IoT avatar: various objects in real space are anthropomorphised as avatars

Yuki Kida*, Toshihiko Chiba and Tetsuro Ogi

Graduate School of System Design and Management,
Keio University,
Yokohama, Japan
Email: ykida@keio.jp
Email: chiba.space@keio.jp
Email: ogi@sdm.keio.ac.jp
*Corresponding author

Abstract: In this paper, we propose an IoT avatar that anthropomorphises various objects in real space as avatars that can communicate with people around them. Each IoT avatar is realised by attaching a small communication device equipped with a camera, microphone, and speaker that is operated by a remote user. In the context of using avatars in real space, robot avatars have attracted attention and have been effectively utilised. However, the use of these robot avatars may require a large number of robots to be deployed at each remote location. In contrast, an IoT avatar has the advantage that the objects themselves can be used as avatars. In this study, we focused on the voice of an IoT avatar with the aim of anthropomorphising various objects as avatars. Therefore, a prototype was developed that implements a voice transformation function and a real-time altered auditory feedback function for a remote user. From the evaluation experiments, it can be said that the proposed system was able to anthropomorphise objects as avatars and communicate with people on-site. It is also suggested that the remote users can communicate with the people on-site as if they were becoming the objects themselves via the proposed system.

Keywords: internet of things; IoT; avatar; anthropomorphism; remote communication.

Reference to this paper should be made as follows: Kida, Y., Chiba, T. and Ogi, T. (2025) 'IoT avatar: various objects in real space are anthropomorphised as avatars', *Int. J. Web and Grid Services*, Vol. 21, No. 1, pp.42–57.

Biographical notes: Yuki Kida is a PhD student at Graduate School of System Design and Management, Keio University. He has been an assistant professor (research encouragement) at the same graduate school since 2022. His research interests include human interface and virtual reality technology.

Toshihiko Chiba is a Master's student at Graduate School of System Design and Management, Keio University. He is mainly interested in human interface and virtual reality technology.

Tetsuro Ogi is a Professor at the Graduate School of System Design and Management, Keio University. He obtained his Master's in Mechanical Engineering from the University of Tokyo in 1986, and joined Mitsubishi Research Institute as a researcher. He obtained his PhD in Mechanical Engineering from the University of Tokyo in 1994 and was named Associate Professor at the University of Tokyo in 1996. He moved to the Department of

Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba in 2004 as an Associate Professor. In 2008, he moved to Keio University. His research interests include human interface, virtual reality, augmented reality, high presence communication, etc.

1 Introduction

Internet of things (IoT) technology has made it possible to exchange data through the Internet with various things in our surroundings. Currently, IoT technology has been integrated into commercially available products and services. Specifically, it has primarily involved using sensors to collect data for analysing and monitoring the thing's surroundings, as well as sending signals from remote locations to control the thing.

On the other hand, in recent research and development related to IoT technology, efforts are not limited to the transmission and reception of sensor data or remote control; there is also a focus on developing effective interactions between IoT devices and the people around them (Alanwar et al., 2017; Rubio-Drosdov et al., 2017; Delgado Rodriguez et al., 2021). A technology that has received particular attention is the voice user interface (Meng et al., 2020; Zhang et al., 2020; Muhamad et al., 2021). For example, by using smart speakers such as Amazon Echo and Google Home, users can interact with surrounding IoT devices through voice commands within their home. In addition, there has been research on IoT devices themselves providing information through voice output, such as studies on multiple IoT speakers (Ogi et al., 2020). This research developed a system in which IoT speakers are installed in various objects within the real space, enabling these objects to speak and present information. Evaluation experiments were conducted, demonstrating the effectiveness of voice-based information presentation using multiple IoT speakers. However, to better understand user needs and provide information flexibly according to the context and situation, it is necessary not only to present information through speakers but also to enable these objects to communicate with the user in a human-like manner.

Therefore, we aim to enable various objects in the real space to serve as avatars, allowing them to communicate with users in a human-like manner. In other words, this paper proposes IoT avatars, which anthropomorphise various objects themselves as avatars.

One of the important factors to consider when anthropomorphising objects as avatars is the avatar's voice. Previously, a system was developed that uses the original voice of a remote user as the voice of an IoT avatar, and an evaluation experiment was conducted in which the IoT avatar communicated with the user on-site (Kida et al., 2024). In this case, it was shown that the users on-site who communicate with the IoT avatar have an awareness that they are communicating with the object itself, and at the same time, they also have an awareness that they are communicating with a remote user. From these results, it was considered that transforming the IoT avatar's voice from the original voice of the remote user to a voice resembling that of a robot or character could be beneficial. This voice transformation (VT) is anticipated to reduce the users' awareness of the remote user and enhance their sense of communicating with the object itself. In addition, to anthropomorphise objects as avatars, it is necessary for the remote user controlling the avatar to have a sense of becoming the object itself. In the context of 'becoming', prior research on a tele-operated robot is relevant (Ogawa et al., 2024). In this study, the remote user's voice is transformed into a voice that matches the robot's appearance, and that voice is output through the robot while simultaneously providing real-time altered auditory feedback to the remote user, thereby demonstrating the effect of 'becoming the robot'.

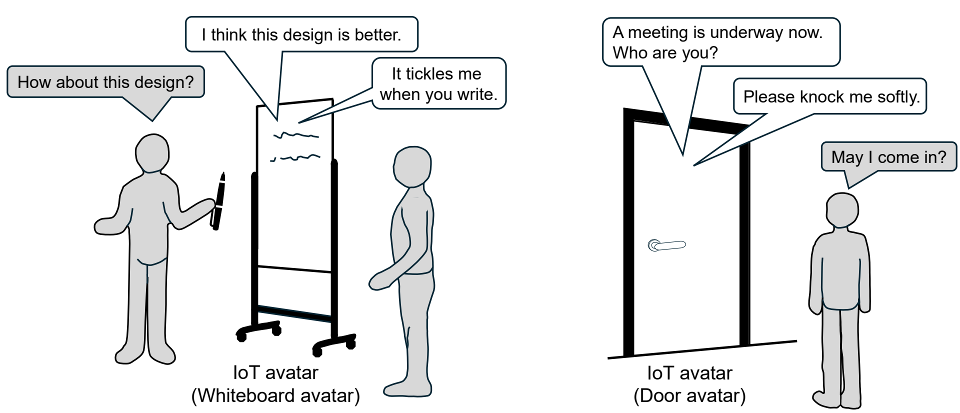
Based on the above, the purpose of this study was to develop a new avatar system, 'IoT avatar', which can anthropomorphise various objects as avatars, and we focused on voice functions as an element of anthropomorphism. Therefore, we developed a prototype that implements VT and real-time altered auditory feedback function for a remote user. In this study, its effectiveness was evaluated and discussed.

In the following Section 2, the concept of the IoT avatar is outlined. In Section 3, related work is described, and the positioning of this study is explained. In Section 4, the system design of IoT avatars is illustrated. In Section 5, the evaluation experiment that verifies the effectiveness of the proposed system is explained, and in Section 6, the discussion is presented. Finally, in Section 7, this study is concluded.

2 Concept of IoT avatar

The concept of the IoT avatar proposed in this study is that various objects in real space themselves are anthropomorphised as avatars, and the objects can communicate with the people around them. Figure 1 presents examples of use cases for this concept. For example, while users are engaged in design work using a whiteboard, the whiteboard could speak, participate in discussions, and offer ideas. Similarly, a door could communicate, providing information about room usage or verifying users. The functionality of this IoT avatar is achieved by attaching an IoT avatar device to various objects, with the avatar itself being controlled remotely by a remote user via the device. While the use cases shown in Figure 1 are just examples, the potential applications of the IoT avatar are broad and can extend to various locations where people gather, such as shopping malls, stations, and airports.

Figure 1 Use cases of IoT avatar



An example of using avatars in real space includes the use of robot avatars (Bremner and Leonards, 2016; Aymerich-Franch et al., 2017; Kida et al., 2023). In robot avatars, a robot with a physical entity serves as an alter ego for remote communication. For example, there are services and research aimed at supporting employment for disabled people by using robots as avatars to perform tasks such as transporting goods and providing customer service remotely (Takeuchi et al., 2020; Barbareschi et al., 2023). Additionally, robots are utilised as avatars in various settings, such as shopping malls (Song et al., 2020), tourist information centres (Glas et al., 2013), and public spaces (Baba et al., 2021). However, the use of robot avatars in various remote locations requires a large number of robots to be deployed at each remote location. In contrast, the IoT avatar has the advantage that the objects themselves in place can be used as avatars.

3 Related work

Anthropomorphism refers to the attribution of human characteristics to non-human entities, and there have been reports of research into the anthropomorphism of objects (Wan and Chen, 2021; Gelman et al., 2022). In terms of the actual anthropomorphism of everyday objects, specifically, systems have been designed and developed that anthropomorphise objects such as a refrigerator, printer, and coffee machine and interact with people (Osawa et al., 2006, 2008; Qiu, 2020). As well as in research, the anthropomorphism of objects is also often depicted in movies and animations. For example, in the Disney movie *Beauty and the Beast*, everyday objects such as a clock, a candelabra, and a teapot appear as characters that have been anthropomorphised. These characters are depicted in the story talking with the protagonists and supporting their actions. The IoT avatar proposed in this study has some similarities to the characters in the movie in that the objects are anthropomorphised and communicate with people. However, it has not been sufficiently investigated whether it is possible to actually anthropomorphise objects in real space as avatars and have them communicate flexibly with people as if they were human. In this study, we actually developed a prototype and investigated it.

Additionally, in terms of the interaction between objects and people, there is a concept called 'Robject' (Rey et al., 2009; Kwak et al., 2017). Robject is an interactive robot that integrates objects found in everyday life with robots. For example, a lamp-shaped robject named 'Lumos' and a microphone-shaped robject named 'Whimbo' have been developed, and they interact effectively with people (Honjo et al., 2023; Hidaki et al., 2023). In other words, it has been shown that effective interaction is possible between people and everyday objects that integrate robots. On the other hand, the IoT avatar proposed in this study anthropomorphises objects around us as avatars rather than robots, and communicates with people on-site. Namely, we aim to enable a remote user to feel as if they become the object itself, and to achieve effective voice communication with the people on-site, such as information presentation and sharing.

Other related work includes 'Pechat', a device that enables stuffed animals to communicate with people (Mitsukuni et al., 2019; Ichikawa et al., 2019). Pechat is a button-type speaker device attached to a stuffed animal, which can be operated using a smartphone and outputs audio as if the stuffed animal were talking. While the main function of Pechat is as a speaker, the IoT avatar device used in this study is equipped with a 360-degree camera, microphone, and speaker. Therefore, remote users can communicate with people on-site using voice while watching images from the IoT avatar's perspective. The details of this system are explained in Section 4.

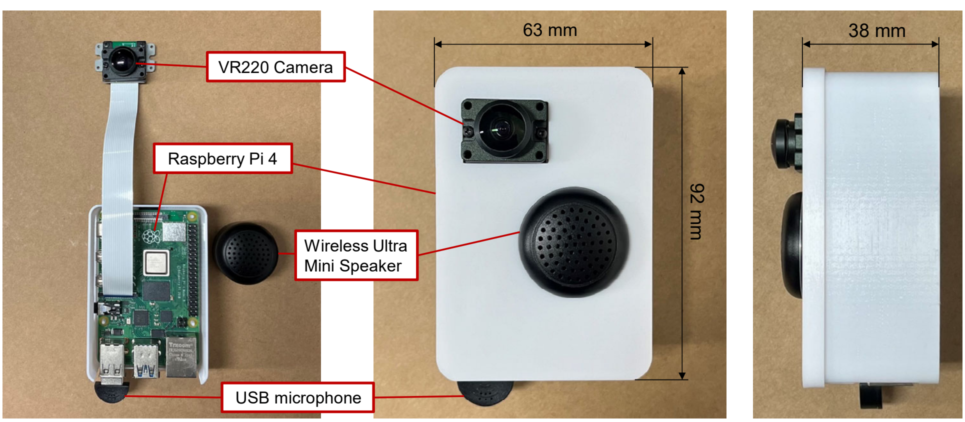
In addition, research has been conducted on IoT avatars that express the underlying behaviours and states internal to IoT devices and IoT-enabled objects with computer graphics (CG) characters using mixed reality (Morris et al., 2020). However, the IoT avatar we propose presents a novel concept of anthropomorphising objects themselves as avatars in real space, rather than using CG to express avatars.

4 Design of IoT avatar

4.1 IoT avatar device

To realise the above IoT avatar concept, we developed an IoT avatar device shown in Figure 2. The device consists of a Raspberry Pi 4 Model B, a mini Bluetooth speaker (Wireless Ultra Mini Speaker), a 360-degree camera (VR 220 Camera), and a mini USB microphone. The body of the IoT avatar device is 92 mm in length, 63 mm in width, and 38 mm in height, and was 3D printed with PLA material. The IoT avatar device can be attached to various objects to make them into IoT avatars. The IoT avatar is controlled by a remote user in an immersive environment using an HMD (Meta Quest 3).

Figure 2 Components of IoT avatar device (see online version for colours)



4.2 Configuration of proposed system

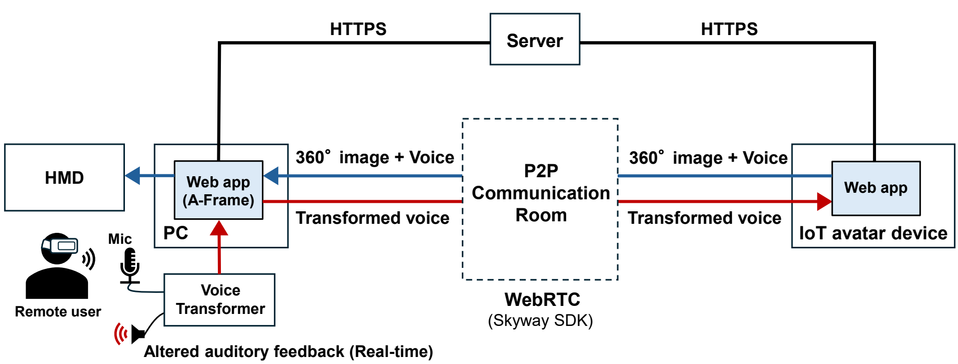
Figure 3 shows the system configuration of the IoT avatar proposed in this study. For real-time remote communication, the system utilises web real-time communication (WebRTC). WebRTC is a technology that enables streaming of images and audio between browsers. In this system, we use the WebRTC platform SDK 'SkyWay' (NTT Communications) to facilitate real-time remote communication between an IoT avatar device and a PC, enabling the transmission and reception of audio and images. To create an immersive environment with an HMD, the PC outputs the received audio and images to the HMD connected to it.

For real-time remote communication between the IoT avatar device and the PC, we developed web applications using the SkyWay SDK. As shown in Figure 3, both the IoT avatar device and the PC open the web application in their respective browsers. The web application programs are stored on a server and are loaded into the browsers via HTTPS communication. To achieve low latency, the communication between the IoT avatar device and the PC is implemented using peer-to-peer (P2P) communication. The IoT avatar device and the PC communicate in real time through a P2P communication room created by the SkyWay SDK in their respective web applications.

In the actual system operation, the voice and 360-degree images captured by the IoT avatar device are sent to the PC's web application via the P2P communication room. The voice is output through the speakers of an HMD connected to the PC. For the images, the system uses A-Frame (a WebVR open-source framework) implemented within the PC's web application to map the received 360-degree images onto a spherical projection, allowing immersive viewing through the HMD.

On the PC side, a voice transformer (Roland VT-4) is used to perform VT and provide altered auditory feedback to the remote user. For the VT, the remote user's voice is captured by a wired microphone connected to the voice transformer via the web application and then transformed. The transformed voice is sent to the IoT avatar device and output through its speakers. Simultaneously, on the PC side, real-time altered auditory feedback is provided to the remote user through wired speakers connected to the voice transformer. The audio of the real-time altered auditory feedback to the remote user is the same as that of VT. Note that the audio of the VT was assumed to be a voice like a robot character, and the pitch and formant of the voice were raised by the voice transformer to make it completely different from the remote user's original voice.

Figure 3 System configuration of proposed IoT avatar (see online version for colours)

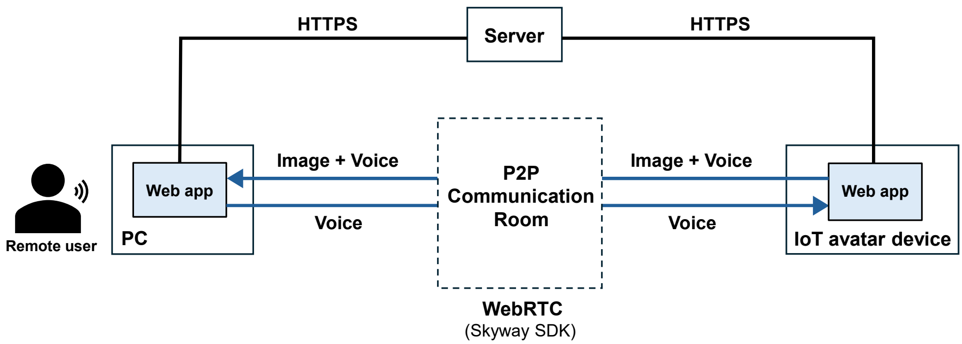


4.3 Configuration of conventionally developed system

Figure 4 shows the system configuration of the conventionally developed IoT avatar for comparison by evaluation experiments. In this system, as in the proposed system in Subsection 4.2, the SkyWay SDK is used for real-time remote communication between an IoT avatar device and a PC. This system uses a Raspberry Pi Camera V2 (angle of view: 62.2 × 48.8°) as the camera of the IoT avatar device, and the images acquired by the camera are sent to the PC and viewed on the PC's monitor. The sound acquired from the IoT avatar device is sent to the PC and output from the PC's speakers. On the PC side, the remote user's voice is acquired by the PC's microphone, sent to the IoT avatar device, and the remote user's original voice is output from the IoT avatar device's speaker.

As described, in this system, a remote user communicates with on-site users and their own original voice via the IoT avatar on the PC.

Figure 4 System configuration of previously developed IoT avatar (see online version for colours)



5 Evaluation experiment

Evaluation experiments were conducted in which a whiteboard and a door, objects that are frequently used by many people within university facilities, were selected to be made into IoT avatars. To evaluate the effectiveness of the proposed system, evaluation experiments were conducted with ten participants, consisting of two parts: experiments 1 and 2.

- Experiment 1: Participants communicate face-to-face with the IoT avatar, which is operated by an experimenter.
- Experiment 2: Participants themselves operate the IoT avatar, communicating with an experimenter on-site.

5.1 Experiment 1

In experiment 1, participants communicated face-to-face with an IoT avatar. As seen in Figure 5, an experimenter (remote user) in another room communicated with a participant on-site via the IoT avatar. In the case of the whiteboard, participants were asked to communicate with the whiteboard while writing freely on the whiteboard. The experimenter communicated with the participants by pretending to be the whiteboard. For example, the experimenter spoke "it tickles me when you write on me with a pen", etc. as if the experimenter were a whiteboard. In the case of the door, participants were asked to communicate with the door by knocking on the door and turning the doorknob. In the same way, the experimenter communicated with the participants while pretending to be the door by saying, for example, "it hurts when you squeeze my hand (doorknob) too hard."

Figure 5 A participant communicating with the IoT avatar (see online version for colours)

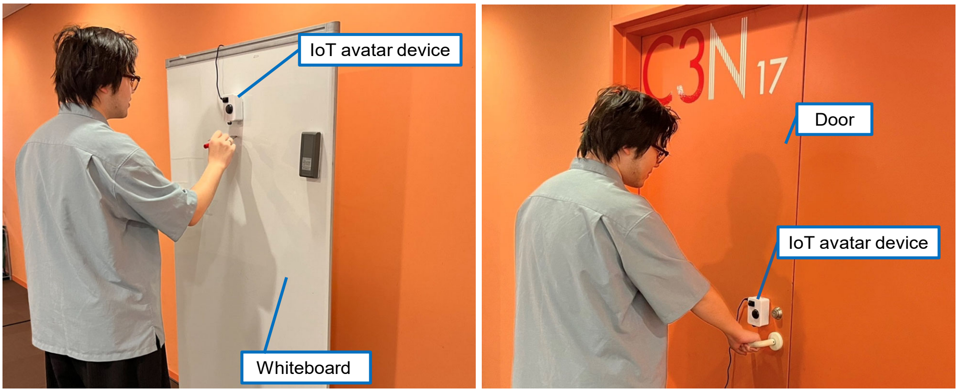


Table 1 presents the experimental conditions for experiment 1. The conditions included VT for the whiteboard, VT for door, no-VT for the whiteboard, and no-VT for door. For the VT conditions, the system described in Subsection 4.2 was used, while the no-VT conditions used the system described in Subsection 4.3. The order of VT and no-VT conditions was randomised for each participant.

Table 1 Conditions for the experiment 1

Voice of IoT avatar	*Object*
VT	Whiteboard, door
no-VT	Whiteboard, door

For the evaluation in experiment 1, participants completed the questionnaire shown in Table 2 after each experimental condition. The questionnaire was rated on a seven-point Likert scale (1: strongly disagree, 2: disagree, 3: slightly disagree, 4: neither, 5: slightly agree, 6: agree, and 7: strongly agree).

Q1 and Q2 were paired questionnaires to compare whether participants felt they were communicating with the object itself (Q1) or with a remote user (Q2). Q3 was a questionnaire asking whether the participants felt humanity toward the object itself, and Q4 was a questionnaire asking whether the participants wanted to use the system.

Table 2 Questionnaire for the experiment 1

Q1	I felt I was communicating with the object itself, not a remote user.
Q2	I felt I was communicating with a remote user, not the object itself.
Q3	I felt humanity towards the object itself.
Q4	I want to use this system.

5.2 Results of experiment 1

Figure 6 shows the evaluation results with and without VT (VT, no-VT), and Figure 7 shows the evaluation results for the types of objects (whiteboard, door). Error bars indicate standard deviations.

Figure 6 Evaluation results of VT and no-VT in experiment 1 (see online version for colours)

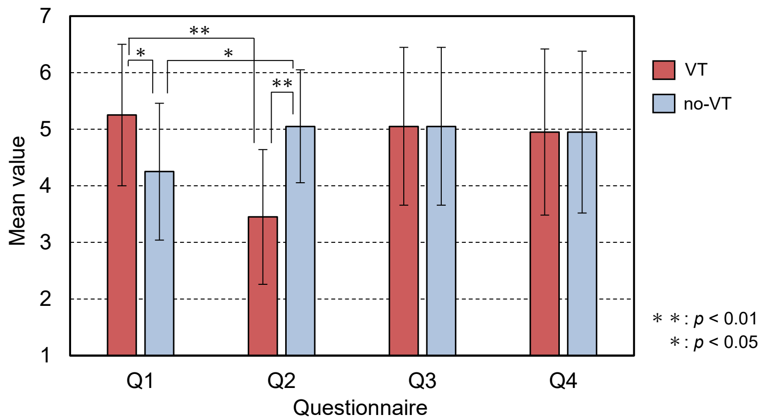
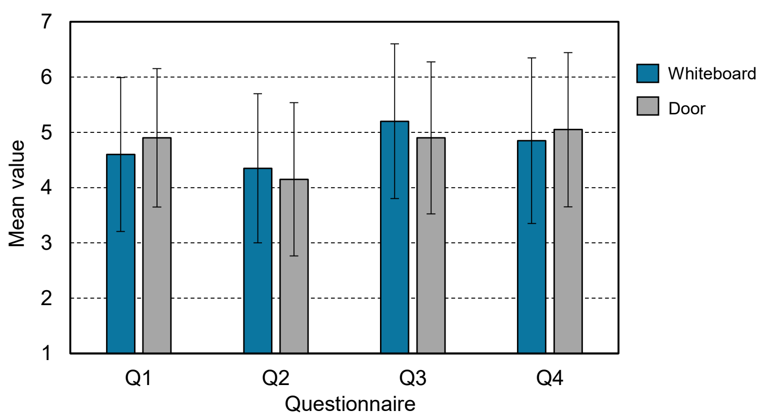


Figure 7 Evaluation results of the whiteboard and door in experiment 1 (see online version for colours)



First, as for Q1, "whether the participants felt they were communicating with the object itself", ANOVA was conducted on the factors of the presence of VT (VT, no-VT) and the type of object (whiteboard, door). The results showed that the factors of the presence of VT were significant ($p = 0.016$). As shown in Figure 6, the value of Q1 (VT) was significantly higher than that of Q1 (no-VT), exceeding a value of 5. On the other hand, there was no significant difference ($p = 0.453$) in the factors of the type of object (whiteboard, door) in Q1 (Figure 7).

Next, as for Q2, "whether the participants felt they were communicating with a remote user", ANOVA was conducted on the factors of the presence of VT (VT, no-VT) and the type of object (whiteboard, door). The results showed that the factors of the presence of VT were significant ($p < 0.001$). As shown in Figure 6, the sense of communicating with a remote user resulted in a significantly higher Q2 (no-VT) than Q2 (VT). This is a reasonable result, since this is the remote user's original voice (no-VT). There was also no significant difference ($p = 0.573$) in the factors of the type of object (whiteboard, door) in Q2 (Figure 7).

Additionally, a t-test was used to compare the results between Q1 (VT) and Q2 (VT) and between Q1 (no-VT) and Q2 (no-VT). Figure 6 shows that there was a significant difference between Q1 (VT) and Q2 (VT) ($p < 0.001$) and between Q1 (no-VT) and Q2 (no-VT) ($p = 0.028$). In short, the results of Q1 and Q2 above suggest that the VT condition is more effective in making the user feel the sense of communicating with the object itself, rather than with a remote user.

As for Q3, "whether the participants felt humanity toward the object itself", ANOVA was conducted on the factors of the presence of VT (VT, no-VT) and the type of object (whiteboard, door). There was no significant difference ($p = 1.000$) in the factors of the presence of VT (between VT and no-VT), and the values exceeded 5 in both VT and no-VT conditions, indicating that the respondents generally felt humanity (Figure 6). There was also no significant difference ($p = 0.508$) in the factors of the type of object (whiteboard, door) in Q3 (Figure 7).

Similarly, As for Q4, "whether the participants wanted to use the system", ANOVA was conducted on the factors of the presence of VT (VT, no-VT) and the type of object (whiteboard, door). It showed no significant difference ($p = 1.000$) between VT and no-VT, and positive results were obtained for both VT and no-VT conditions (Figure 6). There was also no significant difference ($p = 0.673$) in the factors of the type of object (whiteboard, door) in Q4 (Figure 7).

The above results show that the proposed system (VT) is effective in making users feel as if they are communicating with the object itself, and that they also generally feel its humanity. In other words, these results suggest the effectiveness of the proposed system (VT) in terms of anthropomorphism.

5.3 Experiment 2

In experiment 2, participants in another room (remote users) operated the IoT avatar and communicated with an experimenter on-site. In the case of the whiteboard, participants were asked to pretend to be a whiteboard and communicate with the experimenter. The experimenter communicated with the whiteboard while writing freely on the whiteboard. In the case of the door, participants communicated with the experimenter by pretending to be the door. The experimenter communicated with the door while knocking on the door and turning the doorknob.

Table 3 presents the experimental conditions for experiment 2. The experimental conditions included: VT with altered auditory feedback for the whiteboard (VT), VT with altered auditory feedback for the door (VT), no-VT with no altered auditory feedback for the whiteboard (no-VT), and no-VT with no altered auditory feedback for the door (no-VT).

Table 3 Conditions for the experiment 2

IoT avatar system	*Object*
VT+HMD	Whiteboard, door
no-VT+PC	Whiteboard, door

For the VT conditions, the system was the one described in Subsection 4.2 in which participants used an HMD [Figure 8(a)], and for the no-VT conditions, the system was the one described in section 4.3 in which participants used a PC [Figure 8(b)]. The order in which the VT and no-VT conditions were performed was randomised for each participant.

For the evaluation in experiment 2, participants completed the questionnaire shown in Table 4 after each experimental condition. The questionnaire was rated on a seven-point Likert scale (1: strongly disagree, 2: disagree, 3: slightly disagree, 4: neither, 5: slightly agree, 6: agree, 7: strongly agree).

Q1 and Q2 of the questionnaire were paired questions to compare whether the participants felt that the object itself was communicating with the other person (Q1) or that they themselves were communicating with the other person (Q2). Q3 was a questionnaire asking whether participants felt that they were becoming the object, and Q4 was a questionnaire asking whether the participants wanted to use the system.

Figure 8 A participant communicating with an experimenter via the IoT avatar (see online version for colours)

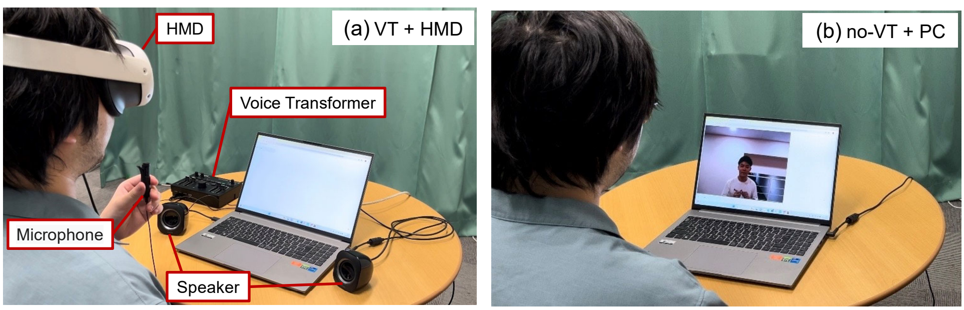


Table 4 Questionnaire for the experiment 2

Q1	I felt the object itself was communicating with the other person, not myself.
Q2	I felt I was communicating with the other person, not the object itself.
Q3	I felt as if I was becoming the object itself.
Q4	I want to use this system.

5.4 *Results of experiment 2*

Figure 9 shows the evaluation results for the IoT avatar system (VT+HMD, no-VT+PC), and Figure 10 shows the results for the types of objects (whiteboard, door). Error bars indicate standard deviations.

First, as for Q1, “whether the participant felt that the object itself was communicating with the other person”, ANOVA was conducted on the factors of the IoT avatar system (VT+HMD, no-VT+PC) and the type of object (whiteboard, door). The results showed that the factors of the IoT avatar system were significant ($p < 0.001$). As shown in Figure 9, the value of Q1 (VT+HMD) was significantly higher than that of Q1 (no-VT+PC), approaching a value of 5. Namely, in the proposed system (VT+HMD), it was shown that the participants felt the sense that the object itself was communicating with the other person. On the other hand, there was no significant difference ($p = 0.914$) in the factors of the type of object (whiteboard, door) in Q1 (Figure 10).

Next, as for Q2, “whether the participant felt that they were communicating with the other person”, ANOVA was conducted on the factors of the IoT avatar system

(VT+HMD, no-VT+PC) and the type of object (whiteboard, door). The results showed the factors of the IoT avatar system were significant (p = 0.005). As shown in Figure 9, the sense of themselves communicating with the other person resulted in a significantly higher Q2 (no-VT+PC) than Q2 (VT+HMD). There was also no significant difference (p = 0.743) in the factors of the type of object (whiteboard, door) in Q2 (Figure 10).

Additionally, a t-test was used to compare the results between Q1 (VT+HMD) and Q2 (VT+HMD) and between Q1 (no-VT+PC) and Q2 (no-VT+PC). There was no significant difference (p = 0.150) between Q1 (VT+HMD) and Q2 (VT+HMD). Q2 (VT+HMD) was evaluated as 'neither' with a value of less than 4.5. On the other hand, there was a significant difference (p < 0.001) between Q1 (no-VT+PC) and Q2 (no-VT+PC). The evaluation of Q2 (no-VT+PC) was significantly higher than that of Q1 (no-VT+PC), exceeding a value of 5, indicating that the participants in the no-VT+PC condition felt the sense that they themselves were communicating with the other person.

Figure 9 Evaluation results of VT+HMD and no-VT+PC in experiment 2 (see online version for colours)

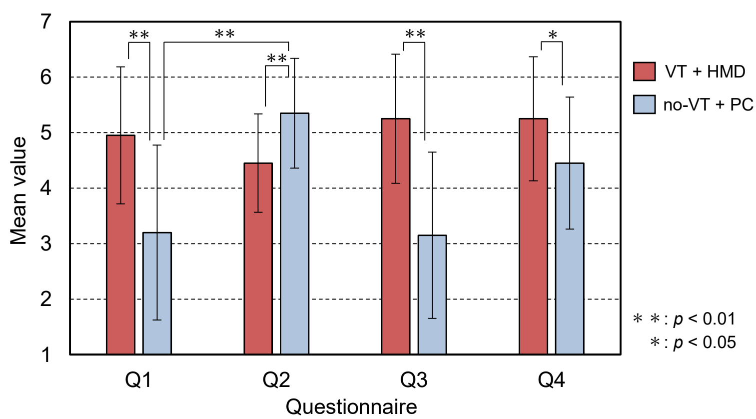
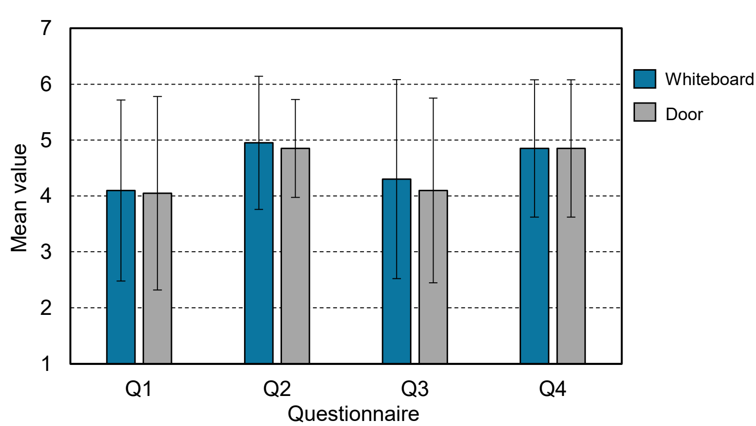


Figure 10 Evaluation results of the whiteboard and door in experiment 2 (see online version for colours)



As for Q3, "whether participants felt that they were becoming the object", ANOVA was conducted on the factors of the IoT avatar system (VT+HMD, no-VT+PC), and the type of object (whiteboard, door). In the VT+HMD condition, the evaluation was significantly higher ($p < 0.001$), with the value exceeding 5. This means that the participants felt a sense of becoming the object itself. There was also no significant difference ($p = 0.648$) in the factors of the type of object (whiteboard, door) in Q3 (Figure 10).

Similarly, As for Q4, "whether the participants wanted to use the system", ANOVA was conducted on the factors of the IoT avatar system (VT+HMD, no-VT+PC) and the type of object (whiteboard, door). The results showed that the VT+HMD condition was significantly higher ($p = 0.040$), indicating a positive result in terms of desire to use the system. There was also no significant difference ($p = 1.000$) in the factors of the type of object (whiteboard, door) in Q4 (Figure 10).

The above results show that the proposed system (VT+HMD) allows remote users to feel a sense that the objects themselves are communicating with the other person. Furthermore, it is clear that this system is effective in terms of becoming the object itself.

6 Discussion and future work

First, from the results of experiment 1, it can be considered that the concept of anthropomorphising objects as avatars has been realised by the proposed system. Specifically, we consider that by transforming the voice of the IoT avatar into a voice as if it were a robot character speaking, rather than the remote user's original voice, the user's awareness of the remote user was reduced, and the user's awareness of speaking with the object itself was increased. Furthermore, we expect that there is potential for additional improvement in the evaluation of the feeling of communicating with the object itself (Q1) and the feeling of humanity toward the object itself (Q3). One of these is the emotional expression of the IoT avatar. In the proposed system, emotions are expressed only by voice, but other ways of expressing emotions are needed. For example, it is considered possible to implement an LED lighting function in the IoT avatar and express emotions by changing the colour of the LED. Therefore, the design and development of technology for emotional expression can be a future work.

Next, from the results of experiment 2, it was found that the proposed system generated a feeling that the remote users were becoming the object itself, and that said object was communicating with the other person. In terms of users becoming avatars and communicating with others, this system is also expected to have the Proteus effect (Yee and Bailenson, 2007; Fox et al., 2013), which is an effect in which the appearance of the avatar influences the user's psychological state and behaviour. For example, Yee and Bailenson (2007) found that users who used attractive-looking human avatars became more intimate in terms of self-disclosure and interpersonal distance than those who used unattractive-looking avatars. In addition, users with taller avatars reported more confident behaviour in a negotiation task than users with shorter avatars. Therefore, a future prospect is to investigate the Proteus effect when using IoT avatars, which could be an advantage for communication using IoT avatars.

7 Conclusions

In this study, we developed an IoT avatar that can anthropomorphise various objects in real space as avatars and communicate with people around them. Specifically, focusing on the voice of the IoT avatar as one of the anthropomorphic elements, we developed a system that implements a voice conversion function and an altered auditory feedback function, and presents immersive images using an HMD. From the evaluation experiments, it can be said that the proposed system was able to anthropomorphise objects as avatars and communicate with people on-site. It is also suggested that the remote users can communicate with the people on-site as if they were becoming the objects themselves via the proposed system.

Acknowledgements

This research was partially supported by JKA promotion funds from KEIRIN RACE (2024M-549).

References

Alanwar, A., Alzantot, M., Ho, B.J., Martin, P. and Srivastava, M. (2017) 'Selecon: scalable IoT device selection and control using hand gestures', in *Proceedings of the Second International Conference on Internet-of-Things Design and Implementation*, pp.47–58.

Aymerich-Franch, L., Petit, D., Ganesh, G. and Kheddar, A. (2017) 'Object touch by a humanoid robot avatar induces haptic sensation in the real hand', *Journal of Computer-Mediated Communication*, Vol. 22, No. 4, pp.215–230.

Baba, J., Song, S., Nakanishi, J., Yoshikawa, Y. and Ishiguro, H. (2021) 'Local vs. avatar robot: performance and perceived workload of service encounters in public space', *Frontiers in Robotics and AI*, Vol. 8, p.778753.

Barbareschi, G., Kawaguchi, M., Kato, H., Nagahiro, M., Takeuchi, K., Shiiba, Y., Kasahara, S., Kunze, K. and Minamizawa, K. (2023) ''I am both here and there' parallel control of multiple robotic avatars by disabled workers in a café', in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp.1–17.

Bremner, P. and Leonards, U. (2016) 'Iconic gestures for robot avatars, recognition and integration with speech', *Frontiers in Psychology*, Vol. 7, p.183.

Delgado Rodriguez, S., Prange, S., Mecke, L. and Alt, F. (2021) 'ActPad – a smart desk platform to enable user interaction with IoT devices', in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp.1–6.

Fox, J., Bailenson, J.N. and Tricase, L. (2013) 'The embodiment of sexualized virtual selves: the Proteus effect and experiences of self-objectification via avatars', *Computers in Human Behavior*, Vol. 29, No. 3, pp.930–938.

Gelman, S.A., Davidson, N.S. and Umscheid, V.A. (2022) 'The role of object features and emotional attachment on preschool children's anthropomorphism of owned objects', *Cognitive Development*, Vol. 62, p.101165.

Glas, D.F., Wada, K., Shiomi, M., Kanda, T., Ishiguro, H. and Hagita, N. (2013) 'Never too old for teleoperation: helping elderly people control a conversational service robot', in *2013 IEEE RO-MAN*, pp.703–710.

Hidaki, R., Honjo, N., Hasegawa, K., Oshima, N. and Okada, M. (2023) 'Microphone-shaped robject 'Whimbo' with minimal hearership', *The Transactions of Human Interface Society*, Vol. 25, No. 3, pp.231–240, in Japanese.

Honjo, N., Hasegawa, K., Oshima, N. and Okada, M. (2023) 'Robject as a novel entity between object and robot and its social acceptability', *The Transactions of Human Interface Society*, Vol. 25, No. 3, pp.203–218, in Japanese.

Ichikawa, J., Mitsukuni, K., Hori, Y., Ikeno, Y., Alexandre, L., Kawamoto, T., Nishizaki, Y. and Oka, N. (2019) 'Analysis of how personality traits affect children's conversational play with an utterance-output device', in *2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*, pp.215–220.

Kida, Y., Chiba, T., Kinoshita, T., Mikami, A., Yakoh, T. and Ogi, T. (2024) 'IoT avatar: turning various objects into avatars', in *The 27th International Conference on Network-Based Information Systems*, pp.398–405.

Kida, Y., Li, S., Ueda, Y., Takayanagi, N., Matsuoka, K. and Ogi, T. (2023) 'Evaluation of user's psychological sense in tele-immersion robot avatar', in *The 26th International Conference on Network-Based Information Systems*, pp.325–333.

Kwak, S.S., Kim, J.S. and Choi, J.J. (2017) 'The effects of organism-versus object-based robot design approaches on the consumer acceptance of domestic robots', *International Journal of Social Robotics*, Vol. 9, pp.359–377.

Meng, Y., Zhu, H., Li, J., Li, J. and Liu, Y. (2020) 'Liveness detection for voice user interface via wireless signals in IoT environment', *IEEE Transactions on Dependable and Secure Computing*, Vol. 18, No. 6, pp.2996–3011.

Mitsukuni, K., Ichikawa, J., Hori, Y., Ikeno, Y., Alexandre, L., Kawamoto, T., Oka, N. and Nishizaki, Y. (2019) 'Analysis of the influence of mothers' traits and behaviors on children's conversational play with an utterance-output device', in *Proceedings of the 7th International Conference on Human-Agent Interaction*, pp.271–273.

Morris, A., Guan, J., Lessio, N. and Shao, Y. (2020) 'Toward mixed reality hybrid objects with IoT avatar agents', in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp.766–773.

Muhamad, W.N.W., Abd Halim, M.A., Sarnin, S.S. and Shuhaimi, N.I. (2021) 'Smart voice user interface office monitoring system via internet of things', in *2021 IEEE 15th Malaysia International Conference on Communication (MICC)*, pp.1–6.

Ogawa, N., Baba, J. and Nakanishi, J. (2024) 'Investigating effect of altered auditory feedback on self-representation, subjective operator experience, and task performance in teleoperation of a social robot', in *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp.1–18.

Ogi, T., Enomoto, K., Akita, S., Miyahashi, T., Akima, T. and Sano, T. (2020) 'Construction of information presentation environment where various objects talk using multiple IoT speakers', in *The 23rd International Conference on Network-Based Information Systems*, pp.432–439.

Osawa, H., Mukai, J. and Imai, M. (2006) 'Anthropomorphization of an object by displaying robot', in *ROMAN 2006 – The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp.763–768.

Osawa, H., Ohmura, R. and Imai, M. (2008) 'Embodiment of an agent by anthropomorphization of a common object', in *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 2, pp.484–490.

Qiu, S. (2020) 'Anthropomorphic design for everyday objects', in *Human-Computer Interaction, Design and User Experience: Thematic Area, HCI 2020, Held as Part of the 22nd International Conference, HCII 2020, Proceedings, Part I*, Vol. 22, pp.137–146.

Rey, F., Leidi, M. and Mondada, F. (2009) 'Interactive mobile robotic drinking glasses', *Distributed Autonomous Robotic Systems*, Vol. 8, pp.543–551.

Rubio-Drosdov, E., Díaz-Sánchez, D., Almenárez, F., Arias-Cabarcos, P. and Marín, A. (2017) 'Seamless human-device interaction in the internet of things', *IEEE Transactions on Consumer Electronics*, Vol. 63, No. 4, pp.490–498.

Song, S., Baba, J., Nakanishi, J., Yoshikawa, Y. and Ishiguro, H. (2021) 'Teleoperated robot sells toothbrush in a shopping mall: a field study', in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp.1–6.

Takeuchi, K., Yamazaki, Y. and Yoshifuji, K. (2020) 'Avatar work: telework for disabled people unable to go outside by using avatar robots', in *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp.53–60.

Wan, E.W. and Chen, R.P. (2021) 'Anthropomorphism and object attachment', *Current Opinion in Psychology*, Vol. 39, pp.88–93.

Yee, N. and Bailenson, J. (2007) 'The Proteus effect: the effect of transformed self-representation on behavior', *Human Communication Research*, Vol. 33, No. 3, pp.271–290.

Zhang, L., Tan, S., Wang, Z., Ren, Y., Wang, Z. and Yang, J. (2020) 'Viblive: a continuous liveness detection for secure voice user interface in IoT environment', in *Proceedings of the 36th Annual Computer Security Applications Conference*, pp.884–896.