Insights from Immersive Learning: Using Sentiment Analysis and Real-time Narration to Refine ASL Instruction in Virtual Reality

Md Shahinur Alam
VL2 Center, Gallaudet University
Washington, DC, USA
shahinur.alam@gallaudet.edu

Joseph Palagano
Educational Neuroscience, Gallaudet University
Washington, DC, USA
joseph.palagano@gallaudet.edu

Lorna Quandt
Educational Neuroscience, Gallaudet University
Washington, DC, USA
lorna.quandt@gallaudet.edu

Abstract

Immersive virtual reality presents a rich opportunity for learning signed languages, given the immersive environment's ability to represent three-dimensional information. We developed a proof-of-concept American Sign Language (ASL) learning in immersive virtual reality (VR), named ASL Champ! Twelve hearing non- or novice signers played one full level of the game, during which they were asked to provide concurrent think-aloud (CTA) commentary, narrating their experience as they played in real time. We conducted a sentiment analysis from recordings of the CTA and subsequent open-ended questions and qualitatively assessed the narrations for salient themes. The analysis revealed specific aspects of the users' experiences that were most likely to lead to positive or negative expressions during the CTA and the question session. The factors that had the most impact on user sentiment were the success of the sign recognition in the game and the extent to which users found the game intuitive or self-explanatory. We also found that users with more technology anxiety were more positive about the game. We also qualitatively examined user comments, revealing their real-time game experiences. This work provides insights into which aspects of an ASL learning VR game are most important for user experiences. We conclude with takeaway recommendations for future virtual or augmented reality sign language learning games.

CCS Concepts

• **Human-centered computing → User studies**; **Usability testing**; **Virtual reality**; Empirical studies in HCI.

Keywords

Sign language, Avatars, Virtual reality, Sentiment analysis, User experience

ACM Reference Format:
Md Shahinur Alam, Joseph Palagano, and Lorna Quandt. 2024. Insights from Immersive Learning: Using Sentiment Analysis and Real-time Narration to Refine ASL Instruction in Virtual Reality. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24), October 27–30, 2024, St. John's, NL, Canada.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3663548.3688503

ASSETS '24, October 27–30, 2024, St. John's, NL, Canada

ACM ISBN 979-8-4007-0677-6/24/10
https://doi.org/10.1145/3663548.3688503

1 Introduction

American Sign Language (ASL) uses gestures, facial expressions, and body movements to convey meaning. Traditional online ASL teaching lacks the necessary immersion. Virtual reality (VR) offers a solution by simulating real-life interactions and providing an immersive learning environment, crucial for effective ASL learning through visual and kinesthetic feedback. [2, 4]. VR also provides the opportunity to learn directly from a simulated signing teacher, who ideally can be animated via motion capture methods to ensure native fluency and fluidity with their sign production [12, 13].

Recent VR advancements have created immersive environments where learners interact with 3D signing avatars and receive instant feedback on their linguistic accuracy, mimicking real-world communication scenarios essential for effective language learning. [15]. Studies emphasize the importance of accurate sign recognition systems integrated within VR that can evaluate and provide corrective feedback to users in real-time [1, 3, 11]. Feedback provided within VR environments closely mirrors that of native ASL instructors, leading to a significant improvement in language acquisition. Research indicates that VR-based learning enhances retention rates compared to non-immersive methods. Key studies found that VR greatly improves learning outcomes over traditional video-based methods. However, high VR setup costs and the need for specialized technical support pose challenges to widespread adoption. [8]. Additionally, Liu et al. investigated the potential of real-time feedback systems within VR that analyze and correct users' sign language gestures [10]. Immediate feedback can enhance learning precision, but technology struggles with nuanced hand movements. Researchers improved VR accessibility for deaf users with visual aids and haptic feedback [6]. Users appreciated the customizable VR experiences, which boosted motivation through collaborative learning. However, issues like VR sickness and disorientation highlight the need for ergonomic improvements [14]. These studies highlight VR's transformative potential in ASL education and the challenges that need to be addressed to fully optimize its educational value.

We conducted a VR user study using "ASL Champ!," where participants practiced ASL with a virtual avatar. Interactions were recorded, and feedback was collected through surveys. This study assessed the VR system's usability, the avatar's teaching effectiveness, and user engagement, focusing on features that most influenced users' feelings during and after use.

2 System Overview

2.1 The ASL Champ! Coffee shop environment

We designed and developed a VR learning game named ASL Champ! The first level of this game is set in a virtual coffee shop where users with VR headsets interact with a 3D-animated avatar proficient in ASL. The avatar was created using 3D software, with human body movements recorded via a Vicon system to animate the avatar. This setting facilitates a simple vocabulary lesson for everyday objects. The avatar guides users, demonstrating signs for virtual items and prompting them to replicate them. The avatar's animations are smooth and realistic and were developed using motion capture technology. A deaf ASL user was recorded with motion capture to create accurate signs.

Figure 1: ASL Champ user experience set up. A user is signing coffee signs displayed on the VR headset mirrored on the computer monitor. The side and front views are shown on the left and right, respectively.

A comprehensive sign recognition system has been implemented to facilitate user interaction and evaluate signing accuracy. This system incorporates the MiVRy Unreal engine plugin along with advanced deep learning algorithms, constituting the backbone of our corrective feedback mechanism. We also developed an advanced deep-learning model for quantitative analysis. The proposed model achieved training, validation, and test accuracies of 90.12%, 89.37%, and 86.66%, respectively. As users perform signs, the VR system records the movements of their hands and fingers, analyzing this data to determine the output accuracy of each sign. Depending on this evaluation, the system prompts the user to retry the sign if inaccuracies are detected or advance them to the next learning segment upon successful demonstration. The technical implementation and the detailed model description can be found in our recent work [3].

2.2 Study Setup

We recruited twelve typically hearing adult non-signers for a user study, including nine women and three men. Nine had 0-1 years of ASL experience, with three having no experience and the rest having 1-3 years. Nine participants knew "very basic" ASL, two knew "basic" ASL, and one knew none. Participants received brief training to ensure they could navigate the VR environment.

Participants wore Meta Quest 2 VR headsets in a simulated coffee shop environment to learn basic ASL during the experiment. Fig. 1. shows the system setup, in which a participant is wearing an Oculus Quest 2 and signing 'coffee' in front of a computer. Two video cameras recorded their movements. Participants used a "Concurrent Think-Aloud" (CTA) technique [5], verbalizing their thoughts and feelings in real time as they interacted with the game. This method provided direct insights into their experiences, decision-making, and emotional responses, helping identify usability issues and patterns in behavior. The real-time feedback from CTA preserved the authenticity of participants' experiences.

Following the interactive session, we invited participants to a structured feedback session. This included a questionnaire designed to assess their experience, querying their familiarity with VR, their inbound proficiency in ASL, their personal impressions of the educational content, the system's efficacy, and any suggestions for improvement.

3 Data Processing and Analyses

The video recordings of the participant sessions were uploaded to a private YouTube channel, and a transcript was automatically generated using YouTube's built-in algorithm. The transcripts for each participant were downloaded, cleaned, and divided into the CTA and question sessions. We conducted two primary analyses to examine the data from the CTA and question sessions. We used the Valence Aware Dictionary and sEntiment Reasoner (VADER) in the NLTK toolkit implemented in Python3 to run a sentiment analysis on each chunk of the transcribed text [7]. To allow for qualitative analyses, we also read the transcribed texts to identify content from the participants, which may provide insights into their thoughts and impressions of the game. We calculated each participant's polarity (e.g., positive, negative, and neutral) sentiments during the CTA (CTA-pos, CTA-neg, and CTA-neut) and the question session (Q-pos, Q-neg, and Q-neut).

We conducted pre- and post-experiment surveys. The pre-survey collected demographic information, including age, hearing status, education level, primary language, prior exposure to sign language, interactions with signing avatars, familiarity with VR technologies, and usage of virtual assistants. Participants also rated their daily technology use and comfort with technology interactions. The post-survey evaluated participants' interactions with ASL Champ, focusing on initial impressions, interface intuitiveness, avatar aesthetics, movement naturalness, instruction clarity, guidance sufficiency, and overall immersiveness.

4 Results

The mean number of words said during the CTA was 650 (SD = 328). A subset of the significant correlations is shown in Fig. 2. A significant negative correlation was found between negative responses in Concurrent Think Aloud (CTA-neg) and the recognition of signs (SignsRecognized), r = -.738, p = .006. This indicates a strong inverse relationship; when participants felt there was better sign recognition, they had less negative sentiment during the CTA session. The effect size, measured by Fisher's z, is -0.945 with a standard error of 0.335. We found a significant positive correlation between positive responses to questions (Q-positive) and a participant's belief that computers are complicated (CompComplicated), r = .588, p = .044. This suggests a moderate positive association, indicating that people who think computers are more complex were likelier to show a positive sentiment during the question session. The effect size is 0.675, with a standard error of 0.329. A significant negative correlation was observed between negative responses to questions (Q-negative) and participants' belief that ASL Champ! was self-explanatory (SelfExplanatory), r = -.586, p = .045. This moderate inverse relationship suggests that as the perceived self-explanatory nature of the game decreases, negative responses to questions increase. The effect size for this correlation is -0.672, with a standard error of 0.329. There was a significant and negative correlation between negative responses to questions (Q-negative) and perceived intuitiveness of the game (Intuitive), r = -.646, p = .023. This strong inverse relationship indicates that lower levels of intuitiveness were associated with increased negative responses during the Q&A session. The effect size, measured by Fisher's z, is -0.769, with a standard error of 0.331. The correlation between negative responses to questions (Q-negative) and the explanatory nature of the game's instructions (InstructionsExplanatory) was significant and negative, r = -.608, p = .036. This suggests a moderate-to-strong inverse relationship, where more negative responses correlated with a participant's perception that the game instructions were not explanatory. The effect size for this correlation is -0.706, with a standard error of 0.33.

Participants expressed largely positive feelings about the signing avatar, as seen in this response from P1: *"I thought [the avatar] was a really cool, realistic-looking avatar. It… felt like a video game, but she looked like a human and could do signs the way I would expect someone to be able to"*. However, the interactions with the avatar were not always self-evident. For example, P2 stated: *"[The avatar] keeps putting her hands in the air like she's waiting for me to do something... Okay, now she's telling me to grab the milk– maybe? I'm sure she's getting really frustrated with me."* The issue of needing more direction or explanation was mentioned by several participants. For instance, P4 said, *"I wouldn't say it was super self-explanatory. I was expecting some text on the screen to tell me what to do instead of just having to interpret her gestures, but that could just be, my biases being a hearing person just waiting for text and not paying attention to the gestures"*. Following up on the generally positive ratings of the environment and the immersive feeling of the game [ref], P9 said, *"The setting feels pretty unique; I thought–you know...the space was very detailed. It looks up-to-par with a lot of the virtual games or virtual environments that I've had experience with."* Related to the immersiveness and semi-realism of the interactions, P12 noted, *"It almost feels like you are having a conversation with someone right there."* These subjective comments from the participants shed light on the elements of the game they noticed–immersion, signing quality–and the game elements that distracted them from learning ASL.

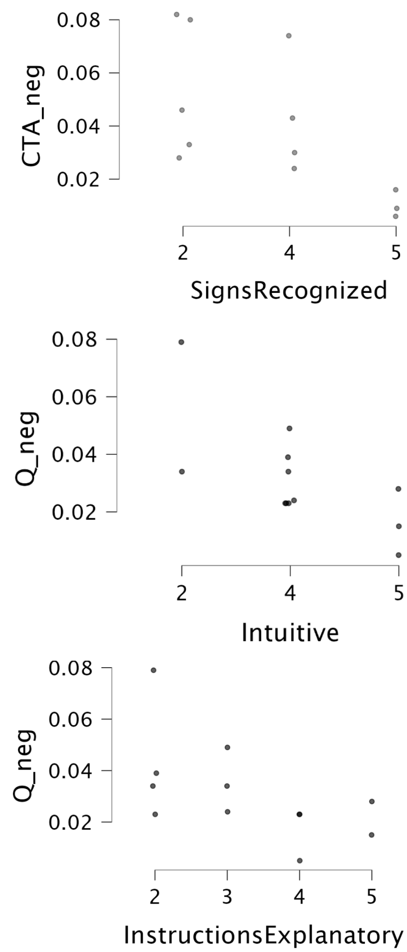

Figure 2: Correlations between sentiment values and quantitative ratings of the users' experience. Each dot corresponds to one user.

5 Discussion

This research aimed to identify key features of interactive learning games that impact user experience. Using quantitative sentiment analysis and qualitative reviews of participant comments, we found five aspects that significantly predicted user sentiments. A key finding was that better sign recognition in ASL Champ! correlated with lower negative sentiments, highlighting the importance of accurate sign recognition. Users responded positively to effective sign recognition but experienced frustration when the system incorrectly detected their signs. This finding highlights the importance of the game's core functionality, wherein the deep-learning-driven sign recognition algorithm determines whether a user's hand shape and movements match [3].

The next finding of this analysis was that individuals who perceived computers as complicated displayed more positive sentiments during the Q&A session. This suggests that people intimidated by technology were pleasantly surprised by the ASL Champ! game, potentially because it exceeded their expectations. This shows that positive VR experiences may counter preconceived negative beliefs about technology. Although the effect size is moderate, future research should explore how attitudes toward technology influence perceptions of emerging technologies like VR.

The three remaining findings closely paralleled one another–statistically significant negative correlations between negative sentiment during the Q&A session and 1) the degree to which the game was self-explanatory, 2) the degree to which the interface was intuitive, and 3) how helpful the game instructions were. Given the close relationship between these three aspects of the game, we consider these findings together. Notably, one participant attempted to physically grab the virtual object. This misinterpretation may potentially lead to user frustration within the virtual environment. It is important here to note that the ASL Champ! system intentionally did not include extensive English instruction, as research has suggested that visual-manual-based instruction (e.g., gesture) for sign language learning can be more effective than speech-print instruction [9].

Overall, our findings have the following impacts on the development of sign language technologies.

(1) Intuitive design and clear instructions are crucial for user experience in VR. Prioritizing user-friendly educational tools with transparent instructions is essential for games involving complex interactions.
(2) Adaptive feedback mechanisms, like the sign recognition tool in ASL Champ!, significantly enhance learning experiences. This suggests that when effective, real-time sign recognition can greatly improve user experiences in other educational technologies.
(3) The research shows VR effectively teaches ASL, offering a scalable solution, especially when in-person classes aren't available or for practice between sessions.
(4) Users who found technology complicated reacted positively when their expectations were exceeded. Intuitive, self-explanatory VR tools can enhance their acceptance, reduce the learning curve, and increase VR usage.

6 Conclusion

In this user study, we demonstrated the significant potential of immersive VR for enhancing ASL learning, particularly for novice signers, by providing interactive, real-time feedback through a 3D-animated avatar. The study emphasized the importance of intuitive design, clear instructions, and adaptive feedback tools like accurate real-time sign recognition to improve the user experience. Users responded positively to intuitive VR designs, which may encourage adoption and reduce learning curves. However, challenges in sign recognition accuracy, especially with nuanced hand movements, and the need for more user-friendly interfaces were identified. Future research should focus on expanding the user base to a wider demographic, exploring advanced sign recognition technologies, and integrating adaptive feedback mechanisms to personalize the learning experience further. These efforts aim to bridge the gap between technology and user needs, ensuring that VR-based language learning tools are accessible, effective, and engaging. Although our study is limited by a small sample size, it highlights the potential of immersive VR in ASL education and the importance of user-centered design for educational technologies.

Acknowledgments

The authors acknowledge funding from the National Science Foundation under grants 1839379 and 2118742 for this work. We are grateful for the efforts of our user study participants.

References

[1] Md Shahinur Alam, Myles de. Bastion, Melissa Malzkuhn, and Lorna C. Quandt. 2023. Recognizing highly variable American sign language in virtual reality. In *IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)* (Rhodes Island, Greece). https://doi.org/10.17605/OSF.IO/9JM3C

[2] Md. Shahinur Alam, Ki-Chul Kwon, and Nam Kim. 2021. Implementation of a Character Recognition System Based on Finger-Joint Tracking Using a Depth Camera. *IEEE Transactions on Human-Machine Systems* 51, 3 (2021), 229–241. https://doi.org/10.1109/THMS.2021.3066854

[3] Md Shahinur Alam, Jason Lamberton, Jianye Wang, Carly Leannah, Sarah Miller, Joseph Palagano, Myles de Bastion, Heather L. Smith, Melissa Malzkuhn, and Lorna C. Quandt. 2024. ASL champ!: a virtual reality game with deep-learning driven sign recognition. *Computers & Education: X Reality* 4 (2024), 100059. https://doi.org/10.1016/j.cexr.2024.100059

[4] Samirah Altukhaim, Daniel George, Kiruba Nagaratnam, Toshiyuki Kondo, and Yoshikatsu Hayashi. 2024. Enhancement of sense of ownership using virtual and haptic feedback. *Scientific Reports* 14, 1, Article 5140 (March 2024). https://doi.org/10.1038/s41598-024-55162-x

[5] Victoria A. Bowers and Harry L. Snyder. 1990. Concurrent versus Retrospective Verbal Protocol for Comparing Window Usability. *Proceedings of the Human Factors Society Annual Meeting* 34, 17 (1990), 1270–1274. https://doi.org/10.1177/154193129003401720 arXiv:https://doi.org/10.1177/154193129003401720

[6] John Dudley, Lulu Yin, Vanja Garaj, and Per Ola Kristensson. 2023. Inclusive Immersion: a review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. *Virtual Reality* 27, 4 (01 Dec 2023), 2989–3020. https://doi.org/10.1007/s10055-023-00850-8

[7] C. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media* 8, 1 (May 2014), 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

[8] Oscar Koller. 2020. Quantitative Survey of the State of the Art in Sign Language Recognition. *ArXiv* abs/2008.09918 (2020). https://api.semanticscholar.org/CorpusID:221265990

[9] Diane Lillo-Martin and Jonathan Henner. 2020. Acquisition of Sign Languages. *Annu Rev Linguist* 7 (Oct. 2020), 395–419.

[10] Ruixue Liu, Lei Wang, Jing Lei, Qiu Wang, and Youqun Ren. 2020. Effects of an immersive virtual reality-based classroom on students' learning performance in science lessons. *British Journal of Educational Technology* 51, 6 (2020), 2034–2049. https://doi.org/10.1111/bjet.13028 arXiv:https://bera-journals.onlinelibrary.wiley.com/doi/pdf/10.1111/bjet.13028

[11] Lorna Quandt. 2020. Teaching ASL signs using signing avatars and immersive learning in virtual reality. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.

[12] Lorna C Quandt, Jason Lamberton, Carly Leannah, Athena Willis, and Melissa Malzkuhn. 2022. Signing avatars in a new dimension: Challenges and opportunities in virtual reality. In *Proceedings of the 7th International Workshop on Sign Language Translation and Avatar Technology (SLTAT 7)*.

[13] Lorna C Quandt, Athena Willis, Melody Schwenk, Kaitlyn Weeks, and Ruthie Ferster. 2022. Attitudes toward signing avatars vary depending on hearing status, age of signed language acquisition, and avatar type. *Frontiers in psychology* 13 (2022), 730917.

[14] Nesse van der Meer, Vivian van der Werf, Willem-Paul Brinkman, and Marcus Specht. 2023. Virtual reality and collaborative learning: a systematic literature review. *Frontiers in Virtual Reality* 4 (2023). https://doi.org/10.3389/frvir.2023.1159905

[15] Zhe (Victor) Zhang. 2020. Learner engagement and language learning: a narrative inquiry of a successful language learner. *The Language Learning Journal* 50 (2020), 378 – 392. https://api.semanticscholar.org/CorpusID:225542849