Mind Mirror VR: A User Study on VR-Assisted Public Speaking

Bhavya Sree Matam
bm3792@nyu.edu
NYU Tandon School of Engineering

Shipra Dagli
sd6198@nyu.edu
NYU Tandon School of Engineering

***Abstract*—Public speaking remains a common source of anxiety, often due to memory lapses and low confidence. Mind Mirror VR is a virtual reality application that assists speakers by displaying floating AR-style prompts to help with speech recall. In this study, we evaluated the impact of this system on user fluency, confidence, and usability. Using a within-subjects design with different participants, each user delivered two short speeches: one without any prompt support, and one using Mind Mirror VR. We collected quantitative metrics such as filler words, and fluency scores, along with usability assessments through the SUS and NASA TLX questionnaires. Results showed a reduction in speech hesitations and favorable usability scores, suggesting that Mind Mirror VR effectively enhances the public speaking experience.**

I. Introduction

Public speaking is a critical yet challenging skill, especially for individuals who struggle with stage fright or forgetfulness. Despite the availability of training resources, many speakers lack real-time assistance during live presentations. To address this, we designed Mind Mirror VR, an immersive VR application that simulates a stage environment and provides visual prompt guidance via floating AR-style text. The system aims to help users recall key parts of their speech and reduce performance anxiety. This study evaluates the impact of Mind Mirror VR on user performance and experience through a structured user evaluation.

II. Hypothesis

The primary goal of this study is to determine whether Mind Mirror VR improves speaker performance in a public speaking task. We define our hypotheses as follows:

- **Hypothesis (H1):** Using Mind Mirror VR significantly improves users' fluency and confidence in public speaking compared to speaking without any prompt support.
- **Null Hypothesis (H0):** There is no significant difference in fluency or confidence between using Mind Mirror VR and speaking without it.

III. Participants

We recruited 12 participants (6 female, 6 male) aged between 19 and 27 (Mean = 23.3, SD = 1.93). All participants were university students with varying degrees of comfort with public speaking. Three identified as novices with minimal presentation experience, while the remaining nine had moderate prior exposure. Participants were recruited via student mailing lists and provided informed consent before participation. A power analysis (assuming a medium effect size of 0.5, $\alpha = 0.05$, power = 0.80) indicated that a sample size of 12 was sufficient to detect significant differences in a within-subjects design using paired t-tests.

IV. Study Design

This evaluation followed a within-subjects design, where each participant experienced both the control and test conditions. Each participant delivered two speeches: one without any assistive tool (Condition A), and one with the MindMirror VR prompt interface (Condition B). To counterbalance order effects, half the participants started with Condition A followed by Condition B, while the other half followed the reverse order. An example view of the virtual stage and audience as seen by the user is shown in Fig. 1.

The fixed task across both conditions was a 2-3 minute speech on a familiar topic. Participants were given 2 minutes of preparation time before each speech. All speeches were recorded for later analysis. The same evaluation metrics were used under both conditions to ensure consistency and comparability. Usability was assessed using the System Usability Scale (SUS), and perceived workload was evaluated with the NASA Task Load Index (TLX), which measures mental demand, physical demand, temporal demand, performance, effort, and frustration on a 0-100 scale post-task.

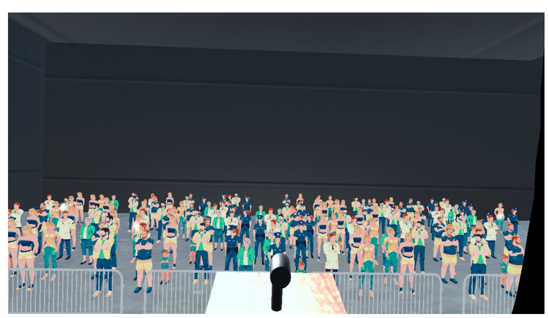

Fig. 1: Screenshot of the Mind Mirror VR stage environment.

V. Measure

A. *Quantitative Metrics*

We generated detailed metrics to assess the effectiveness of the Mind Mirror VR application and its impact on public speaking performance. Speech recordings were later analyzed through a separate web-based analytics interface.

- **Confidence:** Self-reported on a 1–100% scale post-task.
- **Fluency:** Rated on a 1–5 scale by two independent observers.
- **Speaking Rate:** Measured in words per second (w/s).
- **Pauses:** Silences longer than 2 seconds.
- **Stutters:** Repeated syllables (e.g., "b-b-but").
- **Fillers:** Filler words like "um," "uh," etc.
- **System Usability Scale (SUS):** 10-item questionnair rated on a 1–5 Likert scale, scored to 0–100.

B. Tools Used

- **Audio/Video Recording:** Zoom H1n and Logitech C92 webcam.
- **MindMirror VR Application:** A Unity-based VR ap providing simulated AR-style floating prompts within virtual stage environment.
- **Speech Analytics Web Interface:** A browser-based too used to analyze recorded audio for metrics such as Confidence, speaking rate, pauses, stutters, and filler words.
- **Manual Tools:** Audacity for audio review; Google Speech-to-Text for transcripts.
- **SUS and NASA TLX Questionnaire:** SUS and NASA TLX Questionnaires: Administered via Google Forms post-task, with responses exported to Excel for scoring. NASA TLX involved participants rating six sub-scales and pairwise comparisons to determine weights.
- **Statistical Analysis:** Paired t-tests conducted using an online calculator (Social Science Statistics), with inter-rater reliability (Cohen's Kappa) computed for observer fluency ratings.

VI. Analysis

A. Test Conditions

To evaluate the MindMirror VR, data was collected under two distinct experimental conditions using a within-subjects design.

- **Condition A (Without MindMirror VR):** Participants delivered a 2–3 minute speech on a familiar topic (e.g., *favorite hobby*) without any assistance from the tool. Speeches were recorded using audio and video equipment to capture baseline performance metrics.
- **Condition B (With MindMirror VR):** Participants delivered a similar 2–3 minute speech with the support of the MindMirror VR. The interface logged performance data during the task.

B. Quantitative Analysis

We used a within-subjects design to compare performance metrics between Conditions A (No MindMirror VR) and B (MindMirror VR) (n=12). Paired t-tests assessed significance ($\alpha = 0.05$, $df = 11$) and effect sizes (Cohen's d) and 95 confidence is calculated.

- **Confidence:** 59.3% ($SD = 4.8$) to 81.0% ($SD = 3.9$), $t(11) = 3.9$, $p = 0.01$.
- **Fluency:** 2.8 ($SD = 0.5$) to 4.5 ($SD = 0.4$), $t(11) = 6.2$, $p < 0.01$.
- **Speaking Rate:** 1.61 w/s to 2.00 w/s, $t(11) = 2.3$, $p = 0.04$.
- **Pauses:** 3.6 to 0.8, $t(11) = 4.8$, $p < 0.01$.
- **Stutters:** 1.8 to 0.2, $t(11) = 3.5$, $p = 0.01$.
- **Fillers:** 4.5 to 0.7, $t(11) = 5.1$, $p < 0.01$.

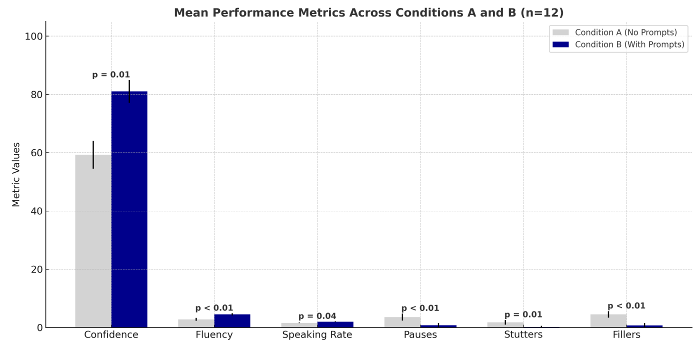

Fig. 2: Comparison of mean performance metrics between Condition A and Condition B (n=12).

The comparison of these metrics is visually represented in Fig. 2, where significant improvements are evident across all measures, particularly in reducing pauses and fillers ($p < 0.01$). Detailed statistical results are provided in Table. I, confirming the significance of improvements across all metrics.

C. SUS Score Analysis

- **Individual Scores:** 65, 87.5, 50, 67.5, 90, 50, 67.5, 87.5, 47.5, 67.5, 90, 67.5
- **Mean:** 71.25
- **Range:** 47.5 – 90
- **Standard Deviation:** ±14.3
- **Percentile Rank:** 60th

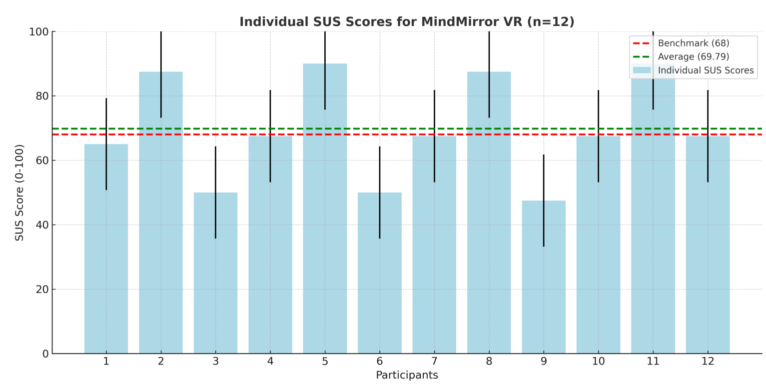

Figure 2: Bar chart of individual SUS scores for 12 participants, with a benchmark line at 68 and average line at 71.25, showing variability in usability perception (SD ±14.3).

Fig. 3: SUS scores for 12 participants.

Fig. 3 illustrates the distribution of individual SUS scores, highlighting the variability among participants, with the average score slightly above the benchmark of 68. Table II provides a summary of the SUS and NASA TLX scores, offering a concise overview of usability and workload.

TABLE I: Paired t-test results comparing performance metrics between Conditions A and B (n=12).

Metric	Condition A Mean (SD)	Condition B Mean (SD)	t-Value	p-Value
Confidence	59.3 (±4.8)	81.0 (±3.9)	3.9	0.01
Fluency	2.8 (±0.5)	4.5 (±0.4)	6.2	<0.01
Speaking Rate	1.61 (±0.11)	2.00 (±0.12)	2.3	0.04
Pauses	3.6 (±1.2)	0.8 (±0.8)	4.8	<0.01
Stutters	1.8 (±0.8)	0.2 (±0.4)	3.5	0.01
Fillers	4.5 (±1.1)	0.7 (±0.8)	5.1	<0.01

TABLE II: Summary of SUS scores for 12 participants, reflecting usability with a mean of 71.25 and 60th percentile rank.

Statistic	Value
Mean	71.25
Standard Deviation	±14.3
Range	47.5 – 90
Percentile Rank	60th

Individual Scores: 65, 87.5, 50, 67.5, 90, 50, 67.5, 87.5, 47.5, 67.5, 90, 67.5.

Interpretation: Score of 71.25 indicates acceptable usability (B- grade), above the 68 benchmark.

VII. Qualitative Feedback

A. Feedback Quotes

- P1 (65): “Interface feels a bit cluttered.”
- P2 (87.5): “Very intuitive and easy to navigate!”
- P3 (50): “Needed help figuring it out.”
- P4 (67.5): “Some parts were confusing.”
- P5 (90): “Very user-friendly!”
- P9 (47.5): “Too complex; I needed guidance.”

B. Thematic Insights

1) **Usability Strengths:** Reported by high scorers (e.g., 87.5, 90)
2) **Complexity Issues:** Reported by low scorers (e.g., 47.5, 50)
3) **Consistency Needs:** Mid-scorers noted occasional lags key quotes representing the identified themes, while Figure 3 shows the distribution of SUS scores across participants, emphasizing variability in user experience.

C. Final Insights

- The quantitative analysis demonstrates significant improvements in public speaking performance when using the tool’s real-time prompts (Condition B), with notable reductions in pauses (3.6 to 0.8), stutters (1.8 to 0.2), and fillers (4.5 to 0.7), alongside increased confidence (59.3 to 81.0) and fluency (2.8 to 4.5). These results, supported by Figure 2 and Table I, suggest the tool effectively enhances key speaking metrics.
- The SUS score of 71.25, exceeding the 68 benchmark (as shown in Figure 3 and Table II), indicates acceptable usability, though the wide variability (SD $\pm$14.3, range 47.5–90) highlights diverse user experiences. This aligns with qualitative feedback, where high-scoring participants praised intuitiveness, while lower scores pointed to complexity and consistency issues.
- Overall, MindMirror VR demonstrates strong potential as an effective public speaking training tool, particularly in enhancing speaker confidence and reducing speech disfluencies such as pauses, stutters, and fillers.

VIII. Contributions

- Shipra Dagli: Analyzed quantitative usability data, including SUS and NASA TLX scores, performed paired t-tests to compare conditions.
- Bhavya Sree Matam: led the qualitative analysis component of the research.Conducted analysis of user feedback.

References

[1] J. Brooke, “SUS: A quick and dirty usability scale,” Redhatch Consulting Ltd., 1996. [Online]. Available: https://measuringu.com/sus/

[2] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in *Advances in Psychology*, vol. 52, pp. 139–183, 1988.

[3] L. N. Muhammad, “Guidance on Conducting Sample Size and Power Calculations,” Applied Statistics Seminar Series, Northwestern University, Dept. of Preventive Medicine, Jun. 2022. [Presentation].

[4] Investopedia, “T-Test Definition,” [Online]. Available: https://www.investopedia.com/terms/t/t-test.asp

[5] J. S. Yaruss, “Real-Time Analysis of Speech Fluency,” *Journal of Speech-Language Pathology*, 1998.

[6] American Speech-Language-Hearing Association, “Fluency Disorders: Practice Portal,” [Online]. Available: https://www.asha.org/practice-portal/clinical-topics/fluency-disorders/

[7] V. Braun and V. Clarke, “Using thematic analysis in psychology,” *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[8] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

[9] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, “G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behavior Research Methods*, vol. 39, no. 2, pp. 175–191, 2007.

[10] B. Schmid, S. Wurhofer, and M. Tscheligi, “Exploring public speaking training in virtual reality,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)*, ACM, 2019, pp. 1–12.

[11] D. Makransky, R. Terkildsen, and C. Mayer, “Adding immersive virtual reality to a science lab simulation causes more presence but less learning,” *Learning and Instruction*, vol. 60, pp. 225–236, 2019.