JengASL: A Gamified Approach to Sign Language Learning in VR

Alex Shaw, Burkhard C. Wünsche, Kevin Mariono, Aiden Ranveer, Manhui Xiao, Ryo Hajika, Yitong Liu
School of Computer Science, University of Auckland, New Zealand
lsha074@aucklanduni.ac.nz, burkhard@cs.auckland.ac.nz, gmar591@aucklanduni.ac.nz,
raid127@aucklanduni.ac.nz, mxia321@aucklanduni.ac.nz, ryo.hajika@auckland.ac.nz,
yliu959@aucklanduni.ac.nz

ABSTRACT

Learning sign language has many advantages ranging from being able to communicate with millions of hearing impaired people, to improving cognitive function and communication skills. Sign language is recognised as an official language in 74 countries, including Germany, Japan, and the UK. Despite that only a small percentage of people attempt to learn sign language.

In this research we investigate how virtual reality and gamification can be used to make learning sign language more enjoyable and motivating. We present JengASL, a gamified approach using 3D hand models, gesture recognition, and interactive gameplay in Virtual Reality to teach American Sign Language. We evaluate this system with a pilot study using eight participants and found that while it is less effective for sign memorisation than traditional 2D image-based learning methods, learning is more, but not significantly more, enjoyable and motivating.

Keywords

Virtual Reality, Sign Language, Teaching, Gesture Recognition, Gamification

1 INTRODUCTION

Learning sign language has numerous advantages ranging from being able to communicate with millions of people suffering from hearing difficulties [MV14] (an estimated 15-20% in America and New Zealand [BLC14, SFB+13]) to improving cognitive function and reasoning, memory, attention span, creativity, and communication skills [CCRV98, MLS06, CCP07].

Despite that sign language is not widely used and most people willing to learn it do so using books, videos, or through community-hosted events [MV14]. However, such teaching media might not always be available and/or provide little or no formative feedback. Furthermore, signs can be viewed differently depending on the angle at which they are observed, and 2D teaching materials cannot provide students with changes in depth and perspective. Virtual Reality overcomes some of these difficulties. By implementing 3D representations of hand gestures in a VR environment such as in [AVCA06], the user will be able to see different perspectives by tilting their head or walking to a different position.

The learning process is often inhibited by a lack of ongoing motivation. Learning a language takes time and effort. Gamification and Serious Games offer significant potential for improving motivation and persistence by combining the learning process with a more enjoyable gameplay activity, that also allows for more avenues of feedback and self-assessment through skill performance influencing performance in the gameplay tasks. Gamification has the potential to improve both user engagement and learning performance [ORCV17, SWL18].

Our research question is thus: Can we improve sign language learning using gamification in a Virtual Reality Environment?

2 RELATED WORK

2.1 Teaching Strategies

Teaching methods have been classified broadly into direct instruction, peer-teaching, and interactive teaching [MR17]. Interactive teaching is based on the idea that students need practical application to fully comprehend study material, motivating students to participate in teaching content and maintain concentration for longer. It also helps teachers to assess how well the student is actually learning the material. Feedback, in particular constructive feedback, is another factor in providing effective interactive learning [Sen18]. Good constructive feedback should be systematic, relevant and encouraging in order to achieve successful teaching [Ova91], and has been associated with increased student confidence and motivation [CR08]. Numerous papers and several systematic reviews have demonstrated the potential of VR, including gamification, in education [KLRWP17].

While numerous games exist for teaching sign language, few of them use VR or evaluate the effectiveness of the approach. "CopyCat" is an educational adventure game to help deaf children improve their language and memory abilities [ZBP+11]. The player interacts with the game's main character using sign language and can use a virtual tutor to learn the correct signs. The research focus of the paper is on the sign recognition system developed as part of the research and no evaluation of the game is provided.

"Sign my World" is a mobile video game for teaching Australian Sign Language (Auslan). The game uses a 2D cartoon like interface and interactive cards which associate words with videos of the Auslan sign for that word [KPN12].

Bouzid et al. developed a memory game where users have to match cards containing words and sign writing notation, which are interpreted by a 3D avatar using gestures [BKEJ16]. The authors performed a user study with 9 participants and report that based on video analysis the majority of users were engaged and enjoyed the game.

2.2 Gesture Recognition for Sign Language Applications

A crucial aspect of any system teaching sign language is the recognition of the sign language gestures in order to assess users' performance. Accurate gesture recognition can also enable a system to provide constructive feedback that targets a specific part of the user's movements as suggested in SignTutor [AAA+09]. Previously developed technologies for gesture recognition can be split into two categories:

Glove-Based gesture recognition involves the user wearing a glove with markers or sensors. The CyberGlove is one such tool, which measures the angles of hand joints and the position of the hand, which can then be used to train a neural network for recognising gestures [WS99, PMS+09, SLC15].

Camera-Based gesture recognition methods use cameras or other optical sensors to gain data from the user by computing gestures using a visual representation of the hands. This includes first determining what needs to be examined, e.g., the hands, and then tracking their movements to determine the gesture being signed [LWLD11, FHA+22].

2.2.1 Object Segmentation

The first step of many tracking methods is foreground/background segmentation. Holden et al. use image sequences from a single colour camera to recognize Australian Sign Language (Auslan) using skin colour detection and active contour models [HLO05]. Object segmentation can be effected by variations in surrounding colours and lighting [HLO05]. SignTutor requires users to wear different coloured gloves to counteract changes in background and lighting conditions and helping with segmentation if fingers/hands overlap [AAA+09]. Keskin et al. used two camera images and non-skin-coloured gloves as markers to separate the hands from complex backgrounds [KEA03].

Depth cameras are often less sensitive to changes in lighting and background. Mo and Neumann used the Canesta camera to estimate the pose of the hands with the assumption that it is the closest object to the camera within the depth threshold, based on finger boundaries that are also calculated from depth values [MN06]. The method failed with non-frontal poses and poses that cannot be modelled due to noise or positioning. Li and Jarvis tried to remove some of the noise that comes with depth mapping using Median filtering and segmenting the hand using a depth histogram method [LJ09]. Histogram binning is also used in SignTutor to determine hand regions, although rather than using depth data it uses the HSV colourspace [AAA+09].

The Leap Motion sensor computes the 3D positions of hands within a certain range of the sensor, but instead of a depth map it dynamically computes a set of hand points (palm, finger positions, hand orientation). This was used by Chuan et al. to recognize 26 letters of the American Sign language alphabet by extracting the position and length the fingers as well as the pose of the palm [CRG14].

2.2.2 Gesture Classification

Gesture classification is usually achieved using Machine Learning. Keskin et al. use a Hidden Markov Models (HMMs) and a 3D Kalman filter to reduce noise [KEA03]. SignTutor uses two Kalman filters, one on each hand to reduce segmentation noise and predict hand trajectories [AAA+09]. Chuan et al. used a k-Nearest Neighbour and Support Vector Machine for alphabet recognition, however, gestures that looked similar were often misclassified, possibly due to mislabelled data in the Leap Motion sensor [CRG14].

Chai et al. used an interesting approach to build a translation application for Chinese Sign Language using Microsoft Kinect [CLL+13]. In their algorithm, the movement trajectory of each word is first aligned to the same sampling point. A match is then performed with existing libraries to determine the gesture. Since movements were tracked in this algorithm, it is possible to determine dynamic gestures and account for varying hand motion speeds between different signers.

More recent work has used deep learning. For example, Kothadiya et al. use two deep learning models for Indian Sign Language to achieve 97% recognition accuracy over 11 signs [KBS+22]. Al-Qurishi et al. present a review of deep learning-based approaches for sign language recognition [AQKS21]. The authors conclude that the presented models are relatively effective for a

ISSN 1213-6972 (print)
ISSN 1213-6964 (online)

range of tasks, but none currently possess the necessary generalization potential for commercial deployment.

The majority of papers we found focused on sign recognition [ABA21]. SignTutor does perform teaching and assesses users accuracy, but is limited by using a 2D display and not having gamification, which means that users must be intrinsically motivated to use the application [AAA+09]. Hence our design will focus on the gamification of sign language learning.

3 DESIGN

Based on the above literature, constructive feedback and interaction plays a large part in whether or not students can effectively learn. JengASL's Design can be split into several components as illustrated in Figure 1, providing both teaching and feedback via Gesture Recognition.

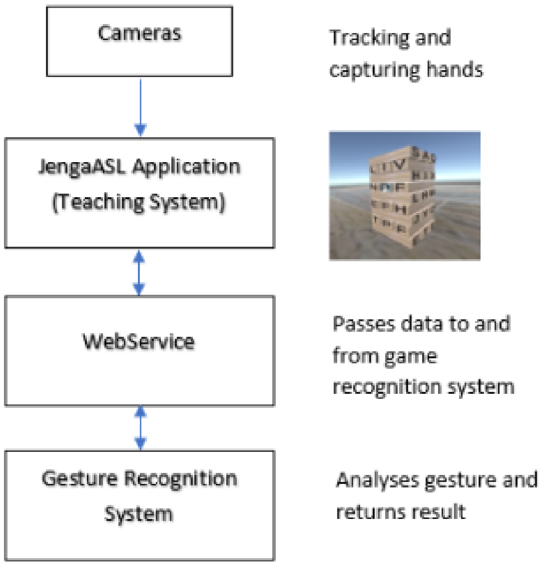

Figure 1: JengASL system architecture overview.

Cameras are accessed from within the JengaASL application to record the users hand gestures. Photos are then parsed in-game into a format suitable for gesture recognition, and data is passed to the gesture recognition system via a web-service.

3.1 JengASL

In order to make learning more interesting and interactive, we have integrated gamification into our VR teaching system. Our game consists of wooden blocks that are stacked as in the well known game "Jenga", each with letters attached to them. Jenga was chosen as a popular game that is both simple and can be played alone or with multiple players. A point system is used to keep track of how well the user is performing, and the game ends once the tower falls.

The user interacts with the system by first selecting a block with an associated letter, as seen in Figure 2. The user then sees an indication of the associated sign and is prompted to replicate it for the camera.

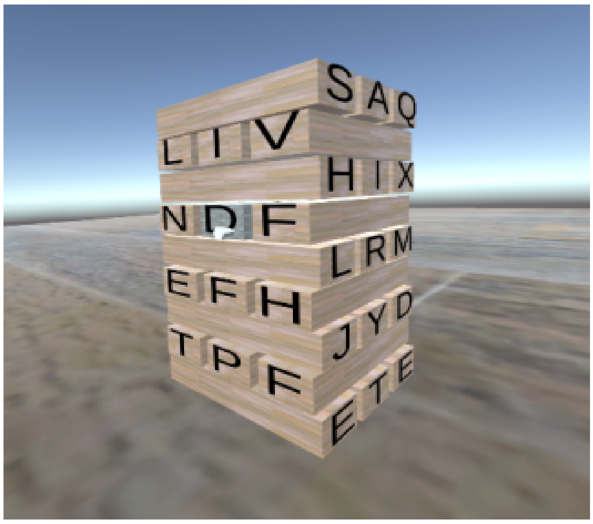

Figure 2: Selecting a block in the game environment.

One of our key contributions to ASL learning is the ability of users to view the gesture at different angles, helping them to learn the gesture as well as assisting them in being able to recognise the gesture in real life applications. When users choose a block they want to move, models of the hand gesture corresponding to the letter on the block are displayed in front of the user. The user is then able to walk around to look at the hand gesture from different view points. To make it easier for the user to see from both the perspective of the signer and signee without having to walk all the way around, both the front and back views of the gesture are shown (Figure 3).

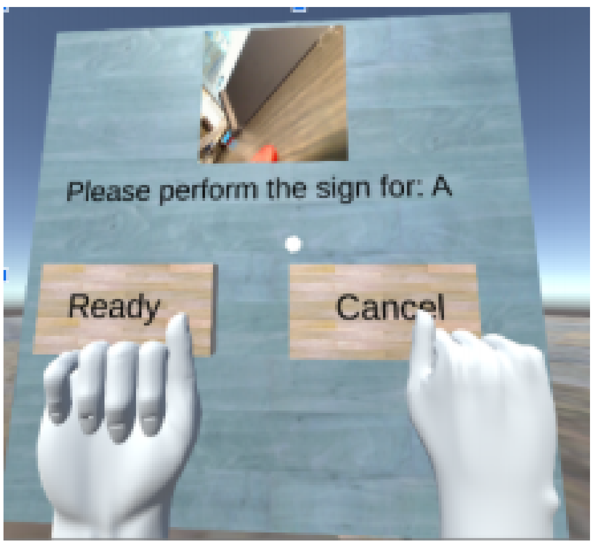

Figure 3: In-game sign demonstration.

In order to effectively teach users, we use active teaching to allow them to practice what they learn. In order to move a block in JengASL, the user must perform the sign language gesture corresponding to the letter on the block. The gesture is captured using a web cam and immediate feedback is given by use of the gesture recognition system, which returns letter/likelihood pairs. For example, the following returned data - '(A, 0.8), (B,0.3), (C,0.5)...' - would indicate that the gesture had an 80% probability of being the letter 'A' and 0.3% probability of being the letter 'B', and so on. The letter with the highest likelihood is shown to the user and they are given an option to try again if it is not the one they intended. In order to motivate users to achieve greater accuracy in their gesturing, a point system is used. Each time a block is removed from the tower, points P are added, with an amount determined by the following equation:

$$
\begin{aligned}
AttemptsLeft &= MaxAttempts - Attempts \\
P &= 100 * |difficulty - accuracy| * AttemptsLeft
\end{aligned}
\tag{1}
$$

Accuracy is obtained from the gesture recognition system as the percentage likelihood of the gesture being the specified letter. Due to it being highly unlikely that any user will ever achieve 100% recognition accuracy for a gesture, we implement a percentage accuracy difficulty level as a cap, which can be modified. In this case, the user only has to achieve, e.g., a 80% gesture accuracy to achieve full points, and higher difficulties would have a higher cap. Doing so can also encourage players to set goals and improve their gesture accuracy, and discourage them from settling for an incorrect gesture to remove the block. Similarly, to motivate the user to become more accurate, the more times the user decides to perform the selected gesture, the fewer points they will be awarded.

An important aspect of game design is to reward the user for doing well, and provide consequences if they do not. In our case the user is rewarded by gaining more points, and penalties are applied by reducing the smoothness with which blocks are removed from the tower. "Jitter" is added to blocks' movement as random vertical shaking. The jitter motion is inversely correlated with gesture accuracy. The height of a block within the tower is taken into consideration as shaking in at the bottom of the tower has a higher chance of toppling it over. The jitter factor is a modifiable constant that depends on the difficulty of the game. If too much jitter occurs (causing instability), the tower will topple over and the game will end. We use a physics system for the jitter, such that if a block is removed with jitter but the tower does not fall, the jitter will have still nudged other blocks in the tower, potentially reducing its stability. This both replicates the real-world version of Jenga and incentivises accuracy on *every* sign. Even if a given sign is not inaccurate enough to knock the tower over, consistently making mistakes will dramatically increase the risk to loose the game. Using probabilities rather than yes/no decisions also adds excitement to the game since, as with real Jenga, the player can never be sure what will happen.

3.1.1 Gesture Recognition

In order to provide correct feedback, we must be able to recognise the accuracy of the users' gestures. Ideally, we would like to provide precise feedback about which finger positions are incorrect. However, this proved difficult with existing technologies such as leap motion, which struggled with capturing motions where fingers overlap. We instead opted for using webcam input and machine learning (CNNs). While we used a very simple model and small training data set, recent publications show that the technology is advancing rapidly and capable of providing increasingly accurate recognition [AQKS21].

In order to reduce size and memory overheads associated with integrating a large trained CNN model into a game engine, we decided to implement a web-service which will be queried in game when the user performs a gesture. The web-service runs the model and returns the result to the game client, which will then parse it into a format suitable for use within the game.

4 IMPLEMENTATION

4.1 JengASL

Our game is implemented using Unity to host the JengASL application. Unity provides in-built Virtual Reality support, and also comes with many assets and game objects which can be used with our game, allowing us to reduce time spent on building the VR environment. The game was built based on the publicly available JengaVR [Ngi17], which implemented the physics required for the blocks to interact with each other or fall, however it had to be extended to include menus, gaze-interaction, webcam access, and the point system. Since we use an older head-mounted display without eye-tracking, gaze interaction uses the user's viewing direction obtained by the HMDs orientation.

A webcam was used in our work to capture users' gesture information by clicking a button in game. Block selection was done using gaze interaction to increase immersion, allowing users to have both hands free as they do their gestures. Communication between the game and the gesture recognition system is done through a TCP connection. During our study the web service for gesture recognition was hosted locally to mimise latency.

Learning Method	Mean	Std. Dev.
Traditional	89.06	20.20
JengASL	66.30	25.68

Table 1: Correctness Rate of the traditional and JengASL learning method (in %)

4.2 Gesture Recognition

For gesture recognition, we used a pre-trained VGG 16 model from Python's Keras library, with 16 total layers including input and output, and configured to classify ASL alphabets.

The dataset chosen for training provided 3000 images for each letter of the ASL alphabet. In this work we chose to use the 8 letters that the model was able to recognise with the highest accuracy.

5 EVALUATION

We conducted a small pilot study with 8 users aged 18–33 to test the effectiveness of our VR learning system, collecting both qualitative and quantitative data. 7 out of 8 users had no experience with ASL, and 6 had no experience with VR. Each user tested both the VR system, as well as the traditional method of looking at ASL gesture images. Demographic information of the users were collected at the beginning of each trial, before they were invited to complete two learning sessions of 8 gestures:

1. Session 1: 2D images of 8 different gestures representing characters different from session 2 (duration: 3 minutes)

2. Session 2: JengASL with 8 different gestures representing characters different from session 1 (duration: 5 minutes)

The additional time given to the VR game was to accommodate for in-game loading times and minor delays associated with interacting via the hardware.

The rate of gesture retention of participants was measured by asking users to perform each of the gestures that they learnt in the preceding session, and recording the correctness.

Data for qualitative analysis was collected using the Intrinsic Motivation Inventory (IMI) questionnaire [RD06, CSD22], with 3 sub-scales - Interest/Enjoyment, Effort/Importance, and Pressure/Tension. Users were asked 17 questions from these sub-scales with a rating from 1 to 7, and for all subscales the average score for its questions was recorded.

6 RESULTS

Table 6 gives the retention rates for both conditions. All users attempted 8 gestures for the traditional method and an average of 7 for the VR method, with the lowest being 4. The performance of users in VR had a moderately strong correlation with their performance in the Traditional method tests (Pearson's r = 0.47). From our data, JengASL performs worse than the traditional method, with a lower mean correctness rate. The differences in correctness is shown to be statistically significant using a two-tailed paired t-test ($p < 0.05$).

Table 2 shows participant responses to the IMI questionnaire subscales. Pressure/tension showed a strong positive correlation to correctness for the VR environment (r = 0.60), and only a weak correlation for traditional (r = 0.28). Effort/importance showed a weak negative correlation for traditional (r=-0.18) but a moderate positive correlation for VR (r = 0.40). Interestingly however, in traditional and VR methods there is a moderate positive correlation between effort/importance and correctness (r = 0.41 and r = 0.50 respectively). In both cases, interest/enjoyment had a moderate negative correlation with gesture correctness.

We can see that on average, users enjoyed the VR method more than the traditional method, made a similar effort for learning, and felt more pressure since a wrong sign could mean loosing the game. However, a two-tailed paired t-test showed no statistically significant difference between the two methods for all subscales.

7 DISCUSSION

JengASL was able to increase user interest and effort, although not at a statistically significant level. The increase in enjoyment was unsurprising, however at a lesser degree as expected. This may have been because although a gamified VR learning environment is a fresh and interesting idea, enjoyment is dependent on the game. Our gesture recognition system required a controlled environment (implemented in this case by placing a black screen behind the user's hand during recognition), and may have made it more awkward or difficult for users to play the game. The amount of jitter implemented in the game also needed to be optimised through user feedback.

Our system failed to increase the retention rate of gestures compared to the baseline. The most likely reason for this is the time restraint implemented during our evaluation: while looking at an image (traditional method), the user is likely to spend the whole time focusing on the image and attempting to memorise the gestures. However, in the VR system gestures are displayed when a block is clicked, and users have only until it disappears to look at and copy the gesture, leading to less time in total to memorise. This means that it

	Traditional Learning		Learning using JengASL	
Sub-Scale	Mean	Trad. Std. Dev.	Mean	Std. Dev.
Interest/Enjoyment	4.09	1.61	4.70	1.17
Effort/Importance	3.3	1.2	3.325	0.93
Pressure/Tension	4.45	1.14	5.325	1.08

Table 2: IMI Results with scores on a 5-level Likert scale from 1 (strongly disagree) to 5 (strongly agree) for traditional learning and learning using our VR tool JengASL.

could take longer to learn gestures using JengASL compared to traditional methods. Note however that taking more time to memorise does not mean our system is less effective, as increased enjoyment means users are likely to be more motivated to put in more time to play the game, or even more motivated to begin learning in the first place.

7.1 Limitations

Our study suffers from order effects since condition 1 (2D images) was always performed first. The reason for this was that we believe that combining three new experiences at once (sign language, game, and VR) might be too demanding. Since we used different sets of characters for each condition, we believe that learning characters for condition 1 should have limited effects on learning characters for condition 2. Some effects might still exits such as getting exhausted or bored after completing condition 2. This might be an additional explanation for the lower retention rates with the VR conditions.

The fact that we used different character sets for each condition might create another problem. The CNN for sign recognition had different accuracies for different characters and we hence chose the 8 characters with the highest accuracy. This meant that for the 2D image condition we had to choose randomly 8 from the remaining characters. The characters used in one condition might be more difficult to learn or more difficult to form using hand gestures, which would effect retention rates and recognition accuracy. For example, the letter "C" is relatively easy to learn and form since it involves making a "C" shape with the hand.

Other limitations are the small size of the user study (n=8) and self-selection bias since participants volunteered and we might have only got students with an interest in research, sign language, or games.

7.2 Design Considerations

Despite the negative results in terms of performance, this work has revealed several key considerations and barriers to be considered when designing tools such as this one for teaching gesture based skills.

One key consideration is the method by which the user interacts with the non-skill elements of the virtual environment. In particular, if the skill requires both hands as in the case of New Zealand Sign Language. In JengASL, users were required to use a controller to identify the Jenga block to be removed from the tower. This was motivated by previous research [ADWW19] and our own observations that gesture-based input with the available technologies was not accurate enough. While it is certainly possible for the user to use the controller to interact with the environment then put it down to perform gestures with both hands, this is an interruption to their activity and a barrier to usability. For this reason, even had training data availability not been a concern, JengASL would still be better suited for teaching American Sign Language (a one-handed language), than New Zealand Sign Language.

Another barrier to use is the availability of training data. While there is training data widely available for the subset of sign language signs that is the alphabet (for many languages), acquiring a dataset that is representative of the larger vocabulary of the language is difficult, not to mention the practical concerns around training a model with high accuracy for a large number of signs.

We also raise the key consideration of feedback and assessment with respect to precision as something to consider when developing a training tool. In JengASL, we provide feedback on accuracy in the form of score and the jitter mechanic. This is important to avoid sloppy use of the taught skill - particularly in the case where the skill in question is for communication and may have many very similar signs.

A final barrier to use for some sign languages is the presence of non-hand gestures in signs. In the example of NZSL, some signs can also include motion (e.g. the sign for "H"), touching the head (e.g. the sign for "Deaf"), and mouthing the associated word. While this is not the case for all sign languages, it is a limiting factor to what signs can be taught with a purely gesture-based system.

We did not have any problems with cybersickness and refer readers to design considerations listed by Shaw et al. [SWL+15] and Yin et al. [YBH+21] (section 3.6) in order to reduce cybersickness and make VR experiences safe.

8 CONCLUSIONS

In this paper, we have presented a tool that integrates sign language teaching into an immersive VR game.

One of the main contributions of our work is to allow learners of sign language to see various gestures from different angles, for easier learning of the different signs. While our system was not able to increase the retention rate of users' sign knowledge, our model does show potential in being more enjoyable, and thus more motivating than learning gestures using traditional methods.

At present there are several significant barriers to the use of VR tools and gesture recognition models for teaching sign language. Some of these can be mitigated with appropriate training data for the gesture recognition system, but the nature of signs in some sign languages and the large possible vocabulary makes such tools currently only suitable for supplementary learning and practice. Further research and development is necessary before these tools are suitable for standalone teaching.

Costs are also a factor for widespread adaption of VR training tools. Our solution uses a simple web cam (20 US$) and a head-mounted display connected to a desktop computer (we used an old Oculus Rift 2, second-hand about 200 US$).

8.1 Future Work

In future work we would like to develop/use more accurate neural networks for sign recognition and test them for different sign languages (ASL, NZSL, Auslan).

We would like to improve the teaching quality of the tool by enabling users to see hand gestures for longer and by providing more informative feedback. For example, a virtual model could be overlayed on the user's hand and the user then has to modify his/her hand gesture to precisely match the model.

Learning is most effective if the material is challenging, but not too difficult. Ideally we would like to measure cognitive load during the learning process [ABC+22] and then either adjust difficulty or provide feedback or visual hints in order to match task difficulty with the learners capabilities. Concepts used in intelligent tutoring systems would also be useful to increase learning [CLW18].

Finally, we would like to make a more extensive user study using randomly assigned characters for each condition with more participants, a longer training phase, and testing both short-term and long-term retention.

In using Jenga, we have taken a game that is traditionally played in a multiplayer form and used it as a solo teaching tool. It would be interesting to investigate how competition and competitiveness factor into motivation, enjoyment, and skill retention in a multiplayer gamified learning environment.

9 REFERENCES

[AAA+09] Oya Aran, Ismail Ari, Lale Akarun, Bülent Sankur, Alexandre Benoit, Alice Caplier, Pavel Campr, Ana Huerta Carrillo, and François-Xavier Fanard. Signtutor: An interactive system for sign language tutoring. IEEE MultiMedia, 16(1):81–93, 2009.

[ABA21] Ibrahim Adepoju Adeyanju, Oluwaseyi Olawale Bello, and Mutiu Adesina Adegboye. Machine learning methods for sign language recognition: A critical review and analysis. Intelligent Systems with Applications, 12:200056, 2021.

[ABC+22] Mohammad Ahmadi, Huidong Bai, Alex Chatburn, Burkhard C. Wünsche, and Mark Billinghurst. PlayMeBack - Cognitive Load Measurement using Different Physiological Cues in a VR Game. In Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology (VRST 22). New York, NY, USA, Article 70, pp. 1–2. ACM, 2022.

[ADWW19] Benjamin J. H. Andersen, Arran T. A. Davis, Gerald Weber, and Burkhard C. Wünsche. Immersion or Diversion: Does Virtual Reality Make Data Visualisation More Effective? International Conference on Electronics, Information, and Communication (ICEIC '19), Auckland, New Zealand, pp. 1–7, 2019.

[AQKS21] Muhammad Al-Qurishi, Thariq Khalid, and Riad Souissi. Deep learning for sign language recognition: Current techniques, benchmarks, and open issues. IEEE Access, 9:126917–126951, 2021.

[AVCA06] Nicoletta Adamo-Villani, Edward Carpenter, and Laura Arns. An immersive virtual environment for learning sign language mathematics. In ACM SIGGRAPH 2006 Educators Program, pp. 20–es. ACM, New York, NY, USA, 2006.

[BKEJ16] Yosra Bouzid, Mohamed Ali Khenissi, Fathi Essalmi, and Mohamed Jemni. Using educational games for sign language learning - a signwriting learning game: Case study. Educational Technology & Society, 19(1):129–141, 2016.

[BLC14] Debra L. Blackwell, Jacqueline W. Lucas, and Tainya C. Clarke. Summary health statistics for us adults: national health interview survey, 2012. Vital and health statistics. Series 10, Data from the National Health Survey, (260):1–160, 2014.

[CCP07] Allegra Cattani, John Clibbens, and Timothy J. Perfect. Visual memory for shapes in deaf signers and nonsigners and in hearing signers and nonsigners: Atypical lateralization and enhancement. Neuropsychology, 21(1):114–121, 2007.

[CCRV98] Olga Capirci, Allegra Cattani, Paolo Maria Rossini, and Virginia Volterra. Teaching Sign Language to Hearing Children as a Possible Factor in Cognitive Enhancement. The Journal of Deaf Studies and Deaf Education, 3(2):135–142, 1998.

[CLL+13] Xiujuan Chai, Guang Li, Yushun Lin, Zhihao Xu, Yili Tang, Xilin Chen, and Ming Zhou. Sign language recognition and translation with kinect. In IEEE Conf. on AFGR, volume 655, page 4, 2013.

[CLW18] Tyne Crow, Andrew Luxton-Reilly, and Burkhard C. Wünsche. Intelligent tutoring systems for programming education: a systematic review. In Proceedings of the 20th Australasian Computing Education Conference (ACE '18). New York, NY, USA. pp. 33–52. ACM, 2018.

[CR08] Mary P. Clynes and Sara E. C. Raftery. Feedback: an essential element of student learning in clinical practice. Nurse Education in practice, 8(6):405–411, 2008.

[CRG14] Ching-Hua Chuan, Eric Regina, and Caroline Guardino. American sign language recognition using leap motion sensor. In Proc. of the 13th International Conference on Machine Learning and Applications (ICMLA '14), pp. 541–544. IEEE, 2014.

[CSD22] CSDT. Intrinsic motivation inventory (IMI), 2022. http://selfdeterminationtheory.org/intrinsic-motivation-inventory/.

[FHA+22] Fahmid Al Farid, Noramiza Hashim, Junaidi Abdullah, Md Roman Bhuiyan, Wan Noor Shahida Mohd Isa, Jia Uddin, Mohammad Ahsanul Haque, and Mohd Nizam Husen. A Structured and Methodological Review on Vision-Based Hand Gesture Recognition System. Journal of Imaging, 8(6):153, 2022.

[HLO05] Eun-Jung Holden, Gareth Lee, and Robyn Owens. Australian sign language recognition. Machine Vision and Applications, 16(5):312, 2005.

[KBS+22] Deep Kothadiya, Chintan Bhatt, Krenil Sapariya, Kevin Patel, Ana-Belén Gil-González, and Juan M. Corchado. Deepsign: Sign language detection and recognition using deep learning. Electronics, 11(11), 2022.

[KEA03] Cem Keskin, Ayse Erkan, and Lale Akarun. Real time hand tracking and 3D gesture recognition for interactive interfaces using HMM. ICANN/ICONIPP, 2003:26–29, 2003.

[KLRWP17] Sam Kavanagh, Andrew Luxton-Reilly, Burkhard C. Wünsche, and Beryl Plimmer. A systematic review of virtual reality in education. Themes in Science and Technology Education, 10(2):85–119, 2017.

[KPN12] Jessica Korte, Leigh Ellen Potter, and Sue Nielsen. Designing a mobile video game to help young deaf children learn Auslan. In Proc. of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers (BCS–HCI '12), pp. 345–350, 2012.

[LJ09] Zhi Li and Ray Jarvis. Real time hand gesture recognition using a range camera. In Australasian Conference on Robotics and Automation, pp. 21–27, 2009.

[LWLD11] Rui Liu, Burkhard C. Wünsche, Christof Lutteroth, and Patrice Delmas, A Framework for Webcam-based Hand Rehabilitation Exercises, Proc. of the International Conference on Computer Vision Theory and Applications (VISAPP '11), Vilamoura, Algarve, Portugal, pp. 1–6, 2011.

[MLS06] Diane C. Millar, Janice C. Light, and Ralf W. Schlosser. The impact of augmentative and alternative communication intervention on the speech production of individuals with developmental disabilities: A research review. Journal of Speech, Language, and Hearing Research, 49(2):248–264, 2006.

[MN06] Zhenyao Mo and Ulrich Neumann. Real-time hand pose recognition using low-resolution depth images. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06), volume 2, pp. 1499–1505. IEEE, 2006.

[MR17] Daniel Muijs and David Reynolds. Effective teaching: Evidence and practice. Sage, 2017.

[MV14] Rachel McKee and Mickey Vale. The vitality of New Zealand sign language project: Report on a survey of the Deaf/NZSL community. Technical report, Victoria University of Wellington, September 2014.

[Ngi17] Ngiani. Jenga–VR, 2017. https://github.com/ngiani/Jenga-VR.

[ORCV17] Margarita Ortiz-Rojas, Katherine Chiluiza, and Martin Valcke. Gamification and learning performance: A systematic review of the literature. In 11th European Conference on Game-Based Learning (ECGBL), pp. 515–522, 2017.

[Ova91] Martha N. Ovando. Constructive feedback: A key to successful teaching. Management, 5(3), 1991.

[PMS+09] Farid Parvini, Dennis McLeod, Cyrus Shahabi, Bahareh Navai, Baharak Zali, and Shahram Ghandeharizadeh. An approach to glove-based gesture recognition. In International Conference on Human-Computer Interaction, pages 236–245. Springer, 2009.

[RD06] Richard M. Ryan and Edward L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55:68–78, 2006.

[Sen18] S. Senthamarai. Interactive teaching strategies. Journal of Applied and Advanced Research, 3(S1):36–38, 2018.

[SFB+13] Gretchen Stevens, Seth Flaxman, Emma Brunskill, Maya Mascarenhas, Colin D. Mathers, and Mariel Finucane. Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. The European Journal of Public Health, 23(1):146–152, 2013.

[SLC15] Neelesh Sarawate, Ming Chan Leu, and Özcelik Cemil. A real-time american sign language word recognition system based on neural networks and a probabilistic model. Turkish Journal of Electrical Engineering & Computer Sciences, 23(Sup. 1):2017–2123, 2015.

[SWL18] Ayoung Suh, Christian Wagner, and Lili Liu. Enhancing user engagement through gamification. Journal of Computer Information Systems, 58(3):204–213, 2018.

[SWL+15] Alex Shaw, Burkhard C. Wünsche, Christof Lutteroth, Stefan Marks, and Rodolphe Callies. Challenges in virtual reality exergame design. Proceedings of the 16th Australasian User Interface Conference (AUIC '15): Volume 162, pp. 61–68, 2015.

[WS99] John Weissmann and Ralf Salomon. Gesture recognition for virtual reality applications using data gloves and neural networks. In Proc. of the International Joint Conference on Neural Networks (IJCNN '99), volume 3, pp. 2043–2046. IEEE, 1999.

[YBH+21] Betty Yin, Samuel Bailey, Emma Hu, Milinda Jayarekera, Alex Shaw, and Burkhard C. Wünsche. Tour de Tune 2 - Auditory-Game-Motor Synchronisation with Music Tempo in an Immersive Virtual Reality Exergame. Proceedings of the 2021 Australasian Computer Science Week Multiconference (ACSW '21). New York, NY, USA, pp. 1–10. ACM, 2021.

[ZBP+11] Zahoor Zafrulla, Helene Brashear, Peter Presti, Harley Hamilton, and Thad Starner. Copycat: An american sign language game for deaf children. In Proc. of the IEEE International Conference on Automatic Face & Gesture Recognition (FG 2011), pp. 647–647, 2011.