'Audience Engagement and Immersion' also showed a strong positive correlation, indicating that how the audience reacts and engages during presentation significantly affects the overall user experience.

VII. ANALYSIS

The comprehensive feature ratings (Fig. 1) from our survey indicate that participants were highly satisfied with almost all the features offered by the public speaking practice app, with each feature receiving a rating close to the maximum of 5. This high level of satisfaction underscores the app's ability to meet user expectations across various functionalities including realism, engagement, navigation, and overall performance.

However, when examining the specific impacts on anxiety levels and public speaking skills (Fig. 3), while the app shows an impressive improvement in public speaking skills with an average rating of 4.37, it appears slightly less effective in reducing anxiety, with a rating of 3.95. This suggests that while the app excels in enhancing speaking abilities, there may be room for improvement in how it addresses and

mitigates user anxiety, pointing to potential areas for further development to enhance its psychological support features.

Our survey data categorizes participants into three groups based on their initial anxiety levels: high, medium, and low. As illustrated in (Fig. 2), participants with high anxiety benefit the most, showing substantial improvements in their skills. Those in the medium anxiety category experience moderate improvements, while participants with low initial anxiety exhibit comparatively smaller gains. This trend underscores the app's ability to provide a supportive environment that is particularly advantageous for users facing significant challenges in public speaking due to anxiety.

In analyzing the correlation between various features of the public speaking practice app and overall user satisfaction, it is evident from (fig. 4) that app performance and audience engagement and immersion are highly correlated with user satisfaction, with coefficients close to 0.9 and 0.8 respectively. This strong positive correlation underscores the critical importance of both smooth app functionality and interactive, immersive audience elements in enhancing user experience. In contrast, the realism of the virtual environment, while positively correlated, shows a weaker link to overall satisfaction, with a coefficient around 0.5. This suggests that although the realism of the setting contributes to user satisfaction, it does not impact user perception as significantly as the app's performance or the quality of audience interaction.

VIII. DISCUSSION

The findings from our survey on the public speaking practice app reveal a nuanced picture of how digital tools can significantly enhance learning and training in human-computer interaction (HCI). The app's effectiveness in improving public speaking skills, especially among those with high initial anxiety, underscores its potential as a critical tool in educational and therapeutic contexts. This capability aligns with current trends in HCI, where personalized digital interventions are increasingly seen as pivotal in addressing individual-specific barriers in learning and professional development.

However, the distinction in the impact on anxiety reduction versus skill improvement invites a broader reflection within the HCI community. It highlights the challenge of designing systems that are not only effective in skill enhancement but also in managing psychological aspects such as anxiety. The differential impact observed across various levels of initial anxiety suggests that while the app is successful in creating a supportive environment, there is a significant opportunity to innovate further by integrating more targeted psychological support mechanisms. The correlations between app functionality, user engagement, and satisfaction also emphasize the importance of seamless performance and immersive user experiences in educational apps. These insights should inspire future research and development efforts to focus on optimizing interactive elements that cater more holistically to user needs, thereby pushing the boundaries of what educational technologies can achieve in real-world settings.

IX. REFLECTION

Reflecting on the current capabilities, audience reactions within the app are predetermined, which limits the dynamism and realism of the feedback provided to users. Our vision to advance this system involves harnessing machine learning (ML) and artificial intelligence (AI) to enable dynamic, real-time audience reactions based on the speaker's performance metrics such as eye contact, body posture, speech relevance and clarity.

We would like to incorporate AI to analyze these aspects of a presentation in real-time to make the virtual audience to react naturally, providing speakers with immediate feedback that is reflective of a real audience. This could dramatically enhance the training effectiveness by simulating a more authentic speaking environment and thus, better preparing users for real-world public speaking scenarios.

Moreover, our current feedback mechanism, while beneficial, offers generic responses that do not necessarily reflect the individual's performance nuances. We would implement advanced AI techniques to provide personalized feedback based on specific performance metrics and implement natural language processing (NLP) to analyze speech content and delivery, providing suggestions on improving articulation, reducing filler words, or enhancing speech structure. Similarly, computer vision could be employed to critique and offer improvements on body language and posture.

If additional time and resources were available, our focus would be on developing these AI-driven functionalities to transition from preset responses to adaptive, insightful feedback that genuinely aids in skill development.

We would like to pursue publishing our work.

Github Link: https://github.com/Anjali0407-git/public$_s$peaking$_a$pp/tree/main − stable

X. BIBLIOGRAPHY

REFERENCES

[1] A. D. Lee, A. F. Da Costa, A. Davis, D. L. Linvill, and L. F. Hodges, "Virtualized speech practice for the college classroom," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Clemson University, 2020.

[2] N. Jinga, A. Moldoveanu, F. Moldoveanu, A. Morar, and O. Mitruț, "Vr training systems for public speaking – a qualitative survey," in *The 17th International Scientific Conference eLearning and Software for Education*. Bucharest, Romania: University POLITEHNICA of Bucharest, 2021.

[3] Y. Glémarec, J.-L. Lugrin, A.-G. Bosser, C. Buche, and M. E. Latoschik, "Conference talk training with a virtual audience system," in *VRST '21*. Osaka, Japan: ENIB Lab-STICC UMR CNRS 6285 and HCI Chair, University of Würzburg, 2021.

[4] H. Siddiqui, H. Irfan, A. S. Lakhani, B. Ahmed, S. Shaikh, M. M. Movania, and M. Farhan, "Manifest: Public speaking training using virtual reality," in *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. Habib University, 2023.

[5] F. Palmas, J. E. Cichor, D. A. Plecher, and G. Klinker, "Acceptance and effectiveness of a virtual reality public speaking training," in *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Garching bei Munchen, Germany: IEEE, 2019.

[6] Y. Glémarec, J.-L. Lugrin, A.-G. Bosser, A. C. Jackson, C. Buche, and M. E. Latoschik, "Indifferent or enthusiastic? virtual audiences animation and perception in virtual reality," *Frontiers in Virtual Reality*, vol. 2, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frvir.2021.666232/full

[7] F. Mostajeran, M. B. Balci, F. Steinicke, S. Kuhn, and J. Gallinat, "The effects of virtual audience size on social anxiety during public speaking," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Hamburg, Germany: Universitat Hamburg and University Medical Center Hamburg-Eppendorf and Max Planck Institute for Human Development, 2020.