

Physical embodiment can produce robot operator's pseudo presence

Kazuaki Tanaka^{1,2}, Hideyuki Nakanishi^{1*} and Hiroshi Ishiguro³

¹ Department of Adaptive Machine Systems, Osaka University, Suita, Japan, ² CREST, Japan Science and Technology Agency, Tokyo, Japan, ³ Department of Systems Innovation, Osaka University, Toyonaka, Japan

Recent studies have focused on humanoid robots for improving distant communication. When a user talks with a remote conversation partner through a humanoid robot, the user can see the remote partner's body motions with physical embodiment but not the partner's current appearance. The physical embodiment existing in the same room with the user is the main feature of humanoid robots, but the effects on social telepresence, i.e., the sense of resembling face-to-face interaction, had not yet been well demonstrated. To find the effects, we conducted an experiment in which subjects talked with a partner through robots and various existing communication media (e.g., voice, avatar, and video chats). As a result, we found that the physical embodiment enhances social telepresence. However, in terms of the degree of social telepresence, the humanoid robot remained at the same level as the partner's live video, since presenting partner's appearance also enhances social telepresence. To utilize the anonymity of a humanoid robot, we proposed the way that produces pseudo presence that is the sense of interacting with a remote partner when they are actually interacting with an autonomous robot. Through the second experiment, we discovered that the subjects tended to evaluate the degree of pseudo presence of a remote partner based on their prior experience of watching the partner's body motions reproduced by a robot. When a subject interacted with an autonomous robot after interacting with a teleoperated robot (i.e., a remote operator) that is identical with the autonomous robot, the subjects tended to feel as if they were talking with a remote operator.

Keywords: teleoperated robot, autonomous robot, videoconferencing, avatar, face-to-face, social telepresence, face tracking

OPEN ACCESS

Edited by:

Javier Jaen,
Universitat Politècnica de València,
Spain

Reviewed by:

Andrej Košir,
University of Ljubljana, Slovenia
Khiet Phuong Truong,
University of Twente, Netherlands

*Correspondence:

Hideyuki Nakanishi,
Department of Adaptive Machine
Systems, Osaka University, 2-1
Yamadaoka, Suita, Osaka 565-0871,
Japan
nakanishi@ams.eng.osaka-u.ac.jp

Specialty section:

This article was submitted to
Human-Media Interaction, a section
of the journal *Frontiers in ICT*

Received: 26 February 2015

Accepted: 28 April 2015

Published: 18 May 2015

Citation:

Tanaka K, Nakanishi H and
Ishiguro H (2015) Physical
embodiment can produce robot
operator's pseudo presence.
Front. ICT 2:8.
doi: 10.3389/fict.2015.00008

Introduction

Currently, we can easily use audio and videoconferencing software. Audio-only conferencing, such as a voice chat, has a problem in that social telepresence decreases. The social telepresence is the sense of resembling face-to-face interaction (Finn et al., 1997). Enhancing social telepresence psychologically makes the physical distance between remote people less and saves time and money on travel. The most common method of enhancing social telepresence is videoconferencing. It had been proposed that live video can transmit the social telepresence of a remote conversation partner (Isaacs and Tang, 1994; de Greef and Ijsselstein, 2001). However, videoconferencing is closer to a situation of talking through a window than face-to-face conferencing due to a display.

To further enhance social telepresence, recent studies have begun on robot conferencing in which people talk with a remote conversation partner through teleoperated humanoid robots. The robots use motion tracking technologies to reflect partner's facial and body motions in real time. The main

features of robot conferencing are to transmit conversation partner's body motions and to present these motions via a physical embodiment. The physical embodiment means the substitution of a partner's body that exists physically in the same place as a user. Thus, it is expected that the user may feel closer to face-to-face interaction. Some studies reported superiorities of robot conferencing to videoconferencing (Morita et al., 2007; Sakamoto et al., 2007). One such study showed that the teleoperated robot, which has a realistic human appearance, enhances social telepresence compared with audio-only conferencing and videoconferencing (Sakamoto et al., 2007). Even so, it is difficult that each user owns a robot with his/her realistic appearance due to the high cost. For this reason, a teleoperated robot that has a human-like face without a specific age or gender is developed (Ogawa et al., 2011). However, there is a question whether such an anonymous robot can produce higher social telepresence compared with videoconferencing in which a user can see the remote partner's motion and appearance.

As the communication medium similar to the robot conferencing, avatar chats are available. Recently, it has become easy and inexpensive to use avatar chats such as avatar Kinect. The avatar chat resembles the robot conferencing in transmitting user's body motions without disclosing the user's appearance, but differs in reflecting these movements onto a computer graphics animation, which does not have a physical embodiment. A lot of studies found positive effects of avatar on distant communication (Garau et al., 2001; Bailenson et al., 2006; Bente et al., 2008; Kang et al., 2008; Tanaka et al., 2013). Several such studies focused on social

telepresence reported that avatar chats are better than audio-only communication (Bente et al., 2008; Kang et al., 2008), but worse than videoconferencing (Kang et al., 2008). Thus, presenting partner's body motion and appearance might contribute to produce social telepresence. If the physical embodiment does not produce social telepresence, the usefulness of humanoid robots would decrease since robots are more expensive than videos and avatars.

In this study, we conducted two experiments to prove the usefulness of humanoid robot. First, it is necessary to demonstrate that the physical embodiment enhance social telepresence independently from the transmitting information, e.g., audio, motion, and appearance. In the first experiment, we investigated how the physical embodiment and transmitting information factors influence the social telepresence (Tanaka et al., 2014). To analyze the effects of the two factors separately, we prepared six communication methods as shown in **Figure 1**. The voice chat, avatar chat, and videoconferencing that do not have a physical embodiment transmit audio-only, audio + motion, and audio + motion + appearance, respectively. The robot conferencing that has a physical embodiment transmits audio + motion, and so it corresponds to the avatar chat as described above. As the method that corresponds to the voice chat, we set an inactive robot conferencing that transmits audio but no motion. Furthermore, we assumed that the face-to-face interaction corresponds to the videoconferencing.

Another method to prove the usefulness is to demonstrate that humanoid robots produce pseudo presence of remote

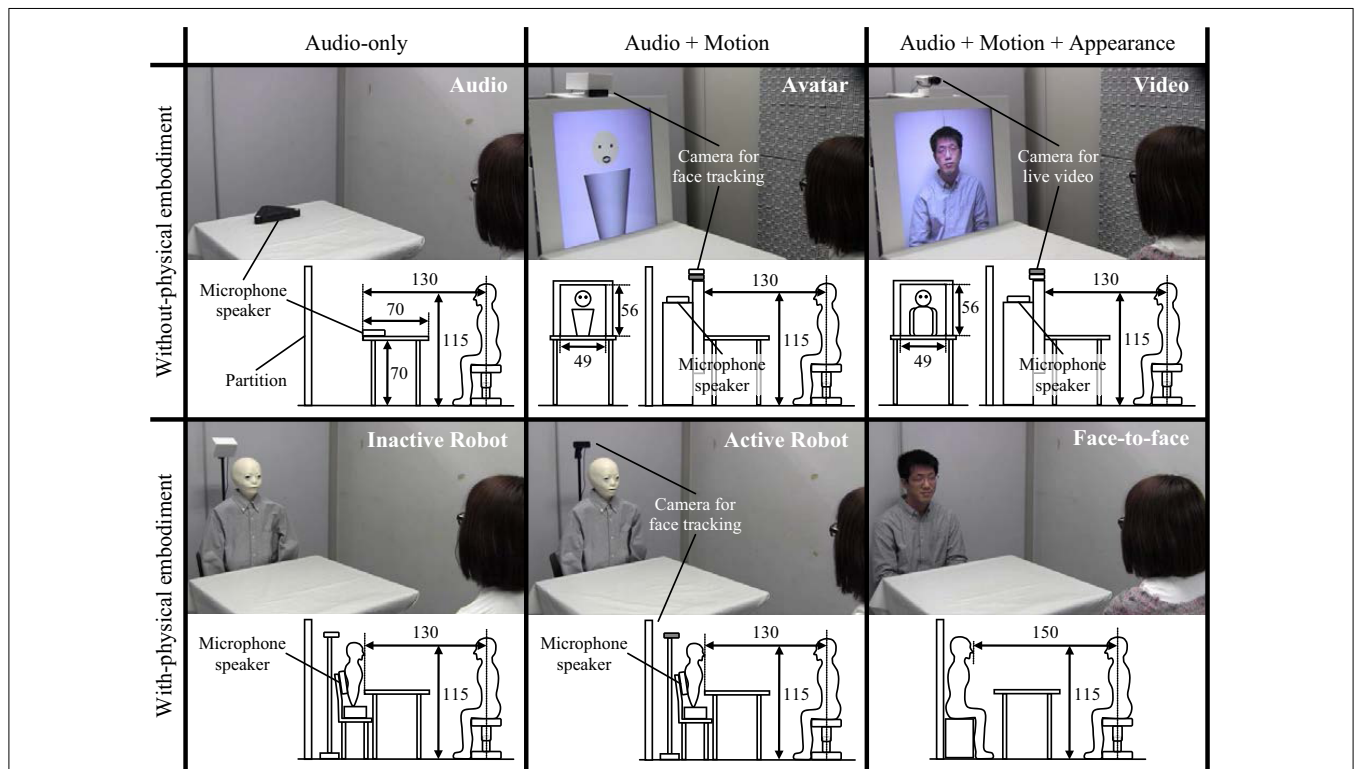


FIGURE 1 | Conditions and the setups of the first experiment (length unit: centimeters): the six communication methods divided into physical embodiment and transmitting information factors.

conversation partner. Pseudo presence means the feeling of interacting with a remote operator when interacting with an autonomous robot. The two main types of humanoid robots are teleoperated and autonomous. Teleoperated robots transmit remote operator's social telepresence by reproducing remote operator's behavior. On the other hand, autonomous robots produce remote operator's pseudo presence by behaving as being controlled by a remote operator. We hence believe that the essential difference between these robots is the presence or absence of a remote operator. If autonomous robots generate human-like behavior comparable to body motions obtained by motion tracking technologies, the user could feel pseudo presence of remote partner even when interacting with the autonomous robot. Thus, to decide presence/absence of conversation partner in interacting with an autonomous robot, prior experience of watching the remote partner's motion reproduced by teleoperated robot whose design is identical as the autonomous robot may be needed.

We expect that the presence of a remote partner in interacting with a teleoperated robot will be recalled in interacting with an autonomous robot by the robot's behaviors and the user might continue to feel the partner's presence. First, a user experiences the teleoperated mode in which the user talks with a remote conversation partner through a robot synchronized to the partner's body motions. After that, the user experiences the autonomous mode in which he/she talks with an autonomous system operating the same robot used in the teleoperated mode. This system generates talking behaviors, e.g., lip motion, from the pre-recorded partner's speech, or nod motions from the user's speech.

There are some applications that a remote operator provides services through a teleoperated robot. An autonomous robot could give the user the same services instead of a remote operator if the robot is able to produce the pseudo presence. In the example of an interaction robot (Ranatunga et al., 2012; Tanaka et al., 2012), if a user who is living alone talked with a remote caregiver through the robot, the user might continue to feel the caregiver's presence even after moving to the autonomous mode. The pseudo presence of remote caregiver may reduce the user's feeling of loneliness more effectively even if the remote caregiver is not talking actually. For a lecture robot (Hashimoto et al., 2011), the students might feel the remote teacher's presence at an autonomous lecture after preliminarily greeting each other in the teleoperated mode. Due to the pseudo presence of remote teacher, the students may pay attention to the lecture even if the lecture is autonomously reproduced. While an autonomous robot interacts with the user, the remote operator does not have to work.

In the second experiment, we compared presence or absence of the prior experience of talking with remote conversation partner through a robot. Our approach that changes the dialog modes of a robot from teleoperated to autonomous utilizes a weakness of anonymous humanoid robot that is the lack of appearance. Therefore, this approach can be applied to media, which do not show partner's current appearance. To confirm the contribution of physically embodied body motion to produce the pseudo presence, we conducted the comparison also in audio-only communication in which the user cannot see the partner.

The paper is organized as follows. The next section presents the related work. Section "Subjects" explains about the subjects

who participated in our experiments. Sections "Experiment 1" and "Experiment 2" explain the methods and the results of the first and second experiment, respectively. The first experiment investigates whether the features of humanoid robot enhance social telepresence. The second experiment investigates the effects of humanoid robot on producing pseudo presence of remote conversation partner. Section "Discussion" discusses the results of these experiments. Finally, Section "Conclusion" concludes the paper.

Related Work

This study is related with the telerobotics and intelligent robotics. In the telerobotics field, many studies have proposed various teleoperated robots that present the operator's facial movements (Kuzuoka et al., 2004; Morita et al., 2007; Sakamoto et al., 2007; Ogawa et al., 2011; Sirkin and Ju, 2012) with a physical embodiment. Several studies reported the superiority of robot conferencing to videoconferencing. One such study showed that the eye-gaze of remote person reproduced by a robot was more recognizable than by a live video (Morita et al., 2007). The study with regard to social telepresence concluded that teleoperated robot transmitted a higher social telepresence of a remote conversation partner than audio-only and videoconferencing (Sakamoto et al., 2007). However, this result seems somewhat obvious, since the teleoperated robot reproduced the whole body of a person, whereas the videoconferencing only showed conversational partner's head. The video image of only a head is harmful to social telepresence (Nguyen and Canny, 2009), so that a superiority of robot conferencing to videoconferencing, which shows the whole body of a person was also not clear. Furthermore, the teleoperated robot that was used in the study had a specific person's appearance, and so it was not clear, which of the factors, the physical embodiment, the appearance, or the ability to present body motions, enhanced social telepresence. To clarify them, we used an anonymous teleoperated robot (Ogawa et al., 2011) that has a human-like face without a specific age or gender, and compared it with partner's life-size video.

In videoconferencing research, it was reported that the remote person's movement that was augmented by a display's physical movement enhanced the social telepresence (Nakanishi et al., 2011). This result implies that the physically embodied body motion enhances social telepresence.

In the intelligent robotics field, there are studies that focused on the effects of the physical embodiment on social presence (Lee et al., 2006; Bainbridge et al., 2011). These studies showed that a humanoid robot produces higher social presence than on-screen agents. These studies evaluated whether people interact with a non-human social agent (i.e., robots and on-screen agent) as if it were an actual human. By contrast, our experiments evaluated whether people feel being with a remote conversation partner in the same room when talking with a humanoid robot. When the teleoperated robot conveyed remote partner's body motions, we estimated the degree of remote partner's social telepresence. On the other hand, when the robot moves automatically, we estimated the degree of the partner's pseudo presence. There is a possibility that the physical embodiment contributes to enhance these presence.

There are some technologies that generate talking behaviors autonomously instead of transmitting remote partner's behaviors. If a robot can generate human-like talking behaviors, a user may believe that it is moving based on the partner's body motions, since the user does not know the partner's current appearance or behavior. Many past studies have proposed algorithms to generate talking behaviors from someone's speech (Cao et al., 2005; Salvi et al., 2009; Lee et al., 2010; Watanabe et al., 2010; Le et al., 2012; Liu et al., 2012). These studies generated human-like talking behaviors that were as natural and various as possible. But no research has investigated whether this approach produces the sense of talking with a remote partner. We predicted that prior experience of watching the remote partner's behavior reproduced by a robot produces the sense when watching talking behavior generated by the same robot. A robot that can be controlled by teleoperated and autonomous modes has been developed (Ranatunga et al., 2012), but the effect of changing these modes on producing remote partner's presence has not been clarified.

Subjects

Thirty-six undergraduates (17 females and 19 males) and 16 undergraduates (9 females and 7 males) participated in our first and second experiments, respectively. We used a recruitment website for part-time workers to collect the subjects who lived near our university campus.

We did not choose students of master's course and upward as subjects to prevent an influence of their expertise on the results. For the same reason, we employed mainly liberal arts undergraduates or science and engineering undergraduates who do not study about robotics. The subjects had never met the experimenter before the experiment.

We recorded the experiments and interviews for the subjects. The subjects were required to sign a consent form that confirmed whether they agree with the recording. The consent form also confirmed whether the recorded movies could be used for presentations, articles, or TV programs. If the subject does not agree with using their movies, he/she could refuse it. We are holding the consent forms and movies under lock and key.

Experiment 1

This section presents the first experiment in which we investigated how the physical embodiment and transmitting information factors influence on social telepresence.

Hypothesis

The main features of robot conferencing are to have a physical embodiment and to transmit conversation partner's body motions. We predicted that these features enhance social telepresence. A previous study showed the superiority of a humanoid robot that has realistic appearance to videoconferencing (Sakamoto et al., 2007). In addition, several previous studies reported that avatars that transmit partner's body motions enhance social telepresence compared with audio-only media (Bente et al., 2008; Kang et al., 2008). Since these findings suggest the contribution of the robot's features on social telepresence, we made the following two hypotheses.

Hypothesis 1: a physical embodiment enhances the social telepresence of the conversation partner.

Hypothesis 2: transmitting body motions enhances the social telepresence of the conversation partner.

Conditions

The hypotheses described in the preceding section consist of these two factors: physical embodiment and transmitting information. The physical embodiment factor had two levels, with/without-physical embodiment, and the transmitting information factor had three levels, audio, audio + motion, and audio + motion + appearance. Thus, to examine the hypotheses, we prepared six conditions of a 2×3 design shown in **Figure 1**.

As described in Section "Introduction," both robot conferencing and avatar chat transmit remote person's body motions without disclosing the person's appearance. We thus supposed that the avatar chat can become robot conferencing by adding a physical embodiment. Similarly, we assumed that the voice chat becomes an inactive robot conferencing, which does not transmit the body motions of a remote person and the video chat can become face-to-face communication by adding a physical embodiment. In terms of the transmitting information, we assumed that the voice chat and inactive robot transmit only audio, the avatar and robot transmit audio and motion, and the video and face-to-face transmit audio, motion, and appearance. These assumptions allowed us to analyze the effect of adding a physical embodiment to existing communication media. The details of each condition are described below.

Active Robot Condition (Transmitting Audio and Motion with a Physical Embodiment)

The subject talked to the conversation partner while looking at the robot. The robot had a three-degrees-of-freedom neck and a one-degree-of-freedom mouth. The head and lips moved at 30 frames per second according to the sensor data sent from face tracking software (faceAPI), which was running in a remote terminal and capturing the conversation partner's movements. The camera for face tracking was set behind the robot. The microphone speaker was set behind the robot. The robot was dressed with the same gray shirt as the conversation partner.

Avatar Condition (Transmitting Audio and Motion but no Physical Embodiment)

The subject talked to the conversation partner while looking at an anonymous three-dimensional computer graphics avatar that reflected the conversation partner's head and lip motions. The avatar consisted of a skin-colored cylindrical head, black lips, black eyeballs, and a gray conical body, which was the same color as the shirt of the conversation partner. In the preliminary experiment, we used an avatar, which had a spherical head and a realistic shirt, which looked like the robot. However, there were some subjects who felt hard to notice facial movements of the avatar. This problem was solved by changing the design of avatar to a cylindrical head. The recognizable facial movements might improve social telepresence, and so we employed the cylindrical head. In addition, we modified its body to a conical shape to standardize the abstraction level of the looks. The diameter of the

head was equal to the breadth of the robot's head (13.5 cm). The conversation partner's head and lip motions were tracked in the same way as on the active robot condition. The head translated and rotated with three degrees of freedom. The lips were transformed based on the three-dimensional positions of 14 markers. The head and lips moved at 30 frames per second. The avatar was shown on a 40" display. The display was set longitudinally on the other side of the desk. The bezel of the display was covered with a white board, so that the true display area was 49 cm by 56 cm. The microphone speaker was set behind the display. There were two cameras on top of the display. One was for face tracking, and the other was for live video. In this condition, the camera for live video was covered with a white box. The camera was used in the video condition described below.

Face-to-Face Condition (Transmitting Audio, Motion, and Appearance with a Physical Embodiment)

The subject talked to the conversation partner in a normal face-to-face environment. The conversation partner wore a gray shirt. The distance from the subject to the conversation partner was adjusted to 150 cm so that the breadth of the conversation partner's head looked the same as the breadth of the robot's head (13.5 cm).

Video Condition (Transmitting Audio, Motion, and Appearance but no Physical Embodiment)

This condition was identical to a normal video chat. The subject talked to the conversation partner while looking at a live video of the conversation partner. The conversation partner wore a gray shirt. The resolution of the camera for live video was 1280 pixels by 720 pixels, and its frame rate was 30 frames per second. The video was shown on the same display that was used in the avatar condition. Thus, the true display area was 49 cm by 56 cm. The horizontal angle of view was adjusted to 87° so that the breadth of the conversation partner's head was equal to the breadth of the robot's head (13.5 cm) on the display. The camera for face tracking that was used on the avatar condition was covered with a white box.

Inactive Robot Condition (Transmitting Audio with a Physical Embodiment)

The subject talked to the conversation partner while looking at the inactive robot. The camera for face tracking that was used on the active robot condition was covered with a white box. The subject was preliminarily informed that the robot did not move in this condition.

Audio-Only Condition (Transmitting Audio but no Physical Embodiment)

This condition was similar to a normal voice chat. The subject talked to the conversation partner through only a microphone speaker that was set on the desk.

In the preliminary experiment, some subjects doubted that the experimenter would be looking at them from somewhere even if the experimental condition required no camera. We hence informed the subjects that the dialog environments of the subject side and the conversation partner side were the same in all the conditions. To make the subjects believe this bi-directionality of

the dialog environments, the subjects were shown a live video of the subjects' avatar, robot, or video, which were seen by the conversation partner on a 7" display before each experiment. At the same time, the subjects confirmed that their avatar and robot reflected their face and lip movements. The subjects also confirmed that the avatar and robot in front of them reflected the conversation partner's face and lip movements by comparing a live video of the conversation partner that was shown on the 7" display with the avatar and robot. The 7" display for these confirmations was removed before the experiments.

Task

In the experiment, we informed the subjects that they were going to talk with a conversation partner who is in another room through six communication methods described above. An experimenter played the role of the partner. To observe the difference in the social telepresence between the conditions, we asked the subject to answer a questionnaire (which is explained in the next section) after the experiment ended. Since body motions in conversation are mainly speaking and nodding, we set the task in which the subject could see the partner's mouth and neck motions.

The subject was asked by the experimenter to talk about the issue and resolution of a certain gadget and requests for a new function on that gadget at the beginning of each condition. Because all the subjects had to experience the six conditions, we prepared six gadgets as conversational topics, i.e., e-book readers, handheld game consoles, smartphones, robotic vacuum cleaners, portable audio players, and 3D televisions. We did not disclose the next topic beforehand, and the experimenter told the subject which gadget to talk about right when the condition began. While the subject was talking, the experimenter gave back-channel responses with an utterance and a small nod of his head. The nod motions of the robot, avatar, and experimenter are shown in **Figure 2**. As the figure shows, the robot and avatar synchronized with the experimenter.

We did not ask the subject to talk for more than a certain duration, so the subject could stop talking anytime. However, since the six gadgets are attracting considerable attention recently, most subjects knew the issue and resolution of the gadgets to a certain level, and their speech was able to last more than 1 min.

The order of experiencing the conditions and the order of the topics were counterbalanced. The subject trained the task in the face-to-face condition in order to familiarize the subject with the task and the experimenter's motion and appearance, before

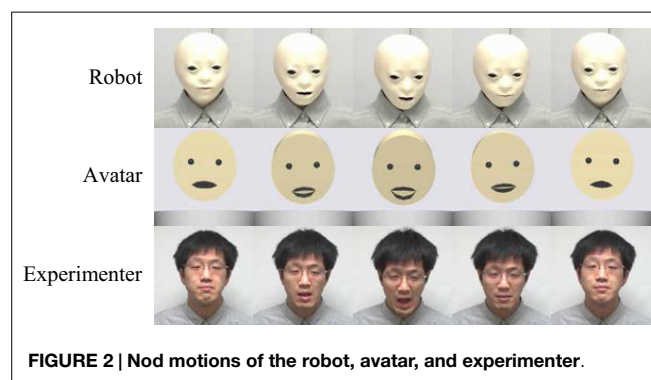


FIGURE 2 | Nod motions of the robot, avatar, and experimenter.

conducting the experiment in the six conditions. The topic of the training was always railway smart cards.

Questionnaire

After experiencing the six conditions, the subjects answered a questionnaire, which asked them to estimate the social telepresence, i.e., the degree of resembling face-to-face interaction (Finn et al., 1997) for each condition. We wanted to obtain the relative comparison of the conditions to avoid a ceiling effect. In the preliminary experiment, we conducted the questionnaire after each condition. However, when a subject marked the highest score for the first condition, the subject was not able to mark higher score for the later conditions even if he/she felt higher social telepresence. In the case of a between-subject design, such a problem will not happen, but another problem here is that there are six conditions, thus a lot of subjects are necessary to enable a between-subjects design.

The questionnaire is shown in **Figure 3**. The questionnaire had six statements that corresponded to the six conditions. The statement was the following: I felt as if I were talking to the conversation partner in the same room. Previous studies showed that the statement which asks a feeling of being in the same room is useful to measure the social telepresence (Nakanishi et al. 2008, 2009, 2011, 2014; Tanaka et al., 2014). The statement was rated on a 9-point Likert scale where 1 = strongly disagree, 3 = disagree, 5 = neutral, 7 = agree, and 9 = strongly agree. The subjects thereby could score the same number on the statements if they felt the same level of social telepresence in the conditions.

The statements were sorted in the order of the conditions and were printed on the questionnaire, with a photo that showed the experimental setup of the corresponding condition. The sort and the photo were good cues to help the subjects remember the feeling of social telepresence in each condition. After conducting

the questionnaire, we interviewed the subjects in order to confirm the reason of scoring. The interview was open-ended. When we received the questionnaire that was against our hypotheses, we asked the subject the reason, e.g., the reason why the avatar condition was higher than the robot condition. Even if the questionnaire followed our hypotheses, we asked the reason, to confirm what point the subject focused on, e.g., physical embodiment, body motion, or appearance.

Result

Thirty-six subjects (17 females and 19 males) participated in our first experiment. The experiment was within-subject design, so each subject experienced all of the six conditions. We did not control the subjects' prior knowledge about the topics of talking with the experimenter. Instead, at the interview following the experiment, we confirmed that their scoring of questionnaire was conducted independently from the difference of topics. There was no subject who mentioned about the topics as the reason of his/her scoring.

Figure 4 shows the result of the questionnaire, in which each point represents the mean value of the scores, and each bar represents the SEM value.

We compared the six conditions to find the effects of the physical embodiment and the transmitting information factors. Since the physical embodiment and the transmitting information factors consisted of two and three levels as shown in **Figure 1** and each subject evaluated all conditions, we conducted 2×3 two-way repeated-measures ANOVA. As a result, we found strong main effects of the physical embodiment factor [$F(1, 35) = 36.955, p < 0.001$] and the transmitting information factor [$F(2, 70) = 279.603, p < 0.001$]. We also found a strong interaction between these factors [$F(2, 70) = 14.794, p < 0.001$]. Regarding this interaction, we calculated the *post hoc* statistical power. First, we calculated the effect size 0.650 from the partial correlation ratio 0.297. Finally, we obtained the sufficiently high-statistical power 0.999 when the significance level was 0.001.

We further analyzed the simple main effects in the interaction with the Bonferroni correction. The physical embodiment significantly improved the social telepresence of the conversation partner, when the transmitting information was audio + motion + appearance [$F(1, 105) = 8.857, p < 0.01$], and audio + motion [$F(1, 105) = 65.470, p < 0.001$]. When the transmitting information was audio only, there was a non-significant tendency for the social telepresence to increase [$F(1, 105) = 3.460, p = 0.086$]. This meant that the subjects felt a higher social telepresence of the conversation partner in the face-to-face condition than in the video condition, and the active robot condition conveyed a higher social telepresence than the avatar. These results support hypothesis 1 that the physical embodiment enhances the social telepresence of the conversation partner. However, the effect of the physical embodiment on the social telepresence was lower in the audio-only communication.

Furthermore, there were significant differences between the three levels of the transmitting information in both cases of without-physical embodiment [$F(2, 140) = 223.095, p < 0.001$] and with-physical embodiment [$F(2, 140) = 107.141, p < 0.001$]. Multiple comparisons showed that the subjects felt a higher social







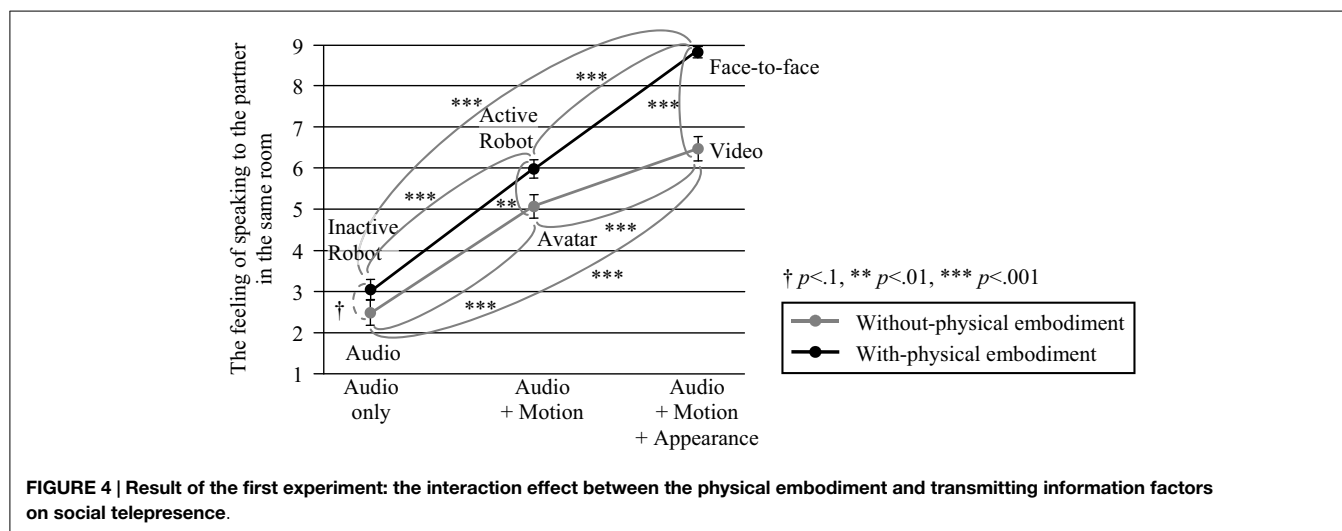
I felt as if I were talking to the conversation partner in the same room.										
	strongly disagree	disagree	neutral	agree	strongly agree					
	1	2	3	4	5	6	7	8	9	
	strongly disagree	disagree	neutral	agree	strongly agree					
Inactive	1	2	3	4	5	6	7	8	9	
	strongly disagree	disagree	neutral	agree	strongly agree					
	1	2	3	4	5	6	7	8	9	
	strongly disagree	disagree	neutral	agree	strongly agree					
Active	1	2	3	4	5	6	7	8	9	
	strongly disagree	disagree	neutral	agree	strongly agree					
	1	2	3	4	5	6	7	8	9	
	strongly disagree	disagree	neutral	agree	strongly agree					
	1	2	3	4	5	6	7	8	9	

FIGURE 3 | Questionnaire to evaluate social telepresence of the six conditions.



telepresence in the face-to-face condition than in the active robot ($p < 0.001$) and inactive robot ($p < 0.001$) conditions, the active robot condition conveyed a higher social telepresence than the inactive robot condition ($p < 0.001$), the video condition conveyed a higher social telepresence than the avatar ($p < 0.001$) and the audio-only ($p < 0.001$) conditions, and the avatar condition conveyed a higher social telepresence than the audio-only condition ($p < 0.001$). These results prove hypothesis 2 that transmitting body motions enhances the social telepresence of the conversation partner. In addition, transmitting appearance also enhanced the social telepresence.

Experiment 2

The first experiment demonstrated that transmitting the remote partner's current appearance as well as a physical embodiment enhances social telepresence. The result that the robot and video conditions were similar level as shown in **Figure 4** might be because the robot condition could not transmit the appearance. This will be discussed in detail in Section "Discussion." Therefore, the clear usefulness of humanoid robots was not demonstrated. To prove the usefulness of humanoid robots, we had to prove the benefit of both the physical embodiment and the absence of remote partner's appearance.

This section presents the second experiment in which we investigated whether the presence of a remote partner in the teleoperated mode produces a sense of talking with the partner while actually talking with the robot in the autonomous mode. To confirm the contribution of a physical embodiment to produce such a pseudo presence, we compared the robot and audio-only conditions.

Since the first experiment suggested that a physical embodiment is effective when a robot moves, the second experiment dealt with a robot as a single condition. We chose audio-only (without-physical embodiment and body motion) as a condition for the baseline. Audio-only media (e.g., voice chats) can also be automated, since a user cannot see the remote partner's current appearance, like with a robot. If a remote partner's presence in audio-only media can produce the sense of talking with the

partner while actually talking with an autonomous reply system, the physical embodiment, which is a robot, is not needed to produce such a pseudo presence.

Dialog Modes

Almost all the studies that proposed algorithms that generate talking behaviors focused on facial movements, e.g., nodding and lip motions. Since they seem to be the most fundamental facial movements while talking, this study also addressed them. Many of the teleoperated robots proposed in past studies have a face that can present these motions (Sakamoto et al., 2007; Watanabe et al., 2010; Hashimoto et al., 2011; Ogawa et al., 2011). In this experiment, we used telenoid that was used in experiment 1. We controlled it in two modes: teleoperated and autonomous.

Teleoperated Mode

In this mode, the robot's head and mouth were synchronized with the remote operator. The method to control the robot was same as the active robot condition described in Section "Conditions."

Autonomous Mode

The roles of the remote partner in the dialogs are listener and speaker. Their behaviors are mainly nod and lip motions, respectively. We constructed a back-channel system that detects the timing of back-channel feedback from user's speech and a lip-sync system that generates a lip motion synchronized with a remote partner's speech. We simplified these systems for the following reason. If our approach that changes the dialog modes makes subjects feel like they are talking with their remote partner even in simple systems, it would obviously also work on systems that generate more natural and various talking behaviors.

Back-channel system

Many methods detect the timing of back-channel responses. Most used prosodic information, including pause (Noguchi and Den, 1998; Takeuchi et al., 2003; Truong et al., 2010; Watanabe et al., 2010), and fundamental frequency (Noguchi and Den, 1998; Ward and Tsukahara, 2000; Truong et al., 2010). Our method used only the speech pause since it is good cue to identify the break or

end of a sentence, which seems to be the appropriate timing of back-channel responses. One study also used only a speech pause, although their algorithm is more complex than ours in order to enable estimating the timing of back-channel earlier (Watanabe et al., 2010).

The detection rule is shown in **Figure 5**. Each box represents an utterance, and the distance between each box is pause duration t_1 . The utterance and pause parts correspond to higher and lower sound pressure, respectively. The system judged t_1 as a target pause if it exceeded 0.6 s. Speech duration t_2 is the elapsed time from the start of the speech to the time at which the target pause was recognized. If t_2 exceeded 2.0 s, the system judged the target pause as the timing of the back-channel response and reset t_2 to 0. This means that the system reproduces the back-channel response when the pause is continued for 0.6 s after the speech continued for more than 2.0 s.

To adjust the parameters of our back-channel system, we conducted preliminary experiments in which a subject evaluated the timing and frequency of robot's backchannel in the same task to the second experiment. We found that back-channels repeated in short time (<2.0 s) decreases the naturalness of the timing. In addition, the backchannel, which was done more than 0.6 s later from the break or end of sentence tended to be felt late, but the pause that is <0.6 s is not enough to judge the break or end of sentence. We therefore set pause duration t_1 and speech duration t_2 to 0.6 and 2.0 s, respectively.

At the back-channel response, the robot made a nodding motion and a pre-recorded acoustic back-channel. In our preliminary experiment, we used only one nodding motion and an acoustic back-channel, but subjects pointed out that the robot's response seemed constant. We therefore prepared three nodding motions that differ in their degree of pitch and speed and two acoustic back-channels that slightly differ in their tone of voice. This problem that subjects feel constant the robot's response was solved by randomly selecting these nodding motions and acoustic back-channels at the timing.

Lip-sync system

Some lip-sync methods generate lip motions from a human's voice to control a robot (Watanabe et al., 2010; Ishi et al., 2012) and a computer graphic avatar (Cao et al., 2005; Salvi et al., 2009; Watanabe et al., 2010). Our method was simpler because controlling a one-degree-of-freedom mouth does not need highly accurate lip-sync methods.

Our lip-sync system measured the acoustic pressure of the human's voice and related the level to the angle of the robot's chin. In other words, the robot's mouth was synchronized with the

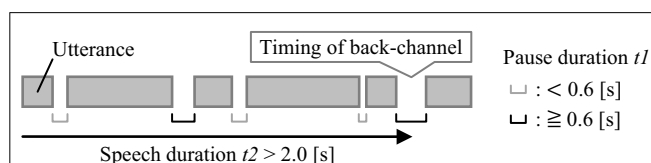


FIGURE 5 | Method to detect timing of back-channel response for the second experiment: the rule to detect the timing when the pause is continued for 0.6 s after the speech continued for more than 2.0 s.

waveform of the human's voice. In our experiments, this system was driven by pre-recorded remote partner's speeches.

Hypothesis

In the first experiment, the active robot condition could convey the same degree of social telepresence as the video condition without transmitting the partner's appearance. We hence thought that physically embodied body motions effectively make the user imagine the remote partner's presence. Due to the experience of imagining the partner's presence, the user might feel the remote partner's pseudo presence when talking with an autonomous robot. On the other hand, when the user talks with an autonomous reply system that uses only speech, it seems harder to feel the partner's presence due to poor information for imagining.

A previous study showed that a teleoperated robot produces higher social telepresence of a remote partner than audio-only communication due to the effect of physical embodiment (Tanaka et al., 2014). We hence predict that physically embodied body motion will improve the sense of talking with a remote partner. The following is the hypothesis of the second experiment.

Hypothesis 3

The user who experienced talking with a remote partner through a teleoperated robot that presents the partner's body motion will feel the sense of talking with the partner even when talking with the same robot that is being autonomously controlled.

Conditions

To examine hypothesis 3, we prepared the four conditions shown in **Figure 6**. The experiment included the experience and autonomous phases. The experience phase was only included in the with-experience conditions. Before experiencing the experience phase, the subjects were told that they would be talking

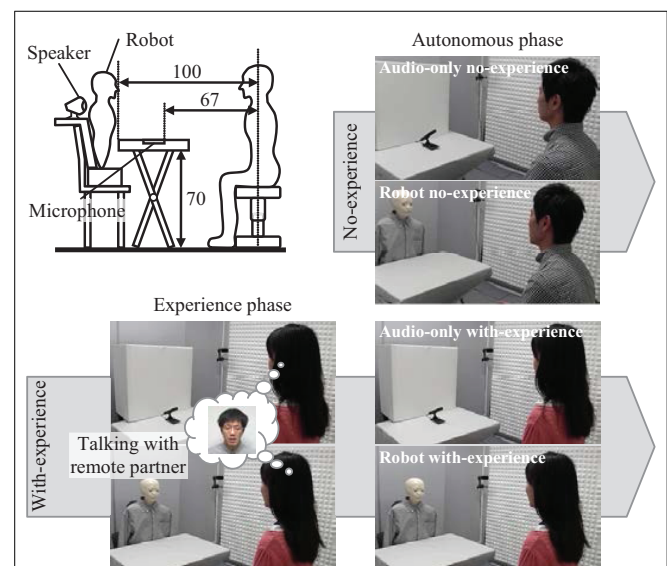


FIGURE 6 | Conditions and the setup of the second experiment (length unit: centimeters): the audio-only and robot conditions were conducted in the presence and absence of prior experience of talking with remote partner.

with a remote partner in the teleoperated mode. However, if an experimenter replied to the subject's speech, the quality of the conversation would differ for each subject. Actually, the experience phase was conducted in the autonomous mode to control the quality of the conversation for each subject. The manipulation check that will be explained in Section "Questionnaire" confirmed that all the subjects believed that the remote partner was listening to their speech through the robot. As described in Section "Autonomous Mode," the back-channel systems proposed by previous works would detect more appropriate timing of backchannel, but our simple algorithm was enough to make the subjects believe the remote partner's presence at a one-turn interaction. Before experiencing the autonomous phase, the subjects were told that they would be talking with an autonomous robot, which autonomously gives back-channel responses. To control the subjects' prior knowledge, we gave them handouts that explained the teleoperated and autonomous modes before the experiment. We also explained our experiments to the subjects.

The figure also shows the experimental setup. In all the experiments, the subject sat in front of a desk. The robot was placed on the other side of it. A directional microphone was embedded in the desk to capture the subject's speech, and the top of it was covered with a cloth to hide the microphone. A speaker was set behind the robot to produce the remote partner's speech.

Task

In the second experiment, the subject was a speaker, and the robot or the system gave a back-channel response to his/her speech as a listener for the following reason. If the subject is a listener, the autonomous system in the audio-only conditions only plays pre-recorded partner's speeches unilaterally from the speaker. In this case, the audio-only conditions seem to have a disadvantage over the robot conditions.

The task was same to the first experiment. The subjects were asked to talk about a gadget at the beginning of each conversation through the robot or the speaker. The lines of asking and the acoustic responses were the pre-recorded voices of a member of our research group. The member greeted the subjects before the experiment to identify the remote partner. The topics in the experience and autonomous phases were portable audio players and robotic vacuum cleaners, and smartphones and 3D TVs, respectively. The order of the topics was counterbalanced.

Questionnaire

After talking about one topic, the subjects were asked to answer manipulation check questions to confirm whether they correctly understand our instructions. For example, after the experience phase, we confirmed that the subjects believed that they were actually talking to a remote partner although it was an autonomous system. The manipulation check consisted of the following two YES/NO statements:

- In the last experiment, your speech was listened to by a remote partner.
- In the last experiment, your speech was recorded instead of being listened to by a remote partner.

After the experiment, the subjects were asked to estimate the pseudo presence that is the remote partner's presence in the autonomous phase. The following was the questionnaire statement:

- I felt as if the conversation partner was listening to me in the same room.

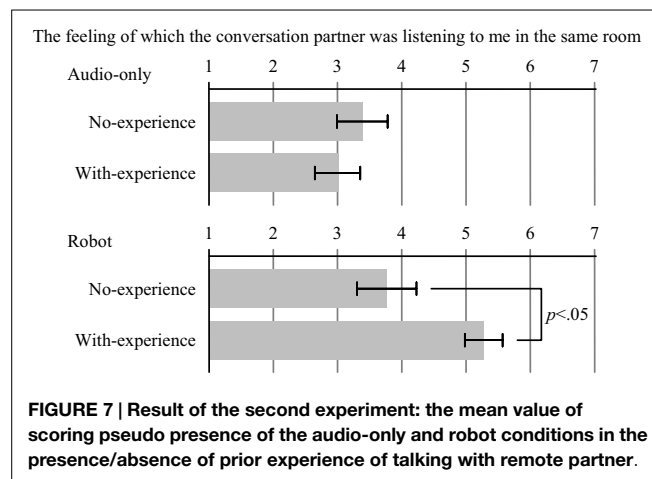
Answers were rated on a 7-point Likert scale: 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = neutral, 5 = slightly agree, 6 = agree, and 7 = strongly agree. We collected open-ended responses to infer what determined the scores. The statement was accompanied with an entry column where the subjects rationalized their scores.

Result

Sixteen subjects participated in our second experiment. Half (five females and three males) participated in the with-experience conditions and experienced both the experience and autonomous phases. At each phase, they talked in both the audio-only and robot conditions. The order of experiencing the audio-only and robot conditions was counterbalanced. The other half of the subjects (four females and four males) participated in the no-experience conditions and only experienced the autonomous phase. According to the manipulation check, we confirmed that all the subjects believed our instruction.

The result of the second experiment is shown in **Figure 7**, where each box represents the mean value of the responses to the statement, and each bar represents the SEM value. The figure compares the no- and with-experience conditions by a between-subjects *t*-test.

In the audio-only conditions, there was no significant difference between the no- and with-experience conditions [$t(14) = 0.664$, *n.s.*]. On the other hand, in the robot conditions, we found a significant difference between them [$t(14) = 2.575$, $p < 0.05$]. This means that the prior experience in which the subjects talked with the remote partner produced the pseudo presence in the autonomous phase when the subjects could see back-channel responses through the robot. However, the experience did not produce pseudo presence in the audio-only



communication. These results proved hypothesis 3 described in Section “Hypothesis.”

In the with-experience conditions, the subjects had the conversations twice, but in the no-experience condition, they only had them once. Therefore, more conversations might improve the sense of talking with a remote partner. In spite of this, in the audio-only conditions, the difference between the no- and with-experience conditions was not significant. We hence considered that physically embodied motion was the significant factor to produce the pseudo presence regardless of the number of conversations.

Discussion

In the first experiment, the physical embodiment enhanced the social telepresence of the conversation partner. In the interviews, 7 of the 36 subjects said that they felt as if they were facing the conversation partner in the active robot condition compared with the avatar condition because there was a physical object in front of them. However, there was no significant difference between the audio-only condition and the inactive robot condition. In the interviews, 3 of the 36 subjects said that the inactive robot condition was not that different to the audio condition because they could not see the conversation partner's reaction. In fact, in the questionnaire, 8 of the 36 subjects rated the same score for the audio and inactive robot conditions. Moreover, 5 of the 36 subjects said that they felt as if the conversation partner was in front of them when the robot moved. These subjective responses support the experimental result that a physical embodiment enhances social telepresence when transmitting body motions. This result indicates the superiority of robots to avatars, which does not have a physical embodiment. Nevertheless, there are some subjects who rated the same or higher score for the avatar condition than the robot condition. Most such subjects mentioned the uncanny appearance of the robot as the reason for their rating, and they tended to prefer the avatar's design. Thus, if the robot's design was more abstracted, the superiority would appear more significantly.

Presence or absence of motion parallax can be cited as one of the differences between physical embodiment and video. When interacting with the robot, the depth from motion parallax could increase visibility of body motions. The lack of the depth information might be the cause of feeling hard to notice facial movements of the avatar used in the preliminary experiment described in Section “Conditions.” A previous study reported that motion parallax generated by the movement of a camera enhances social telepresence (Nakanishi et al., 2009). The visibility of bodily motion improved by the motion parallax may have contributed to enhance social telepresence.

In terms of the transmitting information, the appearance enhanced social telepresence as well as the body motions. This result shows the disadvantage of robots and avatars that do not transmit the partner's appearance. Although the active robot has this disadvantage, the active robot and video conditions seemed to convey the same degree of social telepresence, as shown in **Figure 4**. In the questionnaire, approximately half of the subjects (16 of the 36) rated the same or higher score for the active robot condition than the video condition. We assumed that the

enhanced social telepresence by the physical embodiment offset the decreased social telepresence by the absence of the partner's appearance. Therefore, the reported superiority of the robot in the social telepresence to the video (Sakamoto et al., 2007) could be caused by the robot's realistic appearance.

We did not investigate the conditions that transmit audio and appearance but not motion. Talking through an inactive robot that has a realistic appearance of a partner, and a partner's photo could correspond to such conditions. Watching the partner's photo while talking is a popular situation since many users of instant messengers put their photos in the buddy list. Although the transmitting appearance enhances social telepresence as mentioned above, it has not been clarified whether the appearance works even if the motion is not transmitted. The effect of presenting appearance on the smoothness of speech had already demonstrated (Tanaka et al., 2013, 2015). The previous study showed that presenting partner's avatar increased the degree of the smoothness of speaking to the partner, but partner's photo did not have such an effect. We hence predict that the appearance also does not enhance social telepresence if the motion is not transmitted as is the case with the physical embodiment. To prove this hypothesis is a future work.

Although the subjects who rated the with-physical embodiment condition higher in all level of the transmitting information factor were less than half of all the subjects (14 of 36), there might possibly be a certain bias toward a preference for physical embodiment. A between-subject design avoids such a bias, but there is a problem that requires a lot of subjects to conduct it as described in Section “Questionnaire.” It is a future work to investigate the effect of physical embodiment without the bias.

The first experiment could not show the superiority of humanoid robots to videos, since humanoid robots cannot present the remote partner's current appearance, which can be transmitted by videos. To prove the usefulness of humanoid robot, we had to demonstrate the benefit of both the physical embodiment and the absence of partner's appearance. There are several studies that partly replaced the partner's video with a robot to obtain the positive effects of both of appearance and physical embodiment (Samani et al., 2012; Nakanishi et al., 2014). This is an opposite approach of ours that utilizes one of the features of humanoid robot that is the absence of the partner's appearance. A humanoid robot can pretend as if it is controlled by a remote operator due to not transmitting the appearance. When interacting with the humanoid robot, the user could feel pseudo presence of the remote operator, and the physical embodiment might be able to enhance the pseudo presence as well as social telepresence. The second experiment investigated these predictions.

The second experiment showed that the interaction with a humanoid robot produces the remote partner's pseudo presence that is the feeling of talking with a remote partner when interacting with an autonomous system compared with audio-only interaction. We found that the subjects tended to deduce the presence/absence of a remote partner according to their prior experience with that same remote partner. However, the experience did not work well at the audio-only interaction. We hence considered that the physically embodied body motions might facilitate recalling the partner's presence based on the experience.

We also found that the subjects' deductions of the presence/absence of the remote partner were influenced by their belief about prior experience. In the experience phase, even though the autonomous system gave back-channel responses under the guise of a remote partner, all the subjects believed that the remote partner was listening to their speech. Such a fake experience produced the sense of talking with the partner when talking with an autonomous robot. Nevertheless, the open-ended responses of all the subjects who participated in the with-experience robot condition did not mention the similarity between the experience and autonomous phases. Almost all the subjects focused on whether the back-channel responses were done in appropriate timing. This result implies that the subjects' deductions were subconsciously influenced by the prior experience.

There is a question whether the real experience that a remote partner is actually replying to the user's speech produce higher pseudo presence. Compared with back-channel system, a real partner can give various responses according to the context of conversation. Such a real experience gives a stronger impression that the remote partner is listening, and the impression would effectively produce the pseudo presence. There is another question whether the prior experience produces the pseudo presence when the robot unilaterally speaks to a subject. The user might feel less presence of a remote partner because the robot is unilaterally reproducing talking behaviors and pre-recorded speech like a video message. In this case, it might be difficult to produce pseudo presence, since the factors in determining the presence/absence of the remote partner (e.g., timing of back-channel response) will be less. Answering these questions is future work. In addition, it is also future work to examine whether a user felt the remote partner's pseudo presence through observation data, e.g., observing whether a user replies to the robot's greeting. If he/she felt that the remote partner had been listening/speaking, they might reply to the greeting; if he/she did not feel that way, they might ignore it.

Conclusion

In this study, to prove the usefulness of humanoid robot, we investigated how the features of humanoid robot contribute to produce remote partner's real/pseudo presence. In the first experiment, we compared robot conferencing with existing communication media divided into physical embodiment and transmitting

information factors. As a result, we found that physically embodied body motions enhance the partner's real presence, i.e., social telepresence, although physical embodiment without presenting body motion does not have such an effect. This result shows the superiority of robots to avatars. However, we also found that the partner's appearance, which robots cannot reproduce, enhances social telepresence. Consequently, humanoid robots were comparable to live videos since the positive effect of the physical embodiment offset the negative effect of lacking appearance.

Previous studies have discussed the superiority of humanoid robots to live videos, but our study noted that humanoid robots in the absence of presenting remote partner's appearance do not always have the superiority in social telepresence. Alternatively, this study proposed the utilization of the anonymity of humanoid robot to produce the partner's pseudo presence that is the feeling of talking with a remote partner when interacting with an autonomous robot. In the second experiment, we evaluate whether an autonomous robot produces a similar presence as a teleoperated robot. From the experiment, we found that the prior experience of talking with the remote partner in teleoperation is effective for producing pseudo presence. If a user watched the remote partner's body motion that reproduced by a robot, the user feels the pseudo presence of the partner even while talking with the same robot in autonomous control.

In terms of conveying social telepresence, live videos are more useful than humanoid robots because operating humanoid robots requires higher cost than using displays. We hence conclude that blurring between teleoperation and autonomous control is desirable for effectively utilizing a humanoid robot. Substituting an autonomous system for the remote operator reduces the operator's task, and at the same time, the user could continue to feel the presence of a remote partner also while interacting with an autonomous system.

Acknowledgments

This study was supported by JST CREST "Studies on Cellphone-type Teleoperated Androids Transmitting Human Presence," JSPS KAKENHI Grant Number 26280076 "Robot-Enhanced Displays for Social Telepresence," SCOPE "Studies and Developments on Remote Bodily Interaction Interfaces," and KDDI Foundation Research Grant Program "Robotic Avatars for Human Cloud."

References

- Bailenson, J. N., Yee, N., Merget, D., and Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence* 15, 359–372. doi:10.1162/pres.15.4.359
- Bainbridge, W. A., Hart, J., Kim, E. S., and Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *Int. J. Soc. Robot.* 1, 41–52. doi:10.1007/s12369-010-0082-7
- Bente, G., Ruggenberg, S., Kramer, N. C., and Eschenburg, F. (2008). Avatar-mediated networking: increasing social presence and interpersonal trust in net-based collaborations. *Hum. Commun. Res.* 34, 287–318. doi:10.1111/j.1468-2958.2008.00322.x
- Cao, Y., Tien, W. C., Faloutsos, P., and Pighin, F. (2005). Expressive speech-driven facial animation. *ACM Trans. Graph.* 24, 1283–1302. doi:10.1109/TNN.2002.1021892
- de Greef, P., and Ijsselstein, W. (2001). Social presence in a home tele-application. *Cyberpsychol. Behav.* 4, 307–315. doi:10.1089/109493101300117974
- Finn, K. E., Sellen, A. J., and Wilbur, S. B. (1997). *Video-Mediated Communication*. New Jersey: Lawrence Erlbaum Associates.
- Garau, M., Slater, M., Bee, S., and Sasse, M. A. (2001). "The impact of eye gaze on communication using humanoid avatars," in *Proc. CHI 2001* (Minneapolis: ACM), 309–316.
- Hashimoto, T., Kato, N., and Kobayashi, H. (2011). Development of educational system with the android robot SAYA and evaluation. *Int. J. Adv. Robot. Syst.* 8, 51–61. doi:10.5772/10667
- Isaacs, E. A., and Tang, J. C. (1994). What video can and can't do for collaboration: a case study. *Multimed. Syst.* 2, 63–73. doi:10.1007/BF01274181
- Ishi, C., Liu, C., Ishiguro, H., and Hagita, N. (2012). "Evaluation of formant-based lip motion generation in tele-operated humanoid robots," in *Proc. IROS 2012*. Vilamoura.

- Kang, S., Watt, J. H., and Ala, S. K. (2008). "Communicators' perceptions of social presence as a function of avatar realism in small display mobile communication devices," in *Proc. HICSS 2008*. Waikoloa.
- Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y., and Heath, C. (2004). "Dual ecologies of robot as communication media: thoughts on coordinating orientations and projectability," in *Proc. CHI 2004* (Vienna: ACM), 183–190.
- Le, B. H., Ma, X., and Deng, Z. (2012). Live speech driven head-and-eye motion generators. *IEEE Trans. Vis. Comput. Graph.* 18, 1902–1914. doi:10.1109/TVCG.2012.74
- Lee, J., Wang, Z., and Marsella, S. (2010). "Evaluating models of speaker head nods for virtual agents," in *Proc. AAMAS 2010* (Toronto: IFAAMAS), 1257–1264.
- Lee, K. M., Jung, Y., Kim, J., and Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. *Int. J. Hum. Comput. Stud.* 64, 962–973. doi:10.1016/j.ijhcs.2006.05.002
- Liu, C., Ishi, C. T., Ishiguro, H., and Hagita, N. (2012). "Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction," in *Proc. HRI 2012* (Boston: ACM), 285–292.
- Morita, T., Mase, K., Hirano, Y., and Kajita, S. (2007). "Reciprocal attentive communication in remote meeting with a humanoid robot," in *Proc. ICMI 2007* (Nagoya: ACM), 228–235.
- Nakanishi, H., Kato, K., and Ishiguro, H. (2011). "Zoom cameras and movable displays enhance social telepresence," in *Proc. CHI 2011* (Vancouver: ACM), 63–72.
- Nakanishi, H., Murakami, Y., and Kato, K. (2009). "Movable cameras enhance social telepresence in media spaces," in *Proc. CHI 2009* (Boston: ACM), 433–442.
- Nakanishi, H., Murakami, Y., Nogami, D., and Ishiguro, H. (2008). "Minimum movement matters: impact of robot-mounted cameras on social telepresence," in *Proc. CSCW 2008* (San Diego: ACM), 303–312.
- Nakanishi, H., Tanaka, K., and Wada, Y. (2014). "Remote handshaking: touch enhances video-mediated social telepresence," in *Proc. CHI 2014* (Toronto: ACM), 2143–2152.
- Nguyen, D. T., and Canny, J. (2009). "More than face-to-face: empathy effects of video framing," in *Proc. CHI 2009* (Boston: ACM), 423–432.
- Noguchi, H., and Den, Y. (1998). "Prosody-based detection of the context of backchannel responses," in *Proc. ICSLP 1998*. Sydney.
- Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T., and Ishiguro, H. (2011). Exploring the natural reaction of young and aged person with telenoid in a real world. *J. Adv. Comput. Intell. Inform.* 15, 592–597.
- Ranatunga, I., Torres, N. A., Patterson, R. M., Bugnariu, N., Stevenson, M., and Popa, D. O. (2012). "RoDiCA: a human-robot interaction system for treatment of childhood autism spectrum disorders," in *Proc. PETRA 2012*. Heraklion.
- Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., and Hagita, N. (2007). "Android as a telecommunication medium with a human-like presence," in *Proc. HRI 2007* (Washington, DC: ACM), 193–200.
- Salvi, G., Beskow, J., Moubayed, S. A., and Granstrom, B. (2009). SynFace: speech-driven facial animation for virtual speech-reading support. *EURASIP J. Audio Speech Music Process.* 2009:191940. doi:10.1155/2009/191940
- Samani, H. A., Parsani, R., Rodriguez, L. T., Saadatian, E., Dissanayake, K. H., and Cheok, A. D. (2012). "Kissenger: design of a kiss transmission device," in *Proc. DIS 2012* (Newcastle: ACM), 48–57.
- Sirkin, D., and Ju, W. (2012). "Consistency in physical and on-screen action improves perceptions of telepresence robots," in *Proc. HRI 2012* (Boston: ACM), 57–64.
- Takeuchi, M., Kitaoka, N., and Nakagawa, S. (2003). "Generation of natural response timing using decision tree based on prosodic and linguistic information," in *Proc. Interspeech 2003*. Geneva.
- Tanaka, K., Nakanishi, H., and Ishiguro, H. (2014). "Comparing video, avatar, and robot mediated communication: pros and cons of embodiment," in *Proc. CollabTech 2014, CCIS 460* (Santiago: Springer), 96–110.
- Tanaka, K., Nakanishi, H., and Ishiguro, H. (2015). Appearance, motion, and embodiment: unpacking avatars by fine-grained communication analysis. *J. Concurr. Comput.* doi:10.1002/cpe.3442
- Tanaka, K., Onoue, S., Nakanishi, H., and Ishiguro, H. (2013). "Motion is enough: how real-time avatars improve distant communication," in *Proc. CTS 2013* (San Diego: IEEE), 465–472.
- Tanaka, M., Ishii, A., Yamano, E., Ogikubo, H., Okazaki, M., Kamimura, K., et al. (2012). Effect of a human-type communication robot on cognitive function in elderly women living alone. *Med. Sci. Monit.* 18, CR550–CR557. doi:10.12659/MSM.883350
- Truong, K. P., Poppe, R., and Heylen, D. (2010). "A rule-based backchannel prediction model using pitch and pause information," in *Proc. Interspeech 2010* (Makuhari: UTPublications), 26–30.
- Ward, N., and Tsukahara, W. (2000). Prosodic features which cue back-channel responses in English and Japanese. *J. Pragmat.* 32, 1177–1207. doi:10.1016/S0378-2166(99)00109-5
- Watanabe, T., Okubo, M., Nakashige, M., and Danbara, R. (2010). InterActor: speech-driven embodied interactive actor. *Int. J. Hum. Comput. Interact.* 17, 43–60. doi:10.1207/s15327590ijhc1701_4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Tanaka, Nakanishi and Ishiguro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.