

Article

Maintaining the Sense of Agency in Semi-Autonomous Robot Conferencing

Kazuaki Tanaka ^{1,*}, Kota Takenouchi ², Kohei Ogawa ^{2,3}, Yuichiro Yoshikawa ^{2,3}, Shuichi Nishio ⁴ and Hiroshi Ishiguro ^{2,3}

¹ Information and Human Sciences, Kyoto Institute of Technology, Kyoto 606-8585, Japan

² Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan

³ ERATO, Japan Science and Technology Agency, Osaka 560-8531, Japan

⁴ Institute for Open and Transdisciplinary Research Initiatives, Osaka University, Osaka 560-8531, Japan

* Correspondence: k_tanaka@kit.ac.jp

Received: 18 May 2019; Accepted: 1 July 2019; Published: 3 July 2019



Abstract: In semi-autonomous robot conferencing, not only the operator controls the robot, but the robot itself also moves autonomously. Thus, it can modify the operator's movement (e.g., adding social behaviors). However, the sense of agency, that is, the degree of feeling that the movement of the robot is the operator's own movement, would decrease if the operator is conscious of the discrepancy between the teleoperation and autonomous behavior. In this study, we developed an interface to control the robot head by using an eye tracker. When the robot autonomously moves its eye-gaze position, the interface guides the operator's eye movement towards this autonomous movement. The experiment showed that our interface can maintain the sense of agency, because it provided the illusion that the autonomous behavior of a robot is directed by the operator's eye movement. This study reports the conditions of how to provide this illusion in semi-autonomous robot conferencing.

Keywords: semi-autonomous; teleoperation; telepresence; robot; sense of agency; eye tracking

1. Introduction

Previous studies have proposed various robot conferencing methods in which the user speaks to a person using a teleoperated robot [1–4]. Through the effects of robot conferencing, both the operator of the robot and the conversation partner can feel a greater social telepresence, that is, the sense of being in the same room as the conversation partner [4], than the other remote conferencing media. The physical presence of a robot enhances the sense of talking face-to-face with the operator of the robot [5]. Moreover, the sense of agency (i.e., the feeling that the movement of the robot is the operator's own movement) seems to contribute to the social telepresence. The sense of agency is enhanced when the operator recognizes the synchronization between themselves and the movements of the virtual body [6–9].

Unlike videoconferencing, robot conferencing does not transmit the current appearance of the operator. However, this feature is not always a disadvantage. If a robot is used as an avatar, the operator can act like another person who has a different appearance and/or character than the operator. Furthermore, the body motion of the operator can be modified, because it is indirectly transmitted by a robot. For example, if a robot automatically demonstrates polite manners, the operator may be regarded as a polite person by their conversation partner. We call robot conferencing where not only the operator controls the robot, but the robot itself also moves autonomously, a semi-autonomous robot conferencing. However, it is known that the operator of a robot is sensitive for the robot motion that differs from her teleoperation, as a result of delay and direction gap, and such a discomfort

teleoperation decreases the sense of agency [8]. This study proposes the method so as to maintain a sense of agency in the semi-autonomous robot conferencing.

To incorporate autonomous behavior in a teleoperation while maintaining the sense of teleoperation, we considered that blurring the distinction between teleoperation and autonomous behavior is effective. If an operator believes that the autonomous behavior of the robot is as a result of their unconscious teleoperation, they will not notice a difference. Therefore, we focused on an eye movement that moves voluntarily and involuntarily, and developed an interface in which the operator controls the face of the robot using an eye tracker. While an operator controls the robot through the interface, guiding the operator's eye movement (that is, the visual guidance) towards the autonomous head movement of the robot would make them believe that the behavior is their own involuntarily eye movement.

This paper is organized as follows. The following section describes the interface that allows an operator to remotely control our robot through an eye tracker. The third section presents related work. The fourth section explains our five hypotheses on adding autonomous behavior to teleoperation. The fifth section explains the method we used to examine the hypotheses, along with the experimental settings. The sixth section presents the results, and the seventh section discusses the results. The final section concludes the study.

2. Teleoperation Interface

This section first introduces our teleoperation interface in order to describe our idea to make an operator feel the illusion that their teleoperation is responsible for autonomous behavior. The detailed experimental setup will be described in the Section 5.2. Figure 1 shows our semi-autonomous robot conferencing system. The upper part of the figure shows our teleoperation interface ((a) operator side), which allows an operator to control the facial direction of our robot using an eye tracker. The display presents a live video (1920 × 1080 pixels, 30 fps) of the remote partner, which is captured by a camera included in a motion capture sensor (Microsoft Kinect for Windows) that is located on the remote side. The display was equipped with an eye tracking sensor (Tobii), and the green circle identified the place where the operator is looking. The diameter of the circle was set to 90 pixels, which is enough to contain the face of the partner who is sitting approximately 1 m away from the robot. The operator can look around the remote location on the display using only their eye movement, but the robot has to change its facial direction as well as its eye direction in order to look around the corresponding region in a natural movement. Therefore, the system determines the motor commands of the robot so that the place where the operator is looking matches that of the robot.

The field of view of the web camera was adjusted to face the back of the head of the robot, based on the results of a pre-experiment that tested our interface. The results suggest that the sense of agency would decrease when the operator cannot see the movement of the robot. A headband was set on the head of the robot as a mark, so that the operator can easily understand which direction the robot is facing. Because the operator cannot directly see the eye movement of the robot, its vision line is shown as the yellow arrow. The arrow then moves to follow the green circle. When the view of the operator begins to change from one partner to the other (the operator's eye movement is approximately 970 pixels), the delay before the vision line of the robot reaches its conversation partner is 0.5 s, and the delay before the movement of the robot head is completed is 1.5 s.

When the robot autonomously adjusts its view, the interface guides the operator's eye movement to the view point of the robot by moving the green circle. We considered that the operator will feel the illusion that the operator's teleoperation is responsible for the autonomous head movement, if the interface convinces her to believe that this induced gaze is her involuntary eye movement. It might be difficult to provide such an illusion through visual guidance alone. The movement of the green circle indicates the operator's eye movement, and so there is a possibility that the operator does not notice the autonomous movement of the circle mixed into the teleoperation.

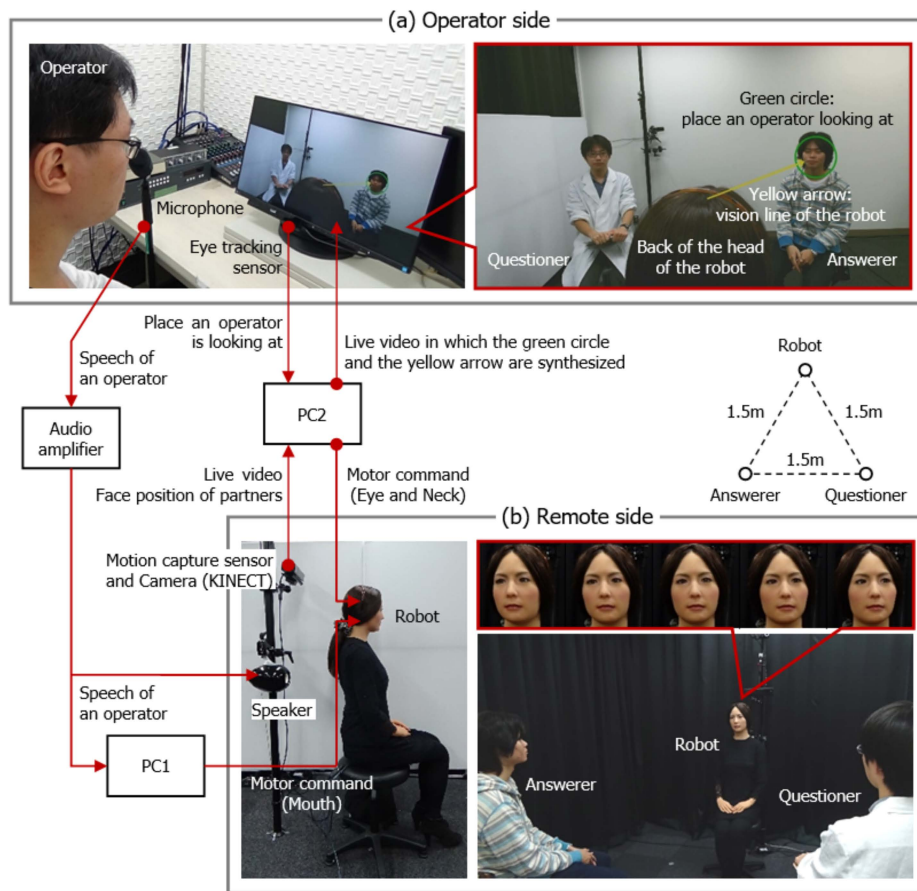


Figure 1. Experimental setup of the operator and remote sides.

3. Related Work

3.1. Robot Conferencing

Previous studies have reported several advantages that robot conferencing has over videoconferencing. First, it is understood that the physical presence of a robot enhances social telepresence [4,10] and clarifies the operator’s gaze [2]. However, it is disadvantageous when a robot cannot directly transmit the appearance of the operator. In order to overcome this disadvantage, the systems that combine an operator’s video feed and a partial robot have been developed. Some studies have proposed systems that physically move the face of an operator in order to clarify facial direction [11–13]. More recent studies have proposed systems that attach a robotic hand to the video feed of an operator in order to point to a remote space [14], or to reproduce handshaking [15]. Compared to these approaches, this study adopts a style of robot conferencing in which the robot does not transmit the appearance of the operator. Instead, we propose a semi-autonomous robot conferencing system that can add autonomous behaviors to the teleoperation.

3.2. Semi-Autonomous Robot

There are many teleoperation methods available to control a humanoid robot [16]. In the method focusing on social telepresence, a robot reproduces the body motions of the operator using master–slave teleoperation, and a head-mounted device reproduces the audio and visual information that are transmitted by a binaural microphone and a stereo camera [17]. The purpose of this method is to make the operator feel as if the robot body is their own by reproducing its body motion and the sensations as accurately as possible. Therefore, this method is not suitable for the incorporation of autonomous

behavior. Thus, we developed the semi-autonomous teleoperation system that guides the teleoperation (i.e., operator's eye movement) according to autonomous behavior.

Several studies have already proposed semi-autonomous systems. The first study involved the development of a handcart-type robot that supported walking motion for elderly people. The results of the experiment reported that the user can obtain a sense of handcart control by adjusting the autonomy of the robot to the walking ability of the user [18]. A study involving teleoperated robots developed a system that allows for autonomous robots to attend to a task, but also allows for an operator to step in and control the robots when they face a difficult problem [19]. This system intends for a small number of operators to control a large number of robots to provide services simultaneously. In this case, the social telepresence and the sense of agency may be not so important. Furthermore, semi-autonomous systems have been proposed for the automatic avoidance of obstacles in the teleoperation of a mobile robot [20], the reduction of an operator's effort in robot-assisted therapy [21], and the adjustment of the differences of movements between an operator and a robot caused by differences in their physicality [22]. These studies have not clarified whether the proposed semi-autonomous systems can maintain the social telepresence and sense of agency. In addition, in the teleoperation of a humanoid robot, the reproduction of physiological movements (e.g., respiration and body oscillation) are typically autonomously reproduced [23], because reproducing such small involuntary movements seems not to be effective for the sense of agency.

As described above, it has been hardly reported heretofore that the sense of agency is maintained, even if a conscious movement related to a conversation is autonomously reproduced. A recent study proposed a semi-autonomous system avoiding collisions between teleoperation and autonomous behaviors so as to reduce the frustration of an operator as a result of the collisions [24]. This previous study is closely related to ours, but the approach to avoiding the collisions is different. While it adjusts the priority of autonomous behaviors according to teleoperation, our study guides teleoperation according to the autonomous behavior. The approach of the previous study may not contribute to the sense of agency, because the execution of the autonomous behavior is clear to the operator.

In this study, we adopted an autonomous behavioral pattern in which a robot pays attention to the face of a speaker. This action is performed so as to simulate the social behavior that is consciously performed in multi-party conversations. We investigated whether our proposed interface, which guides the operator's eye movement, provides the operator with the illusion that they are continuing controlling the robot.

4. Hypothesis

Several previous studies reported that being conscious of the discomfort of the motion delay [9], the appearance [25], the predictability of behaviors [26], and the synchrony with the virtual body [7,8] decreases the sense of agency. Therefore, it is a concern that the sense of agency may decrease when the operator notices autonomous behavior that differs from their eye movement.

On the other hand, many previous studies demonstrated that visuomotor synchrony and/or synchronous visuo-tactile stimulation (e.g., the rubber hand illusion technique) enhance the sense of agency. The illusion that a virtual body is one's own body occurs, even if the virtual body has a different age [27], size [27], gender [28], and race [29] from the real body. We expected that the synchronization between the guided eye movement and the autonomous vision line movement of the robot also enhances the sense of agency. We hence set the following hypotheses.

Hypothesis 1. *Guiding the operator's eye movement to synchronize with the autonomous vision line movement of the robot enhances the sense of agency.*

There is a previous study that attempted to produce the feeling of talking to an operator when talking to an autonomous robot [30]. This study mentioned that a psychological contrivance that blurs the line between teleoperated and autonomous states is effective at producing that feeling. We

considered that mixing the visual guidance into teleoperation that uses the operator's eye movement blurs the line between teleoperated and autonomous states, and maintains the sense of agency. Thus, we set the following hypothesis.

Hypothesis 2. *Even when the visual guidance mixed into the teleoperation that uses the operator's eye movement, the same level of the sense of agency is maintained as that which is seen in continuous teleoperation.*

It is known that social telepresence decreases depending on the inconvenience of teleoperation [31]. Therefore, it is considered that social telepresence will decrease when the operator is conscious that the robot is autonomously moving. However, along with the sense of agency, we expected that the visual guidance maintains social telepresence in our semi-autonomous robot conferencing. Thus, we added the following hypothesis.

Hypothesis 3. *Even when autonomous behavior is incorporated into teleoperated behavior, the visual guidance maintains the same level of social telepresence as that seen in continuous teleoperation.*

5. Experiment

5.1. Task

In this experiment, the subject spoke to two experimenters—a questioner and an answerer. The subjects and the experimenters had never met before the experiment. Before the conversation, the subjects were told that the experimenter (whose role is an answerer) is also a subject.

In the conversation, the questioner asked the answerer and the subject questions such as, "Please tell me what your favorite animal is and why?", and then they answered the question. They continued the conversation in this fashion and answered three more questions. The answerer used a script that was prepared ahead of time, but he responded as though he was answering for the first time. When the subject was answering the question, the two experimenters gave vocal backchannel responses and nodded. When the answerer was answering the question, the questioner gave the backchannel responses, but the subjects behaved freely. To establish the roles of the experimenters, the questioner wore a white coat. During the conversation, the number of turn-taking was nine times, and the number of responses provided by the experimenter was approximately twelve.

5.2. Setups

Figure 1 shows the experimental setup of the operator (i.e., subject) and the remote sides. On the operator side (a), our teleoperation interface, the mic capturing speech from the operator, and the speaker that reproduced the sound from the remote side, were positioned. The speaker was positioned behind the display.

On the remote side (b), the robot and two chairs for the experimenters were placed at the apices of an equilateral triangle with a side length of 1.5 m. The speaker that reproduced the speech of the operator and the motion capture sensor that acquired the facial positions of the experimenters were positioned behind the robot. The mic that acquired the sound from the remote side was positioned behind the experimenters.

As shown in the figure, the audio amplifiers and the Windows machines PC1 and PC2 connected the operator and remote sides. The eye, neck, and mouth motions of the robot were generated by pneumatic actuators. The control system of the robot adjusted the air pressure of the pneumatic actuators according to the motor command. The control system communicated with PC1 and PC2 by serial communication (RS232C). The microphones and speakers in the both sides were connected to the audio amplifiers (the speaker of the operator side and the mic of the remote side are omitted for simplicity). The mic of the operator side was also connected to the PC1 that generated the mouth motion (one degree-of-freedom) of the robot, based on the speech of the operator [32].

The eye tracking sensor, display, and motion capture sensor were connected to the PC2 that generated the eye and neck movements of the robot. The figure ((b) remote side) also shows an example of robot movement when the operator's view changes from one experimenter to the other. The robot reproduced the movement of the vision line of the operator through neck movement (pitch, roll, and yaw) and eye movement (yaw). PC2 synthesized the green circle and the yellow arrow, that is, the place where the operator is looking at, and the vision line of the robot, onto the live video, and it showed on the display.

5.3. Conditions

To confirm the three hypotheses, we set the following three conditions.

Teleoperated (tele) condition: The subject controlled the robot by using our interface. The robot reproduced the movement directed by the vision line of the subject at all times, so it did not move autonomously. This condition is hereafter referred to as the "tele" condition.

Autonomous (auto) condition: This condition also used the same interface as the tele condition, but the subject did not control the robot. The robot autonomously paid attention to the speaker. The rules for the autonomous behavior of the robot are defined as follows.

Rule 1: When the subject begins to answer the question, the robot looks at the face of the questioner.

Rule 2: When the questioner or the answerer begins to speak, the robot looks at their face.

Rule 3: When the answerer provides a backchannel response to the voice of the subject, the robot looks at the answerer's face for 1.0 s.

Rule 4: When the questioner provides a backchannel response to the voice of the answerer, the robot looks at the questioner's face for 1.0 s.

It goes without saying that looking at the face of someone when they begin to speak is normal social behavior. In addition to Rule 1, we considered that taking a glance at a person who provided a backchannel response is also reflective of social behavior. After Rules 3 and 4, the robot turned its attention to the face of the questioner or answerer, based on Rules 1 and 2, respectively. The recognition of the speaker can be automated by a microphone array, but the experimenter (who is different from the questioner or the answerer) executed these rules to prevent erroneous recognition.

In order to guide the subject's gaze toward the eye-gaze position of the robot, the green circle on the interface moved according to the autonomous vision line movement of the robot. This condition is hereafter referred to as the "auto" condition.

Semi-autonomous (semi) condition: Not only did the robot move according to the teleoperation of the subject, but also autonomously moved based on Rules 1–4. As in the auto condition, the green circle on the interface moved to synchronize with the autonomous vision line movement of the robot in order to guide the subject's gaze. After the autonomous behavior was performed, the teleoperation was immediately enabled. In this condition, the duration of all of the autonomous movement was 1.5 s (the time it takes to complete the robot eye-movement (0.5 s) + its gaze time (1.0 s)), based on Rules 1–4. This condition is hereafter referred to as the "semi" condition.

5.4. Experimental Design

To confirm the hypotheses, we compared the tele condition to the auto- or semi-conditions. Hypothesis 1: Determine whether the auto condition can produce the same level of sense of agency as the tele condition. Hypotheses 2 and 3: Determine whether the semi condition can maintain the sense of agency and social telepresence, compared to the tele condition.

The pre-experiment indicated that the difference between the tele- and semi-conditions was hardly recognizable compared to the difference between the tele- and auto-conditions. If the tele-, auto-, and semi-conditions are directly compared, there is a possibility that the subjects will focus on only the differences between the auto condition and the two other conditions. For this reason, a comparison of

the conditions was separately conducted; Experiment 1 compared the tele- and auto-conditions, and Experiment 2 compared the tele- and semi-conditions.

5.5. Measures

5.5.1. Gaze Rate

To determine the degree to which the interface succeeded at guiding the subject’s eye movement, we calculated the rate at which the subject looked at the face of the target (i.e., the questioner and answerer) when turn-taking or a back-channel response occurs. In this study, this rate is called the gaze rate. When turn-taking occurs (Rules 1 and 2), looking at the face of the speaker is a natural behavior, so the subjects might look at the face without visual guidance. Compared with this, when someone gave a back-channel response (Rules 3 and 4), looking at their face is not always done. Therefore, we separately calculated the gaze rate in these timings.

Figure 2 shows an example of the movement of the view point of one subject at the semi-condition for the first 60 s, and the method to calculate the gaze rate. The blue and red lines represent the horizontal movement of the view point and the autonomous movement of the green circle, that is, visual guidance. The dotted lines show the timing of turn-taking (TT) and the back-channel response (BR). First, we counted the number of successful visual guidance. When the subject’s eye-gaze position was included in the circle with a diameter of 180 pixels centered on the facial position of the target within 1.5 s from the visual guidance, we determined that the visual guidance was successful. In the example of the figure, TT1 and TT2, and BR2 and BR4 satisfied this condition (i.e., had successful visual guidance). Then, the gaze rate was obtained by dividing this number by the total number of turn-taking or back-channel responses.

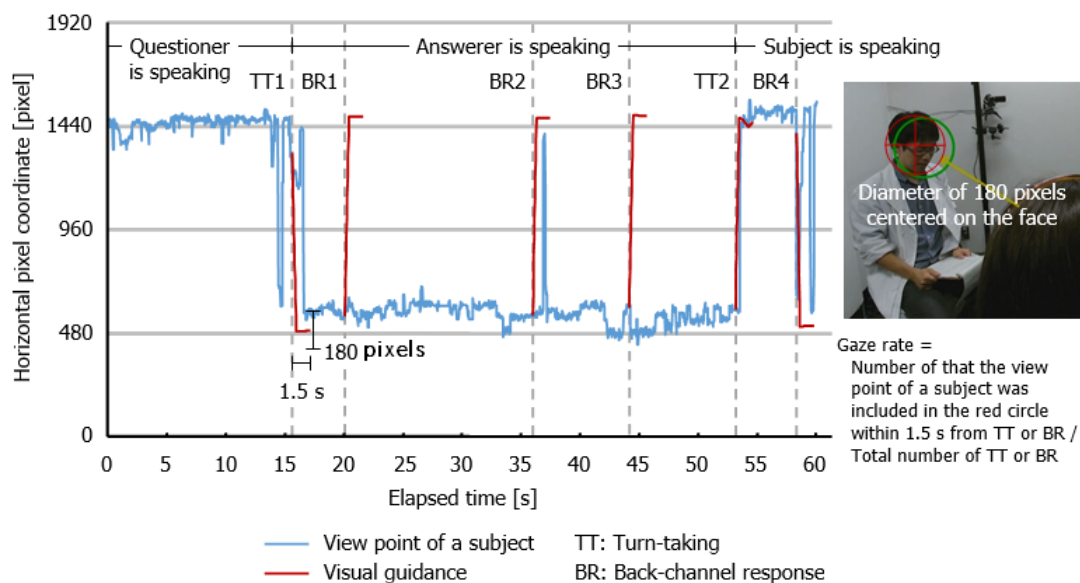


Figure 2. Example of the movement of view point at the semi condition, and the method to calculate the gaze rate.

5.5.2. Questionnaire

To estimate the following operator feelings, the subjects were asked to complete a questionnaire. Their answers were rated on a nine-point Likert scale, as follows: 1 = strongly disagree, 5 = neutral, 9 = strongly agree.

Sense of agency: In reference to previous studies, we set the following two items.

- I felt as if I were controlling the head of the robot by myself.
- I felt as if the movement of the robot head were my own.

Social telepresence: Because determining the feelings of being in the same room is useful for measuring the social telepresence of the remote conversation partner [8,12,13,29], we used the same representation to measure the social telepresence and set the following two items.

- I felt as if I were listening to the partners in the same room.
- I felt as if I were speaking to the partners in the same room.

Comfort level: In addition to the above feelings, we measured how much the subjects felt comfortable during the robot conferencing. We expected that the subject might feel comfortable, as the autonomous behavior substitutes the subject's teleoperation, and set the following four items.

- I felt nervous while listening to the partners.
- I felt nervous while speaking to the partners.
- I was able to listen to the partners comfortably.
- I was able to speak to the partners comfortably.

As other indicators of comfort level, we recorded how often the subject remembered the partner's answer. We considered that substituting the teleoperation with autonomous behavior might enable the operator to focus on the conversation. The method used to conduct this memory test will be described in the procedure section.

Intention: When the operator did not notice the intervention of autonomous behavior during teleoperation, they may feel that the behavior is their intentions. Furthermore, when autonomous behavior was convenient for the subjects, they may hardly notice it. To confirm these predictions, we added the following two items.

- I felt that the robot head was moving according to my intentions.
- I felt that the robot head was looking in the appropriate direction, depending on the situation.

The questionnaire including these two items was completed after showing a recorded video of the conversation of the subject (the view angle is shown by the bottom right picture of Figure 1) was shown to them. This was done so that the subject could evaluate these feelings objectively.

5.6. Subjects

A total of 24 students participated in the experiments. Six females and six males participated in Experiment 1, and six different females and six different males participated in Experiment 2. All of the subjects were undergraduate students at our university. No one was previously acquainted with any of the experimenters, and no one participated in both experiments.

5.7. Procedure

First, we presented a live video feed of a remote site (the view angle is shown by the picture on the bottom right of Figure 1) to the subject, and explained that they would remotely control the head of the robot by using their eye movement. Then, we conducted the following procedure.

5.7.1. Calibration

To obtain the subject's eye movement through the interface as accurately as possible, two actions were carried out, namely: the calibration of the eye tracking sensor, and a test confirming that the acquired viewpoint fell within the 90-pixel error range. Subsequently, the subject was presented with a screen of the interface (as shown in the upper part of Figure 1) and was told that the green circle is their eye gaze position, and that the yellow arrow indicates the direction of the line of view of the robot.

5.7.2. Instructions

The subject was informed that they would have a practice conversation before the two test conversations. Before conducting the conversations, the subject was informed about how the robot head moves. For the practice conversation and the tele condition, the robot head moves to see where the subject is looking. For the auto condition, the robot head moves autonomously. For the semi condition, the robot head basically moves to see where the subject is looking, but sometimes it moves autonomously.

5.7.3. Conversation

The conversation task was conducted three times. A practice conversation was conducted first so as to practice the task and the teleoperation, and the conversation involved only one topic. The second and third conversations incorporated the experimental conditions that represented the tele- and auto-conditions in Experiment 1, or the tele- and semi-conditions in Experiment 2. The conversations consisted of three topics. We then prepared two groups of topics, as follows: group 1 included topics pertaining to animals, music, and movies, and group 2 included topics involving books, food, and sports. The order of the conditions, the order of the topic groups, and their combination were counterbalanced. The topic of the practice conversation was a historical personage every time. The conversation time for the practice and conditional conversations were approximately 1.5 min and 3 min, respectively.

5.7.4. Questionnaire

After the conversations, the subject was asked to complete the questionnaire in about 30 min. Subsequently, the subject took the memory test that measured how well they remembered the answers of the answerer. The memory test included seven questions. Examples of the types of animal questions used are, "What is his favorite animal?", "How many animals has he owned so far?", and "How old is his pet?".

After the memory test, the subject watched a recorded video of their own conversation, in which the robot they controlled was speaking to the experimenters, and then answered the second questionnaire. The second questionnaire was focused on intention.

Finally, the subjects were interviewed to discuss their answers. This interview was open-ended.

6. Results

The results of Experiments 1 and 2 are shown in Figures 3–7. Each box represents the mean values and each bar represents the standard error of the mean. Figure 3 compares the gaze rate by using a 2×2 , two-way repeated-measures analysis of variance (ANOVA) with "situation" (turn-taking vs. back-channel) and "condition" (tele vs. auto or semi) factors, followed by Tukey's HSD (honestly significant difference) correction. Figures 4–7 compare the scores of the questionnaire using a repeated-measures *t*-test. The following sections will report the results of the statistical test with effect sizes (*r*). For all items with statistically significant differences, their *r* values were larger than 0.5, and the values of the other items were less than 0.5. Generally, an *r* value larger than 0.5 is considered to be enough high.

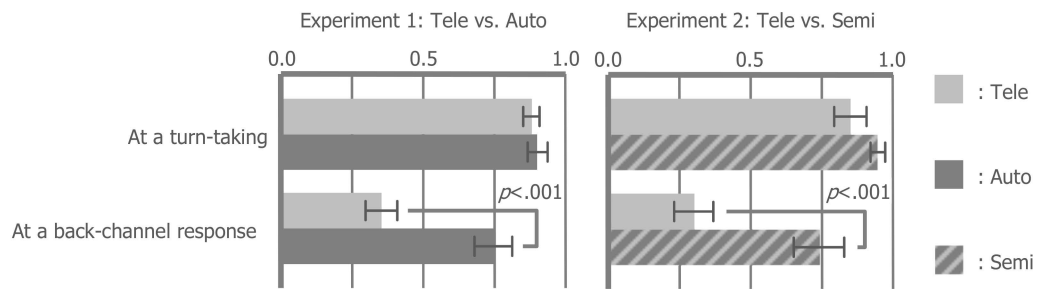


Figure 3. Results of gaze-rate measurement.

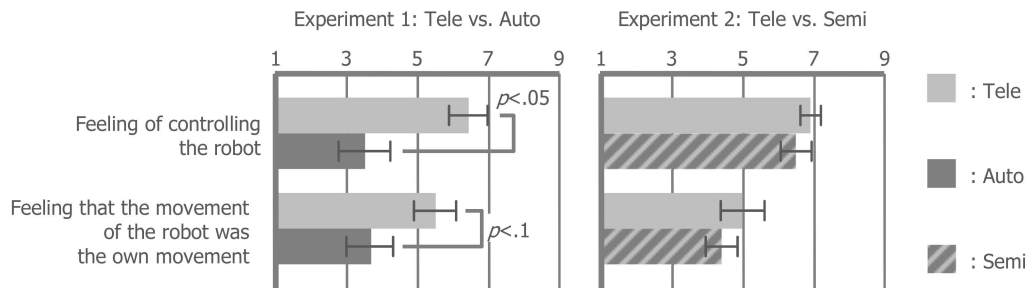


Figure 4. Results of measuring the sense of agency.

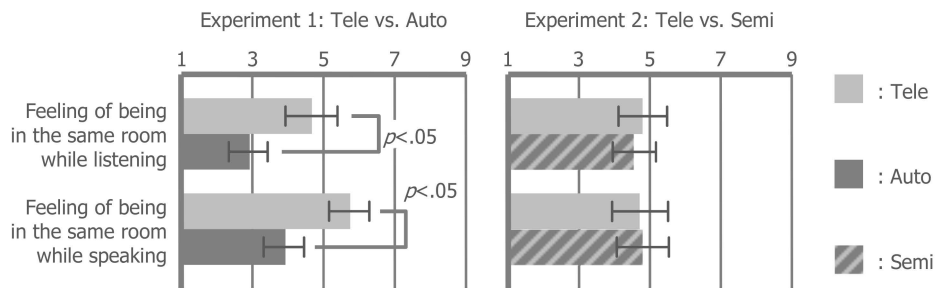


Figure 5. Results of measuring social telepresence.

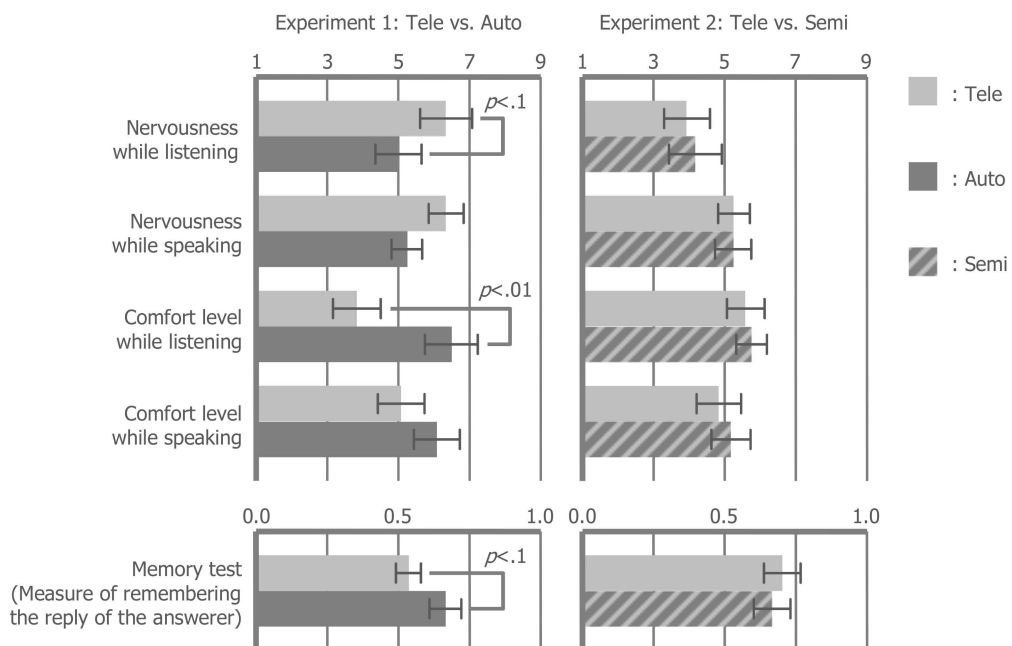


Figure 6. Results of measuring comfort level.

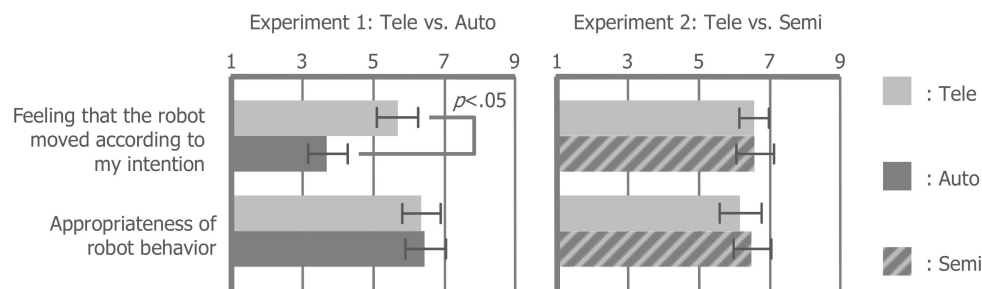


Figure 7. Results of measuring intention.

6.1. Gaze Rate

We found strong main effects of the situation factor (Experiment 1: $F(1, 11) = 64.400, p < 0.001, r = 0.924$; Experiment 2: $F(1, 11) = 34.930, p < 0.001, r = 0.872$) and the condition factor (Experiment 1: $F(1, 11) = 15.490, p < 0.01, r = 0.765$; Experiment 2: $F(1, 11) = 17.490, p < 0.01, r = 0.784$). We also found a strong interaction between these factors (Experiment 1: $F(1, 11) = 32.070, p < 0.001, r = 0.863$; Experiment 2: $F(1, 11) = 15.590, p < 0.01, r = 0.766$). We then analyzed the simple main effects in the interaction. During turn-taking, the auto- and tele-conditions were higher than the tele condition (Experiments 1 and 2: $p < 0.001$). In both of the tele conditions of Experiments 1 and 2, the gaze rate at a back-channel response was only approximately 30%. Compared with this value in the auto- and semi-conditions, the gaze rate increased up to 70%. This result indicates that the auto- and semi-conditions similarly succeeded in the visual guidance at a back-channel response. However, during turn-taking, the gaze rates of both of the tele conditions of Experiments 1 and 2 were already sufficiently high, and so the effect of visual guidance did not appear.

6.2. Sense of Agency

The feeling of controlling the robot decreased in the auto condition compared with the tele condition ($t(11) = 2.519, p < 0.05, r = 0.604$). Similarly, the feeling that the movement of the robot was the own movement tended to decrease in the auto condition compared with the tele condition ($t(11) = 1.795, p = 0.100, r = 0.476$). In contrast, these feelings did not statistically decrease in the semi condition compared to the tele condition, regardless of the intervention of autonomous behavior during teleoperation (controlling: $t(11) = 1.047, p = 0.318, r = 0.301$; own movement: $t(11) = 1.246, p = 0.239, r = 0.352$). These results mean that the visual guidance according to the autonomous vision line movement of the robot can maintain the sense of agency (Hypothesis 2) when mixing the autonomous vision line movement of the robot into the teleoperation. However, only the visual guidance is not effective for maintaining the sense of agency. Therefore, Hypothesis 1 was not supported.

6.3. Social Telepresence

The feeling of being in the same room with another person while listening ($t(11) = 3.022, p < 0.05, r = 0.674$) and speaking ($t(11) = 2.421, p < 0.05, r = 0.590$) decreased in the auto condition compared with the tele condition. As with the sense of agency, these feelings did not statistically decrease in the semi condition compared with the tele condition (listening: $t(11) = 1.393, p = 0.191, r = 0.387$; speaking: $t(11) = -0.561, p = 0.586, r = 0.167$). These results mean that robot conferencing performed through an autonomous robot decreases social telepresence, but that autonomous behavior mixed into teleoperation does not decrease social telepresence (Hypothesis 3).

6.4. Comfort Level

When listening to the partner speak, the auto condition was more comfortable ($t(11) = -3.545, p < 0.05, r = 0.730$) and less nervousness-inducing ($t(11) = 2.035, p = 0.06, r = 0.414$) than the tele condition. On the other hand, when speaking to the partners, there was no significant difference in

comfort level (listening: $t(11) = -0.456, p = 0.658, r = 0.136$; speaking: $t(11) = -0.938, p = 0.368, r = 0.272$) and nervousness (listening: $t(11) = -0.491, p = 0.633, r = 0.146$; speaking: $t(11) = 0.000, p = 1.000, r = 0.000$) between the auto- and tele-conditions. In addition, the result of the memory test showed that the subjects tended to remember the answers of the answerer in the auto condition better than in the tele condition ($t(11) = -2.030, p = 0.067, r = 0.522$). However, the difference between the tele- and semi-conditions was not significant ($t(11) = 0.583, p = 0.572, r = 0.173$). These results indicated that completely replacing a teleoperation with autonomous behavior improves the comfort level while listening to the partner speak, but this effect is not attained when an operator is speaking and when partially replacing a teleoperation.

6.5. Intention

The feeling that the robot moved according to the intention of the operator was decreased in the auto condition compared with the tele condition ($t(11) = 2.253, p < 0.05, r = 0.562$), but there was no significant difference between the semi- and tele-condition ($t(11) = 0.000, p = 1.000, r = 0.000$). This result means that visual guidance when incorporating the autonomous behavior of the robot into the teleoperation made the subject feel that the behavior reflected their intentions. In addition, there was no significant difference in the appropriateness of the behavior of the robot between the conditions (Experiment 1: $t(11) = -0.137, p = 0.894, r = 0.041$; Experiment 2: $t(11) = -0.411, p = 0.689, r = 0.123$), so the subjects felt that the autonomous behavior of the robot was as appropriate as their teleoperation. However, the interview indicated that the appropriateness of the behavior of the robot differs depending on the preference of the subject. This will be discussed in detail in the next section.

7. Discussion

7.1. Visual Guidance and Intention

The experiment indicated the probability that the semi-autonomous robot conferencing system can guide the subjects' eye movement without making them notice the visual guidance. In the auto- and semi-conditions, the degree of improvement of the gaze rate from that seen in the tele condition means the degree of success in the visual guidance. The subjects looked at the face of the experimenters without visual guidance when they began to speak, because looking at the face of a person while they are speaking is a natural social behavior. The effect of visual guidance was remarkable with regard to the backchannel response. The gaze rate was greater than 70% in both the auto- and semi-conditions, which is a 40% increase from the tele condition. In the auto condition, although the subjects' eye movement was guided to the green circle when there was visual guidance, it was clear for the subjects that the head movement of the robot was autonomous. In contrast, in the semi condition, one half of the subjects did not notice the visual guidance and the autonomous head movement of the robot despite being informed that the robot will change its line of view autonomously. Furthermore, the other half of the subjects noticed the visual guidance only when the robot moved based on Rule 4, and they noticed it only one or two times, even though it was conducted seven times in the conversation. Thus, no subject noticed the visual guidance when the robot moved based on Rules 1–3.

The opinions of the subjects appear to support these results. According to the interview regarding the intention items, almost all of the subjects thought that the following were natural social behaviors: looking at a questioner when the subjects answer (Rule 1), looking at a partner who is speaking (Rule 2), and taking a glance at a partner who gives a backchannel response to the subjects' speech (Rule 3). However, several subjects said that taking a glance at a partner who gives a backchannel response after another answerer spoke (Rule 4) is not necessarily a natural social behavior. Therefore, it is considered that when the behavior is acceptable for the operator, the teleoperation of an operator is easily guided to the autonomous behavior of a robot, unconsciously.

In the semi condition, most subjects felt that the head movement was their intentions, even after comparing the recorded videos of the tele- and semi-conditions. As described above, visual guidance

increased the gaze rate by the same degree in the auto- and semi-conditions. However, in the auto condition, the subjects felt that the head movement of the robot was different from their intentions. Thus, it is considered that mixing the visual guidance corresponding to the autonomous behavior into teleoperation makes the operator postdict that the behavior is their intention.

7.2. Sense of Agency

The semi condition seems to maintain the sense of agency, but the auto condition decreased it from the tele condition. As described in the previous section, the degree of success in visual guidance was at a similar level in both the auto- and semi-conditions, but the consciousness of the subjects during visual guidance tended to be higher in the auto condition. This consciousness might decrease the sense of agency. In the semi condition, the total intervention time of the autonomous behaviors was approximately 30 s (20 times of 1.5 s) in approximately 180 s of teleoperation, but it was difficult for the subjects to notice the intervention. We considered that the visual guidance made the subjects believe that their guided gaze is an involuntary eye movement. Consequently, this belief might blur the border between the teleoperated and autonomous states. However, in the auto condition, because of the absence of ambiguity between these states, visual guidance would not work effectively.

In the evaluation of the feeling of controlling the robot, the three subjects who noticed the visual guidance rated the semi condition lower than the tele condition. In the interview, they said that the behavior of the robot when reacting to a backchannel response to the speech of another answerer is unnatural every time. We thought that this is one of the social behaviors seen in multi-party conversation, but desirable social behaviors vary from person to person. If the autonomous behavior is not desirable to the operator, leading the teleoperation to this behavior would decrease the sense of agency.

In the interview on the appropriateness of the behavior of the robot, several positive opinions on the semi condition were obtained. Three of the twelve subjects said that the robot seemed to give attention to both of the partners at a more appropriate timing than in the other conditions. Furthermore, one subject said that she was shy with strangers, and so it was difficult to look at the face of her partner. However, she also said that the facial movement of the robot seemed to imitate her behavior when she talks to her close friends. Thus, there is a probability that the operator can behave desirably without decreasing the sense of agency by using the semi-autonomous robot conferencing system.

We expect that continuing semi-autonomous teleoperation makes the operator conduct desirable behavior even in a normal teleoperation and face-to-face conversation. Future work will examine this learning effect of the semi-autonomous robot conferencing system. Furthermore, examining the duration of the continuation of the sense of agency in the semi-autonomous robot conferencing is also a possible direction for future work.

In the experiment, as both the audio and the live video feed were directly played without going through the network, there was almost no perceptible deviation or delay. On the other hand, there was a delay in the neck movement of the robot (1.5 s), as described in the Section 2. However, none of the subjects pointed out the motion delay of the robot. This is considered to be because if the neck movement of the robot is completely synchronized with the eye movement of the operator, the neck movement is too fast. Basically, an asynchrony between the movements of an operator and a robot is harmful for the sense of agency. Therefore, there is a possibility that this asynchrony that the eye movement of the operator reproduced as the delayed neck movement of the robot decreased the sense of agency. In order to produce the higher sense of agency, it is necessary to apply our proposed method that guides the behavior of the operator to the teleoperation interface, which synchronizes the views of an operator and a robot using a head-mounted display. For example, guiding the head movement of the operator using the hanger reflex may be effective. The development of a semi-autonomous robot conferencing system that yields a higher sense of agency is for a future work.

7.3. Social Telepresence

Similar to the sense of agency, social telepresence was ranked lower in the auto condition than it was in the tele condition. The interview on the social telepresence items suggested that the consciousness of the subject, where the robot is autonomously behaving, decreased the social telepresence. Three of the twelve subjects commented that during the auto condition, they felt as if they were looking at the conversation occurring between the three people while standing behind the robot. Although the interface of the tele- and semi-conditions similarly presented the back of the head of the robot, none of the subjects had such comments for these conditions. This might mean that the subjects of the auto condition felt that the robot is not behaving as their avatar. In fact, the scoring of the feeling that the movement of the robot was the own movement was also ranked lower in the auto condition than in the tele condition.

In the evaluation of social telepresence, three of the twelve subjects rated the semi condition lower than the tele condition in either one of the social telepresence items (when speaking or listening). The reason for this is that the subjects mentioned that they noticed the several autonomous behaviors during the conversation. This also indicated that their awareness of the consciousness of the autonomy of the robot decreased the social telepresence. Therefore, desirable autonomous behavior accompanied by the visual guidance could maintain social telepresence as well as the sense of agency, because such behavior is hardly noticed in the semi-autonomous robot conferencing.

The social telepresence scores in the tele- and semi-conditions are approximately five, and so they were not high enough. Presenting the green circle, the yellow arrow, and the back of the head of the robot on our interface for visual guidance may decrease the sense of reality. If an interface guides teleoperation to the autonomous behavior of the robot by a more seamless method, social telepresence might be maintained. For example, the method using the hanger reflex described in the previous section may produce a higher level of social telepresence.

7.4. Comfort Level

The subject felt more comfortable in the auto condition than in the tele condition. In the interview, eight of the twelve subjects said that they could more easily focus on the voice of their partner in the auto condition, because they did not need to care about their attention as much as they did in the tele condition. In fact, the subjects tended to more clearly remember the speech of their partner in the auto condition than in the tele condition. These results mean that the operator feels more comfortable eliminating the cognitive load imposed by teleoperation when listening. However, this effect did not appear while the subject was speaking, or in the semi condition.

In the semi condition, the subjects hardly noticed the timing for when the teleoperation was replaced with autonomous behavior, and so the effect of reducing the cognitive load of teleoperation was not obtained. Furthermore, while the subject was speaking, the subjects seemed to focus on thinking about what to talk about rather than where to look. For this reason, the operator's cognitive load while they were speaking was not improved, but replacing teleoperation with autonomous behavior is useful for them to focus on thinking.

It is expected that robot conferencing would become more useful if the autonomy of the robot is adjusted according to the intentions of the operator or the needs of the situation. For example, when the operator speaks, they may place emphasis on social telepresence and the sense of agency, which places the robot in a semi-autonomous state. When the conversation partner speaks, they will focus on the speech of the partner and change the robot into a fully autonomous state.

8. Conclusions

This study aimed to provide an operator with the illusion that the autonomous behavior of a robot is directed by their teleoperation in semi-autonomous robot conferencing. For this purpose, we developed an interface that allows an operator to control the gaze of a robot using their eye movement.

The interface guides the operator's eye movement toward the autonomous gaze movement of a robot. Because human eyes move voluntarily and involuntarily, we expected that the operator would believe that the guided eye movement is actually their own involuntary eye movement, and feel that they are continuing the teleoperation, regardless of the interruption of the autonomous gaze movement.

As a result of the experiment, in which subjects performed robot conferencing through our interface, we found several conditions under which to provide the illusion. (1) When mixing the autonomous behavior into the teleoperation, and (2) when the autonomous behavior is acceptable for the operator, the visual guidance facilitates the illusion. (3) However, visual guidance could not work when the robot autonomously moves at all times. The illusion makes the operator hardly notice the autonomous behavior, and therefore the sense of agency and social telepresence are maintained.

We also confirmed the possibility that replacing teleoperation with autonomous behavior decreases the cognitive load of the operator with regard to social behavior. Therefore, the operator could focus on the speech of the partner. This effect is obtained in the conversation through a full-autonomous robot, but not in semi-autonomous robot conferencing.

For these findings, we considered that adjusting the autonomy of the robot according to the intentions of the operator would improve the usefulness of semi-autonomous robot conferencing. The semi-autonomous state will be used so that the operator can behave more socially while obtaining the sense of agency and social telepresence, and it will be changed to the fully autonomous state when the operator wants to focus on listening to their partner. This idea might be also applied to when adding task oriented autonomous behaviors. In the case of a lecturer robot, the teleoperation interface detects students who are interested and/or bored with the lecture, and guides the gaze of the remote lecturer to pay attention to them without decreasing the sense of agency and social telepresence. If the lecturer needs to listen to and evaluate the presentation of a student, the full autonomous listening behaviors of the robot would be useful for decreasing the cognitive load of the lecturer.

This study argues the possibility that adopting autonomous behaviors to a teleoperated robot by our proposed method increases its usefulness, without decreasing its original benefits. We hope that this study diminishes the distance between the social robotics and telerobotics fields, and facilitates in the study of semi-autonomous robot conferencing.

Author Contributions: Conceptualization, K.O., Y.Y., S.N., and H.I.; data curation, K.T. (Kazuaki Tanaka); funding acquisition, K.O., Y.Y., and H.I.; investigation, K.T. (Kazuaki Tanaka) and K.T. (Kota Takenouchi); methodology, K.T. (Kazuaki Tanaka), K.O., Y.Y., S.N., and H.I.; project administration, Y.Y. and H.I.; resources, K.O., Y.Y., and H.I.; software, K.T. (Kazuaki Tanaka) and K.T. (Kota Takenouchi); supervision, H.I.; validation, K.T. (Kazuaki Tanaka); visualization, K.T. (Kota Takenouchi); writing (original draft), K.T. (Kazuaki Tanaka) and K.T. (Kota Takenouchi).

Funding: This work was supported by JSPS KAKENHI, grant number JP25220004.

Acknowledgments: We thank Naoki Koyama and Kodai Shatani for helping the experiment of this study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hashimoto, T.; Kato, N.; Kobayashi, H. Development of educational system with the android robot SAYA and evaluation. *Adv. Robot. Syst.* **2011**, *8*, 51–61. [[CrossRef](#)]
2. Morita, T.; Mase, K.; Hirano, Y.; Kajita, S. Reciprocal attentive communication in remote meeting with a humanoid robot. In Proceedings of the 9th International Conference on Multimodal Interfaces, New York, NY, USA, 12–15 November 2007; pp. 228–235.
3. Ogawa, K.; Nishio, S.; Koda, K.; Balistreri, G.; Watanabe, T.; Ishiguro, H. Exploring the natural reaction of young and aged person with telenoid in a real world. *Adv. Comput. Intell. Inform.* **2011**, *15*, 592–597. [[CrossRef](#)]
4. Sakamoto, D.; Kanda, T.; Ono, T.; Ishiguro, H.; Hagita, N. Android as a telecommunication medium with a human-like presence. In Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Arlington, VA, USA, 9–11 March 2007; pp. 193–200.

5. Finn, K.E.; Sellen, A.J.; Wilbur, S.B. *Video-Mediated Communication*; Lawrence Erlbaum Associates Inc.: Mahwah, NJ, USA, 1997.
6. Alimardani, M.; Nishio, S.; Ishiguro, H. Humanlike robot hands controlled by brain activity arouse illusion of ownership in operators. *Sci. Rep.* **2013**, *3*, 2396. [[CrossRef](#)] [[PubMed](#)]
7. Banakou, D.; Slater, M. Body Ownership Causes Illusory Self-Attribution of Speaking and Influences Subsequent Real Speaking. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 17678–17683. [[CrossRef](#)] [[PubMed](#)]
8. Kalckert, A.; Ehrsson, H.H. Moving a rubber hand that feels like your own: A dissociation of ownership and agency. *Front. Hum. Neurosci.* **2012**, *6*, 40. [[CrossRef](#)] [[PubMed](#)]
9. Ogawa, K.; Taura, K.; Nishio, S.; Ishiguro, H. Effect of perspective change in body ownership transfer to teleoperated android robot. In Proceedings of the 2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication, Paris, France, 9–13 September 2012; pp. 1072–1077.
10. Tanaka, K.; Nakanishi, H.; Ishiguro, H. Physical embodiment can produce robot operator’s pseudo presence. *Front. ICT* **2015**, *2*, 8. [[CrossRef](#)]
11. Misawa, K.; Ishiguro, Y.; Rekimoto, J. LiveMask: A Telepresence surrogate system with a face-shaped screen for supporting nonverbal communication. *Proc. AVI2012* **2012**, *8*, 394–397.
12. Sirkin, D.; Ju, W. Consistency in physical and on-screen action improves perceptions of telepresence robots. In Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, Boston, MA, USA, 5–8 March 2012; pp. 57–64.
13. Tachi, S.; Kawakami, N.; Nii, H.; Watanabe, K.; Minamizawa, K. Telesarphone: Mutual telexistence master-slave communication system based on retroreflective projection technology. *SICE J. Control Meas. Syst. Integr.* **2008**, *1*, 335–344. [[CrossRef](#)]
14. Onishi, Y.; Tanaka, K.; Nakanishi, H. Embodiment of Video-mediated Communication Enhances Social Telepresence. In Proceedings of the Fourth International Conference on Human Agent Interaction ACM, Biopolis, Singapore, 4–6 October 2016; pp. 171–178.
15. Nakanishi, H.; Tanaka, K.; Wada, Y. Remote Handshaking: Touch Enhances Video-Mediated Social Telepresence. In Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 2143–2152.
16. Goodrich, M.A.; Jacob, W.C.; Emilia, B. Teleoperation and beyond for assistive humanoid robots. *Rev. Hum. Factors Ergon.* **2013**, *9*, 175–226. [[CrossRef](#)]
17. Fernando, C.L.; Furukawa, M.; Kurogi, T.; Kamuro, S.; Minamizawa, K.; Tachi, S. Design of TELESAR V for transferring bodily consciousness in telexistence. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura, Portugal, 7–12 October 2012; pp. 5112–5118.
18. Wasson, G.; Gunderson, J.; Graves, S.; Felder, R. An assistive robotic agent for pedestrian mobility. In Proceedings of the Fifth International Conference on Autonomous Agents, Montreal, QC, Canada, 28 May–1 June 2001; pp. 169–173.
19. Shiomi, M.; Sakamoto, D.; Kanda, T.; Ishi, C.T.; Ishiguro, H.; Hagita, N.A. Semi-autonomous communication robot: A field trial at a train station. In Proceedings of the 2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Amsterdam, The Netherlands, 12–15 March 2008; pp. 303–310.
20. Storms, J.; Tilbury, D. A New Difficulty Index for Teleoperated Robots Driving through Obstacles. *J. Intell. Robot. Syst.* **2018**, *90*, 147–160. [[CrossRef](#)]
21. Zarak, A.; Wood, L.; Robins, B.; Dautenhahn, K. Development of a semi-autonomous robotic system to assist children with autism in developing visual perspective taking skills. In Proceedings of the 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), Nanjing, China, 27–31 August 2018.
22. Muratore, L.; Laurenzi, A.; Hoffman, E.M.; Baccelliere, L.; Kashiri, N.; Caldwell, S.G.; Tsagarakis, N.G. Enhanced tele-interaction in unknown environments using semi-autonomous motion and impedance regulation principles. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 21–25 May 2018.
23. Lee, J.K.; Stiehl, W.D.; Toscano, R.L.; Breazeal, C. Semi-autonomous robot avatar as a medium for family communication and education. *Adv. Robot.* **2009**, *23*, 1925–1949. [[CrossRef](#)]

24. Okuoka, K.; Takimoto, Y.; Osawa, M.; Imai, M. Semi-autonomous telepresence robot for adaptively switching operation using inhibition and disinhibition mechanism. In Proceedings of the 6th International Conference on Human-Agent Interaction, Southamton, UK, 15–18 December 2018; pp. 167–175.
25. Tsakiris, M.; Carpenter, K.; James, D.; Fotopoulou, A. Hands only illusion: Multisensory integration elicits sense of ownership for body parts but not for non-corporeal objects. *Exp. Brain Res.* **2010**, *204*, 343–352. [[CrossRef](#)] [[PubMed](#)]
26. Hamasaki, S.; An, Q.; Murabayashi, M.; Tamura, Y.; Yamakawa, H.; Yamashita, A.; Asama, H. Evaluation of the Effect of Prime Stimulus on Sense of Agency in Stop Operation of the Object in Circular Motion. *J. Adv. Comput. Intell. Intell. Inform.* **2017**, *21*, 1161–1171. [[CrossRef](#)]
27. Banakou, D.; Groten, R.; Slater, M. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 12846–12851. [[CrossRef](#)] [[PubMed](#)]
28. Slater, M.; Spanlang, B.; Sanchez-Vives, M.V.; Blanke, O. First Person Experience of Body Transfer in Virtual Reality. *PLoS ONE* **2010**, *5*, e10564. [[CrossRef](#)] [[PubMed](#)]
29. Kilteni, K.; Bergstrom, I.; Slater, M. Drumming in immersive virtual reality: The body shapes the way we play. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 597–605. [[CrossRef](#)] [[PubMed](#)]
30. Tanaka, K.; Yamashita, N.; Nakanishi, H.; Ishiguro, H. Teleoperated or Autonomous? How to produce a robot operator's pseudo presence in HRI. In Proceedings of the The Eleventh ACM/IEEE International Conference on Human Robot Interaction, Christchurch, New Zeland, 7–10 March 2016; pp. 133–140.
31. Riley, J.M.; Kaber, D.B.; Draper, J.V. Situation Awareness and Attention Allocation Measures for Quantifying Telepresence Experiences in Teleoperation. *Hum. Factors Ergon. Manuf.* **2004**, *14*, 51–67. [[CrossRef](#)]
32. Ishi, C.; Liu, C.; Ishiguro, H.; Hagita, N. Speech-driven lip motion generation for tele-operated humanoid robots. In Proceedings of the Auditory-Visual Speech Processing, Volterra, Italy, 1–2 September 2011; pp. 131–135.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).