Necessary Requirements of Avatars for Remote Communication in Real Space

Yuki Kida[1], Kei Matsuoka[2], Tetsuro Ogi[1]
[1] Graduate School of System Design and Management, Keio University, Japan
ykida@keio.jp, ogi@sdm.keio.ac.jp
[2] Department of Multidisciplinary Design Science, Shonan Institute of Technology, Japan
matsuoka@md.shonan-it.ac.jp

Abstract. The purpose of this study was to evaluate requirements for avatars in hybrid communication environments, namely the appearance of the avatar and the presence of a physical entity for remote communication in real space. Therefore, we constructed remote communication systems using three types of avatars, taking into account the appearance of the avatar whether it is the person himself or not and whether it has a physical entity or not. We then conducted evaluation experiments comparing communication with said avatars to communication with a real human. As a result, it was found that it is necessary for the avatar to have the same appearance as the remote user in order to enhance the sense of being in the same room, the sense of collaborating with the specific remote user, etc.

1 Introduction

In recent years, remote communications in a hybrid environment have become commonplace, combining offline meetings in real space with online video meeting tools that allow people in remote locations to participate. In this case, participants in the same physical meeting space can have natural conversations and collaborate with each other, but it is difficult for those participants to feel that remote users are in the same space. Therefore, one of the important research issues in remote communication in a hybrid environment is how to express the presence of remote users who are not in the physical meeting space.

As a method to share the presence of remote users, there has been active research on remote communication using AR avatars, which are virtual embodiments of remote users in real space, or robot avatars, which are physical embodiments of remote users in real space [1-4]. However, the necessary requirements for real-space avatars to express the presence of remote users have not yet been fully explored. In the case of an AR avatar, there is no physical entity, but its appearance can be freely changed to communicate. On the other hand, in the case of a robot avatar, its appearance cannot be changed freely, but there is a physical entity present. Based on these specific characteristics, the proposed viewpoints are a) the avatar has the same virtual appearance as the remote user, b) the avatar is a physical entity, but it does not have the same appearance as the remote user, and c) the avatar is neither a physical entity nor has the same virtual appearance as the remote user.

Therefore, for this study on remote communication using avatars in real space, we prepared three types of avatars. The first is an “AR real avatar with the same apparent virtual presence as the remote user”. The second is a “Robot avatar with only a physical entity that does not have the same appearance as the remote user”. The third is an “AR robot avatar without the same apparent virtual presence as the remote user and is not a physical entity”. The purpose of this study is to compare how people communicate with those three types of avatars as well as directly with a human in the same physical space, and to make a comparative evaluation regarding the necessary requirements for avatars, such as the appearance of the avatar and the presence of a physical entity. For this purpose, we constructed a remote communication system in real space using the three types of avatars described above and conducted evaluation experiments.

2 System architecture

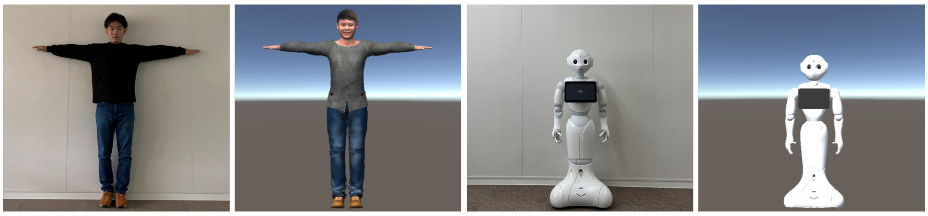
2.1 Avatars used

In this study, we used an “AR real avatar with the same apparent virtual presence as the remote user” (Fig.1(b)), a “Robot avatar with only a physical entity that does not have the same appearance as the remote user” (Fig.1(c)), and an “AR robot avatar without the same apparent virtual presence as the remote user and is not a physical entity” (Fig.1(d)). We then compared communicating with them to communication with a physically present human experimenter (Fig.1(a)).

First, the “AR real avatar” treated in this study is an avatar made to look like a real human, and in this study, it was made to look like the experimenter’s appearance. The AR real avatar was created based on the face of the experimenter using *Character Creator3,* a software package provided by Reallusion Inc. The AR real avatar was given the ability to lip-sync through Unity so that it could move its mouth in conjunction with vowels (a, i, u, e, o), as shown in Fig.2. In addition, the height of the AR real avatar was adjusted to be displayed in real space at about 168 cm in order to match the appearance of the experimenter.

The robot avatar was a communication robot, Pepper [5], provided by Softbank Robotics Co., Ltd. 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. There have been several studies on human interaction with Pepper [6,7] which led to this study choosing Pepper’s use as an avatar.

Regarding the AR robot avatar, in order to make a comparison as an “AR robot avatar without the same apparent virtual presence as the remote user and is not a physical entity”, the conditions were aligned so that we created a virtual model of Pepper with the same appearance and height as the real Pepper and used it as an AR robot avatar.



(a) Real human (b) AR real avatar (c) Robot avatar (d) AR robot avatar

Fig. 1. Real human and the three types of avatars used

Fig. 2. AR real avatar with lip-sync

2.2 Remote communication system using avatars

(1) AR real avatar

Figure 3 shows an overview of the remote communication system using an AR real avatar in real space. In this study, we constructed a system in which a remote user appears as a virtual AR avatar in front of a subject in another room and communicates with the subject while performing movements linked to the remote user. The devices used were the HoloLens2 for presenting the AR avatar, an Azure Kinect kit for tracking the remote user's body movements, a control PC to make the avatar move and output sound using Unity, and a pair of communication PCs.

The system function is described below. First, the remote user is body-tracked when standing in front of the Azure Kinect connected to the control PC. At the same time, the avatar is moved by a system for real-time motion of the avatar in linkage with the body tracking constructed on Unity in the control PC. Then, the avatar is displayed on the HoloLens2 from the control PC, which is worn by the subject in another room, using the Holographic Remoting Player function of HoloLens2. The avatar is displayed face-to-face with the subject wearing the HoloLens2. The remote user's voice is output to the HoloLens2 via Unity from the control PC's microphone. The voice of the AR real avatar was the voice of the experimenter. The voice from the subject's side in another room was sent from the subject's communication PC to the remote user's communication PC using a video meeting tool. The viewpoint image of the subject wearing the HoloLens2 could be sent from the HoloLens2 to the remote user's communication PC, and the remote user could confirm the movement of the avatar who was facing the subject.

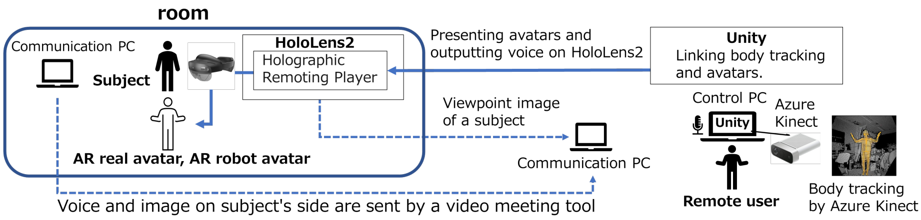


Fig. 3. Remote communication system using the AR avatar

(2) Robot avatar

Figure 4 shows an overview of the remote communication system using a robot avatar in real space. In this study, we constructed a system in which a remote user, as a robot avatar, faces a subject in another room and communicates with the subject by moving the robot avatar. The devices used were a physical Pepper robot used as the avatar, a server, a PC for robot avatar operation, and a communication PC. Regarding the control of the robot avatar, since Pepper is designed to output voice from text data and cannot output the remote user's voice as it is, a system was needed to convert the remote user's voice to text data and output it as Pepper's voice. As for the motions of the robot avatar, the robot avatar's motions were programmed in advance, and the system was designed to perform the programmed motions by inputting and operating the motion information. Therefore, the control of this robot avatar is done through a web application on the PC for the robot avatar operation. The web application implemented a function for inputting the remote user's voice and a function for inputting motion commands to operate the robot avatar.

The system function is described below. First, for voice input by the remote user, the remote user presses the Start Voice Input button on the web application and speaks into the PC. The input remote user's voice is converted to text by the Web Speech API and sent to the server via HTTP communication. For example, if the remote user says "Hello" to the PC, it is converted to text data as "Hello" and sent to the server. As for the input of motion commands to operate the robot avatar, the remote user presses a button such as point to lower right on the web application, and the motion instruction is sent to the server via HTTP communication. The voice text and motion command information sent to the server is read by the robot application in Pepper, and the voice and motion are output from Pepper. The audio and motion on the subject's side were confirmed through communication between the subject's PC for communication and the remote user's PC for operating the robot avatar, using a video meeting tool.

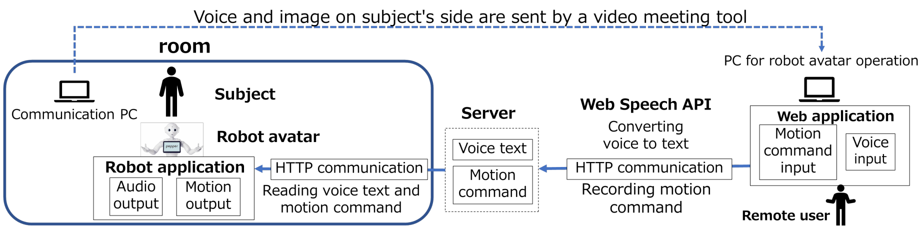


Fig. 4. Remote communication system using the robot avatar

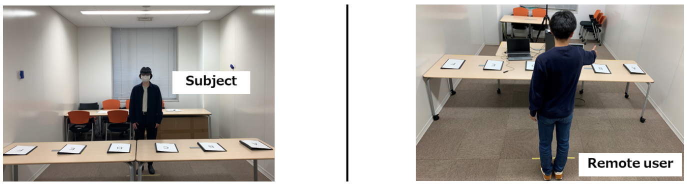
(3) AR robot avatar

The remote communication system using the AR robot avatar in real space is the same as the system described in (1) AR real avatar (Fig.3). As for the AR robot avatar, the range of motion of the AR robot avatar's head and arms was limited on the Unity control PC in order to match the range of motion of the actual Pepper's head and arm joints. For the voice of the AR robot avatar, we played the pre-recorded voice of Pepper.

3 Evaluation experiment

3.1 Experimental environment and procedures

In the evaluation experiment, subjects performed the tasks of collaborating with an avatar and evaluated their experience after the tasks by filling in a questionnaire. The collaborative tasks consisted of having the subject take three files in the direction indicated by the avatar's hand. The experimental environment consisted of two rooms, one with a single subject and the other with a single remote user (experimenter) (Fig.5). The rooms on the subject's side and on the remote user's side were created in the same environment so that the direction pointed by the remote user's hands and the direction pointed by the avatar's hands in the room on the subject's side were the same. In the experimental environment, subjects and avatars faced each other across a 70 cm-high table with six files on it. The avatar is presented in the same orientation as the remote user, facing the subject. The distance from the subject to the file was 1 m, the distance from the file to the avatar was 1 m and the distance between the subject and the avatar was 2 m. The six files were placed 50 cm apart on the table. Similarly, in the room on the remote user side, six files were placed on a table at intervals of 50 cm, with a distance of 1 m from the files to the remote user.



(a) Room on the subject's side (b) Room on the remote user's side

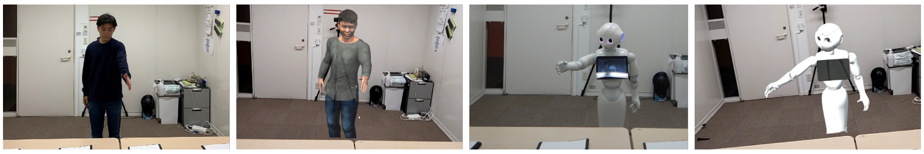
Fig. 5. Experimental environment

The specific experimental procedure is described below. First, the subject and avatar begin the experiment facing each other across the table. The avatar says, "Please take it", and points to the first file with the avatar's hand at random. The subject takes the file in the direction of the avatar's hand and places it on the table. Again, the avatar says, "Please take it", and points to another file with the avatar's hand at random. The subject takes the file in the direction of the avatar's hand and places it on the table in a stack. This procedure was repeated, and the subject was asked to take three files (out of six) in the direction indicated by the avatar's hand and stack them (Fig.6). The instruction of the voice was unified to "Please take it", and the voice of the AR real avatar was the voice of the experimenter, and the voice of robot avatar and AR robot avatar was the voice of Pepper. In terms of movement, the AR real avatar and AR robot avatar could point to a file by hand with the same motion linked to the remote user. As for the robot avatar, motions were programmed to point to the six files by the robot avatar's hand, and the motions were controlled by button operations on the operation PC.

Fig. 6. How to take files and stack them

3.2 Evaluation method

As for the evaluation method, we first conducted an experiment in which the experimenter himself and the subject performed collaborative tasks (Fig.7(a)), without using an avatar. The evaluation that the subject felt when the experimenter gave the instructions in the same room was used as the standard stimulus 10, and the subject evaluated the subjective evaluation of the three types of avatars (Fig.7(b), (c), (d)) by answering a questionnaire on a numerical scale by magnitude estimation [8]. For each of the three types of avatars, the subjects were asked to answer a questionnaire at the end of the collaboration tasks. The order of the three types of avatar experiments was random for each subject. The subjects performed two sets of experiments, and the second set of three avatar experiments was performed after the first set. The experiment was conducted on 10 subjects in their 20s to 50s in age. The questionnaire items are shown in Table1.



(a) Experimenter (b) AR real avatar (c) Robot avatar (d) AR robot avatar

Fig. 7. Collective tasks with an experimenter, three types of avatars

Table 1. Questionnaire items

Q1. Did you understand the instructions?
Q2. Did you feel as if the remote user (experimenter) was in the same room?
Q3. Did you feel as if the avatar was in the same room?
Q4. Did you feel as if you were doing the collaborative tasks with the remote user (experimenter)?
Q5. Did you feel as if you were doing the collaborative tasks with the avatar?
Q6. Did you find it easy to communicate with the remote user (experimenter)?
Q7. Did you find it easy to communicate with the avatar?
Q8. Did you have a favorable impression towards the remote user (experimenter)?
Q9. Did you have a favorable impression towards the avatar?
Q10. Were you able to trust the remote user (experimenter)?
Q11. Were you able to trust the avatar?

4 Results and discussion

Figure 8 shows the results of the evaluation of each response to the questionnaire. First, the paired evaluation items of Q2 and Q3, Q4 and Q5, Q6 and Q7, Q8 and Q9, and Q10 and Q11 for remote users and avatars were evaluated by t-test. Figure 8 shows that the evaluation of each item (Q3, Q5, Q7, Q9, Q11) for the avatar is significantly higher than the evaluation of each item (Q2, Q4, Q6, Q8, Q10) for the remote user. In other words, it was confirmed that the sense of being in the same room, the sense of working on tasks collaboratively, the ease of communication, holding a favorable impression, and having a sense of trust were evaluated higher with the avatar in front of the subject compared to with the remote user. These results indicate that it is necessary to enhance the sense of being in the same room, the sense of working collaboratively on tasks, etc., with the remote user.

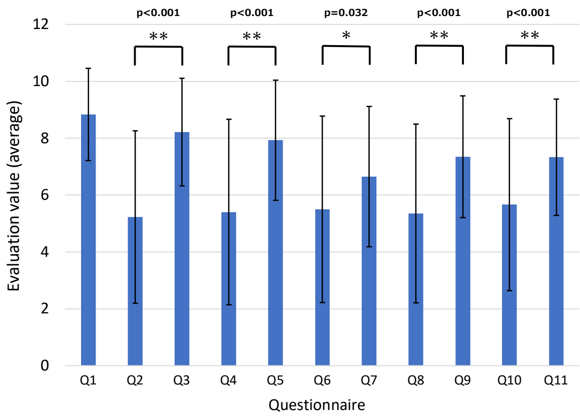


Fig. 8. Questionnaire Results

Next, in order to examine the necessary requirements of avatars for remote communication in real space, an analysis of variance was conducted from the evaluation experiment using the presence or non-presence of the physical entity of the avatar and the appearance of the avatar whether it is the person himself or not as factors. Figure 9 compares the responses according to the presence or non-presence of the physical entity of the avatar and Figure 10 compares the responses according to the difference in the appearance of the avatar whether it is the person himself or not.

As shown in Figure 9, there were significant differences in the evaluation of the entity for Q3 (p=0.001), Q5 (p=0.002), Q7 (p=0.039), and Q9 (p=0.004), and the evaluation was higher when the avatar entity was present. Therefore, it became clear that the presence of a physical entity for an avatar increases the sense of being in the same room, the sense of collaborating for tasks, the ease of communication, and holding favorable impressions with the avatar in front of the subject rather than with remote users. In addition, regarding Q11, there were no significant differences in sense of trust towards the avatar in front of the subject. In this regard, for example, when a person has a sense of trust in a certain person, information such as what kind of person he is influences the sense of trust and is considered important. Therefore, the requirement of the presence of the physical entity of the avatar did not have much of an effect on the feeling of trust, and there may have been no significant difference in the evaluations of Q11.

In addition, as shown in Figure 10, for the apparent factor, there were significant differences in Q2 ($p<0.001$), Q4 ($p<0.001$), Q6 ($p<0.001$), Q7 ($p<0.001$), Q8 ($p<0.001$), and Q10 ($p<0.001$) with the avatar's appearance as the remote user himself being more highly rated. The results indicate that the avatar's appearance as the remote user himself enhances the sense of being in the same room, the sense of collaborating for tasks, the ease of communication, favorable impressions, and the sense of trust, with the remote user. Regarding the ease of communication in Q6 and Q7, it can also be considered that the lip-sync of the AR real avatar affected and enhanced the ease of communication resulting the avatar's appearance as the remote user himself was rated higher in both Q6 and Q7, and the difference was significant.

In addition, there was no significant difference in the understanding of the instructions given by the avatar in Q1 for both the entity and the appearance of the avatar, confirming that the presence or non-presence of the physical entity of the avatar and the difference in the appearance of the avatar whether it is the person himself or not have no effect on the understanding of the instructions.

Therefore, from Figures 9 and 10, it was found that in the case of remote communication in real space, which requires the sense of being in the same room, the sense of collaborating for tasks, the ease of communication, favorable impressions, and the sense of trust, with the remote user, it is effective for the avatar to have the same apparent appearance as the remote user. On the other hand, in the case of remote communication in real space, where there is no need to feel the sense of being in the same room and the sense of collaborating for tasks with a specific remote user (for example, when anonymity is required), even an avatar with only a physical entity can be considered effective. From this, it can be said that it is important to properly use the requirement of the presence or non-presence of the physical entity of the avatar and the appearance of the avatar whether it is the person himself or not according to the necessity. For example, for remote communication that requires the sense of being in the same room, the sense of collaborating for tasks, the ease of communication, favorable impressions, and the sense of trust, with the specific remote user, such as a meeting or an interview, it is effective to use an avatar that has the same appearance as the remote user. On the other hand, in the case of remote communication, such as customer service to an unspecified number of people, where remote user information is not required, it is effective to use an avatar that has only a physical entity.

Based on the above, it became clear that it is very important for an avatar to have the same appearance as the remote user in order to enhance the sense of being in the same room, the sense of collaborating for tasks, etc., with the remote user, not the avatar in front of him or her in a real space. Therefore, it is considered necessary to make the avatar look more similar to the remote user in order to more greatly enhance the sense of being in the same room, the sense of collaborating for tasks, etc. with the remote user. In this study, we used an "AR real avatar" as an avatar that has the same appearance as the remote user but expected one way to make the avatar look more like the person himself/herself is to use a video avatar [9], which uses the video image of the person himself/herself as the avatar. Therefore, the development and comparative evaluation of the video avatar in remote communication in real space will be an important issue in the future.

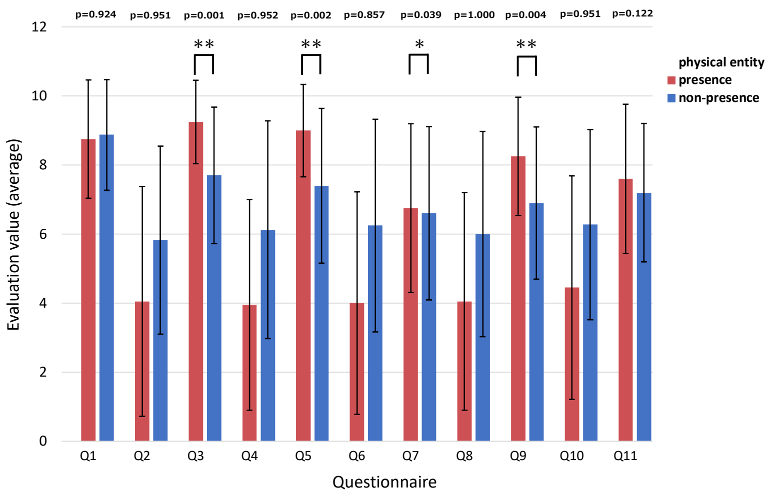


Fig. 9. Comparison of the presence or non-presence of the physical entity of the avatar

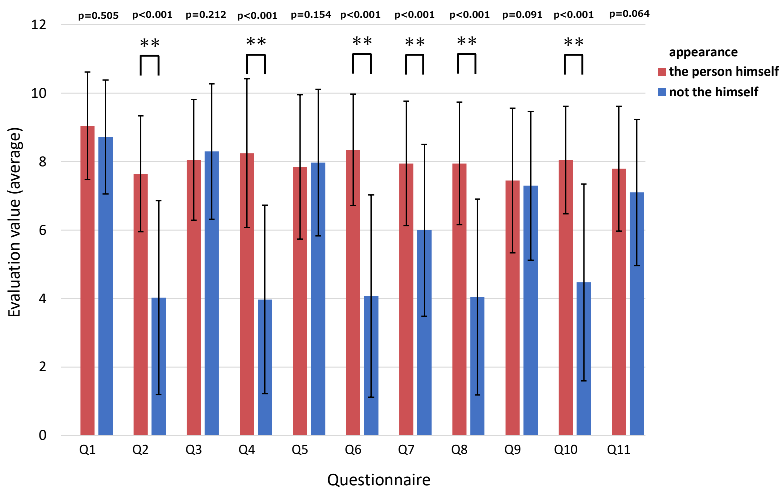


Fig. 10. Comparison of the difference in the appearance of the avatar

5 Conclusion and future work

In this study, in order to examine the necessary requirements of avatars for remote communication in real space, we constructed remote communication systems using three types of avatars, taking into account whether it has a physical entity or not and the appearance of the avatar whether it is the person himself or not, and conducted evaluation experiments comparing them with communication with a real human. As a result, it was found that in the case of remote communication in real space, which requires the sense of being in the same room, the sense of collaborating for tasks, the ease of communication, favorable impressions, and the sense of trust, with the remote user, it is effective for the avatar to have the same apparent appearance as the remote user. On the other hand, in the case of remote communication in real space, where there is no need to feel the sense of being in the same room or for the sense of collaborating for tasks with the specific remote user, even an avatar with only a physical entity can be considered effective when anonymity is required. In summary, it was found that it is important for the avatar to have the same appearance as the remote user in order to enhance the sense of being in the same room with the remote user, the sense of cooperating for tasks with the remote user, etc. Therefore, it is suggested that one way to make the avatar look more like the person himself/herself is to use a video avatar, which uses the video image of the person himself/herself as the avatar. The development and comparative evaluation of these systems are future issues.

Acknowledgments. This research was partially supported by JKA promotion funds from KEIRIN RACE (2022M-272). This work was supported by JST SPRING, Grant Number JPMJSP2123.

References

1. Yu, K., Gorbachev, G., Eck, U., Pankratz, F., Navab, N., Roth, D.: Avatars for Teleconsultation: Effects of Avatar Embodiment Techniques on User Perception in 3D Asymmetric Telepresence. IEEE Trans. Vis. Comput. Graph. 27 (2021) 4129–4139
2. Wang, X., Wang Y., Shi, Y., Zhang, W., Zhang, Q.: AvatarMeeting: An Augmented Reality Remote Interaction System With Personalized Avatars. Proceedings of the 28th ACM International Conference on Multimedia (2020) 4533–4535
3. Schwarz, M., Lenz, C., Rochow, A., Schreiber, M., Behnke, S.: NimbRo Avatar: Interactive Immersive Telepresence with Force-Feedback Telemanipulation. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2021) 5312–5319
4. Morita, T., Mase, K., Hirano, Y., Kajita, S.: Reciprocal attentive communication in remote meeting with a humanoid robot. Proceedings of the 9th international conference on Multimodal interfaces (2007) 228–235
5. Pandey, A.K., and Gelin, R.: A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. IIEEE Robotics & Automation Magazine Vol.25, 3 (2018) 40–48
6. Di Nuovo, A., Varrasi, S., Lucas, A., Conti, D., Mcnamara, J., and Soranzo, A.: Assessment of Cognitive skills via Human-robot Interaction and Cloud Computing. Journal of Bionic Engineering 16 (2019) 526–539
7. Barakeh, Z.A., Alkork, S., Karar, A. S., Said, S., Beyrouthy, T.: Pepper Humanoid Robot as a Service Robot: a Customer Approach. 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART) (2019) 1–4
8. Han, S.H., Song, M., Kwahk, J.: A systematic method for analyzing magnitude estimation data. International Journal of Industrial Ergonomics, Vol.23, 5-6 (1999) 513–524
9. Ogi, T., Yamada, T., Tamagawa, K., Kano, M., Hirose, M.: Immersive telecommunication using stereo video avatar. Proceedings IEEE Virtual Reality 2001 (2001) 45–51