September 1-2, 2023, Hiyoshi Campus, Keio University, Japan & Online
International Conference on Design and Concurrent Engineering 2023
& Manufacturing Systems Conference 2023

Regular Paper: No. 47

Evaluation of conveying spatial information by pointing gestures of a tele-immersion robot avatar

Yuki KIDA*, Kei MATSUOKA** and Tetsuro OGI*
*Graduate School of System Design and Management, Keio University
4-1-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa 223-8526, Japan
E-mail: ykida@keio.jp
**Department of Multidisciplinary Design Science, Shonan Institute of Technology
1-1-25 Tsujido, Nishi-kaigan, Fujisawa, Kanagawa 251-8511, Japan

Abstract
Robot avatars are attracting attention as a means of remote communication. One of the elements required when using a robot avatar to work and communicate with a partner in another location is the conveyance of information by pointing in order to share the location of objects and other information. However, it has not yet been fully evaluated to what extent pointing gesture by a robot avatar, which is treated as its own alter ego, is able to convey spatial information. In this study, we developed a robot avatar that realizes the tele-immersion of communication and evaluated the degree to which pointing information by the robot avatar can be conveyed to the other person through an evaluation experiment. The results showed that the information conveyed by the pointing gesture of the robot avatar using a humanoid communication robot was generally communicated and effective. In the case of the robot avatar that can be programmed to point in the correct direction, it was found that information is conveyed to some extent by pointing with the hand without using the index finger.

Key words : Avatar, Robot, Pointing gesture, Tele-Immersion, Communication

1. Introduction

In recent years, the use of robot avatars that act as one's alter ego to communicate with remote person has been attracting attention. Examples include "newme" by avatarin Inc. and "temi" by iPresence Ltd. robot avatars that combine displays and mobile robots. These robot avatars are effectively used as avatars by moving and displaying remote user's face on the display via video images. The characteristic of the communication of these robot avatars is the utilization of only video and audio on the display. This makes it difficult for the remote user to share the environment such as pointing to the location of objects in the environment where the communication partner is located, and to accurately convey the information to the partner and perform the collaborative work. Therefore, the authors have aimed to realize a robot avatar that enables remote users to communicate with a communication partner in real space with a high sense of presence, as if they were actually working in the same place, by utilizing a humanoid communication robot that can behave like a person. In other words, we aim to realize the tele-immersive communication in which the remote user can work at the same place as the communication partner by seeing the information captured by the robot avatar's eyes and expressing the remote user’s intention to behave by the robot avatar's gesture. One of the elements required when using a robot avatar to work and communicate with a partner in another location is the conveyance of information by pointing in order to share the location of objects and other information. For example, when people usually do design work together in the same location, they use conversation and pointing gesture to refer to what they are designing. However, it is difficult for a person in a remote location to share the spatial information with a local person. By giving visual instructions such as pointing gesture through the robot avatar, the remote person can provide clearer information to the local person, which is expected to support remote communication.

Regarding pointing gesture, previously, studies have been conducted on the accuracy of human pointing, and it has been reported that some errors occur in perceiving the location pointed to by the other party (Imai, et al., 2004). In addition, it has been reported that there is a gap between the location where the subject thinks they are pointing and the location illuminated by the laser pointer attached to their finger, and that the subject is not pointing to the object with pinpoint accuracy (Saito and Ogi, 2016). On the other hand, it has not yet been fully evaluated to what extent pointing gesture by a robot avatar, which is treated as its own alter ego, is able to convey spatial information.

Therefore, in this paper, we developed a robot avatar that realizes the tele-immersive communication and evaluated the degree to which pointing information by the robot avatar can be conveyed to the other person through an evaluation experiment.

3. Tele-immersion robot avatar
3.1 The robot and software used

In this study, Pepper (Pandey and Gelin, 2018), a humanoid communication robot provided by Softbank Robotics Co., Ltd., was used as a robot avatar (Figure 1). Pepper is a 121cm-tall humanoid robot that can communicate with people by voice. Pepper uses a large number of actuators, including on its head, shoulders, elbows, hips, knees, and wrists, allowing it to perform a wide variety of human-like behaviors and gestures. Pepper also has a camera mounted on its forehead, allowing the user to view the video images from Pepper's point of view. There have been several studies on human interaction with Pepper (Carros, et al., 2022) (Hsieh, et al., 2020) (Mou, et al., 2020).

Considering the above features of Pepper, such as its ability to interact with humans and to perform various human-like behaviors and gestures, we used Pepper as a robot avatar in this study. As for the control of Pepper, we developed an application software using Choregraphe.

Fig. 1 Humanoid communication robot Pepper

3.2 System design

Figure 2 shows an overview of the tele-immersion robot avatar system developed in this study. As shown in Figure 2, Pepper operates as a robot avatar by running the application software on the robot developed using Choregraphe. The robot avatar was designed as a system that outputs motion and audio from the robot by inputting motion commands and voice from a remote user on a PC on the remote location side. Specifically, motion commands and voice input on the PC on the remote location side are sent to the server as text data. Then, the application on the robot avatar reads the data and outputs motions and audio from the robot avatar.

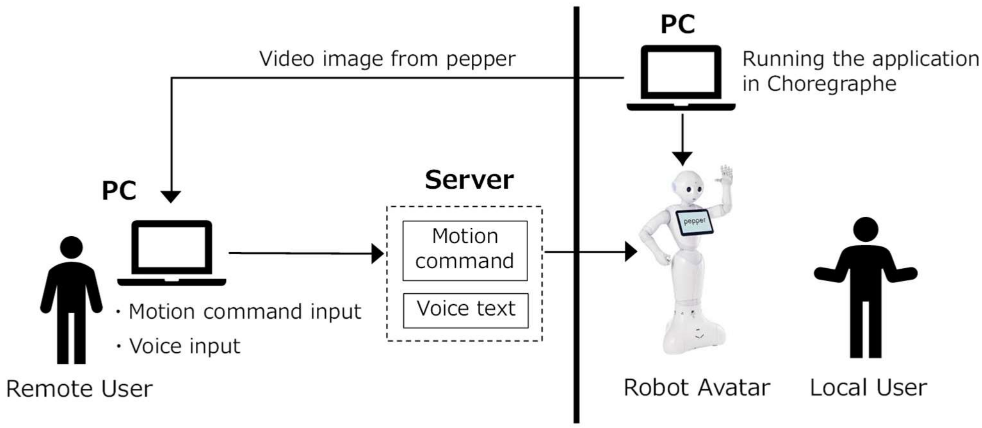


Fig. 2 Tele-immersion robot avatar system

The remote user can see local conditions and people through the video images captured by the camera on Pepper's forehead. For the video images sharing, the video image viewed from the Pepper's point of view was displayed on the screen on the PC placed near Pepper and it is shared with the PC on the remote location side using Quick Assist, which is an application for windows.

4. Evaluation experiment
4.1 Experimental conditions

In this study, a prototype system was used to evaluate the degree to which the pointing information of the robot avatar can be conveyed to its counterpart. Specifically, a prototype of robot avatar was used to point to the nine numbered circular objects shown in Figure 3 with the right hand. The nine numbered objects were placed on the wall in three rows and three columns with numbers 1 through 9 equally spaced at 50 cm. The height of the 3rd row at the bottom of the horizontal row was 50 cm from the floor. The diameters of the nine numbered circles were 10 cm.

Figure 4 shows Pepper's placement when pointing to the nine numbered objects. As for the placement of Pepper, Pepper's right shoulders with the 1st column from the left (1, 4, 7) of the arranged nine numbered vertical columns, as shown in Figure 4(a). We then created a motion gesture on Choregraphe to point to every nine numbered objects with the robot avatar's right hand at the 1m (Figure 4(b)) and 2m (Figure 4(c)) distance between Pepper and the wall where the numbered objects are placed, respectively, and implemented this motion in the application for the robot avatar.

In this experimental system, the pointing motion was performed by pointing to the target number with Pepper's arms straight out, as shown in Figure 4(b) and (c). For the method of creating the pointing motion, the pointing motion was created by attaching a laser pointer to Pepper's right hand; the pointing motion was defined as the laser pointer's light pointing in a straight line to the target number when Pepper's arm was straightened. Since Pepper's fingers can only be controlled to open or close, we decided to point with the open hand instead of using the index finger. The head motion was also controlled in the pointing motion so that the head moves along with the pointing. For the head motion, we defined the head direction as the state that is directed when the number is viewed in the center of the video image captured by the camera on Pepper's forehead. By using this method, remote user can point to the object while looking at the target through the robot camera.

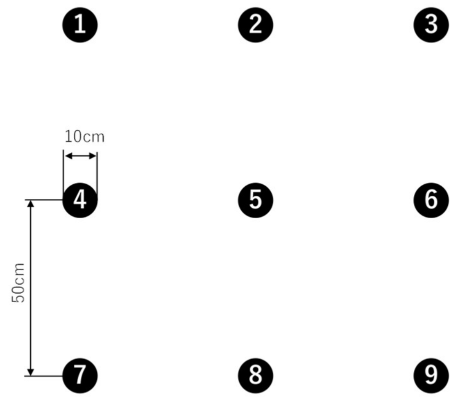
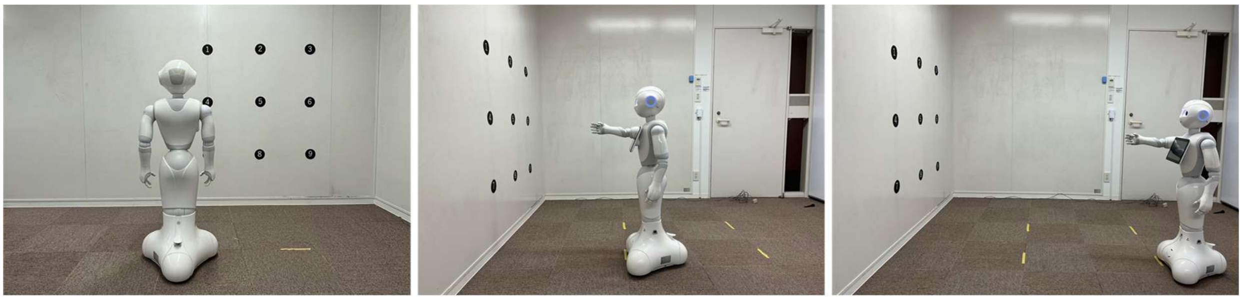


Fig. 3 Nine numbered objects



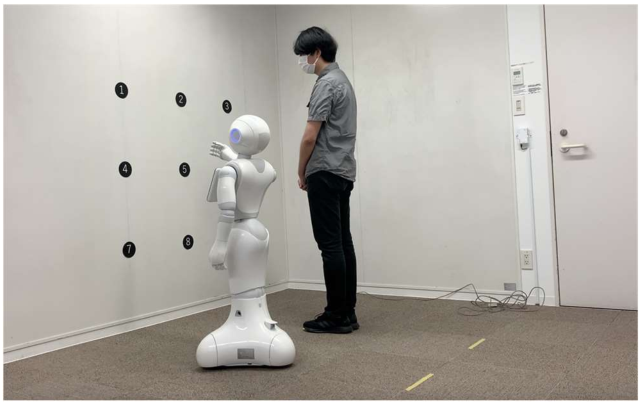
(a) A view from behind (b) 1m (c) 2m

Fig. 4 Pepper's placement when pointing to the nine numbered objects

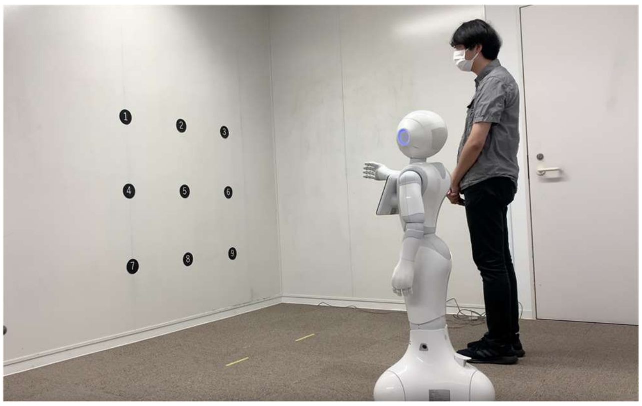
4.2 Experimental Task

For the evaluation experiment, the robot avatar and the subject were placed side by side as shown in Figure 5, and the pointing evaluation experiment was conducted at a distance of 1 m and 2 m from the wall on which nine numbered objects were placed, respectively. The subject stood on the right hand side of the robot avatar. The evaluation experiment was conducted on 10 subjects (5 males and 5 females).

In the experiment, first, the robot avatar pointed to one of nine numbered objects with its right hand, and then the subjects answered aloud the number that the robot avatar was pointing to. This trial was performed 18 times at 1 m and 18 times at 2 m for each subject. Each number from 1 to 9 was included twice in the 18 trials. At this time, the order of the numbers pointed to by the robot avatar was randomized each time. The experimenter controlled the robot avatar from a distance in the same room where the subject and the robot avatar were present and recorded the subject's responses. In this experiment, no voice instructions were given by Pepper in order to evaluate information conveyed only by pointing gesture. With regard to the subject's answers, the subject sometimes restated their answers, in which case the restated number was evaluated as their answer. In addition, subjects were not given feedback on whether they answered correctly or incorrectly after each answer.



(a) 1m



(b) 2m

Fig. 5 Evaluation experiments

4.3 Result and discussion

First, the graph showing the total number of incorrect answers for each number is shown in Figure 6. The feature that the number of incorrect answers at numbers 2, 5, and 8 is relatively high is shown. The numbers 2, 5, and 8 are the numbers in the 2nd column of the numbers arranged in three rows and three columns. In other words, we found that there were differences in the accuracy of pointing gestures in vertical columns and horizontal rows.

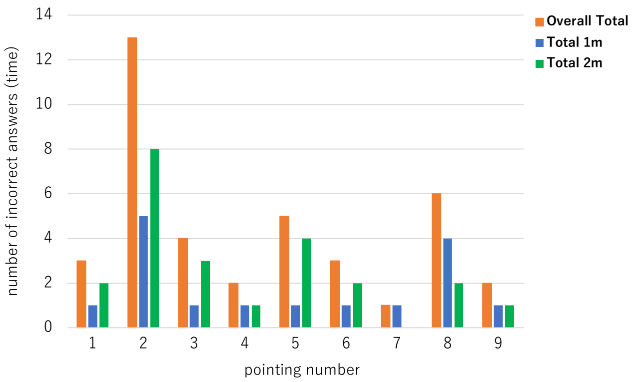


Fig. 6 Total number of incorrect answers for each number

Therefore, we conducted an analysis of variance for the correct answer on the factors of subject, distance, vertical columns, and horizontal rows, where the correct answer is 1 and the incorrect answer is 0. The results showed that there was no significant difference (p=0.169) in the factor of distance, and significant differences in the factors of subject (p<0.001), vertical column (p<0.001) and horizontal row (p=0.014).

Regarding the factor of distance, the average percentage of correct answers at 1m and 2m were generally high, 91% at 1m and 87% at 2m. The results show that for a 10cm diameter objects placed on a wall in three rows and three columns at 50cm intervals, the conveyance of information by pointing gesture by a robot avatar is effective at distances of 1 m and 2 m from the object.

The reason for the significant differences for the subjects is thought to be related to the fact that each subject has a different viewpoint and angle of view due to differences in height. Next, since there were significant differences for the vertical column and horizontal row factors, multiple comparisons were conducted. The results of the multiple comparisons for the vertical column are shown in Figure 7, and the results of the multiple comparisons for the horizontal row are shown in Figure 8.

The results in Figure 7 show that the accuracy in the 2nd column (2,5,8) were the lowest, and significant differences were confirmed between the 1st column (1,4,7) and 2nd column (p<0.001) and between the 2nd column and 3rd column (3,6,9) (p<0.001). This result is considered to be caused by the arrangement that the 2nd column has other columns (1st and 3rd) on both sides. On the other hand, the results in Figure 8 show that the 1st rows (1,2,3) have the lowest accuracy, and significant differences were confirmed between the 1st row and the 2nd row (4,5,6) (p=0.043) and between the 1st row and the 3rd row (7,8,9) (p=0.023). These results indicate that there is some error in the conveyance of pointing position in the 1rst row and it should be improved.

The possible improvement is the state of the robot avatar's arms during the pointing motion. In this study, the robot avatar pointed to the target with its arms straight outstretched. Usually, when we point to the target, we move our arms by flexing our elbows, shoulders, and other joints. In this experiment, the robot avatar pointed to the target with its arms straight, which may have caused confusion on the row and column lines. When people convey information by pointing gesture, they flexibly change their elbows, shoulders, and other joints, and this action is considered to be an important act in conveying information by pointing. Therefore, the development of a robot avatar that can point to the target while freely bending its elbows, shoulders, and other joints, like human joint movements will be future work.

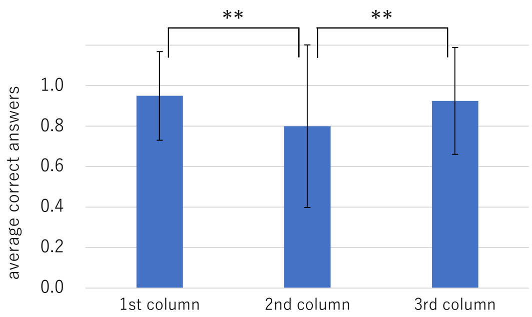


Fig 7 Multiple comparisons (columns)

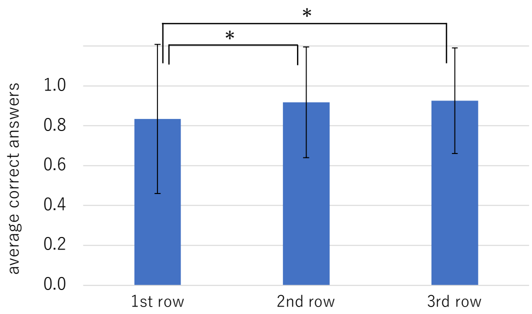


Fig. 8 Multiple comparisons (rows)

As described in previous studies in section 1, it has been reported that in the case of humans, there is a gap between the place where the humans think they are pointing and the place illuminated by the laser pointer attached to their finger. But in this study, the robot avatar was programmed to make accurate pointing gestures so that the laser pointer attached to its hand pointed to the target object (Figure 9). As a result, it can be said that spatial information could be conveyed with a certain degree of accuracy using only the direction of the head and the direction of the hand, without using the index finger.

Fig. 9 Pointing gestures for people and the robot avatar

5. Conclusion and future work

In this study, we aim to realize a robot avatar that enables tele-immersive communication. Namely, it is a robot avatar that allows remote users to communicate with a partner in a different location with a high sense of presence, as if they were working in the same place. One of the elements necessary to achieve this is the transmission of information by pointing gesture. In this paper, we developed a tele-immersion robot avatar system and evaluated the degree to which pointing information can be conveyed by the robot avatar to the partner. The results showed that the information conveyed by the pointing gesture of the robot avatar using a humanoid communication robot was generally communicated and effective. In the case of the robot avatar that can be programmed to point in the correct direction, it was found that information is conveyed to some extent by pointing with the hand without using the index finger.

Future work includes the evaluation of pointing gestures by a robot avatar that moves in conjunction with the movements of a person's wrist, elbow, shoulder, rather than the programmed straight arm pointing gesture. Another future work is to propose a method for resolving the variation in pointing accuracy of the robot avatar due to the effects of the individual subject's height and field of viewpoint. The effectiveness of the pointing of the robot avatar at 1m and 2m was confirmed in this study, and next verification at various distances and angles from the object is needed, taking into account that the robot avatar can move. In addition, more effective conveyance of spatial information can be expected if the robot avatar performs pointing gestures while moving on the floor and verification of this is also necessary.

Acknowledgements

This research was partially supported by JKA promotion funds from KEIRIN RACE (2022M-272).

References

Carros, F., Schwaninger, I., Preussner, A., Randall, D., Wieching, R., Fitzpatrick, G. and Wulf, V., Care workers making use of robots: Results of a three-month study on human-robot interaction within a care home, In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (2022), pp.1-15.

Hsieh, W. F., Sato-Shimokawara, E. and Yamaguchi, T., Investigation of robot expression style in human-robot interaction, Journal of Robotics and Mechatronics, 32(1), (2020), pp.224-235.

Imai, T., Sekiguchi, D., Kawakami, N., and Tachi, S., Measuring accuracy of nonverbal information perception of humans: measurement of pointing gesture perception, Transactions of the Virtual Reality Society of Japan, (2004), Vol.9 Vol.1, pp.89-95 (in Japanese).

Mou, W., Ruocco, M., Zanatto, D. and Cangelosi, A., When would you trust a robot? a study on trust and theory of mind in human-robot interactions, In 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN) (2020), pp.956-962.

Pandey, A. K. and Gelin, R., A mass-produced sociable humanoid robot: Pepper: The first machine of its kind, IEEE Robotics & Automation Magazine, 25(3), (2018), pp.40-48.

Saito, M., and Ogi, T., Development of Information Guide System Based on Natural User Interface, The 26th Design and System Division Conference, Japan Society of Mechanical Engineers, (2016), pp.1-5 (in Japanese).