Exploring the Potential of Immersive Virtual Environments for Learning American Sign Language

Jindi Wang[1] [0000-0002-0901-8587], Ioannis Ivrissimtzis[1] [0000-0002-3380-1889], Zhaoxing Li[1] [0000-0003-3560-3461], Yunzhan Zhou[1] [0000-0003-1676-0015], and Lei Shi[2] [0000-0001-7119-3207]

[1] Department of Computer Science, Durham University, Durham, UK
[2] Open Lab, School of Computing, Newcastle University, Newcastle upon Tyne, UK
{jindi.wang, ioannis.ivrissimtzis, zhaoxing.li2, yunzhan.zhou}@durham.ac.uk
lei.shi@newcastle.ac.uk

Abstract. Sign languages enable effective communication between deaf and hearing people. Despite years of extensive pedagogical research, learning sign language online comes with a number of difficulties that might be frustrating for some students. Indeed, most of the existing approaches rely heavily on learning resources uploaded on websites, assuming that users will frequently consult them; however, this approach may feel tedious and uninspiring. To address this issue, several researchers have started looking into learning sign language in a game-based environment. However, the majority of the existing work still relies on website-based designs, with only a very few proposed systems providing an immersive virtual environment, and there are no user studies comparing website-based and immersive virtual environments. In this paper, we present an immersive environment for learning numbers 0–9 in American Sign Language (ASL). Our hypothesis is that an immersive virtual environment can provide users with a better learning experience and that users will show a higher level of engagement compared to website-based learning. We conducted a questionnaire-based user survey, and our initial findings suggest that users prefer to learn in an immersive virtual environment.

Keywords: Human Computer Interaction (HCI) · Interaction Design · Empirical Studies in Interaction Design · ASL Learning

1 Introduction

The World Health Organization (WHO)[3] predicts that, by 2050, about 2.5 billion people will have some degree of hearing loss, and at least 700 million of them will need some sort of hearing rehabilitation. The rehabilitation training

[3] https://www.who.int/zh/news-room/fact-sheets/detail/deafness-and-hearing-loss

procedures for people with hearing loss include the use of sign language and various alternative sensory techniques such as voice reading, writing with fingers on the palm of the hearing-impaired, and vibration sensing. While the most common form of communication for those who are deaf is sign language, most people without hearing loss have never studied sign language, making communication between these two groups challenging. Thus, the learning of sign language has become a key topic in educational research to break down communication barriers between diverse groups.

In the past years, face-to-face teaching of sign languages has been severely affected by the COVID-19 restrictions, with alternative online approaches filling in the gap. The majority of the latest approaches to teaching sign languages [10, 11, 20] employ website-based tools, while approaches based on immersive virtual reality (VR) technology [29] are more sparse in the literature. The following are some of the few studies on the VR-based approach. Adamo *et al.* [1] proposed the use of an immersive 3D learning environment to increase the mathematical skills of deaf children by teaching mathematical concepts and ASL's math terminology through user interaction with fantasy 3D virtual signers and environments. Schioppo *et al.* [24] proposed a sign language recognition method using features extracted from data acquired by a Leap Motion controller from an egocentric view. The method was tested on the 26 letters of the ASL alphabet. In a related development, Phan *et al.* [21] used motion tracking to trial a number of different methods for providing user feedback in a sign language learning system.

For the purposes of our study, we created a learning environment employing widely used VR technology to provide users with an immersive environment for learning numbers 0-9 in American Sign Language (ASL). Several issues with existing website-based and immersive approaches were identified and addressed in the design of our system, including small datasets for training the gesture recogniser, a lack of real-world settings, and most importantly, a lack of user satisfaction for sustained engagement with the system. To improve user experience, inspired by the ASL Sea Battle [7], a sign language game that was created by Bragg *et al.* to facilitate the gathering of user data, we developed and introduced into the system a Whack-a-Mole type of game.

To the best of our knowledge, no previous user studies are focusing on comparing website-based with VR-based systems, especially for learning 0-9 in ASL. Hence, we conducted a user study based on the survey scheme proposed by Schrepp *et al.* [26], aiming at comparing differences in user experience between an immersive and a website-based environment. The website-based environment was developed by Quizlet [4]. To make fair comparisons, the design aimed at maximising the consistency between the two environments.

Summarising, the main research question motivating our work: ***"Does an immersive ASL learning environment provide better user experience to learn ASL compared to a web-based environment?"***, was looked into within the context of an immersive environment for learning numbers between 0 and 9 in ASL. Our main contributions are as follows:

[4] https://quizlet.com/560702085/asl-numbers-0-9-flash-cards/

1. We implemented a novel immersive virtual environment for ASL learning with a Whack-a-Mole type of game.
2. We provide initial evidence that immersive virtual environments can enhance users' learning experience and engagement when compared to website-based learning environments for ASL. This suggests that incorporating immersive elements into ASL education may be a promising direction for improving learning outcomes and user satisfaction.

2 Related Work

First, we briefly review sign language recognition, which is a crucial technology for computer-assisted sign language learning and an important technical part of our system. Then, we go over some current research on web-based sign language learning, concentrating on the issues of effectiveness and usability of the website's features. Finally, we review research on sign language as a communication tool in general, going beyond the learning of the language.

Sign language recognition: Bheda *et al.* [4] proposed a method based on deep convolutional neural networks (CNNs) to recognize images of gestures of the letters and digits in ASL. Kim *et al.* [16] proposed a novel sign language recognition method, which employs an object detection network for a region of interest (ROI) segmentation to preprocess the input data. Battistoni *et al.* [3] described a method for ASL alphabet recognition based on CNNs, which allows for monitoring the users' learning progress. Jiang *et al.* [13] proposed a novel fingerspelling identification method for Chinese Sign Language via AlexNet-based transfer learning and evaluated four different methods of transfer learning. Camgoz *et al.* [8] introduced a novel transformer-based architecture that jointly learns Continuous Sign Language Recognition and Translation while being trainable in an end-to-end manner. Zhang *et al.* [31] proposed MediaPipe Hands, a real-time on-device hand tracking pipeline to compute hand landmark positions from a single RGB camera frame for AR/VR applications. Goswami *et al.* [12] created a new dataset for ASL recognition and used it to train a CNN-based model for hand gesture recognition and classification. Finally, Pallavi *et al.* [18] presented a deep learning model based on the YOLOv3 architecture, reporting high recognition rates on the ASL alphabet.

Having reviewed the existing work on sign language recognition, we concluded that Mediapipe is the most suitable tool for the purposes of this paper, and thus, we used it for sign language recognition, benefiting from its highly accurate, real-time detection of hand landmark points. Moreover, as an open-source hand gesture detection framework from Google, it is well-documented and supported.

Website-based sign language learning: Kumar *et al.* [17] proposed a sign language translation system based solely on visual input, employing deep learning for accurate translation. Joy *et al.* [15] proposed SignQuiz, a finger-spelt sign learning application for Indian sign language (ISL), utilizing automatic sign language recognition techniques. Vaitkevičius *et al.* [28] presented a system capable of learning gestures using the data from the Leap Motion device and classifying them with Hidden Markov Classification (HMC). Bird *et al.* [5] used a late fusion approach for sign language recognition from multi-modal data. They significantly improved the overall accuracy compared to single-modality image classification (88.14%) and Leap Motion data classification (72.73%). John *et al.* [14] proposed a system aiming at handling regional variations in vocabulary and grammar through a common vision-based platform. Empe *et al.* [10] developed a smartphone app called SimboWika to help deaf primary school kids learn Filipino Sign Language. They were motivated by the observation that communication between those who have hearing loss and those who do not can be challenging, especially when considering the unique perspective of persons with hearing loss language, and the lack of proficiency in sign language by those without hearing loss. Estrada *et al.* [11] proposed a web tool aimed at kids with and without hearing difficulties. They used a game to help children learn sign language through play. Patricks *et al.* [20] developed the interactive website application Sign2Sign, integrating real-time sign language detection AI, dynamic 3D avatars, and conversation-focused sign language instruction to support sign language education.

We note that, despite the significant size of some of the research projects behind these proposed methods and systems, there is no systematic evaluation of how users are affected by their interactions with these systems and, ultimately, how well the sign language students are progressing. Thus, we argue that user research is necessary, and the questionnaire interview format seems to be appropriate for an initial investigation.

Sign language applications: Bantupalli *et al.* [2] presented a vision-based application for translating sign language to text, thus aiding communication between signers and non-signers. Their recognition model takes video sequences and extracts temporal and spatial features from them. Schnepp *et al.* [25] argued that an animated sign language dictionary is a valuable resource for caregivers learning to communicate with residents who use sign language. They developed such a tool using a human-centred design methodology. Samonte [23] developed an e-tutor system, assisting instructors with course delivery and assessment. Economou *et al.* [9] developed a Serious Game (SG) aiming at closing the communication gap between the able hearing people and those with hearing impairment. Their tool facilitates sign language learning, specifically targeting the adult population. Wang *et al.* [30] developed a gamified sign language environment with characteristics that users could personalise. They found that gamification improved user experience. The research we reviewed on sign language applications shows that dictionary searches and gamification can improve users' motivation, which informed our choice to include such features in our system design.

Our survey of the literature verified that, as we have already noted, the majority of the existing work utilizes web-based solutions for the creation of sign language learning interfaces. This is often the most convenient solution, making it easy for the user to access all the learning materials. However, on the other hand, such an approach can lead to users developing a sense of repetitiveness and feelings of boredom. This is an observation supported by user feedback on such systems, and in fact, some users would drop out of the learning process altogether due to a lack of motivation [19].

Thus, the main objective of our work is to investigate whether virtual environments can provide a better user experience in sign language learning. We built a VR-based system that provides users with an immersive experience and added a quiz and a small game to stimulate users' interest and improve their experience. We note that recent advances in VR technology mean that there is a more general trend of migration of online tasks from web-based systems on the 2D planar screen to the immersive 3D space [22]; our system does something similar for sign language learning. Moreover, noting the absence of user research on the subject, in order to determine if an immersive learning environment can indeed improve the user experience when learning a sign language, we invited two groups of users to undertake an interview survey. Our aims were to evaluate our system and compare the user experience of learning 0-9 in ASL with a website-based learning environment.

3 User Interface of Immersive Environment and Website

This section provides an overview of the key components of the proposed immersive environment and the main features and user interfaces to compare with the web-based learning environment, as shown in Fig. 1.

3.1 The Immersive Learning Environment

Fig. 1(a) shows the learning resources and the game in the immersive learning environment, from the user's perspective 45 degrees to the left. The whole scene was created in Unity (2020.3.32f1). Regarding the user's interaction with the system, we track their eye position with the HTC Vive Pro's eye-tracking feature and enable the functions of clicking or picking an object after 3 seconds of fixed attention by the user.

The image acquisition was done by an integrated camera, linked to a PC using openCV (3.4.2) [6]. Regarding gesture recognition, Mediapipe was used to detect the user's hand and extract a sequence of 21 feature points (p_0, p_1, p_2,...,p_{20}), corresponding to landmarks on the detected hand. We set p_0, the point at the bottom of the palm near the user's wrist, as the origin of the frame's coordinate system. Let (x_i, y_i) be the coordinates of the point p_i. They are normalized by

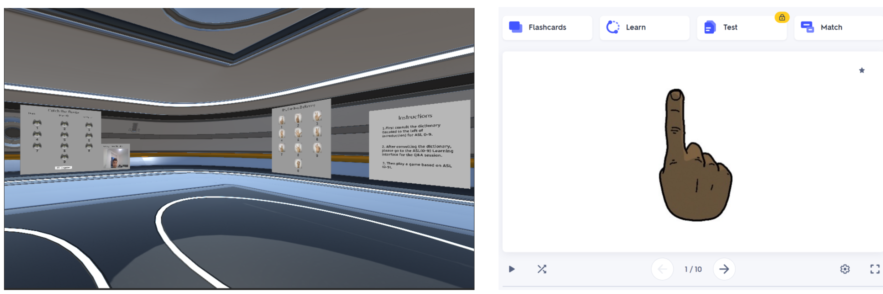
$$x_i = \frac{x_i - x_0}{x_{max}}, \quad y_i = \frac{y_i - y_0}{y_{max}}, \qquad i = 1, 2, \ldots, 20. \tag{1}$$

where

$$\begin{aligned} x_{max} &= max(\,|x_1 - x_0|, |x_2 - x_0|, \ldots, |x_{20} - x_0|\,) \\ y_{max} &= max(\,|y_1 - y_0|,\ |y_2 - y_0|,\ \ldots, |y_{20} - y_0|\,) \end{aligned} \tag{2}$$

and are then fed as a feature vector to the classifier. The classifier is a multilayer perceptron consisting of three fully connected layers, implemented in Python 3.6 and Tensorflow 2.6.0. We trained the classifier on a commodity PC with an RTX3080 GPU. The obtained recognition accuracy rates were above 90%, a result that was deemed sufficient for the purposes of this study as it is expected to support an overall smooth user experience.

The implemented user interfaces comprise four different modules: **Instructions**, **Sign Language Dictionary**, **Quiz**, and **Whack-a-Mole Game**, respectively. Each module is described in more detail below.



(a) (b)

Fig. 1. (a) The implemented immersive environment for 0-9 ASL learning. (b) The website-based 0-9 ASL learning environment.

Instructions In order to make the system easy to use, we created an **Instructions** interface, which also serves as the point of entry when the user logs into the system. In this interface, the user is introduced to the following three basic steps of the learning process, see also Fig. 2 (right).

1. Consulting the numbers 0–9 in an ASL dictionary (located to the left of the **Instructions**) to familiarise themselves with ASL 0–9 for about 5 minutes.
2. Users can self-assess their study of ASL 0–9 by accessing the **Quiz** module.
3. Users can engage in a game of Whack-a-Mole.

Sign language dictionary We created an ASL dictionary for users to search. Fig. 2 shows illustrations of how the user can sign/express the numbers 0 to 9 (to the left of the Instructions).

Quiz To improve the efficacy of the learning process, we integrated into the system a question-answer module that allows the users to assess their level of competence and, at the same time, exercise their signing skills by responding to a series of questions generated at random from a data bank. In the example

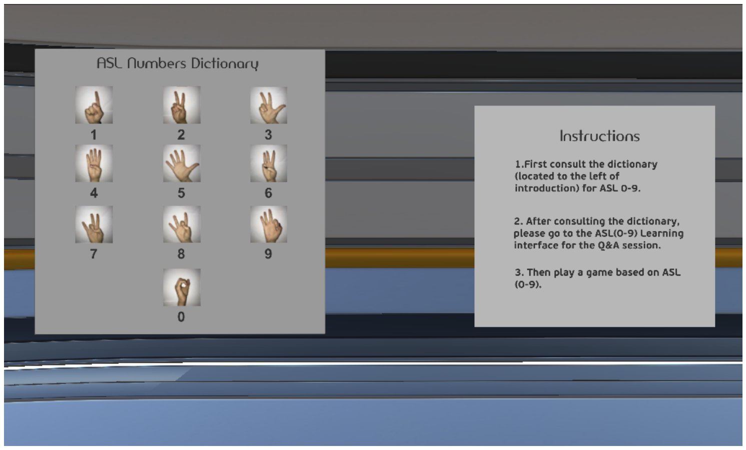


Fig. 2. The user's initial perspective of 0-9 ASL **Dictionary** and **Instructions**.

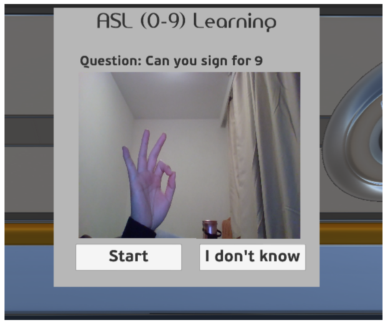
shown in Fig. 3(a), the system pulled from the data bank the question "Can you sign for 9?". After perhaps consulting the dictionary, and if they have developed the appropriate skill, the user can sign the word "9". Alternatively, they can select the "I don't know" option, and the system will demonstrate to them the appropriate expression. In that case, the user will also be advised to continue exercising until they indicate that they feel comfortable signing that digit by pressing the relevant button.

Whack-a-Mole game We adopted the Whack-a-Mole game and implemented a sign language-based version of it, aiming to engage users and improve their learning experience. In our game, as shown in Fig. 3(b), each location is marked by a unique identifier. If the user signs correctly the current position of the gopher, one point is added; otherwise, no point is awarded. By default, if the user does not sign the gopher's location within 3 seconds, a new gopher will appear. The total duration of the game is 30 seconds.

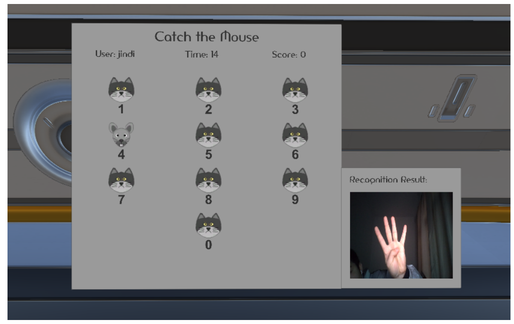
3.2 Website Environment

Fig. 1(b) shows the user interface of learning 0-9 ASL in a website environment, which includes **Flashcards**, **Learn** and **Match**.

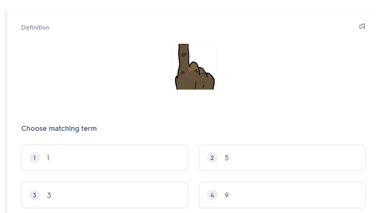
The **Flashcards** corresponds to the user interface of the sign language dictionary in the VR environment so that users can quickly become familiar with the representation of 0–9 numbers. Besides, the **Learn** module, like the Quiz interface, is used to further enhance the user's familiarity with the 0-9 ASL as shown in Fig. 3(c). Moreover, the **Match**, module as shown in Fig. 3(d), like the Whack-a-Mole game, is a match game used to enhance the entertainment of learning 0-9 in ASL.



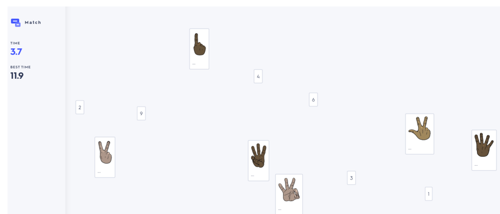
(a)



(b)



(c)



(d)

Fig. 3. (a) The 0-9 ASL learning quiz module in a VR environment. (b) The 0-9 ASL learning game in a VR environment. (c) The 0-9 ASL learning quiz module in a website environment. (d) The 0-9 ASL learning game in a website environment.

4 Methodology

With the intention of collecting user feedback, which would serve as the study's data source, we invited 15 users (M = 8, F = 7) to test the immersive sign language learning environment we developed, and 15 different users (M = 8, F = 7) to test the website-based ASL learning environment. Most users have minimal to no prior knowledge of ASL or other sign languages.

Users of the immersive environment first read the Instructions interface to become familiar with the learning procedure and then moved to the 0-9 ASL dictionary to spend 5 minutes studying the representations of the 0-9 digits. Users were encouraged to use the Quiz interface and take quizzes after they felt comfortable with them, and to use the Whack-a-Mole interface to play the game when they considered themselves familiar with them. In the website-based environment, users first queried the **Flashcards** for 5 minutes to get familiar with 0-9 in ASL, then they were encouraged to click the "Learn" button to take quizzes that contain multiple choice questions, and finally to click the "Match" button to enter the match game.

We adopted the user survey scheme proposed by Schrepp *et al.* [26], which is commonly used for evaluating the user experience in interactive systems, comprising six *scales*, each one representing a distinct user experience aspect: **Attractiveness**, **Efficiency**, **Perspicuity**, **Dependability**, **Stimulation**, **Nov-**

elty. Each scale is divided into either six or four more specific items, as shown in Table 1. Following the recommended protocol, we evaluated the user experience, on each of the 26 items, on a 7-point Likert scale ranging from -3 (fully agree with a negative term) to +3 (fully agree with a positive term). After the completion of the questionnaire, qualitative feedback was collected through follow-up interviews with the users.

Table 1. User experience questionnaire.

Attractiveness	**Perspicuity**
A1: annoying / enjoyable **A2**: good / bad **A3**: unlikable / pleasing **A4**: unpleasant / pleasant **A5**: attractive / unattractive **A6**: friendly / unfriendly	**P1**: not understandable / understandable **P2**: easy to learn / difficult to learn **P3**: complicated / easy **P4**: clear / confusing
Efficiency	**Dependability**
E1: fast / slow **E2**: inefficient / efficient **E3**: impractical / practical **E4**: organized / cluttered	**D1**: unpredictable / predictable **D2**: obstructive / supportive **D3**: secure / not secure **D4**: meets expectations / does not meet expectations
Stimulation	**Novelty**
S1: valuable / inferior **S2**: boring / exciting **S3**: not interesting / interesting **S4**: motivating / demotivating	**N1**: creative / dull **N2**: inventive / conventional **N3**: usual / leading edge **N4**: conservative / innovative

5 Result Analysis

In this section, we present the analysis of the data collected from the two groups, aiming at a comparison of user experience between the VR and the Web-based environments. Fig. 4 shows the distribution of scores for each question in the two learning environments.

Fig. 4 shows mean value scores for each item in the two learning environments. Clearly, the distribution of the mean scores of VR is more favourable than that of the web-based environment, indicating higher user satisfaction when using the former. We also notice that all mean scores for the VR are positive, while those for the Web are mixed. From qualitative user feedback, we found that the items with negative mean scores are mostly because users felt bored during the Web learning sessions.

To further compare the user experience in the two learning environments, we studied user feedback against the benchmark proposed in [27]. In that paper, the authors analysed a large database of questionnaire responses and proposed the benchmark intervals in Table 2.

- **Excellent**: In the range of the 10% best results.
- **Good**: 10% of results better, 75% of results worse.
- **Above average**: 25% of results better, 50% of results worse.

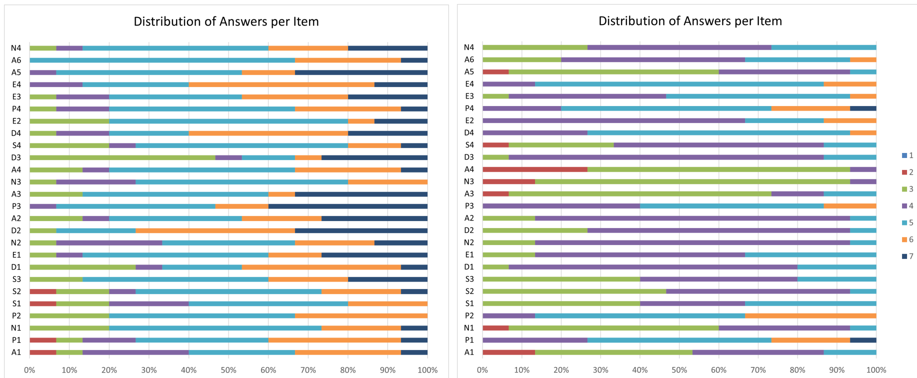


Fig. 4. The score distributions for the VR (left), and the Web (right) environments.

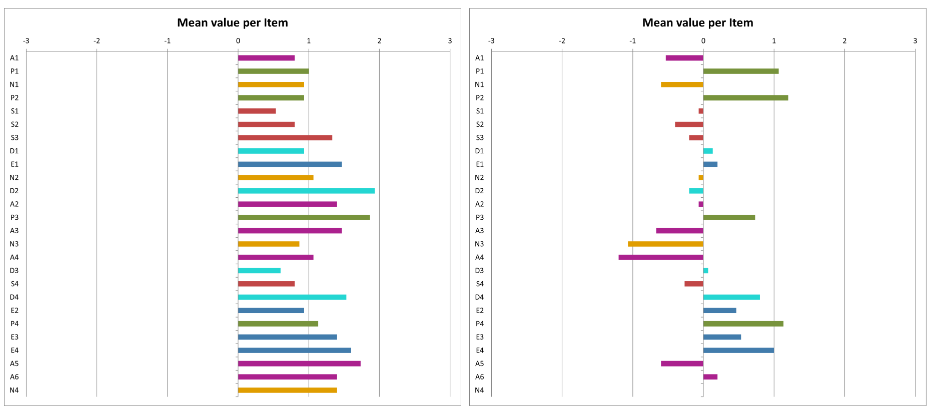


Fig. 5. Mean value scores for each item, in VR (left), and the Web (right) environments.

– **Below average**: 50% of results better, 25% of results worse.
– **Bad**: In the range of the 25% worst results.

Fig. 6 depicts the distribution of six scales in Benchmark intervals, in which different colours represent different benchmark intervals. We can observe that the mean values of six scales except **Stimulation** in the VR environment are above average; this indicates that a large number of users can accept learning ASL 0-9 in VR, while in the Web environment, the mean values of six scales are almost below the average, indicating that users have a low acceptance of learning ASL 0-9 on website. The six scales in the two learning environments are analysed below.

Attractiveness: the average score in VR is 1.311 (SD = 0.791) in the "above average" category, while for Web it is -0.478 (SD = 0.511) in the "Bad" category. This suggests that users find the VR environment more attractive for them

Table 2. Benchmark intervals for the user experience scales.

	Attractiveness	**Perspicuity**	**Efficiency**	**Dependability**	**Stimulation**	**Novelty**
Excellent	≥ 1.75	≥ 1.78	≥ 1.90	≥ 1.65	≥ 1.55	≥ 1.40
Good	[1.52, 1.75)	[1.47, 1.78)	[1.56, 1.90)	[1.48, 1.65)	[1.31, 1.55)	[1.05, 1.40)
Above average	[1.17, 1.52)	[0.98, 1.47)	[1.08, 1.56)	[1.14, 1.48)	[0.99, 1.31)	[0.71, 1.05)
Below average	[0.70, 1.17)	[0.54, 0.98)	[0.64, 1.08)	[0.78, 1.14)	[0.50, 0.99)	[0.30, 0.71)
Bad	<0.70	<0.54	<0.64	<0.78	<0.50	<0.30

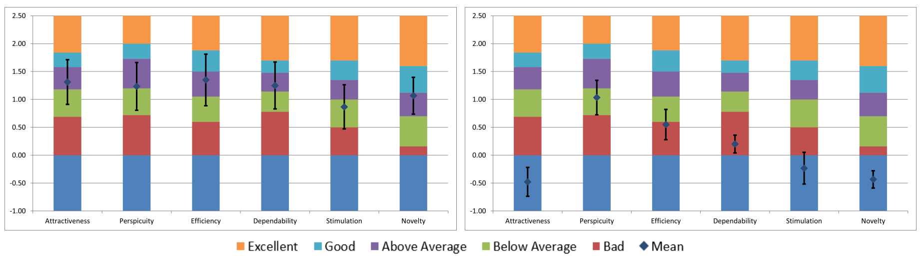


Fig. 6. Benchmark intervals for the six scales, in VR (left), and the Web (right).

to learn sign language. However, it received low scores from several users in terms of "annoying/enjoyable", as shown in Fig. 4. This could be due to the fact that the user needs to wear a headset which increases their learning burden. Besides, for the web-based environment, it received low scores in terms of "unpleasant/pleasant", as shown in Fig. 4. This may be because the content they were learning is not interesting enough.

Perspicuity: the average score for VR is 1.233 (SD = 0.848) in the "above average" category, while for Web it is 1.033 (SD = 0.611) in the "Below average" category. The two platforms have similar scores on this scale and are both around the average, indicating that the perspicuity of both platforms is well-received by users.

Efficiency: the average score for VR is 1.350 (SD = 0.915) in the "above average" category, while for Web it is 0.550 (SD = 0.536) in the "Bad" category. Although the score of VR is higher than that of the Web, the standard deviation of VR is also larger than that of the Web, indicating that VR fluctuates greatly under the efficiency scale. The possible reason is that some users have used VR equipment before and did not need to re-learn the use of VR equipment; while some users had no experience in using VR before and needed to learn the use of VR, which might increase their learning load.

Dependability: the average score in VR is 1.250 (SD = 0.829) in the "above average" category, while for Web it is 0.200 (SD = 0.316) in the "Bad" category. Although the standard deviation of VR on this scale is large, it means that some users feel that VR needs to be strengthened in terms of Dependability. Several users thought they executed the right gesture but were judged as being wrong. The back-end gesture recognition algorithm may be underperforming due to inadequate training data for particular gestures, resulting in low recognition accuracy rates. Future developments are planned to train recognition models to better generalize to user hand physiological variances.

Stimulation: the average score in VR is 0.867 (SD = 0.784) in the "Below average" category, while for Web it is -0.233 (SD = 0.563) in the "Bad" category. The score of VR on this scale is higher than that of the Web, because according to the feedback of users, learning sign language on the Web was only a selection with a mouse click, while in VR, sign language could be used to express numbers in an immersive way, which is more attractive to them.

Novelty: the average score in VR is 1.067 (SD = 0.651) in the "above average" category, while for Web it is -0.433 (SD = 0.306) in the "Bad" category. Today's users are acclimated to searching and learning content on the web, hence the Web's score on this scale is low. VR scores higher than the Web because many people have never learned in VR and find it novel.

In [27], the scales of the user experience questionnaire are grouped into *pragmatic quality* (Perspicuity, Efficiency, Dependability), and *hedonic quality* (Stimulation, Novelty), while Attractiveness is a pure valence dimension. Pragmatic quality describes task-related quality aspects, and hedonic quality describes the non-task-related quality aspects. Table 3 shows the mean scores for the two environments over these grouped scales, and the corresponding p-values. We note that, in all cases, we have statistically significant differences, especially regarding attractiveness and hedonic quality.

Table 3. Means and p-values for the three groups of scales.

Pragmatic and Hedonic Quality	VR	Website	P-Value
Attractiveness	1.31	-0.48	5.73216E-20
Pragmatic Quality	1.28	0.59	1.86984E-09
Hedonic Quality	0.97	-0.33	1.23737E-18

6 Discussion

We used VR technology to create an immersive environment for learning numbers in ASL and conducted a user study on it. We investigated whether immersive environments can provide a better user experience. Our hypothesis that users prefer to learn ASL 0-9 in an immersive, rather than a website-based environment, was tested with a questionnaire-based study on two groups of users.

We employed six assessment scales — **Attractiveness**, **Efficiency**, **Perspicuity**, **Dependability**, **Stimulation**, **Novelty** — on which users judged the system's design. The results show that our immersive ASL learning system performs effectively and can essentially meet user requirements for ASL learning. Thus, we were able to answer affirmatively the research question of whether an immersive ASL learning environment provides a better user experience to learn ASL. The analysis of environmental experience reveals that most users prefer to learn sign language in an immersive VR environment, probably because it is more experiential, while only very few users prefer to learn ASL via the web-based learning environment, primarily because they find it convenient and user-friendly to search for sign language information on web pages.

We note that, even though most users prefer immersive VR learning environments, the technology is still in its early stages of development. The system needs to be further streamlined and tweaked for specific functions to enhance the user experience. In particular, qualitative user feedback revealed the following limitations of the immersive VR learning environment.

6.1 Limitations

For each assessment scale, the main limitations of the system's design and implementation were as follows. **Attractiveness**: some users complained that there weren't any animated hints when sign language was correctly deciphered. The user's experience could be improved by including animated hints. **Efficiency**: some users have complained that our interface is not sufficiently automatic because playing the sign language game requires them to press the start button actively. **Perspicuity**: some users have reported that they did not know how to move around in the scene. **Dependability**: some users reported that they had made the correct sign language while playing the game, but the system judged that it was wrong, resulting in a lower score for the user. **Stimulation**: from the analysis of the stimulation scale results, a small number of users gave low scores, indicating that our system needs to be designed more creatively. **Novelty**: a small number of users thought that the system was not innovative enough, perhaps because they perceived the learning model as too easy.

Additionally, there were several methodological issues with our study that need to be considered. Firstly, with only 30 people taking part in the user survey, the sample size is relatively small and may not give an accurate representation of the intended audience. Secondly, as our study included only Chinese participants, the findings might not apply to people from different cultural backgrounds. To ensure broader validity of the findings, future studies should be based on larger and more varied samples.

7 Conclusion

We developed a virtual setting that gives users an immersive experience for learning ASL. We created four user interfaces corresponding to four distinct types of functionality, enabling users to easily comprehend the system's workflow and each phase of ASL learning: an instruction module, an ASL dictionary, a quiz module, and an ASL game based on Whack-a-Mole. To determine the acceptability of our UIs and evaluate user satisfaction with the virtual environment design, the findings of a user questionnaire were analysed ($N = 30$). The results indicated that users were generally satisfied with the virtual environment we built, and they preferred it against the website-based learning mode. Overall, the outcome supports our initial hypothesis that immersive virtual environments can improve users' experience of learning ASL.

In the future, we plan to include in our immersive environment more interactive elements such as backdrop movement, scene changes, and animation prompts. We also plan to implement various automatic settings to minimize the need for user interaction for the control of the system, and a follow-through user interface will be created so that the user can comprehend how to move the items in the scene. Furthermore, a more robust sign recognition model will be created, allowing us to include more sophisticated sign language learning material.

References

1. Adamo-Villani, N., Carpenter, E., Arns, L.: An immersive virtual environment for learning sign language mathematics. In: ACM SIGGRAPH 2006 educators program, pp. 20–es (2006)
2. Bantupalli, K., Xie, Y.: American sign language recognition using deep learning and computer vision. In: Proc. ICBD. pp. 4896–4899. IEEE (2018)
3. Battistoni, P., Di Gregorio, M., Sebillo, M., Vitiello, G.: Ai at the edge for sign language learning support. In: Proc. HCC. pp. 16–23. IEEE (2019)
4. Bheda, V., Radpour, D.: Using deep convolutional networks for gesture recognition in american sign language. arXiv:1710.06836 (2017)
5. Bird, J.J., Ekárt, A., Faria, D.R.: British sign language recognition via late fusion of computer vision and leap motion with transfer learning to american sign language. Sensors **20**(18), 5151 (2020)
6. Bradski, G., Kaehler, A.: Opencv. Dr. Dobb's journal of software tools **3**, 120 (2000)
7. Bragg, D., Caselli, N., Gallagher, J.W., Goldberg, M., Oka, C.J., Thies, W.: ASL Sea Battle: Gamifying Sign Language Data Collection. In: Proc. CHI-HFCS. pp. 1–13 (2021)
8. Camgoz, N.C., Koller, O., Hadfield, S., Bowden, R.: Sign language transformers: Joint end-to-end sign language recognition and translation. In: Proc. CVPR. pp. 10023–10033 (2020)
9. Economou, D., Russi, M.G., Doumanis, I., Mentzelopoulos, M., Bouki, V., Ferguson, J.: Using Serious Games for Learning British Sign Language Combining Video, Enhanced Interactivity, and VR Technology. Journal of Universal Computer Science **26**(8), 996–1016 (2020)
10. Empe, N.A.A., Echon, R.C.L., Vega, H.D.A., Paterno, P.L.C., Jamis, M.N., Yabut, E.R.: SimboWika: A Mobile and Web Application to Learn Filipino Sign Language for Deaf Students in Elementary Schools. In: Proc. R10-HTC. pp. 1–6. IEEE (2020)
11. Estrada-Cota, I., Carreño-León, M.A., Sandoval-Bringas, J.A., Leyva-Carrillo, A.A., Quiroz, H.X.C.: Design of a Web tool for teaching–learning of states and capitals of México through the Mexican sign language. In: Proc. ICITE. pp. 174–179. IEEE (2021)
12. Goswami, T., Javaji, S.R.: Cnn model for american sign language recognition. In: ICCCE 2020, pp. 55–61. Springer (2021)
13. Jiang, X., Hu, B., Chandra Satapathy, S., Wang, S.H., Zhang, Y.D.: Fingerspelling identification for chinese sign language via alexnet-based transfer learning and adam optimizer. Scientific Programming **2020** (2020)

14. John, A., Krishnan, R.H., Vinitha, A.M.: Language Recognition System: An Application Based Study with Special Reference to Sociolinguistics and Computational Linguistics (2021)
15. Joy, J., Balakrishnan, K., Sreeraj, M.: Signquiz: a quiz based tool for learning fingerspelled signs in indian sign language using ASLR. IEEE Access **7**, 28363–28371 (2019)
16. Kim, S., Ji, Y., Lee, K.B.: An effective sign language learning with object detection based ROI segmentation. In: Proc. IRC. pp. 330–333. IEEE (2018)
17. Kumar, S.S., Wangyal, T., Saboo, V., Srinath, R.: Time series neural networks for real time sign language translation. In: Proc. ICMLA. pp. 243–248. IEEE (2018)
18. Pallavi, P., Sarvamangala, D.: Recognition of sign language using deep neural network. Int. J. of Advanced Research in Computer Science **12**, 92–97 (2021)
19. Park, J.H., Choi, H.J.: Factors influencing adult learners' decision to drop out or persist in online learning. Journal of Educational Technology & Society **12**(4), 207–217 (2009)
20. Patricks, A.: Developing an accessible learning application for sign language (c) (2022)
21. Phan, H.D., Ellis, K., Dorin, A., Olivier, P.: Feedback strategies for embodied agents to enhance sign language vocabulary learning. In: ACM-IVA. pp. 1–8 (2020)
22. Reisoğlu, I., Topu, B., Yılmaz, R., Karakuş Yılmaz, T., Göktaş, Y.: 3d virtual learning environments in education: A meta-review. Asia Pacific Education Review **18**, 81–100 (2017)
23. Samonte, M.J.C.: An assistive technology using fsl, speech recognition, gamification and online handwritten character recognition in learning statistics for students with hearing and speech impairment. In: Proc. ICFET. pp. 92–97 (2020)
24. Schioppo, J., Meyer, Z., Fabiano, D., Canavan, S.: Sign language recognition: Learning american sign language in a virtual environment. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. pp. 1–6 (2019)
25. Schnepp, J., Wolfe, R., Brionez, G., Baowidan, S., Johnson, R., McDonald, J.: Human-centered design for a sign language learning application. In: Proc. PETRAE. pp. 1–5 (2020)
26. Schrepp, M., Hinderks, A., Thomaschewski, J.: Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: International Conference of Design, User Experience, and Usability. pp. 383–392. Springer (2014)
27. Schrepp, M., Thomaschewski, J., Hinderks, A.: Construction of a benchmark for the user experience questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence **4**(4), 40–44 (2017)
28. Vaitkevičius, A., Taroza, M., Blažauskas, T., Damaševičius, R., Maskeliūnas, R., Woźniak, M.: Recognition of American sign language gestures in a virtual reality using leap motion. Applied Sciences **9**(3), 445 (2019)
29. Wang, J., Ivrissimtzis, I., Li, Z., Zhou, Y., Shi, L.: Developing and evaluating a novel gamified virtual learning environment for ASL. In: INTERACT 2023. LNCS, Springer (2023)
30. Wang, J., Ivrissimtzis, I., Li, Z., Zhou, Y., Shi, L.: User-defined hand gesture interface to improve user experience of learning american sign language. In: International Conference on Intelligent Tutoring Systems. pp. 479–490. Springer (2023)
31. Zhang, F., Bazarevsky, V., Vakunov, A., Tkachenka, A., Sung, G., Chang, C.L., Grundmann, M.: Mediapipe hands: On-device real-time hand tracking. arXiv:2006.10214 (2020)