Sign Language Learning Support System using HMD with Hand-tracking Function

Ryota Akaba, Shinsuke Shikama

20m801ar@edu.setsunan.ac.jp, shikama@ele.setsunan.ac.jp
Setsunan University, 17-8 Ikedanakamachi, Neyagawa city, Osaka 572-8508, JAPAN
Keywords: Virtual Reality, Hand-tracking, Sign Language, Learning Support

ABSTRACT

We have developed learning support application of Japanese sign language using stand-alone HMD with hand-tracking function. The application is based on a unique hand-shape recognition algorithm. Regarding the application program, we have conducted subjective evaluation regarding to the understandability of the VR-learning system. In addition, we have compared the learners' favorability among three teaching materials (VR-system, textbooks, videos) regarding sign language learning items.

1 Introduction

As shown in Fig. 1, there are about 340,000 hearing and language impaired person in Japan as of 2016, and sign language is indispensable for communicating with them. Videos and illustrations are used in learning sign language. However, these causes problems such as the dominant hand being inverted by a mirror image. These problems reduced learning efficiency for beginners. Therefore, we tried to solve the issue by implementing a software that utilizes the characteristics of a hand-tracking HMD.

2 Related Works

There are several previous studies on software that tracks human movements with sensors and assists in learning body movements. In the research of Nagai et al. [1], an application has been developed to support fingerspelling learning by using LeapMotion (Fig. 2), a device specialized for hand recognition. In their study, the effects of having learners compare their hand shapes with a 3DCG model to confirm the correctness of the three-dimensional shapes and having learners repeat the learning process as if it were a game were reported. In addition, Shibata et al. [2] proposed a system to support learning of first aid using a head-mounted display. In their study, it is reported that the learning method using HMD is effective in acquiring skills because it is easy to understand the position and movement of the hands and also allows practice with spatial continuity. They also show hand-sign examples which employ 3DCG displayed on the HMD. In addition, unlike fingerspelling, the sign language to be learned is expressed with the entire body and uses a wide space, so it is thought that the advantages of the HMD, such as the ease of understanding spatial movements, can be fully utilized.

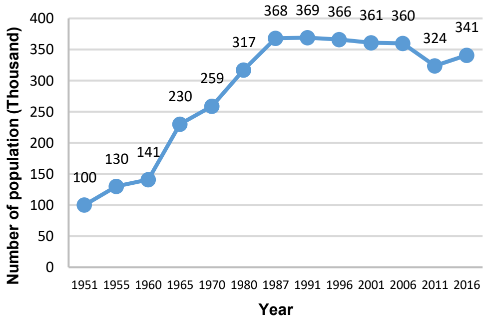


Fig. 1 Changes in the number of hearing and language impaired person in Japan [3]

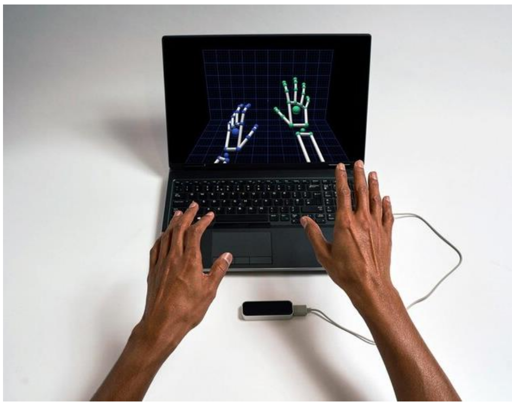


Fig.2 Example of hand recognition using LeapMotion [4]

3 HMD and Processing Algorithm

In this study, the learners observe the model and move their own hands to learn. Therefore, the hardware required is a HMD that provides a first-person view and a sensor section that recognizes the learner's hand shape. Usually, HMDs are connected to a PC and require external sensors, which can affect the user's environment and posture. However, the Oculus Quest (Facebook) employed in this research has the advantage of being able to be used in any environment because it has a built-in hand tracking sensor and the HMD operates as a stand-alone device. In addition, the development environment used was Unity, a game engine.

3.1 Processing overview

The flow of the system is shown in Figure 3. In this system, the position and rotation of the user's hands and fingers are acquired by the infrared camera sensors built into the four corners of the HMD. The acquired information is processed inside the HMD to display a 3D model of the hand in the virtual space at the same position as the user's hands in the real space. Figure 4 shows an example of an experimental scene using this system. The left side of the figure is an image of real space. The position, rotation, and shape of the user's hands are tracked by the HMD and synchronized with the 3D model of the hands in the virtual space on the right side of the figure. The learning flow of the application is shown in Figure 5. When the learner selects the desired sentence and starts sign language learning, a model is displayed in front of the learner (Figure 5(2)). The learner observes it from any angle and uses it as a reference to move his or her hand and superimpose it on the model (Figure 5(3)). When the learners' hand shape matches the shape of the 3DCG model, a model image of the next word is presented. This process is repeated, and if all the words in the sentences are finished, one learning process is completed (Figure 5(4)). The above is the flow of learning, and this is repeated several times to make the learners' understanding sign language.

3.2 Hand shape recognition algorithm

In sign language, not only the bending and stretching of the fingers, but also the position and rotation of the hand in relation to the body are important factors. The posture and shape of the hand were categorized into nine elements in order to identify the learner's sign language and match them with the model. An example of sign language classification is shown in Figure 6. The posture information acquired on Oculus Quest is managed by a decimal number of position values and rotation values. However, if this is used for classification as it is, it takes much time to detect matches though the measurement accuracy is high. And it also increases the burden of creating teaching material data. We created an array of integer values that simplify the states of the nine classified elements. The test is passed when the learner's array matches the model's array.

In the 9-digit numerical array, the first 5 digits represent the three states of stretching, bending, and folding of the fingers of one hand (Figure 6(a)). The sixth digit indicates the rotation of the hand (Figure 6(b)). The seventh and eighth digits indicate the vertical and horizontal position of the hand relative to the body (Figure 6(c)). The ninth digit indicates the left and right tilt of the hand (Figure 6(d)).

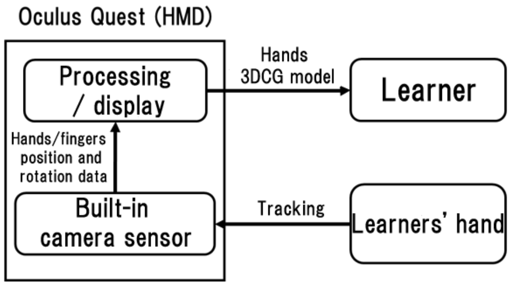


Fig. 3 System flow

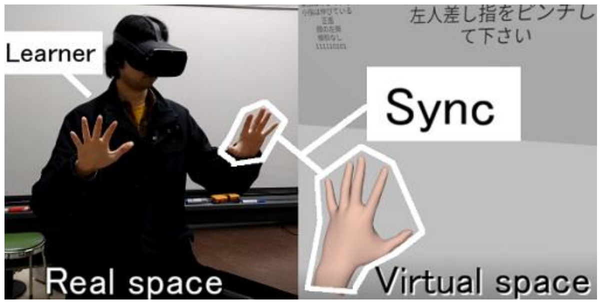


Fig. 4 Correspondence between real space and virtual space

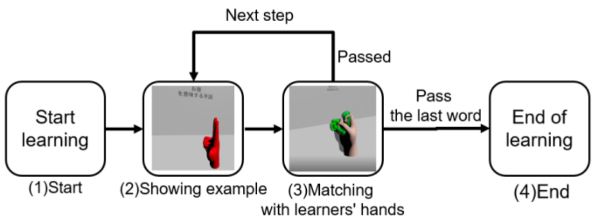
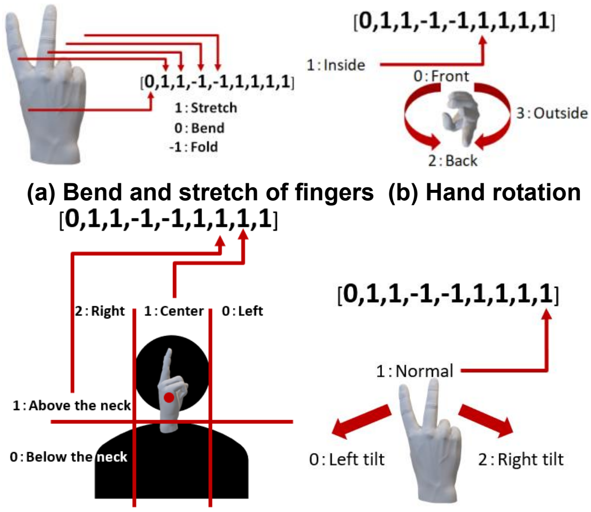


Fig. 5 Learning flow



(a) Bend and stretch of fingers (b) Hand rotation

(c) Hand position (d) Tilt of palm

Fig. 6 Trainee's hand shape/posture recognition Method

4 Experiments

In order to evaluate the understandability and usability of the learning support system created in this study, we conducted an experiment to compare it with conventional teaching materials.

4.1 Teaching Materials

In order to evaluate the characteristics of this VR system in use, we will explain the three conditions of the textbook and video materials and the VR materials, which are the comparison objects. These were assigned different sentences each, considering the learning effect when the same sentence is learned.
The textbook materials used were excerpts from the official textbook for the Sign Language Proficiency Test Levels 5, 6, and 7 (NPO Sign Language Proficiency Test Association), printed on paper. The content is a sign language that means "please wait a moment" and consists of about four steps. An illustration of a face-to-face person and arrows are used to represent the movement, and text is used to supplement the explanation of the procedure. Video materials used were videos made available on the Internet by the Division of Health and Welfare for Persons with Disabilities, Welfare Bureau, Health and Welfare Department, Hokkaido. The content is a sign language of about four steps, meaning "What is your name?" : A 33-second video recording of a face-to-face person performing a series of actions and explaining individual words. This video was played back on a laptop computer with a 14-inch display (Modern-14-B4MW, MSI). Regarding to the VR material, the sign language is about four steps, which means "Nice to meet you,". We gave about two-minutes preliminary explanation and preparatory work on HMD operation and content control.

4.2 Evaluation procedure

Prior to the experiment, the content of the experiment and the handling of personal information were explained, and consent was obtained before the experiment was conducted. Since it is important to confirm the stereoscopic function of the collaborators, we conducted a binocular stereoscopic function test using the “three-rods test [5]” prior to the experiments using the HMD. We also conducted a pre-questionnaire asking about age and sign language experience. Considering the bias in the number of people, the subjects were assigned to three conditions to study: textbook, video, and VR. The learning time was set to 5 minutes for each condition. After the study was completed, the learners were asked to answer a questionnaire about the understandability of the study, demonstrate the sign language they had learned, and write down the procedure.
After the learners experienced the other two conditions, we asked the subjects to answer a comparative questionnaire on learning experiences using three types of teaching materials.

4.3 Questionnaires

The questionnaire for the subjects to answer in the course of the experiment is composed as follows. The questionnaire was divided into two parts: the first part for the evaluation of the initial conditions and the second part for the comparison of all conditions.
The purpose of the first part is to survey the usability of the first-time experience condition, and to have the subjects describe the procedure and check if they understand it correctly. After that, the subjects were asked to answer the following four questions about the understandability using a 5-point scale ((1) I don't think so at all, (2) I don't think so much, (3) I can't say either, (4) I think so a little, (5) I think so a lot.).

#1: Was the shape of the hand easy to understand?
#2: Were the hand movements easy to understand?
#3: Was the procedure easy to understand?
#4: Was the positioning of the hands easy to understand?

The second half of the questionnaire is a comparative questionnaire regarding the preference of each item of the three types of teaching materials. Subjects were asked to respond to the above four items using the ranking method. The subjects were also asked to describe the experiences they had on the experiment.

5 Results

5.1 Subjective evaluation of understandability

The results of the subjective evaluation of the understandability of each material are shown in Figure 7. The histogram shows the average value for each condition for each item, and the error bars indicate the standard deviation. The number of subjects for each condition is 4 for textbook materials, 4 for video materials, and 3 for VR materials. At present, the number of evaluators is not sufficient. Therefore, the results are considered to be a provisional reference data.
As shown in Fig. 7, the VR and video materials tended to be highly rated in all items.
The textbook materials are rated lower than the other two in all items, and the understandability of the motions is rated particularly low.

5.2 Comparison of favorability between each teaching material

After the subjects had experienced all three types of materials, they were asked to rank which materials were easier to understand for each item. Table 2 shows the number of people who ranked their choices for each material in each item. The table highlights the materials with the highest number of people in each item and each rank.
In terms of the understandability of the shape, 9 out of 11 subjects chose the VR material as the first place, 2

chose the video material, and no subject chose the textbook.
In terms of the understandability of the motion, all 11 subjects chose the VR materials as the first place, and none of them chose the video or textbook materials. In terms of the understandability of the process, 8 out of 11 subjects chose the VR material as the first place, 3 out of 11 subjects chose the video material, and no subject chose the textbook material.
In terms of the understandability of the position, 8 out of 11 subjects chose the VR material as the first place, 3 out of 11 subjects chose the video material, and no subject chose the textbook material.

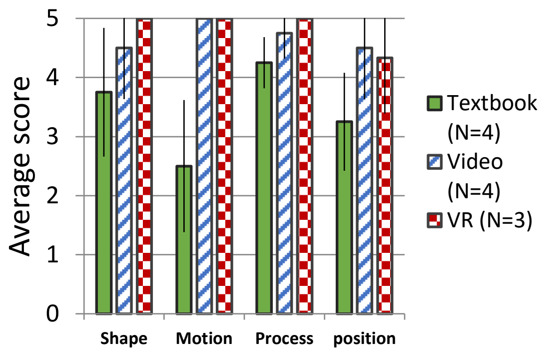


Fig. 7 Evaluation results on understandability of learning materials

Table 1 Favorability ranking among each teaching material (N=11)

	Shape			Motion		
	Textbook	Video	VR	Textbook	Video	VR
1st	0	2	9	0	0	11
2nd	0	9	2	0	11	0
3rd	11	0	0	11	0	0
	Process			**Position**		
	Textbook	Video	VR	Textbook	Video	VR
1st	0	3	8	0	3	8
2nd	2	7	2	0	8	3
3rd	9	1	1	11	0	0

6 Discussion

As a result of the experiment, the subjective evaluation of understandability (Fig. 7) shows that the VR materials are as easy to understand as the video materials in each item.
In addition, the comparative data of favorability among the three materials (Table 1) shows that the VR materials tend to be easier to understand than the video materials in terms of shape, motion, process, and position.

The subjects commented that it was easy to understand the shape of the hands and how to move them, and that they felt they were doing it correctly compared to other teaching materials. It was found that the advantage of these VR materials was that they could express movement and that they could actually superimpose their own hands on the model.

7 Conclusions

In this paper, we implemented an application to support learning sign language using the VR-HMD with hand tracking function, and conducted evaluation experiments. Through each item(i.e. shape, motion, process, and position) of the evaluation experiment, the evaluation score was as good as or better than the conventional teaching materials. It is thought that the elements that this system was initially designed for, such as the ability of the model to move and the ability to observe from any angle, worked well. In addition, the fact that the learners not only can see, but also can move their own hands help them to be aware that they are understanding correctly.

A future task is to increase the number of subjects in the experiment and improve the reliability of the evaluation data. In addition, the application contains a lot of extra visual objects that may confuse the users who are not familiar with VR. Therefore, it is necessary to improve the application to make it more user-friendly by using auditory information. After that, we would like to proceed to measure the learning effect numerically.

References

[1] Atsushi Nagai, et al., A Study on Learning Support Application of Fingerspelling using Motion Sensor. Computer Vision and Image Media (CVIM), 2016, 2016.5: 1-8.

[2] Takashi Shibata; Miyako Harada. Educational Tool for First Aid Training Using an Optical See-through Head Mounted Display. Japanese journal of educational media research, 2017, 23.2: 35-45.

[3] Mitsubishi UFJ Research and Consulting Co.,Ltd, Toward the success of employee with hearing disabilities (in Japanese), 2019, https://www.murc.jp/wp-content/uploads/2019/05/seiken_190510_1.pdf

[4] Ultraleap for Developers - Leap Motion https://developer.leapmotion.com/

[5] Matsuo, Toshihiko, et al., Correlation between depth perception by three-rods test and stereoacuity by distance randot stereotest. Strabismus, 2014, 22.3: 133-137.

[6] Oculus Integration Asset Store : https://assetstore.unity.com/packages/tools/integration/oculus-integration-82022 (Reference October 2021)