2018 Joint 10th International Conference on Soft Computing and Intelligent Systems and 19th International Symposium on Advanced Intelligent Systems

Comparison of Gesture Inputs for Robot System Using Mixed Reality to Encourage Driving Review

Yuya Aikawa*, Masayoshi Kanoh*, Felix Jimenez†, Mitsuhiro Hayase‡, Takahiro Tanaka§ and Hitoshi Kanamori§
* Chukyo University,
101-2 Yagoto Honmachi, Showa-ku, Nagoya 466-8666, Japan
† Aichi Prefectural University,
1522-3, Ibaragabasama, Nagakute-shi, Aichi, 480-1198, Japan
‡ Toyohashi Sozo University,
20-1 Matsushita, Ushikawa-cho, Toyohashi 440-8511, Japan
§ Nagoya University,
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

***Abstract*—We have been developing a robot system that encourages people to review their driving at home in order to improve it. Mixed reality technology is used to implement a function in the system for reviewing one's driving. Unintuitive input interface used in an existing mixed reality system makes it difficult to operate the system. In this paper, we propose two gestures. One uses a swiping motion, and the other acts like a laser pointer. The gesture that imitates a laser pointer performed best in experiments.**

I. Introduction

In recent years, the number of traffic accidents has been decreasing, but the traffic accident rate of elderly drivers has been increasing year by year. To this problem, we think that it is necessary for elderly people to review their own driving and have opportunities to do so. Currently, they can go to a driving school and review their own driving by receiving lectures on driving. However, to improve their driving, it is necessary for them to repeatedly learn about the mistakes they make when driving. Therefore, although it is necessary to receive a course a number of times, they have to go to the driving school at a fixed time, so the burden placed on them is large and a course takes a long time. Also, when they learn about the mistakes they make while driving, the examiner need to instruct them each time, so the cost of instruction will also increase. For these reasons, we need a system that enables elderly people to review their driving alone without going to driving school. Here, although driving is improved by instruction, its effect is temporary, and there is a possibility that they will immediately return to how they drove before instruction [1]. In training methods [2] based on coaching theory, it is said that it is necessary to repeat the process of having a driver recognize, analyze, and improve their driving. If this process can be presented by a robot, not by an examiner at driving school, elderly people could improve their own driving by themselves. In our research, we are developing a robot system that promotes the improvement of driving at home [3]. When reviewing driving, we believe that it is possible for elderly people to become aware of their own driving by using their own driving records. Furthermore,

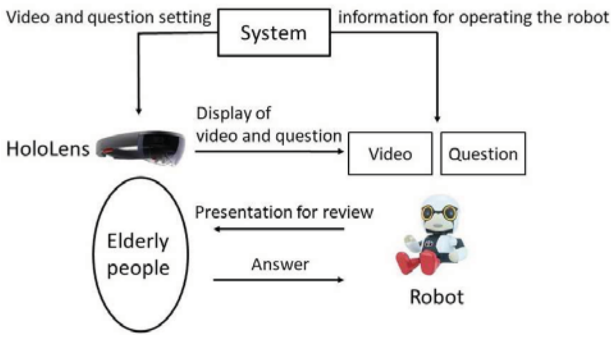

Fig. 1. System overview

it is expected that driving can be effectively improved by receiving instructions from a robot that asks questions that make one analyze their driving from these driving records. For this reason, we will implement a function in this system for recording the driving of elderly people and allowing them to review their driving by using the obtained data. This system consists of three functions, a robot, video, and interface, but in this paper, we discuss the interface.

In this system, mixed reality technology is used to present video taken during driving. Mixed reality is a technique for displaying a virtual space on lenses such as eyeglasses [4]. By using this technology, elderly people can review their driving while watching a robot and video in the same field of view. What is important here is the input interface. Instead of mixed reality, a method that displays video taken while driving on a television screen would be possible. However, in addition to a robot, it would be necessary to prepare a networked television, and there is a possibility that users will be confused by how to operate the remote control if it is unfamiliar to them. In addition, although it is conceivable to have an interface for talking directly with a robot by voice, there is a possibility that a high recognition accuracy could not be achieved depending on the environment in which the robot system is being used. Therefore, attention is paid to an input interface that uses

978-1-5386-2633-7/18/$31.00 ©2018 IEEE
DOI 10.1109/SCIS-ISIS.2018.00020

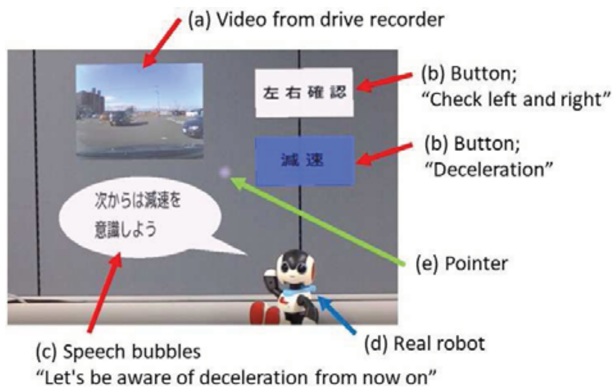

Fig. 2. Image from perspective of elderly people

gestures and mixed reality. To reduce the cognitive load of reviewing driving to as little as possible, simpler gestures are required. To use conventional gestures, a special input method must be adopted, so smoothly operating our system could be difficult. In this paper, we propose and evaluate two kinds of gestures used when using mixed reality in the robot system.

II. System overview

Figure 1 shows an overview of the system. The system transmits both video of a scene judged as dangerous that was taken by the drive recorder and questions that ask how one should react in such a scene to a mixed reality apparatus called "HoloLens." HoloLens displays these video and questions. In addition, in order to encourage intuitive understanding of displayed video and questions, information for operating the robot is transmitted from the system to the robot. Elderly people look at the video displayed on the HoloLens, receive instructions from the robot, and respond to the questions to review their driving.

Figure 2 shows an image of what elderly people see when viewing the robot and video from their perspective. (a) Video of driving, (b) buttons, and (c) speech bubbles are presented in mixed reality. These are created with computer graphics. Users can review their driving by using the buttons displayed in (b) to answer the robot's question while looking at the video presented in (a). The speech bubbles of (c) are used to present auxiliary information when elderly people cannot hear what the robot is says. The robot in (d) shares the video, utterance contents, and states which button has been pressed.

A. Robot

Regarding the functions of the robot, the robot is set up by deciding what it converses about, how it operates, and how it aids in reviewing driving. For reviewing driving, we adopt a review method with the goal of making it pleasurable and fun for those using the system. The robot developed in this research is classified as partner-type robot that learns together with learners. One problem with this type of robot is that learners sometimes feel that the behavior of the robot is uniform, so they gradually get bored with this joint-learning. To solve this, emotional expressions for robots have been proposed to make learners feel sympathy with the robot [5]. By using robot's emotional expressions, elderly people can enjoy reviewing their driving by sharing their emotions with the robot. In addition, to make them feel the effect of reviewing driving, Jimenez et al. investigated whether the learning methods of learners will change as those of robots change as learning progresses and found that the learning effect does improve [6]. By introducing this method, we will realize a learning effect suitable for each elderly person and enhance the continuity of learning.

B. Video analysis

With the video-analysis function, video taken by the drive recorder is analyzed, and driving situations that are more different than usual are extracted. Deep learning is used to detect these situations [7].

C. Interface

During the review, even if users were to receive information on a particular driving situation in spoken form from the robot, users may not be able to recall the situation. Therefore, we present situations in video form by using mixed reality during conversations with the robot. Furthermore, to learn how to drive appropriately while checking one's driving in the presented situation, the robot asks a question in problem form and converses with the user as they answer. Buttons are arranged within sight as an interface for answering. Users can answer questions by pressing these buttons with gestures.

In the next section, we explain how to press the buttons.

III. Pressing Buttons

We propose two gestures for pressing buttons in mixed reality. In this section, we first explain a gesture, which we call the "pinch gesture," that comes installed as standard with HoloLens. Next, we explain the proposed gestures, the "swipe gesture" and "laser and grasp gesture."

A. Conventional Gesture: Pinch

When the HoloLens is being worn, the pointer is set in the direction in which the wearer is facing as shown in Fig. 2. When moving the device, that is, the head, this pointer moves in conjunction with it. To press a button, the pointer is first moved to a button by moving the head. Next, the user pinches their thumb and forefinger in sight of the HoloLens. There are two problems with this. First, it is difficult to manipulate the pointer. To push a button, the head must be fixed in place on the button after moving. Since this is not done in everyday life, there is a possibility that it may take time to get used to. Second, it is necessary to keep raising one's hand to a certain height after adjusting the pointer to the button. Every time you press a button, you have to raise your hand and risk fatigue.

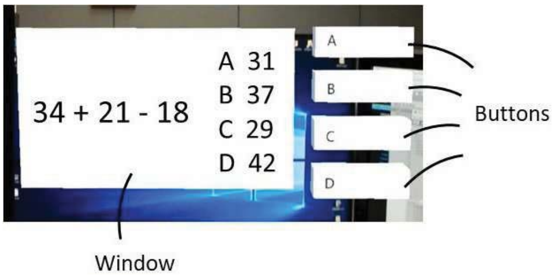

Fig. 3. Experimental environment

Fig. 4. Window displayed first

B. Conventional Gesture: Swipe

This gesture is based on the swiping used for input on tablets, which has become more common as tablets increase in use. It is intuitive because buttons can be touched directly. To press a button with this gesture, the user holds their right hand in front of the button that they want to select. Then, they swipe their arm to the right.

C. Conventional Gesture: Laser and Grasp (LnG)

This gesture imitates a laser pointer. A laser pointer can be used to control the pointer in the manner of finger-pointing. Thus, we think that a pointer can be easier to use than moving one's head. With this gesture, the laser pointer appears as if it is being emitted from the user's arm. To press a button, the pointer is aimed at the button, and a fist is made.

To create the two proposed gestures, we use MYO, which is armband-type device. MYO is attached to the forearm of the user and can detect arm movement. Tilting of the arm is measured by a gyro sensor, and the myoelectric potential of the arm is measured by using an inner electrode pad.

IV. Experiment

With our system, users review their driving by watching video displayed on a window created with computer graphics and answering the robot's questions. For an experiment, we constructed an environment similar to this situation and evaluated each of the three gestures. Figure 3 shows computer graphics displayed in the experimental environment. The participant solves the arithmetic question displayed in the window and selects the correct answer from among four options. In the example of Fig. 3, since the answer is "37," B is the correct answer. The participant answers by pressing the B button next to the window. When answering, the button is displayed in white, and the button turns light blue when pressed. How a button is pressed depends on the gesture. When pinching, the pointer has to face in the direction of the head. Swiping involves holding the right hand over the button. With LnG, when the pointer is aimed at the button, the button can be pressed. Also, after pressing the button, it is displayed in blue to indicate that it has been pressed correctly. After that, the button is erased and redisplayed in white.

The experimental procedure was explained beforehand. Participants were told about the flow of the experiment, the

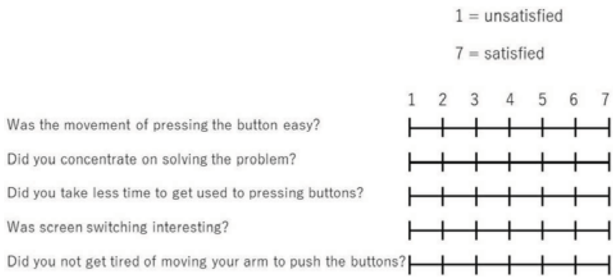

Fig. 5. Questionnaire

equipment to be used, and the contents of the experiment, as shown in Fig. 3. After that, they were given an explanation on how to press the buttons and use each gesture. The order of using the gestures was random for each participant, and a counterbalance was taken. Twenty windows were displayed in total per gesture. Figure 4 shows the window displayed first.

The first to third windows were used as practice windows, and participants pressed A, B, or C buttons according to the instructions. The fourth window was the last practice window, and pressing the D button displayed a arithmetic question window. The fifth to the nineteenth windows were for arithmetic questions, so there was a total of 15 questions to solve. The twentieth window showed the end of the experiment and instructed participants to remove the HoloLens. A questionnaire was conducted after trying each gesture.

Figure 5 shows the questionnaire. There were three questions related to button pressing and two questions for evaluating the system, and 7-point Likert scales were used for evaluation. The participants were 20 college students.

We also measured the following four types of time during the experiment as an index of the time required to press the buttons with each gesture.

First was the time required to answer all 15 arithmetic questions (total time), second was the duration of time for which the buttons were available to be pressed for each gesture (pressable time), third was the time during which the participant was facing the direction of the window during the experiment (window-facing time), and fourth was the time during which the participant was facing the direction of the buttons during the experiment (button-facing time).

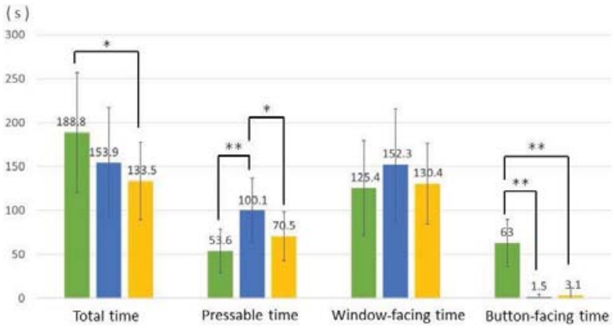

Fig. 6. Measurement results

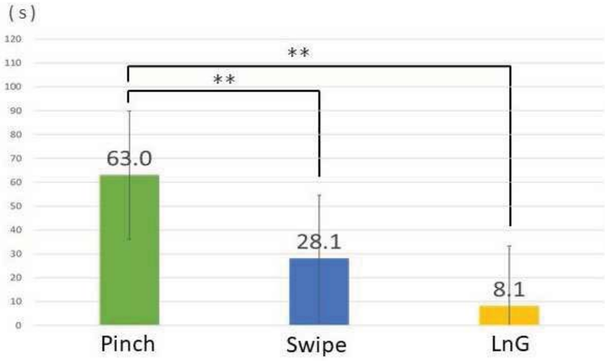

Fig. 7. Total button-pressing time

TABLE I
AVERAGE CORRECT ANSWER RATE

Pinch	Swipe	LnG
92%	72%	95%

V. RESULTS AND DISCUSSION

Figure 6 shows the measurement results. The vertical axis is time (seconds). The bar displayed for each item shows Pinch, Swipe, LnG from the left. For total time, a significant difference was observed between Pinch and LnG with a significance level of 5%. For pressable time, a significant difference was observed between Pinch and Swipe with a significance level of 1% and between Swipe and LnG with a significance level of 5%. There was no significant difference in window-facing time. For button-facing time, a significant difference of 5% was observed between Pinch and Swipe and between Pinch and LnG.

First, let us look at the button-facing time. Pinching requires facing the direction of the button whenever pressing the button. On the other hand, the two proposed gestures require only eye gazing in order to push a button. In fact, the measurement results were very short for the two gestures, compared with the average for Pinch, which was 63 seconds. From this, it can be said that, with the two gestures, it is possible to press a button without directly looking at it. From the above, the window-facing time of proposed gestures contains the time required to solve all of the arithmetic questions and the time to operate the button. Therefore, we first set the window-facing time (125.4 seconds) for Pinch as the reference time, and subtracted the reference time from the window-facing time for Swipe and LnG. Each of the times was 26.9 and 5.0 seconds. Second, we added the time to the button-facing time, respectively. We called the time as total button-pressing time, and compared the total button-pressing times of each gesture.

Figure 7 shows the total button-pressing time. Pinch was 63 seconds, Swipe was 28.1 seconds, LnG was 8.1 seconds, and a significant difference was observed between Pinch and Swipe and between Pinch and LnG with a significance level of 1%. From this, it can be said that the two proposed gestures can press buttons in a significantly shorter amount of time compared with Pinch.

From the results of the total button-pressing time and the total time, the gesture that could press buttons in the shortest amount of time was LnG.

Next, let us look at the pressable time and compare the pushability of the buttons. A significant difference was found between Pinch and Swipe and between Swipe and LnG in the pressiable time. By checking the video during the experiment, it was confirmed that more than half of the participants continued raising their hands after pushing the buttons. Since Pinch requires fixing the pointer on a button, even if you are raising your hand, it does not affect the pressable time. On the other hand, for LnG, as long as the laser hits the button, it is counted as the pressable time. Here, participants did not raise their hands for the reason to keep the laser on the button but raised them to reduce the time required to press the button the next time. Therefore, the position of the laser slightly deviated from the button because they did not concentrate on targeting it, so the pressable time did not increase greatly. Similarly, they answered arithmetic questions by keeping their hands to be raised. However, since the area judged as the button pressed state by the palm is wider than the range of LnG, it is considered that the pressable time significantly increased. From the above, it can be said that the significant difference in the pressable time is irrelevant to evaluating of pushability of the button rather than the one caused by the easy pushing of the button, because the one is merely to recognize the time of continuing to hold the hand.

Next, to evaluate mistakes made when pressing the buttons, the correct answer rates were compared. Table I shows the average rates. In this experiment, since only simple arithmetic questions with four possible answers were given, it is hard to assume that the correct answer rate will be significantly low. Nevertheless, the correct answer rate for Swipe was low, so button mistakes occurred. Pinch and LnG had nearly the same correct answer rate, and LnG could be used to press the buttons accurately.

Finally, Fig. 8 shows the questionnaire results. There was a significant difference between the Pinch and proposed LnG for the two items of “Did it take you less time to get used to pressing buttons?” and “Did not you get tired of moving

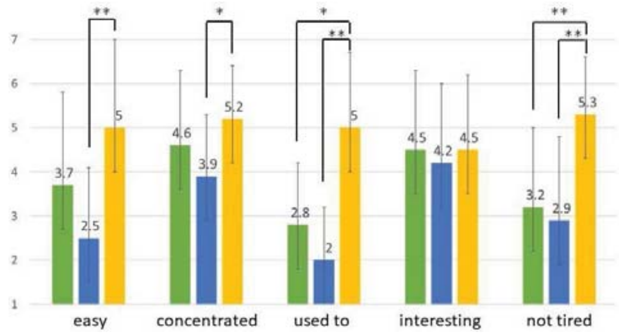

Fig. 8. Questionnaire result

your arm to push the buttons?", which were items concerning pressing the buttons.

From the above objective and subjective results, it can be said that the gesture most suitable for pressing buttons is LnG.

VI. Conclusion

We reported a robot system for improving driving while at home. Our idea is that, by presenting questions on videos taken when driving for reviewing driving, driving can be effectively improved. Therefore, the system we developed involves robots teaching users while users watch videos. To display a robot and video within the same field of view, mixed reality was used. Focusing on how to answer questions, we proposed two gestures. The experimental results showed that the gesture simulating a laser pointer was the most suitable one.

In the future, we will investigate the influence of review when using mixed reality in the robot system. We will also add controls to accurately guide the content [8] and gaze of conversations aimed at improving the acceptability [9] of robot functions.

Acknowledgements

This work was supported by a driving support project using agent, Nagoya University.

References

[1] D.Parker and S.Stradling: Influencing driver attitudes and behaviour, DETR Road Safety Research Report, 2001.

[2] I.Edwards: Can drivers really teach themselves, SMA Support Inc, 2014.

[3] T.Tanaka et al: Driver agent for encouraging safe driving behavior for the elderly, 5th International Conference on Human-Agent Interaction, 2017.

[4] Y.Ohta and H.Tamura: Mixed reality - merging real and virtual worlds, Ohmsha & Springer-Verlag, 1999.

[5] F.Jimenez, T.Yoshikawa, T.Furuhashi and M.Kanoh: Effects of a novel sympathy-expression method on collaborative learning among junior high school students and robots, Journal of Robotics and Mechatronics, 2018.

[6] F.Jimenez and M.Kanoh: Robot That Can Promote Learning by Observing in Collaborative Learning, IEEE International Conference on Systems, Man and Cybernetics, 2013.

[7] E.Hinton, S.Osindero and Y.Teh: A fast learning algorithm for deep belief nets, Neural Computation, 2006.

[8] T.Miyamoto, D.Katagami and Y.Shigemitsu: Improving relationships based on positive politeness between humans and life-like agents, 5th International Conference on Human-Agent Interaction, 2017.

[9] N.Mukai, S.Takashima, B.Yamanaka, K.Fuzikake, T.Tanaka, and H.Kanamori: Clustering for eye-gaze tracking data of drivers supported by drive assist agent, The 16th Asia-Pacific Forum, 2018.