Signing Avatars in a New Dimension: Challenges and Opportunities in Virtual Reality

Lorna C. Quandt, Jason Lamberton, Carly Leannah, Athena Willis, & Melissa Malzkuhn
Gallaudet University
800 Florida Ave NE Washington, DC 20002
{lorna.quandt, jason.lamberton, carly.leannah, athena.willis, melissa.malzkuhn}@gallaudet.edu

Abstract

With improved and more easily accessible technology, immersive virtual reality (VR) head-mounted devices have become more ubiquitous. As signing avatar technology improves, virtual reality presents a new and relatively unexplored application for signing avatars. This paper discusses two primary ways that signed language can be represented in immersive virtual spaces: 1) Third-person, in which the VR user sees a character who communicates in signed language; and 2) First-person, in which the VR user produces signed content themselves, tracked by the head-mounted device and visible to the user herself (and/or to other users) in the virtual environment. We will discuss the unique affordances granted by virtual reality and how signing avatars might bring accessibility and new opportunities to virtual spaces. We will then discuss the limitations of signed content in virtual reality concerning virtual signers shown from both third- and first-person perspectives.

Keywords: signing avatars, virtual reality, motion capture

1. Introduction

Immersive virtual reality (VR) continues to become more popular, with almost 10 million VR devices shipped in 2021 alone (Alsop, 2022). Along with this proliferation of new technology comes the possibility of new ways of communicating, socializing, or learning in virtual spaces. Likewise, interest in technology-supported sign language instruction is growing. Unlike spoken language, which can be taught and evaluated using smartphones or computers, the three-dimensional nature of signed languages and facial expression's impact on meaning has created a severe barrier to technology-based sign language instruction. In-person classes are expensive and difficult to access in many areas. The other available options include books, videos, or smartphone apps that cannot fully demonstrate the highly spatial nature of signed language or provide real-time feedback. Emerging technologies like mixed and virtual reality allow the development of three-dimensional interactions in immersive environments. By taking advantage of the three-dimensional (3D) nature of immersive VR, it may be possible to create immersive learning experiences to engage learners' bodies and minds more effectively and enhance learning. In this paper, we will discuss the possibility of signing avatars in a VR environment, considering both the opportunities afforded by the current technology and the limitations.

Our work focuses on American Sign Language (ASL), but these considerations may also apply to other signed languages. We direct our attention primarily toward using VR for supporting sign language learning (Quandt et al., 2020). However, sign language in VR is also relevant for entertainment, gaming, and socialization in virtual spaces.

While developers are designing many different types of learning experiences in VR, the applications of VR for learning signed languages are particularly encouraging. A fundamental theory in learning science, called embodied learning, posits that greater involvement of conceptually-aligned movement and action during learning can lead to better understanding and higher recall (Kontra et al., 2012; Kontra et al., 2015; Lindgren & Johnson-Glenberg, 2013; Weisberg & Newcombe, 2017). The immersive and spatially rich nature of VR allows for the possibility of embodied learning. Learning signed languages in VR may represent a step toward the potential far-reaching application of embodied learning through signed languages.

Many new ASL learners use online two-dimensional videos to learn introductory signs, but these pre-recorded videos have no interactive features and may not engage all learners (Shao et al., 2020). By contrast, immersive VR creates a powerful experience wherein people feel as if they are physically present in a 3D virtual space (Bailenson, 2018; Lindgren & Johnson-Glenberg, 2013). This immersive, spatially rich environment is particularly well-suited to the highly spatial nature of ASL, in which space is used as a core feature of the language. In one study, interaction with a signing avatar in augmented reality (e.g., the avatar is overlaid upon the real-world view) led to improved ASL learning outcomes compared to learning by video or book (Shao et al., 2020). Throughout our work on signing avatars, a critical guiding goal has been to ensure that the movements of the animated signing are as natural (i.e., human-like) as possible. It is crucial to ensure that animated sign language accurately delivers the nuances and inflections of the original linguistic content, rather than relying on automated animations which do not include smooth transitions or natural movements (Quandt et al., 2022). Even a slight error in the synchronization of the animation can affect the interpretation of a signed production. Motion capture enables the highest quality animation (Joerg, Hodgins, & O'Sullivan, 2010), although it does come at a high processing cost.

Our group has established the feasibility of an immersive VR ASL learning environment populated with high-quality signing avatars. In this work, we created a prototype for teaching ASL in immersive VR (Quandt et al., 2020), in which both a virtual Teacher avatar and the VR user are present in the virtual environment (Fig. 1). This scenario

Figure 1: Two types of signing present in an immersive virtual reality environment.

encapsulates the two types of signing which may present in the virtual space: 1) Third-person Perspective, wherein one or more signing avatars are present in front of the VR user in space; and 2) First-person perspective, wherein the VR user's own hands can be seen in the virtual space and can potentially be animated with real-time signing based on hand-tracking the user's actual signing (Fig. 1). These two signers in virtual space bring about different challenges and opportunities, which we will discuss in this paper.

2. The State of Signing Avatars in VR

Signing avatars are not commonly seen in virtual spaces yet, but as VR becomes more affordable, and research development continues in this area, we are becoming more familiar with what is possible regarding signing in VR. The newest publicly available models of VR head-mounted devices have better-performing hardware and software, which makes the representation of ASL in VR more feasible. For example, newer VR headsets are wireless, which allows for better head and body mobility. They also include better video resolution and built-in cameras to aid in hand-tracking. Recent software updates have further improved the hand tracking capabilities of some devices (Henry, 2022). We expect this trajectory to continue as VR becomes more mainstream. Particularly relevant to signing in virtual spaces, the Oculus Quest 2 contains built-in hand-tracking cameras. Currently, some publicly available software (e.g., Waltz of the Wizard; Hand Physics Lab) use hand-tracking to control user interfaces or as an integral part of the gameplay, while many programs still rely on controller-based commands. We use the built-in hand tracking of the head-mounted Oculus Quest 2 device in our current work, but other options may be commercially available, and developers regularly release new hardware with updated capabilities. Some external hardware could enhance hand-tracking capabilities (e.g., Kinect, depth sensors), however, users much prefer a fully wireless experience, especially if they are moving their hands around to learn and produce signs (Quandt et al., 2020). Keeping the equipment manageable and avoiding physically burdensome add-ons is an intentional design choice.

Socialization and community-building are growing activities within VR, allowing users to connect with others in a natural, immersive environment (Li, Vinayagamoorthy, Williamson, Shama, & Cesar, 2021). Virtual avatars have already been hacked to communicate in sign languages for casual conversation and social interaction. VRChat is a community accessible to any VR user wherein people can virtually navigate a built environment, inhabiting a character that they customize. Users can chat and form online communities with other users. An emergent sign-language using community has emerged in VRChat, including drop-in sign language chats and informal sign language lessons in several different signed languages (Davis, 2019). Since not all users have devices that can track hand movements, VRChat users cannot sign naturally with their hands. Instead, they use controllers to produce signs. Some controllers give the user the ability to make certain handshapes, with the thumb and the index and middle fingers, and the ability to open and close your fist. Within those limitations, a user can sign in a modified way, involving a limited number of moveable fingers and opening and closing their fists. The hands can move freely in the space around the user, allowing for sign location to be represented reasonably well. This emerging signing community in VRChat demonstrates interest in the casual use of signed languages for socialization and learning in VR and highlights the adaptations that communities come up with to work around technological limitations.

One significant limitation of signed communication in VR is the difficulty animating natural facial expressions, especially for real-time communication as in VRChat described above. In ASL, and all signed languages, facial expressions, including the mouth, eye, cheek, and eyebrow movements, are intricate and nuanced, adding and changing the meaning and structure of signs produced by the hands. To successfully capture facial expressions in VR, the capture technology must pick up on the slightest differences and changes in the face that accompany the hand movements of ASL. The two distinct types of signers in VR each present different opportunities and challenges, which we will discuss below.

3. Third-person perspective signing

3.1 Opportunities

Third-person perspective signing—in which the user views a signing avatar in front of them (Fig. 1) is the more straightforward representation of sign language in VR. This scenario is similar to animating a signing avatar outside of VR. Developers create the avatar using development pipelines the same way they do for non-VR use. The 3D avatar file is then placed in the VR environment. The 3D nature of the virtual environment means that a user can see the avatar's movements with rich spatial detail. For instance, in VR, a signing avatar can be seen from all angles in ways that accurately represent signing movements in space.

3.2 Challenges

With the third-person signing avatar, the primary challenge is creating avatars that are not too resource-heavy, since typical animations are made up of too many polygons and become a burden on the VR platform. Polygon count is a critical consideration when developing and populating virtual environments. Essentially, the more polygons, the more computing power is needed. There is a tradeoff between quality and the ability of the VR platform to handle the torrent of data efficiently. To ensure real-time interactivity, we ensure that the system maintains stability by keeping the avatar's animation within reasonable limits for polygon count.

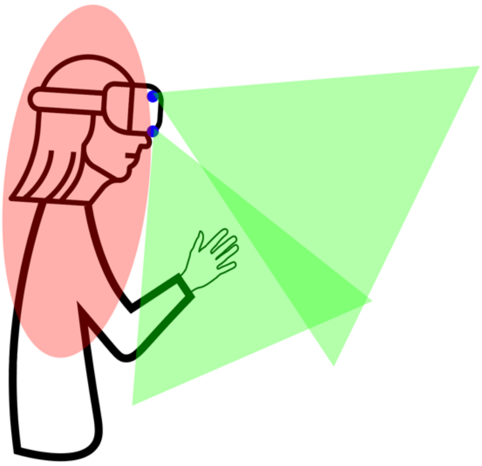

Figure 2. The built-in hand-tracking cameras on a VR device (located at the blue dots) can capture movements in certain locations well, as shown in the green areas. However, areas on the user's head and body are not easily captured by the cameras, as shown in red. This schematic is generic and not specific to any specific device.

In the past, our team used motion capture markers on the face. However, the markers only captured a subset of facial movements, overlooking other possible facial expressions, and could not track eye movements. The Faceware system (Faceware Tech, Austin, TX, USA) has proved to be effective in capturing a broader range of a signer's facial expressions and eye movements, resulting in an avatar that portrays ASL facial expressions accurately. Our project requires the use of a customized Faceware helmet camera which allows for natural movements of the hands near the signer's face (Quandt et al., 2020). One remaining issue is that whenever the hands cover the signer's face during recording, there are gaps in the facial data, which must be hand-animated in later in the development pipeline.

4. First-person perspective signing

4.1 Opportunities

When hand-tracking in enabled on a VR device, the user can see his or her own hands moving in the virtual world. Seeing one's own hands moving in VR provides a strong sense of embodiment—especially if the virtual hands correctly represent the users' real-life movements. If and when hand-tracking technology develops sufficiently to accurately track signed language handshapes and movements, users will be able to sign while wearing VR devices and will be able to see their signs in the virtual space. The user's signing will also be visible to other online users, as in the example of VRChat in Section 2. Popular VR devices have recently improved their hand-tracking capabilities (Henry, 2022), deploying updates remotely to all users. These updates have mitigated some of the major limitations of hand-tracking, but some issues still remain (see Section 4.2).

In our ongoing research (Quandt et al., 2020), we have evaluated which ASL signs are best captured by existing hand-tracking technology. Most of our team members are deaf, which affords us the unique opportunity to self-evaluate the representation of signs in VR. For example, we evaluated a list of potential signs and decided whether different signs would work well with the Oculus Quest 2 and modified the signs as needed. Because there are often different signs for the same word, we track what variations of a sign are most compatible with the device's current hand tracking capabilities. With a Deaf-centric team, the ability to make quick informed decisions to make the whole system work well is an advantage.

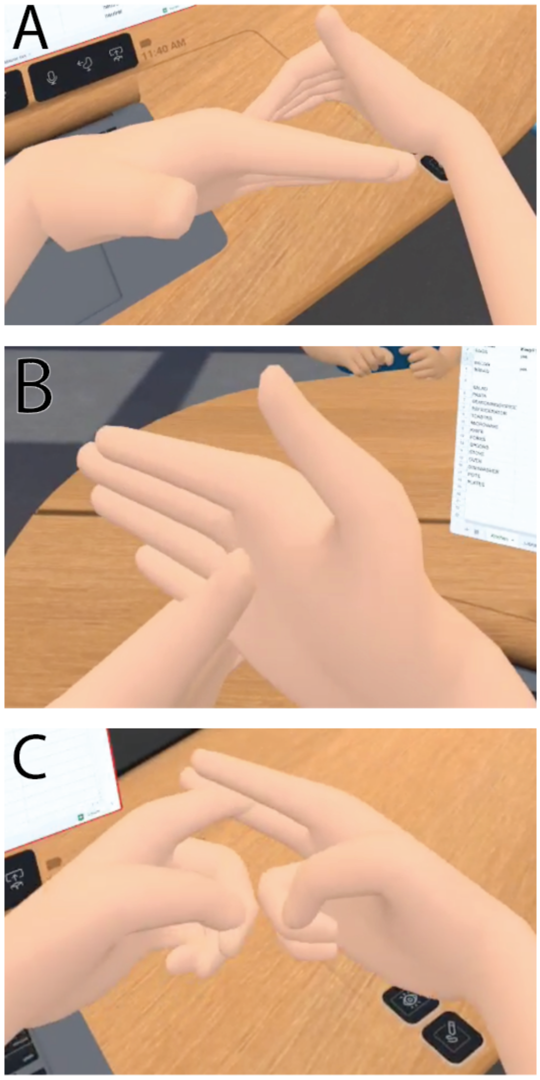

Figure 3. ASL signs as captured by Oculus Quest 2 (v38). A) the sign BREAD is represented well, with no occlusion or disfigurements. B) the sign WOOD resulting in unnatural overlap of the hands. C) the sign EGGS resulting in overlap and inaccurate handshape.

4.2 Challenges

Animating hands in real-time as a camera tracks a signer's hands is a significant challenge. Our team has identified several specific issues remaining before real-time ASL can be well represented in VR. All currently available VR headsets protrude several inches away from the bridge of the nose and eyes. Headsets with built-in hand-tracking capabilities have cameras embedded that look outward, and each camera has a cone of view that expands as the distance from the camera increases (Fig. 2). Close to the cameras, there are significant blind spots. Additionally, as the user moves their head to look around in space, the field of view which the cameras can see changes. Thus, the space in which the device can sense signs is inherently limited and changes depending on where the user is looking. Signs located outside the space in front of the signer are poorly captured by the cameras. This limited field of view causes technological limitations in recognizing three key visuospatial parameters of ASL: 1) handshapes, 2) physical location, and 3) facial expressions. These parameters are necessary to convey communication accurately and effectively in ASL and other signed languages (Friedman, 1976).

A crucial parameter of ASL is the physical location in which signed words are produced. In ASL, location in signing space is inherent to each sign's meaning. However, the many signs which are located near the body present a challenge for representation in VR. For example, the common sign for PARENT uses the "5" hand shape with all five fingers extended, touching the lower cheek, then touching the upper cheek. Because this sign includes touching the face near where the device rests, the normal production of the sign is prohibited, and the cameras cannot capture the sign.

Another challenge is the representation of ASL handshapes in VR headsets. While some current VR devices allow for improved hand-tracking, the technology still has limitations with recognizing certain handshapes, especially handshapes that require fingers crossing one another. For instance, the ASL handshape R involves the middle finger crossing over the index finger and is often not well tracked by current devices. Occlusions can also happen with two-handed signs if the hands or fingers cross one another, as with the word EGG (Fig. 3). In ASL, EGG is signed as the index and middle finger on both hands together, each hand forming the "H" handshape and tapping once, then moving downwards slightly away from each other. These shapes and movement tend to produce a great deal of occlusion when tracked by VR headset cameras.

To address the challenges related to sign location and occlusion, our current work focuses on signs that the hardware cameras can most accurately capture. When hands or fingers are placed on top of each other, it is difficult for the built-in cameras to see the hidden hand or fingers. The software interpolates the missing information, and often the resulting visualization is distorted (Fig. 3). Signs that avoid those handshapes and movement patterns are better represented in current VR devices.

Lastly, current hardware cannot capture a user's facial expressions. While it appears that developers are testing various approaches to capturing users' facial and eye movements while wearing a VR device (Wen et al., 2022), no options are commercially available at the time of writing. Naturally, given the importance of facial expression to signed languages, this still constitutes a major challenge for the progress of ASL in VR.

5. Conclusion

There is much room for improvement and undoubtedly, developers are racing to produce sophisticated hardware with better hand tracking, resolution, and capture for signing in virtual spaces. However, VR devices continue to be an obstacle given that in natural signed communication, many signs touch the face and body. We expect that advances in artificial intelligence will help solve some of the computer-vision related problems in this area. Signing in VR remains novel but brings much potential for learning, teaching, and interacting in virtual environments. Our research group is pursuing signing in VR in both the first- and third-person perspectives, and while the representation of signed languages improve in those two dimensions, we continue to identify remaining problems. Fluency and clarity of signing are essential and cannot be compromised without harming communication. Without the accurate representation of sign language, researchers risk compromising the representation of deaf people in virtual spaces.

6. Acknowledgements

The authors acknowledge the support of National Science Foundation grants #2118742, 2012924, and 2123626 to LQ and MM. We also thank Sarah Miller, Yiqiao Wang. Jianye Wang, Myles de Bastion, and Heather Smith for their contributions to this work.

7. Bibliographical References

Alsop, T. (2022). AR/VR headset shipments worldwide 2019-2023. Statista. https://www.statista.com/statistics/653390/worldwide-virtual-and-augmented-reality-headset-shipments/

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. WW Norton & Company.

Davis, R. N. (2019). *Deaf VRChat players are inventing their own sign language*. https://www.gamerevolution.com/news/617632-deaf-vrchat-players-asl-sign-language-index-vr

Friedman, L. A. (1976). *Phonology of a soundless language: phonological structure of the American sign language*. University of California, Berkeley.

Henry, J. (2022). *Meta rolls out hand tracking update on Oculus Quest 2*. Tech Times. https://www.Techtimes.com/articles/274566/20220421/meta-rolls-out-hand-tracking-update-oculus-quest-2-here.htm

Jörg, S., Hodgins, J., & O'Sullivan, C. (2010, July). The perception of finger motions. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pp. 129-133.

Kontra, C., Lyons, D. J., Fischer, S. M., & Beilock, S. L. (2015). Physical experience enhances science learning. *Psychological Science*, 26(6), 737-749.

Kontra, C., Goldin-Meadow, S., & Beilock, S. L. (2012). Embodied learning across the life span. *Topics in Cognitive Science*, 4(4), pp. 731-739.

Li, J., Vinayagamoorthy, V., Williamson, J., Shama, D. & Cesar, P. (2021). Social VR: A New Medium for Remote Communication and Collaboration. In *CHI EA '21: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 81, pp. 1-6.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher*, 42(8), pp. 445-452.

Quandt, L. C., Lamberton, J., Willis, A. S., Wang, J., Weeks, K., Kubicek, E., & Malzkuhn, M. (2020). Teaching ASL signs using signing avatars and immersive learning in virtual reality. *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '20)*, 98, 1-4.

Quandt, L. C., Willis, A. W., Schwenk, M., Weeks, K., &

Ferster, R. (2022). Attitudes toward signing human avatars vary depending on hearing status, age of signed language acquisition, and avatar type. *Frontiers in Psychology*, 13.

Shao, Q., Sniffen, A., Blanchett, J., Hillis, M. E., Shi, X., Haris, T. K., Liu, J., Lamberton, J., Malzkuhn, M., Quandt, L. C., Mahoney, J., Kraemer, D. J. M., Zhou, X., & Balkcom, D. (2020). Teaching American Sign Language in mixed reality. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, 4, 152.

Weisberg, S. M., & Newcombe, N. S. (2017). Embodied cognition and STEM learning: Overview of a topical collection in CR: PI. *Cognitive Research: Principles and Implications*, 2(1), pp. 1-6.

Wen, L., Zhou, J., Huang, W., & Chen, F. (2022). A Survey of Facial Capture for Virtual Reality, in *IEEE Access*, vol. 10, pp. 6042-6052.