Reducing Selection Bias in the Training Data of ASL Champ! To Improve the Sign Language Recognition (SLR) System

By: Nushla Pradhan and Laine Silverman

Abstract

American Sign Language (ASL) is a natural language that is critical for effective communication within the Deaf community and also to bridge the gap between hearing and Deaf or Hard-of-Hearing individuals. Conventional methods of ASL learning *apart* from in person classroom instruction provide foundational knowledge but often lack the immersive and interactive elements. People often opt to learn ASL through textbooks and videos due to the limited availability of proficient ASL instructors, lack of other educational resources and limited time. This creates challenges of replicating real-life conversational scenarios and lack of real time feedback.

To address these limitations, Virtual Reality (VR) technology has emerged as a promising tool for ASL education by offering an immersive learning environment where users can practice ASL, a 3D language in a three dimensional space to enhance their learning as well as retain the language. The motivation behind this study is to explore the efficacy of the VR-based ASL learning platform, ASL Champ! and its potential to revolutionize the way ASL is taught and learned. Additionally, this study will also evaluate its Sign Language Recognition (SLR) system by addressing the research question: "How does reducing the selection bias in the training data of ASL Champ! affect the accuracy of its Sign Language Recognition (SLR) system?"

Literature Review

Background

American Sign Language (ASL) is a visual language with its own set of linguistic parameters such as handshape, palm orientation, movement, location, and non-manual markers. ASL is the third most common language in the US and is used primarily by Deaf or Hard of Hearing individuals (World Health Organization (n.d.)). Language acquisition is a crucial aspect for cognitive and social development and is typically a natural process passed down from parents to children. However, for most Deaf individuals, this process differs significantly as 9 out of 10 Deaf children are born to hearing parents. Hearing parents often have no prior experience with ASL, creating barriers and missed opportunities for language learning and cultural integration for Deaf children. Therefore, learning ASL is vital to not just communicate with Deaf and Hard of Hearing individuals but to also learn about the Deaf community and their rich culture. It can also have cognitive benefits especially enhancing visual spatial reasoning among ASL users.

Because ASL is its own language with its own grammar separate from English, learning ASL and English leads to being bilingual. Bilingualism has many cognitive benefits; such as enhanced memory, attention, reasoning and problem solving. This is just one of the many benefits of learning ASL. Not to mention that without a communication barrier, stronger bonds can be formed between Deaf and hearing people, and help bridge the two worlds together. Outlining all the benefits and importance of learning ASL, it becomes even more crucial to drive innovation in ASL education and language learning technologies.

ASL in VR

The use of virtual reality (VR) platforms for teaching ASL presents significant potential for enhancing the learning experience. VR can influence various educational outcomes, including memory, attention, visuospatial skills, performance, cognitive and linguistic development, communication, and social skills (Ghoul & Othman 2022). ASL is a complex language, that utilizes various upper body movements, including hand gestures, facial expressions, lip-reading, head nods, and body postures to convey information (Gupta & Rajan, 2020; Chowdhry et al., 2013; Cheok et al., 2019; Butt et al., 2019). Therefore, this complexity continues to challenge intelligent systems to create effective recognition systems that accurately interpret the variations in sign languages to provide a comprehensive learning experience.

Early ASL recognition systems primarily focused on translating signs into English text but often failed to capture the full complexity of ASL. These systems struggled to incorporate aspects such as facial expressions and body movements, which are crucial in ASL. Especially the fact that ASL is a 3D language and needs to be reflected that way (Adeyanju 2021). While there have been developments in depth-based hand pose estimation for sign language detection, integrating high precision pose estimation for immersive virtual reality remained a challenge (Supančič et al., 2018). For instance, recognizing hand shapes and positions even when hands interact with nearby objects is necessary. Therefore, it's crucial to consider variability in motion position, hand shape, and position of body parts while developing automatic sign language recognition systems (Adeyanju 2021).

In traditional ASL learning methods, it’s important to have face to face interaction with the teacher as this allows to truly capture the essence of the language. Additionally, learning from a Deaf person is of the utmost importance. If you would like to learn a language, it is best to learn from someone who primarily speaks or uses that language themselves. They can offer their students not only real-time feedback in an authentic way, but also deeper linguistic knowledge of the language.

Oftentimes, hearing people who are not capable of teaching ASL believe that they are. They self-report that they are fluent or qualified however, they lack the experience necessary to effectively teach it. It’s also important to take Pidgin Signed English and Signed Exact English into account. Neither of these are real languages and are often mistakenly categorized as ASL. Knowing this distinction is critical to understand in the academic context of ASL. Hearing people are more likely to teach PSE or SEE instead of true ASL. This can lead to learning ASL in a way that is not genuine or respectful of the Deaf community. Lastly, teaching ASL should primarily be Deaf individual's responsibility, because of their credible knowledge of the language. If a hearing person is selected as a teacher of ASL, that clearly undermines the expertise of Deaf educators and is also inequitable.

The linguistic components of ASL including non manual markers such as facial expressions and body movements make it extremely difficult to represent ASL in a 2D environment. Similar to how changing tone changes the meaning of what is being said, in ASL there are different ways to sign something that can change the meaning. Because these intonations are visual as opposed to vocal, they need to be captured in a different way. Thus comes the obstacle of representing ASL and its 3D nature.

ASL Champ

ASL Champ is an ASL learning platform in a VR environment that facilitates immersive interaction and real-time feedback for learners. The game uses a signing avatar that facilitates the interactive learning process powered by an ASL recognition system using deep learning in the

VR environment. The game will provide five different learning environments including coffee shop, maker space, library, outside and kitchen.

Advanced motion-capture technology powers an expressive ASL teaching avatar within an immersive three-dimensional environment. The teacher demonstrates an ASL sign for an object, prompting the user to copy the sign. Upon the user's signing, a third-party plugin executes the sign recognition process alongside a deep learning model. Depending on the accuracy of a user's sign production, the avatar repeats the sign if the learner was incorrect or introduces a new one. We gathered a 3D VR ASL dataset from fifteen diverse participants to power the sign recognition model. The proposed deep learning model's training, validation, and test accuracy are 90.12%, 89.37%, and 86.66%, respectively. The functional prototype can teach sign language vocabulary and be successfully adapted as an interactive ASL learning platform in VR(Alam 2024).

Usage and Target Audience

This project is mainly targeting hearing parents of Deaf children and learners who are interested in learning ASL. Most of the time, these children are the first Deaf person that their parents meet. It is vitally important for parents and children to be able to communicate with each other and this project can eliminate that communication barrier. You might ask, "why not just go to a class?" or "why not watch online videos?". Let's say hypothetically that parents have just had a child who is Deaf or has hearing loss. New parents do not have the time to go to a class; additionally there is no guarantee that a class is offered in their area. Therefore, going to a class might not be possible. Secondly, watching videos doesn't give real-time feedback and can lead to incorrect signing. Both of these problems are solved by using ASL Champ!. These new parents can put on the VR headset while her newborn is sleeping for a few minutes everyday and learn some new signs intermittently when they have the time.

Methodology

Figure 1: PI Dr.Lorna Quandt testing out VR technology

The technology that we are using is the Oculus Quest 2 VR device, MiVRy Unreal Engine plugin and a deep learning model for sign detection. The deep learning model is the first of its kind when it comes to VR understanding sign language. This model allows us to train the AI to be able to recognize different signs from different signers. From native signers to later in life bilinguals, across all races, genders, and ethnicities, as well as left-handed and right-handed signers. In this way, the system can recognize whether or not a sign is correct based on the movement and facial expressions and not because of a bias towards a certain kind of signer.

Because this technology is so new, the amount of vocabulary is limited; however, with more time and advances in technology, we hope the vocabulary will become more robust. After building up a list of vocabulary, we are hoping to include full sentences and grammar learning into our system so that this tool can be used most applicably.

Through watching this research unfold, we noticed that the system itself had a right handed bias. This is most likely due to the fact that most of the participants who trained the system were right-handed, approximately 20%. In the future, we hope to recruit more left handed participants so that we can get rid of this bias.

Next, is the design of the system and the avatar. The system design incorporated input from ASL educators, Deaf ASL users, human-computer interaction engineers and digital 3D designers. In the past simulation of the coffee shop, the 3D design was extremely intentional given that the team wanted the user to feel like they were truly in the coffee shop. This intention will be consistent with the maker space after we input the data collected from this study.

The avatar is based on the motion capture recording of a native Deaf signer. This ensures authenticity of the avatar being based on a user of this language natively. The look wanted for the avatar was to have it seem natural, aesthetically pleasing, appearing human-like but not hyperrealistic. Mainly, the goal of the avatar is to provide the ideal ASL teacher; someone who is knowledgeable, competent, kind, and professional. This avatar uses repetition as many language teachers use to ensure that their students understand. It also introduces signs in 3 groups with 3 signs each (9 total for each environment) in order to break down the amount of material the user is learning and not have it be too overwhelming. And lastly and most importantly, the avatar is able to give real-time feedback such as “yes, good job” or “no, try again”. During usability testing, users said that they would be interested in learning ASL this way.

Procedure

The data collection procedure involves recruiting 11 participants to train the data set within a VR environment specifically designed for this study. The VR environment, set up in the maker space, included nine different ASL signs. Each participant was required to produce 20 renditions of each sign using the VR headset. The data gathered from these sessions trains the sign language recognition model through deep learning.

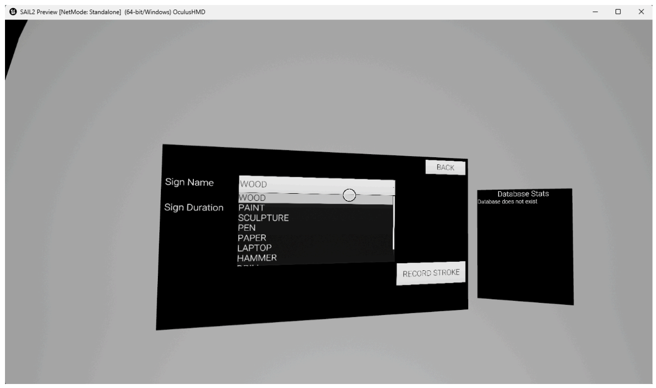

Figure 2: What the user sees in the oculus quest 2 headset when they are training the dataset.

Results

Demographics

62% of our participants have used VR before, and 37.5% had never used VR before. We found that the people who had used VR before completed the study faster because they were used to being in a virtual environment. 71.4% of our participants were female while 28.6% of the participants were male. 42.9% were Deaf while the rest were hearing. We were very happy to have a good variety of hearing status in order to represent all signing groups. 86% of the participants rated themselves at the expert level of ASL while 14% said that they were intermediate. Throughout data collection, many participants struggled to answer this question as to rating their ASL fluency. Many were not confident that they were an expert but felt that they were a higher level than intermediate. Regardless of their rating, all participants had at least 5-10 years of sign language experience.

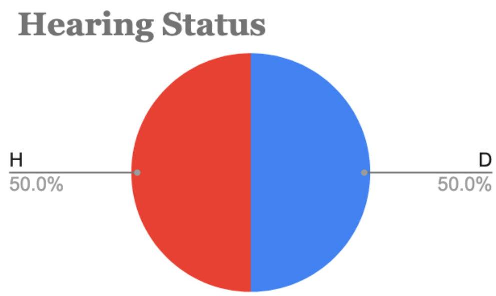

Figure 3: Demographic information: hearing status.

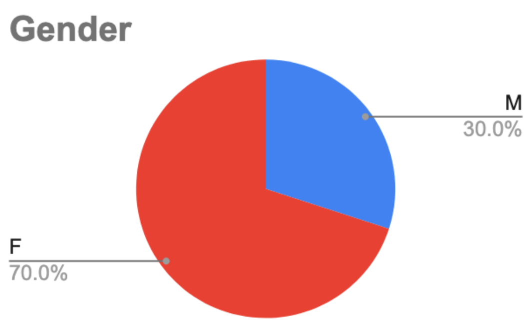

Figure 4: Demographic information: gender.

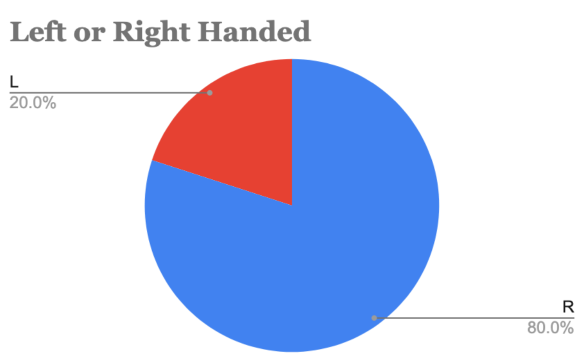

Figure 5: Demographic information: left vs right-handedness.

Quantitative Data: Accuracy of the Model

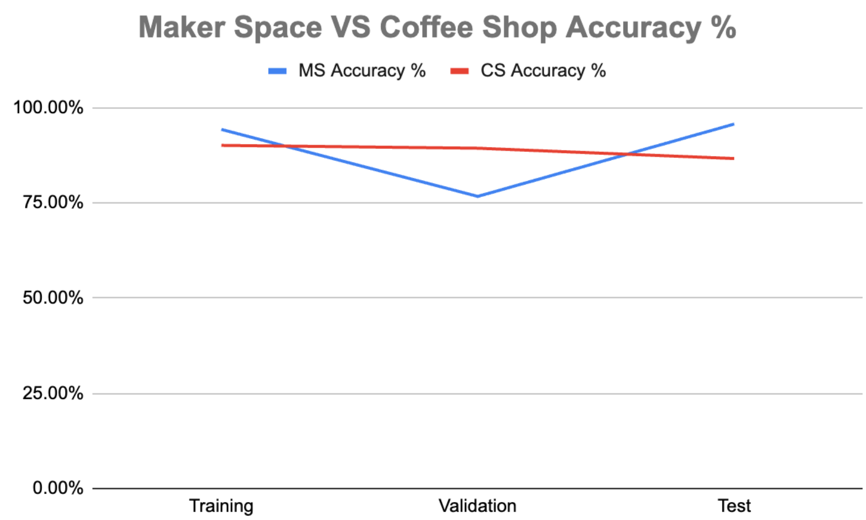

Figure 6: Maker Space and Coffee Shop accuracy percentages compared.

We split the data 80% to 20% to train and validate the model. Our past virtual environment, the coffee shop, had 90.12% training, 89.37% validation, and 86.66% test. Whereas our new virtual environment, the makerspace, had 94.31%, 76.72%, and 95.72% respectively. This shows improvement in both the training and test in the current model.

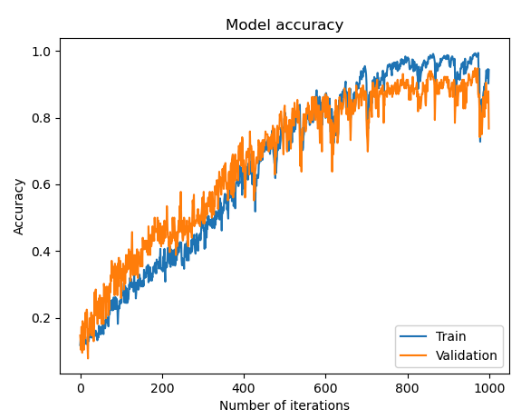

Figure 7: Model accuracy for Maker Space

As seen here, the model was trained 1000 times. Initially, the accuracy was close to 0.

But overtime, the accuracy increased. After around 700 iterations, the accuracy plateaus won't see more improvement.

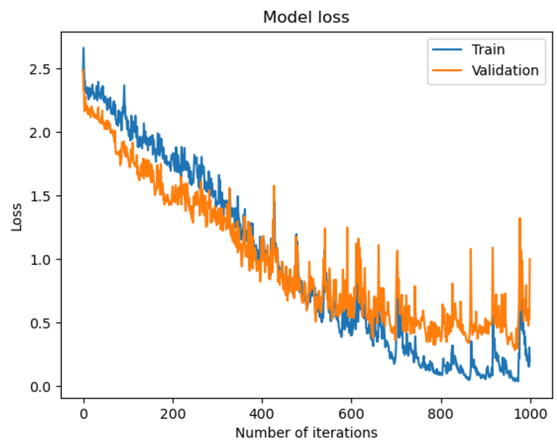

Figure 8: Model loss for Maker Space

Model loss calculates the difference between the actual value and the predicted value. Initially, the loss is 2.5 and overtime the loss is near 0. Similar to the accuracy, the function flattens after about 700 times.

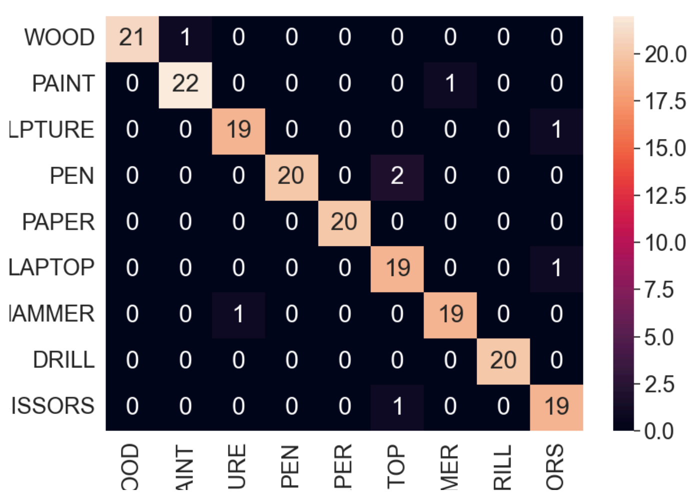

Figure 9: Confusion matrix of Maker Space Model

The confusion matrix is a prediction summary in matrix form that shows the prediction accuracy of each of the nine signs. The rows represent the predicted value whereas the columns represent the actual value. As shown, the diagonal represents the highest accuracy. For example, out of 22 times, WOOD was recognized correctly 21 times. However, it was once incorrectly categorized as PAINT. Next, out of 23 times, PAINT was correctly categorized 22 times, but was incorrectly labeled as HAMMER once. SCULPTURE was correctly recognized 19 out of 20 times and incorrectly categorized one time as SCISSORS. Out of 22 times, PEN was correctly labeled 20 times, and was mistakenly labeled as LAPTOP twice. LAPTOP was correctly recognized 19 out of 20 times, and labeled incorrectly once as SCISSORS. HAMMER was correctly seen 19 out of 20 times, only once being mistaken for SCULPTURE. SCISSORS was correctly labeled 19 out of 20 times and was once mistaken for LAPTOP. Both PAPER and DRILL were correctly labeled flawlessly!

Qualitative Feedback

Throughout conducting our study, many of our participants gave feedback as to improving the VR experience, ranging from the UI experience to personal feedback. A participant asked us to clarify a sign and we realized that our videos were contradictory for PEN/PENCIL. PEN only requires doing the sign once, whereas PENCIL requires the signer to sign PEN twice. A few of our participants felt self-conscious about their production of signs saying things like “you know when you say a word over and over again and it starts to feel like it’s not a real word anymore? That’s how I feel right now”. Additionally, a few participants were concerned about the VR causing dizziness or motion sickness due to neurological disorders, etc. However, both of them reported it not being an issue with our VR.

On the UI side of things, participants complained that our scroll bar was too thin and was hard to easily use. In order to combat this, we have since made the scroll bar thicker, therefore easier to grab and use. Another comment participants made was to move the most recent sign counter to the top of the list on the right side of the screen. Once the participant has signed enough of the nine signs, the list gets long and they can't see the counter which defeats the whole purpose of having a counter in the first place. Coding the program to move the most recent to the top of the list makes it easier for the user so that they will never need to scroll on the right side of the screen. Lastly, our second participant's counter started at 20 as a direct result of following up our first participant. This caused confusion because the participant needed to sign each sign until the counter got to 40. Whereas if the counter started at 0 brand new for each participant, this wouldn't be a concern.

Discussion

Based on our results on the accuracy comparison between the Maker Space and the Coffee shop simulation, there are clear inconsistencies between the two functions. Despite the high test accuracies in the Maker Space data, there is high variability between Validation Accuracy(76.72%) and Test Accuracy(95.72%). In an ideal model, the training, test and validation accuracies should be similar to demonstrate consistency. This inconsistency implies that the model requires more data to determine the holistic accuracy of the system. However, the difference in accuracy of the signs cannot be directly compared as both environments had completely different signs.

Therefore, the Model loss function was also used to evaluate the prediction accuracy of the maker space data set which demonstrates a wider gap between the validation and training accuracy functions as the number of iterations increase. The validation loss is greater than the training loss in this scenario which may indicate that the model is underfitting. This occurs when the model is unable to accurately model the training data, and consequently creating larger margin of errors.

To address these issues, it is important to consider the differences in the sample size and variability of signs as well as diversity of participants who trained the data set. Additionally, previous data sets in the coffee shop simulation did not include data from left handed signers. With 25% left handed signs in the data set, the model performance could have potentially degraded due to this change. Geographic limitations due to limited access to signers outside of the campus community also contributes to homogenous datasets. Therefore, it is important to include a more diversified data set for improved model performance.

Future Directions: Usability Study

In our new usability study, we aim to test user experience with learning ASL in VR. We will be building on our past usability study, which focused on testing the user experience with participants who are hearing and new to signing. This was in order to evaluate if a hearing non-signer can use SAIL 2's features rather easily. After the VR experience with the signing avatar, the users did a survey where they rated their experience with VR, sign recognition, immersion in the VR and aesthetics of the coffee shop simulation. Previously, we noted that participants who had some exposure to VR or avatars understood the instructions better, while some required further clarification. The previous usability study also focused more on quantitative data on individual ratings to understand user experience.

Considering the previous study, we plan to introduce new methodologies and make some alterations to the current structure. In order to evaluate whether VR learning is the most effective way of learning ASL, we will conduct a comparative analysis between ASL Champ! and traditional video based learning. This comparison study will involve two groups of participants: one group will use ASL Champ! and the other will learn through instructional videos. We will also explore new ways to engage users by including interactive tasks in both ASL Champ! and the videos for a fair comparison.

Additionally, we are also interested in learning about their retention of signs as another metric to evaluate the effectiveness of VR based learning. We plan to send out an online quiz to the participants a month after learning these signs and compare the retention rates among both groups. In practice, this quiz will test both expressive and receptive understanding. For example, a video of someone signing PAINT will pop up and the user will need to correctly select from 4 options which word it is. Another example could be the word ‘paint’ appearing on the screen and the user must correctly sign PAINT.

By addressing past challenges and adjusting our approach, we aim to enhance the accuracy and reliability of our usability testing outcomes by understanding the user’s comfort with the VR setup and ASL learning tools.

Conclusion

Language education is now able to take on a whole new level especially while using VR technology such as ASL Champ to provide a fun learning experience for users to learn ASL. This study focused on the impact of reducing selection bias in training data on the accuracy of ASL Champ!'s Sign Language Recognition (SLR) system. The methodology included using a deep learning model within the VR space in order to train the system. The overall test accuracy reached up to 95.72% which shows improvement of the model and its ability to successfully recognize the signs. But the difference between the validation score and test accuracy shows that the data itself was not very diverse. Secondly, the difference between the validation and the test accuracy shows that we need more data and diversity in those datasets to get a holistic picture of the SLR system's performance. Recruiting more left-handed signers is needed to avoid using those data points as outliers and to get more accurate insights about the model's performance.

In the future, we would like to conduct a usability study that compares VR-based learning vs video-based learning of ASL. To do this ,we will send out a survey to participants to see how many signs they have retained over the time of about a month. ASL Champ! certainly has huge potential to revolutionize ASL education and bridge the gap between Deaf, Hard of Hearing and Hearing communities by making language learning more accessible. Therefore, it becomes even more important to address the biases within the SLR system, acknowledge the need for improvements and continue to adapt with emerging technologies within the VR space.

References

Adeyanju, I. A., Bello, O. O., & Adegboye, M. A. (2021, December 8). *Machine learning methods for sign language recognition: A critical review and analysis*. Intelligent Systems with Applications. https://www.sciencedirect.com/science/article/pii/S2667305321000454

Al-Hammadi, M., Muhammmad, G., Alsulaiman, M., Abdul, W., Bencherif, M. A., Alrayes, T. S., Mathkour, H., & Mekhtiche, M. A. (n.d.). *Deep Learning-Based Approach for Sign Language Gesture Recognition with Efficient Hand Gesture Representation*. IEEE Xplore. https://ieeexplore.ieee.org/

Cabanillas-Carbonell, M., Cusi-Ruiz, P., Prudencio-Galvez, D., & Salazar, J. L. H. (n.d.). *Development of a mobile application for learning sign language ...* iJIM. https://find-testing.sdl.edu.sa/EdsRecord/edseee,edseee.10372883?sid=134279

Economou, D., Russi, M. G., Doumanis, I., Mentzelopoulos, M., Bouki, V., & Ferguson, J. (n.d.). *Using Serious Games for Learning British Sign Language Combining Video, Enhanced Interactivity,andVRTechnology*.Uclan.https://clok.uclan.ac.uk/35229/1/35229%20jucs_26_08_0996_1016_economou.pdf

Ghoul, O. E., & Othman, A. (n.d.). *Virtual Reality for Educating Sign Language Using Signing Avatar: The Future of Creative Learning for Deaf Students*. IEEE Xplore. https://ieeexplore.ieee.org/Xplore/home.jsp

I., Hauptman, S., Kosec, P., & Debevc, M. (1970a, January 1). *Assessing the efficiency of using augmented reality for learning sign language*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-319-20681-3_38

I.A. Adeyanju, O.O. Bello, M.A. Adegboye, Machine learning methods for sign language recognition: A critical review and analysis, Intelligent Systems with Applications, Volume 12, 2021, 200056, ISSN 2667-3053, https://doi.org/10.1016/j.iswa.2021.200056.

Md Shahinur Alam, Jason Lamberton, Jianye Wang, Carly Leannah, Sarah Miller, Joseph Palagano, Myles de Bastion, Heather L. Smith, Melissa Malzkuhn, Lorna C. Quandt, *ASL champ!: a virtual reality game with deep-learning driven sign recognition, Computers & Education: X Reality*, Volume 4, 2024, 100059, ISSN 2949-6780, https://doi.org/10.1016/j.cexr.2024.100059 .

Metaxas, D., Dilsizian, M., & Neidle, C. (n.d.-a). *Scalable ASL Sign Recognition using Model-Based Machine Learning and Linguistically Annotated Corpora* . NSF Public Access. https://par.nsf.gov/

Ozkul, A., Kose, H., Yorganci, R., & Ince, G. (n.d.). *Robostar: An Interaction Game with Humanoid Robots for Learning Sign Language*. IEEE Xplore. https://ieeexplore.ieee.org/

Paudyal, P., Lee, J., Kamzin, A., Soudki, M., Banerjee, A., & Gupta, S. K. S. (n.d.). Learn2Sign:ExplainableAIforsignlanguagelearning.https://explainablesystems.comp.nus.edu.sg/2019/wp-content/uploads/2019/02/IUI19WS-ExSS2019-13.pdf

Quandt, L. (2020). Teaching ASL signs using signing avatars and immersive learning in virtual reality. *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. https://doi.org/10.1145/3373625.3418042

Quinto-Pozos, D. (n.d.). *Teaching American Sign Language to Hearing Adult Learners*. Proquest. https://www.proquest.com/

Rho, E., Chan, K., Varoy, E. J., & Giacaman, N. (n.d.). *An Experiential Learning Approach to Learning Manual Communication Through a Virtual Reality Environment*. IEEE Xplore. https://ieeexplore.ieee.org/

Schioppo, J., Meyer, Z., Fabiano, D., & Canavan, S. (n.d.). *Sign Language Recognition: Learning American Sign Language in a Virtual Environment*. USC. http://library.usc.edu.ph/ACM/CHI2019/2exabs/AACRcour02.pdf

Shaw, A., Wunsche, B. C., Mariono, K., Ranveer, A., Xiao, M., Hajika, R., & Liu, Y. (n.d.). *JengASL: A Gamified Approach to Sign Language Learning in VR*. WSCG Conference on Computer Graphics, Visualization and Computer Vision. https://www.wscg.cz/

Supančič, J.S., Rogez, G., Yang, Y. et al. Depth-Based Hand Pose Estimation: Methods, Data, and Challenges. Int J Comput Vis 126, 1180–1198 (2018). https://doi.org/10.1007/s11263-018-1081-7

Wang, J., Ivrissimtzis, I., Li, Z., Zhou, Y., & Shi, L. (1970b, January 1). *Developing and evaluating a novel gamified Virtual learning environment for ASL*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-031-42280-5_29

Wang, J., Ivrissimtzis, I., Li, Z., Zhou, Y., & Shi, L. (1970c, January 1). *Exploring the potential of immersive virtual environments for learning American sign language*. SpringerLink. https://link.springer.com/chapter/10.1007/978-3-031-42682-7_31

Wang, J., Iyrissimtzis, I., Li, Z., & Shi, L. (n.d.). *The Impact of 2D and 2D Gamified VR on Learning American Sign Language*. Arxiv. https://arxiv.org/pdf/2405.02512

World Health Organization. (n.d.). *Deafness and hearing loss*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss