Development Of A First-Person View-Based Korean Sign Language Education System Using VIVE Hand Tracking

Chongsan Kwon, Changhui song and Dong Hyun Kim*

College of Software Convergence. Dongseo University, Busan, 47011, Korea.
*Corresponding author. Tel.: +82-10-3696-8572; Email address: pusrover@dongseo.ac.kr

Abstract: In this study, using VIVE and Tracking, we developed a 'first-person view-based Korean sign language education system' that guides sign language handshapes using a 3D virtual hand from the user's point of view and then evaluates the accuracy of the handshapes that learners perform. The learning effects of a Korean sign language education system that guides sign language handshapes using 3D virtual hands in a virtual space and a system that only learns with a teacher's lecture video in a virtual space were compared and analyzed. The number of times the two groups of participants tried to make the sign language correctly was measured. Then, the measurement results were compared and analyzed and interviews also conducted. As a result of the experiment, the 'first-person view-based Korean sign language education system' demonstrated that some sign language handshapes, which are not easy for participants to make, were implemented faster and more accurately than traditional learning methods. Therefore, if this system is used, it is expected that sign language learners will be able to learn more intuitively through the first-person perspective. Furthermore, it is expected that learning effects can be improved through the ability of being able to judge whether users are making the correct shapes with their hands. In addition, since it is possible to evaluate the accuracy of the hand shapes, it is believed that single-person learning, which was not easy in the existing sign language education paradigm, will become possible. However, for the sign language examined in the experiment, some showed significant differences while others were less meaningful. These were analyzed because the difficulty of the learning contents used in the experiment was not high. In future research, in order to more accurately verify the effectiveness of the sign language education system using hand tracking and 3D virtual hand guidance, we intend to conduct an experiment implementing learning content with more complex sentences

Keywords: Sign language education, Handshape recognition, Gesture recognition, Virtual reality, First-person view, VIVE hand tracking

1. Introduction

Sign language is a language used by people who are deaf and hard of hearing and is a method of communication using gestures and signs [1,2]. However, sign language is not only necessary for communication between deaf people, but also necessary for the relationship between people without hearing difficulties and deaf people. Recently, interest in sign language is increasing around the world. However, if sign language teaching materials and training services are insufficient, unqualified sign language teachers might be produced and this will cause problems. In Finland, for example, a large number of deaf schools teachers are unqualified sign language teachers; consequently, there have been cases in which a solution was sought [3]. However, as social awareness about the need for sign language education has recently improved, the number of places that teach sign language in government offices and universities is gradually increasing. Nevertheless, it is still not easy to learn sign language systematically, and education conditions remain poor. Thus there are many difficulties in sign language education [4].Under the awareness of the lack of an overall sign language education system and lack of teachers, various learning methods incorporating new media and technologies such as YouTube, mobile smartphones, and Virtual Reality (VR) have recently been attempted as shown in Figure 1 [e.g., 5-7]. However, classroom field learning that follows the shape of the hand of the teacher in front has been moved to online, smartphone, and virtual spaces. However, the advantages of the medium have not effectively utilized yet. Therefore, it is difficult for learners to judge whether they are making the correct shapes with their hands.

(a)   (b)   (c)

*Corresponding author: Dong Hyun Kim
College of Software Convergence. Dongseo University, Busan, 47011, Korea.
Email address: pusrover@dongseo.ac.kr

Figure 1 Existing sign language education methods using various media technologies: (a) online video learning method [5], (b) mobile app method [6], and (c) VR learning method [7]

In addition, there are various studies on how to recognize sign language using gesture recognition or data globes. There are studies on how to recognize sign language in real time, such as the Korean sign language recognition system based on elementary components [8], a study on continuous Korean sign language (KSL) recognition using color vision [9], and a system that recognizes Korean sign language using a pair of data-gloves and then translates it into Korean text [10]. These studies are mainly aimed at real-time interpretation between deaf people and those without hearing impairment and are not aimed at learning sign language.

In sign language, accurate hand shapes are very important because the wrong hand shape can lead to misunderstandings in communication with other people. In the example of Figure 2, Figure 2 (a) is an accurate sign language expression that means mountain, but if it is expressed incorrectly as in Figure 2 (b), the other person may misunderstand it as a form of swearing. Therefore, an accurate evaluation and correction system is particularly important in a one-person education system that is not always a necessary condition when a teacher can correct the learner's hand shapes in offline education.

Figure 2 Problems with incorrect sign language: (a) an example of an accurate sign language expression meaning 'mountain' [11], (b) An example misrepresented as swearing through incorrect sign language [11]

2. Materials and methods

2.1 System Overview

In this study, a sign language education system was developed that maximizes learning effects by learning from a first-person perspective according to the guidance of a 3D hand in a virtual space with a learner wearing a Head Mounted Display (HMD). The advantage of VR that differentiates it from other media is that it does not simply look at the screen. Instead the user can directly enter a virtual space to experience and interact with the content in three dimensions. On a smartphone or monitor, the only way is to simply watch and follow the video; however, in VR, the user can intuitively learn the shapes through a virtual hand that appears three-dimensionally from the user's point of view.

As shown in Figure 3, VIVE Pro was used as an HMD that provides a virtual space [12], and the Vive Hand Tracking SDK was used as an API that evaluates the accuracy of the hand shape by measuring the angle of the finger and showing the hand shape in a three-dimensional manner [13]. VIVE Hand Tracking SDK basically supports the function of holding objects using one or two hands, and the user can also hold objects remotely using ray casting. The VIVE Hand Tracking Engine supports positional tracking of each hand. As shown in Figure 3(b), positional tracking consists of 21 points, four per finger and one at the wrist. The Plugin contains three modes for rendering hands as skeletons or single points for 2D/3D modes. VIVE Hand Tracking SDK supports Pre-defined gesture classification, Pinch Detection, and Hand Positions Mode. There are 6 pre-defined gestures in VIVE Hand Tracking SDK: Point, Fist, OK, Like, Five, Victory. Pinch is defined by tapping the

thumb and index tips. The position of the other three fingers does not matter. Hand Positions Mode provides a function that allows users to freely define and use the shape of their hands [13].
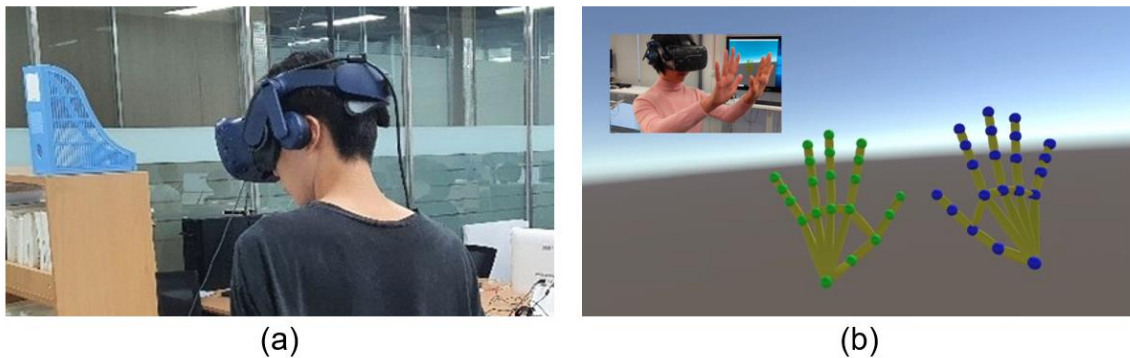

(a)   (b)

Figure 3 HMD and Handshape Recognition Plugin used for 'First-Person View-Based Korean Sign Language Education System' development: (a)VIVE Pro (b) VIVE Hand Tracking SDK [13]
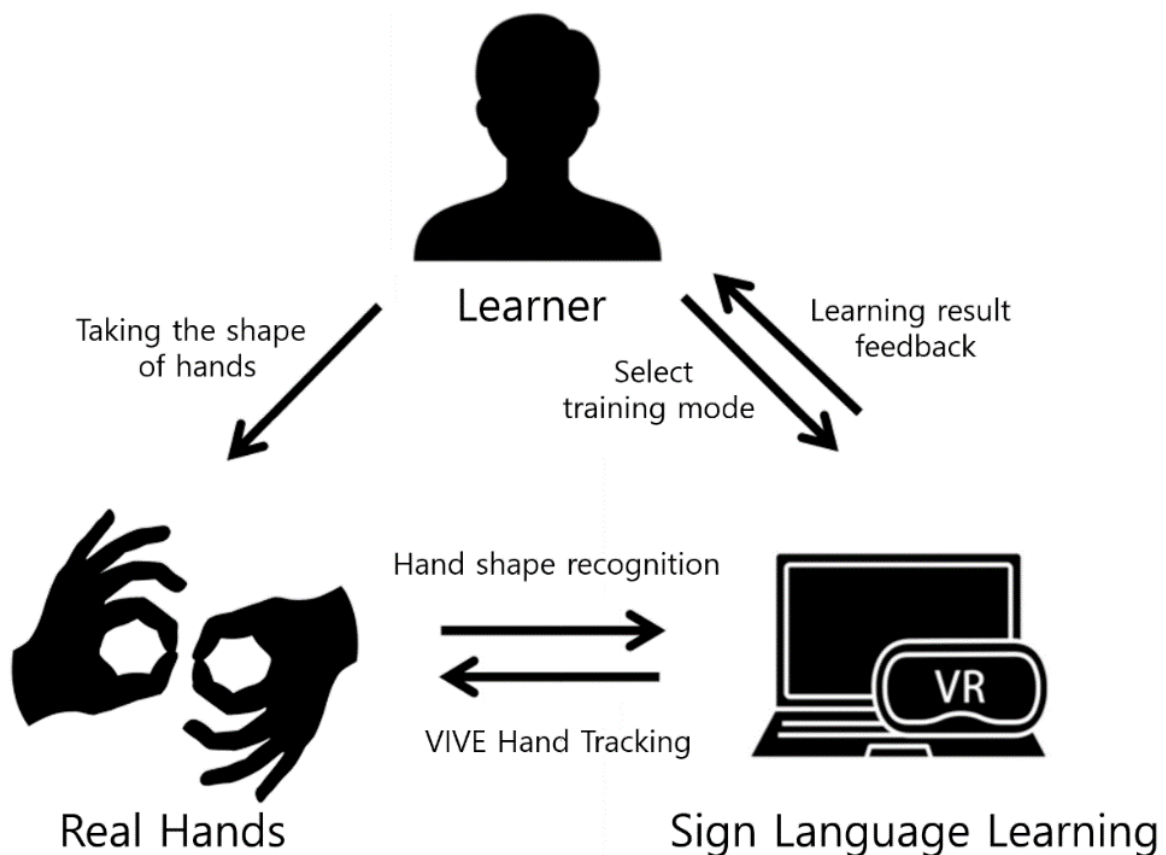

Figure 4 System configuration diagram

In this study, in order to more accurately and correctly define sign language handshapes, the 'first-person view-based Korean sign language education system' was implemented using the Hand Positions Mode of VIVE Hand Tracking SDK. In addition, a guide hand was implemented using the hand model of the VIVE Hand Tracking SDK. The accuracy of the handshapes made by the learner could then be automatically evaluated in real time. Unity was used as the game engine, and the minimum GPU specifications were NVIDIA GTX1060 and AMD RX480. Thus we implemented the system on NVIDIA GTX1060 and conducted experiments. The system configuration diagram is shown in Figure 4.

In general, language learning involves learning about consonants and vowels, followed by basic words, short sentences, long sentences, and conversations. In this study, the learning contents were composed of a total of 14 Korean consonants, the most basic of steps. When a learner wears an HMD as shown in Figure 5 (a) and carries out the 'first-person view-based Korean sign language education system', the Korean consonant sign language curriculum is conducted in a virtual space under the guidance of a sign language teacher as shown in Figure 5

(b). When the learning game starts, a virtual helper dialog is created, and the learner can see that his or her hands are reproduced with virtual hands in the virtual space. With this hand, the learner can interact with objects in the game. After the learner checks the "Start the game" guide, they then touch the next button to move to the next scene. When the learner touches the next button after confirming the "video is played" guidance, the video is created and played. A 3D guide hand appears, allowing the learner to see the shape of the teacher's hand in the video in front of him in three dimensions. The more accurately the learner matches his or her hand in this 3D guide hand, the higher the probability of success in learning. If the learner makes a shape different from the 3D guide hand shape, it does not proceed to the next step. When the learner makes as close the same shape as the 3D guide handshape as possible, a congratulatory message appears. Learners can proceed to the next step of learning by touching the Next button. The video used for learning was a Korean consonants YouTube lecture by sign language interpreter Kim Hyun-cheol. This video was used only for the purpose of this experiment, not for commercial purposes [14]. VIVE Hand Tracking can control the accuracy of the handshape by adjusting the degree of bending for each of the five fingers. In this study, for the purpose of the study, the value of the degree of bending was strict. Thus the user must makes the correct hand shape for it to be recognized as the correct answer.
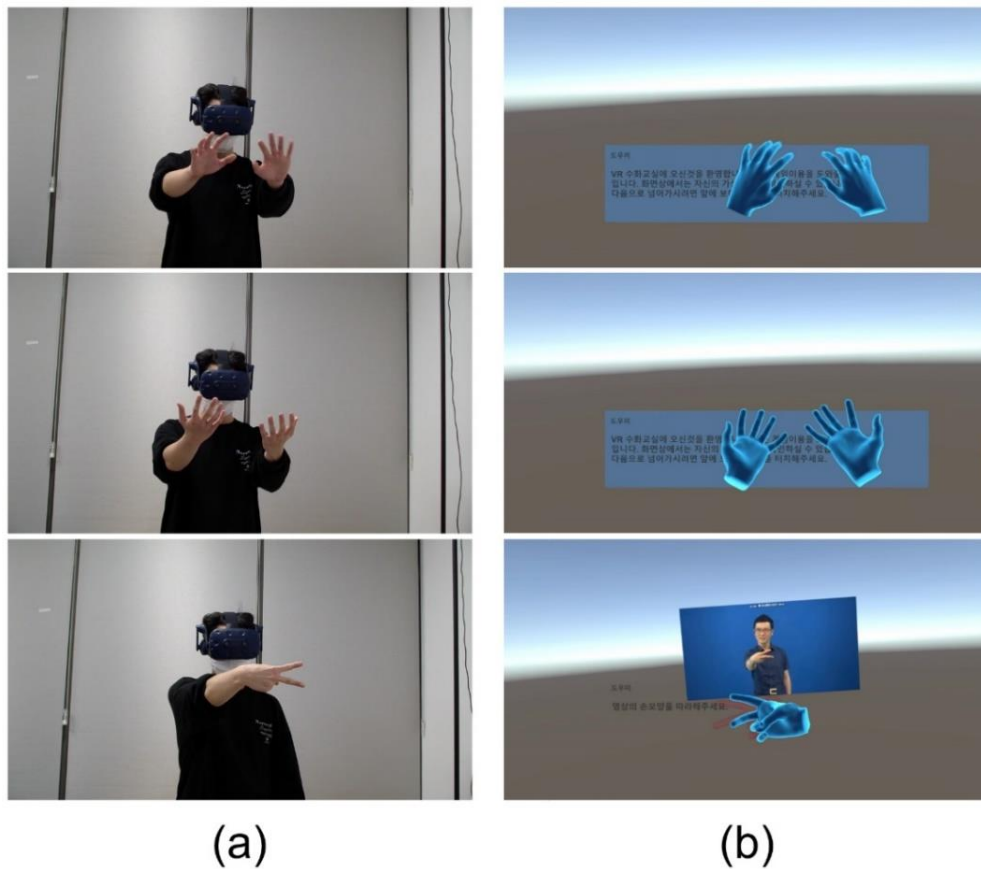

Figure 5 Learning through play: (a) the learner playing the game (b) the game play screen seen from the learner's field of view

2.2 Experiment
2.2.1 Experiment procedure and participants

In this study, VR was used in order to overcome the limitations of the existing media-based learning method, which transferred the learning in the sign language classroom that follows the shape of a teacher's hand to a digital education system. We developed a system that can intuitively learn the shape of the sign language and evaluate it in real time through a virtual hand that appears three-dimensionally from the user's point of view.

Figure 6 Images of (a) Korean alphabet consonants; (b) Images of learners playing the game; (c) Learning the Korean sign language education system' without 3D virtual hand guidance; (d) Learning the Korean sign language education system' with the guidance of 3D virtual hands; (e) Learner's sign language handshapes projected by 3D hand modeling in a virtual space.As shown in Figure 6, the effectiveness was verified. In order to verify that this 'first-person view-based Korean sign language education system' is more effective than the existing learning method by observing and following the teacher, a version of the traditional learning method without 3D virtual hand guidance as shown in Figure 6(b) was implemented separately. The two Korean sign language education systems were then compared. The two systems have all the same conditions, and only the presence or absence of the guidance of the 3D virtual hand is different. If the two systems are compared and analyzed, it is considered that the influence of the 3D virtual hand guidance and real-time evaluation system, the core elements of the learning system developed in this study, can be effectively analyzed.

Due to COVID-19, it was difficult to recruit a large number of experiment subjects. Thus the experiment was conducted by recruiting a minimum number of personnel. The experiment was conducted at a university in Busan, South Korea from August 3 to 14, 2020. The purpose of the experiment was explained to college students, and after obtaining consent, the experiment was conducted. A total of 10 people (9 male, 1 female) participated in the experiment. One person a day participated in the experiment, so there were no encounters between the participants. In addition, the risk of COVID-19 was minimized by participating in the experiment with both the experiment participant and the research director wearing a mask, and avoiding physical contact between the experiment participant and the research director running the experiment. Due to the small number of participants in the experiment, each participant was allowed to experience the two systems alternately. Five people (5 male, 0 female) first experienced the learning contents of the Korean sign language education system without 3D virtual hand guidance, and 3 hours later, experienced the Korean sign language education system with 3D virtual hand guidance. And the remaining 5 people (4 male, 1 female) first experienced the learning contents of the Korean sign language education system with 3D virtual hand guidance, and then 3 hours later experienced the Korean sign language education system without 3D virtual hand guidance. In this manner, all the participants of the experiment took turns experiencing each educational content, and a total of 10 experimental data were collected for each Korean sign language education system.

2.2.2 Measures

To verify the effectiveness, when each learner made a Korean consonant sign language shape presented in a virtual learning space, the number of repetitions it took for the system to recognize it and acknowledge it as the correct answer was measured and compared and analyzed. The number of samples in the group was 10, which did not satisfy normality, so the Mann-Whitney U test was performed as a statistical analysis method. The Mann-Whitney U test is a non-parametric statistical method that analyzes differences between two groups when the sample size of the group is relatively small and normality is not satisfied. In addition, after the experiment, an interview was conducted and used as a reference to interpret the statistical analysis results.

3. Results and Discussion

The number of samples in each group was less than 30. As a result of the normality test, the values of "Kolmogorov-Smirnov" and "Shapiro-Wilk" for all variables were less than 0.05 ($p < 0.05$), which did not satisfy the normality. Therefore, the Mann-Whitney U test was conducted.

Table 1 Results of the Mann-Whitney U test analysis for the number of attempts to succeed in the sign language handshape for each consonant

No.	Variable	Korean Sign Language Education System	N	Mean Rank	Sum of Rank	Mann-Whitney U	Wilcoxon W	Z	P
1	ㄱ	3D virtual hand included	10	9.25	92.50	37.50	92.50	-1.045	0.30
		3D virtual hand not included	10	11.75	117.50				
2	ㄴ	3D virtual hand included	10	9.60	96.00	41.00	96.00	-0.841	0.40
		3D virtual hand not included	10	11.40	114.00				
3	ㄷ	3D virtual hand included	10	9.20	92.00	37.50	92.00	-1.158	0.25
		3D virtual hand not included	10	11.80	118.00				
4	ㄹ	3D virtual hand included	10	9.60	96.00	41.00	93.00	-0.844	0.40
		3D virtual hand not included	10	11.40	114.00				
5	ㅁ	3D virtual hand included	10	10.10	101.00	46.00	101.00	-0.398	0.7
		3D virtual hand not included	10	10.90	109.00				
6	ㅂ	3D virtual hand included	10	10.05	100.50	45.50	100.50	-0.449	0.65
		3D virtual hand not included	10	10.95	109.50				
7	ㅅ	3D virtual hand included	10	10.05	100.50	45.50	100.50	-0.449	0.65
		3D virtual hand not included	10	10.95	109.50				
8	ㅇ	3D virtual hand included	10	9.65	96.50	41.50	96.50	-0.796	0.43

|		3D virtual hand not included	10	11.35	113.50				
9	ㅈ	3D virtual hand included	10	8.65	86.50	31.50	86.50	-2.007	0.05*
		3D virtual hand not included	10	12.35	123.50				
10	ㅊ	3D virtual hand included	10	8.60	86.00	31.00	86.00	-1.775	0.08
		3D virtual hand not included	10	12.40	124.00				
11	ㅋ	3D virtual hand included	10	9.25	92.50	37.50	92.50	-1.115	0.27
		3D virtual hand not included	10	11.75	117.50				
12	ㅌ	3D virtual hand included	10	10.05	100.50	45.50	100.50	-0.449	0.65
		3D virtual hand not included	10	10.95	109.50				
13	ㅍ	3D virtual hand included	10	9.65	96.50	41.50	96.50	-0.796	0.43
		3D virtual hand not included	10	11.35	113.50				
14	ㅎ	3D virtual hand included	10	10.50	105.00	50.00	105.00	0.000	1.00
		3D virtual hand not included	10	10.50	105.00				

*p<0.05
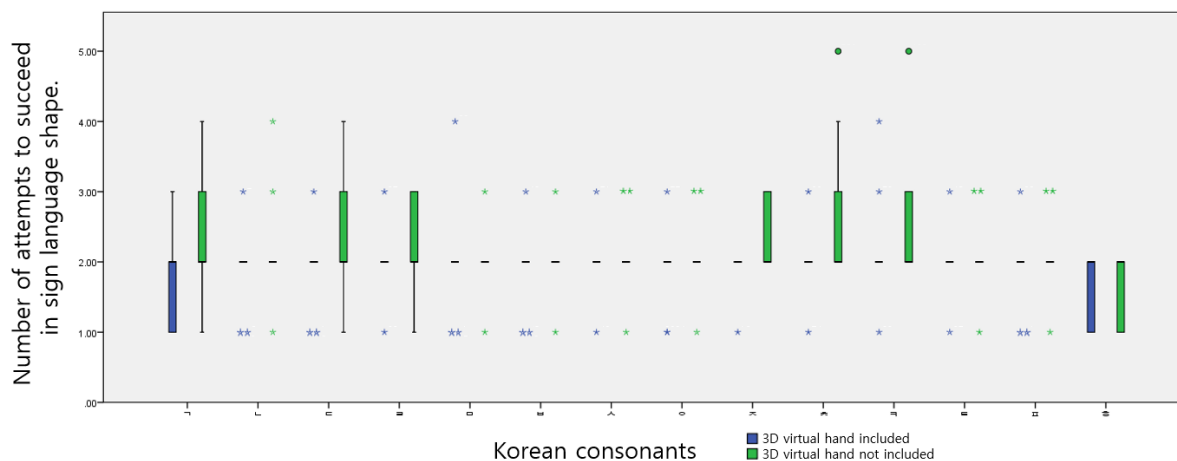
Figure 7 Comparison of Boxplots for number of attempts to succeed in the sign language handshape for each consonant

As shown in Table 1, only variable No. 9 showed significant results (Mann-Whitney U = 31.50, Wilcoxon = 86.50, p<0.05), and for all other variables there was no significant difference in the number of times it took to succeed in the handshape of the sign language between the Korean sign language education system with a 3D virtual hand and the Korean sign language education system without a 3D virtual hand. Figure 7 shows this result visually. For half of the variables (No. 2, No. 5, No. 6, No. 7, No. 8, No. 12, No. 13) out of a total of 14 consonants, most of the participants in the two groups fully implemented consonants in sign language in only two attempts. However, six of the remaining variables (No. 1, No. 3, No. 4, No. 9, No. 10, No. 11) show differences between the two groups. Variables No.1, No. 3, No. 4, No. 10, and No. 11 are not significant, but in the system without the 3D virtual hand, the participants had to try repeatedly to make the sign language shape accurately. In No. 9, it can be seen that the experiment participants implement the sign language handshape faster and more accurately in the Korean sign language education system with 3D virtual hands. However, in the case of variable No. 14, it can be seen that there is no difference between the two groups.

This result is believed to have occurred for the following reasons. First, in this study, learning contents were implemented by basic consonant learning rather than sentences connecting complex words. Therefore, it is understood that the presence or absence of a 3D virtual hand from the first-person view did not have a significant effect on the experimental results when an adult, like a college student, was learning. From the interviews carried out after the experiment was completed, many students said that the sign language was not difficult, and that they could fully follow the video in front of them. However, in the case of consonants No. 1, No. 3, No. 4, No. 9, No. 10, and No. 11, the wrist must be bent somewhat dramatically and the fingers must maintain an uncomfortable position in order to make the correct sign language shape. It seems that it would have

been easier to make a shape by matching the handshape to the 3D virtual hand rather than just watching the video. This result can also be seen from the fact that the IQR (Interquartile Range) of the Korean sign language education system without the 3D virtual hand is higher than that of the Korean sign language education system with the 3D virtual hand, as shown in Figure 7. Nevertheless, the reason the result is not significant is that creating a simple handshape at each learning stage is not such a complicated process. Thus, even without 3D virtual hand guidance, it was possible to quickly create the shape shown in the image after one or two attempts.

As mentioned earlier, there was a significant difference in the case of variable No. 9, and even though the case of variable 10 is not significant, the IQR (Interquartile Range) of the Korean sign language education system without the 3D virtual hand is higher than that of the Korean sign language education system with the 3D virtual hand, showing a difference. This result is believed to have taken place because the difficulty of implementing these handshape is higher than that of other consonants. Although the handshape of these two variables is similar to the shape of consonants No. 1 or No. 7, it seems that it would not have been easy to make the handshapes in the sense that several more fingers had to be extended. In the interview results, one participant, who succeeded in consonants No. 9 and No. 10 after a total of 5 attempts, said that it was not easy to make these shapes because of his lack of flexibility, and that he felt a cramp in his hand. However, he said that when learning through the Korean sign language education system with 3D virtual hand guidance, he made a shape by matching the shape of his hand to the 3D virtual hand, and it was easier than the learning method by simply watching the video. Thus he was able to succeed in making the shape faster.

In the case of variable No. 14, most of the participants in the two groups correctly created the handshape on their first attempt, and only one person in each group made the handshape correctly on their second attempt. This result is considered to be because, in the case of consonant No. 14, not only is it easy for learners to make the shape, but also because the degree to which all the fingers must be bent is large and easily recognized by the hand tracking system.

As a result of the overall analysis, the difficulty of the learning content was not high, so it was not possible to derive a meaningful result. However, when using the 'first-person view-based Korean sign language education system', it was found that some sign language handshapes which are not easy for experiment participants to make can be implemented faster and more accurately than traditional learning methods. Therefore, if this system is used, Korean sign language learners are expected to be able to learn more intuitively through the first-person perspective, and it is expected that the learning effect can be improved as they can immediately judge whether they are making the correct shape with their hands. In addition, since it is possible to evaluate the accuracy of the hand shape, it is expected that single-person learning, which was not easy in the existing sign language education, will be possible.

4. Conclusions

In this study, the 'first-person view-based Korean sign language education system' which guides the sign language shape from the user's point of view with 3D virtual hands using VIVE Hand Tracking, was developed and its effectiveness verified. As a result of comparing it with a system that simply sees and learns images under the same conditions, when learning a sign language shape that is difficult to make, the 3D virtual hand of this system serves as guidance, helping learners to make an accurate shape more quickly. Therefore, if this system is developed and used for practical purposes, it is expected that learners can learn more intuitively through the first-person perspective and judge whether they are making an accurate shape with their hands.   Thus the learning effect can be improved.   In addition, since the accuracy of the handshape can be evaluated immediately, it is expected that single-person learning, which was not easy in the existing sign language education, will be possible.

However, in all learning contents except for one, the learning effect was not significantly improved. The reason for this can be found in the difficulty of the learning content. In this study, learning contents were implemented with Korean consonants, the basis of language learning. However, since individual consonants are rather simple, it is judged that the presence or absence of a 3D virtual hand that serves as a guide when learners make handshapes is less affected. Therefore, in future research, we intend to conduct further experimental research by implementing interactive learning contents composed of sentences with complex structures in order to complement the effectiveness of hand tracking and 3D virtual hand for sign language.

5. Acknowledgements

6. References

1. Nidcd.nih.gov [Internet]. [Bethesda (MD)]: National Institute on Deafness and Other Communication Disorders; c2019-03 [updated 2019 May 8; cited 2020 Jul 2]. American Sign Language. Available from: https://www.nidcd.nih.gov/health/american-sign-language

2. Lexico.com [Internet]. [place unknown]: Meaning of sign language in English; [cited 2020 Aug 9]. Available from: https://www.lexico.com/definition/sign_language

3. De Weerdt D, Salonen J, Liikamaa A. Raising the profile of sign language teachers in Finland. Kieli, koulutus ja yhteiskunta. 2016;15.

4. McLaughlin J. Sign language interpreter shortage in California: Perceptions of stakeholders. Doctoral dissertation, Alliant International University, Shirley M. Hufstedler School of Education, San Francisco. 2010.

5. EBSCulture [Internet]. [Seoul (Korea)]: EBSCulture; c2013-04 [cited 2020 Jul 15]. Korean sign language to learn together. Available from: https://www.youtube.com/watch?v=C0lItcMX9M0&list=PLsxXTlmm_bBvrpybjt4-_fM9VLU_LgqO9

6. Jeonbuk Education Research Information Center. [Internet]. [Jeonbuk (Korea)]: c2018-01 [cited 2020 Jul 15]. Sign Language Class [cited 2020 Jul 15]. Available from: https://play.google.com/store/apps/details?id=com.kt.android.JBedu2&hl=ko

7. Valentin R. [Internet]. [place unknown]: c2019-08 [cited 2020 Jul 15]. Yujin teach us learn KSL (Korean Sign Language) in vrchat. Available from: https://www.youtube.com/watch?v=FVXpyKR1IL0&feature=youtu.be

8. Lee CS, Bien Z, Park GT, Jang W, Kim JS, Kim SK. Real-time recognition system of Korean sign language based on elementary components. In Proceedings of 6th International Fuzzy Systems Conference; 1997;3:1463-1468. IEEE.

9. Kim JB, Park KH, Bang WC, Bien ZZ. Continuous gesture recognition system for Korean sign language based on fuzzy logic and hidden Markov model. In 2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No. 02CH37291); 2002;2:1574-1579. IEEE.

10. Kim JS, Jang W, Bien Z. A dynamic gesture recognition system for the Korean sign language (KSL). IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 1996;26(2):354-359.

11. Park SH. (2018). [Internet]. [place unknown]: c2018-08 [updated 2018 Oct 19; cited 2020 Aug 3]. Video of deaf-muteness correcting sign language that the people use incorrectly. HUFFPOST. Available from: https://www.huffingtonpost.kr/entry/story_kr_5bc8118fe4b055bc947d5f77

12. VIVE. [Internet]. [place unknown]: VIVE Pro HMD. [cited 2020 Aug 4]. Available from: https://www.vive.com/us/support/vive-pro-hmd/category_howto/about-the-headset.html

13. VIVE Developers. [Internet]. [place unknown]: VIVE Hand Tracking SDK [cited 2020 Aug 4]. Available from: https://developer.vive.com/resources/vive-sense/sdk/vive-hand-tracking-sdk/

14. Kim HC. [Internet]. [Seoul (Korea)]. c2014-10 [cited 2020 Aug 8]. Korean fingerprint 'consonant'. Available from: https://www.youtube.com/watch?v=OOnuoUl4gP4