



Having Different Dialog Roles in Telecommunication by Using Two Teleoperated Robots Reduces an Operator's Guilt

Reina Nozawa¹ · Kazuki Sakai¹ · Megumi Kawata¹ · Hiroshi Ishiguro¹ · Yuichiro Yoshikawa¹

Accepted: 11 March 2024 / Published online: 10 April 2024
© The Author(s) 2024

Abstract

In recent years, applications of social robots as the operator's avatar have been widely studied for remote conversation with rich nonverbal information. Having another side-participant robot beside the avatar robot of the operator was found to be effective for providing long-lasting backchannels to the interlocutor. The side-participant robot is also expected to play a role in assisting human participation in multiparty conversations. However, such a focus has not been applied to remote conversations with multiple robots. Here, we propose a multiple-robot telecommunication system with which the operator can use a side-participant robot to assist conversation that is developed by the operator through the main speaker robot to verify its effectiveness. In the laboratory experiment where the subjects were made to feel stressed by being forced to provide rude questions to the interlocutor, the proposed system was shown to reduce guilt and to improve the overall mood of operators. The result encourages the application of a multi robot remote conversation system to allow the user to participate in remote conversations with less anxiety of potential failure in maintaining the conversation.

Keywords Multiple robots · Tele-communication robot · Cybernetic avatar · Tele-communication support · Assisting robot operator

1 Introduction

In recent years, there has been an increase in remote communication between different locations due to various factors, including globalization. However, conventional telecommunication devices such as telephones and videoconferencing systems do not sufficiently convey the social presence [6] of the remote speaker (i.e., the sense of being together with the remote speaker) because nonverbal information such as a gaze to establish eye contact is transmitted in a different

format from that in face-to-face communication [32] and because the physical space cannot be shared [25]. Social presence has been the focus of studies of computer-mediated communication [9, 24], and the lack of social presence is known to decrease the willingness to continue the dialog [43] and social conformity [28]. In contrast, several studies have been conducted by using robots as communication media that can transmit nonverbal information and can allow their operator to share physical space with their interlocutors via its body [4, 35, 40]. Takeuchi et al. [40] reported that a physically handicapped person may serve customers from his or her home by operating an avatar robot. Veronica et al. [35] also reported that students who are unable to physically attend school due to illness may participate in the classroom by operating a robot.

However, it is difficult for an operator to operate a robot rapidly and precisely to sufficiently convey desired nonverbal information due to the limitations in its degree of freedom and human skill. Therefore, methods that allow the robot to be semiautonomously operated have been studied. Ishii et al. [23] and Sakai et al. [36] developed a semiautonomous robot that can autonomously produce its lips and head movements based on the operator's voice input. However, due to

✉ Yuichiro Yoshikawa
yoshikawa@irl.sys.es.osaka-u.ac.jp

Reina Nozawa
nozawa.reina@irl.sys.es.osaka-u.ac.jp

Kazuki Sakai
sakai.kazuki@irl.sys.es.osaka-u.ac.jp

Megumi Kawata
kawata@irl.sys.es.osaka-u.ac.jp

Hiroshi Ishiguro
ishiguro@sys.es.osaka-u.ac.jp

¹ Graduate School of Engineering Science, Osaka University,
Toyonaka, Japan

sensing errors or other factors, semiautomatic teleoperated robots sometimes generate actions that look unrelated to the conversation and interfere with it. Furthermore, it is possible that the movement directly given by the operator and the one automatically generated by the system may interfere with each other. On the other hand, Arimoto et al. [2] improved the social presence of the tele-operator by introducing another autonomous robot to be a side-participant of the conversation and by making it autonomously produce necessary nonverbal information. The adoption of multiple robots resolved the interference between movements generated by the operator and the system by assigning the role of responding to the interlocutor only to the side-participant robot.

Even in a face-to-face conversation where participants can feel the interlocutor's social presence, just saying something is often not an easy task in some situations, such as cross-cultural conversations [41] and group conversations [10]. In other words, improving social presence cannot be a sufficient solution to support communication. In social psychology, the communication participation style scale (COMPASS) [12] was developed as a scale of communication participation style, i.e., how people usually participate in communication, and is used to evaluate participants' attitudes in collaborative learning [34]. The scale contains four factors: passive participation, which shows a tendency not to engage in the conversation; receptive participation, which shows a tendency to listen to others and to try to understand what they are saying; conversation management, which shows a tendency to manage and spark conversation like the chairperson; and proactive participation, which shows a tendency to describe ideas without hesitation and to claim their opinions. There are some studies in which robots or agents were implemented to have the style of conversation management to assist communication [7, 22, 31, 39, 45]. Matsuyama et al. [31] showed that the behavior of facilitation robots can be accepted as a participant's behavior and can produce a feeling of group connectedness. Yamaguchi et al. [45] also showed that a presenting robot can stimulate conversation by introducing a funny story. However, in their study, the robot was controlled by a Wizard of Oz (WoZ) method, and its speech contents were limited to one implemented in advance. On the other hand, it is not clear how a robot can assist in conversation with the style of proactive participation.

In this study, therefore, we propose a remote dialog system that allows an operator to have different dialog roles by using two tele-operated robots: One robot is used for conveying the operator's voice and the other for assisting conversation. Mehmood et al. [29] proposed the use of multiple teleoperated robots to alternately convey the operator's utterance and to instill in the operator stronger perceptions of the right to talk and social support. Although the two robots employed in the previous study shared the speaker role to produce the utterance of the operator, the second robot in

the current study functions as an assistive robot that participates in conversation semiautonomously and proactively with the goal of helping the operator produce the utterance via the main robot. In the proposed system, specifically, two roles are shared by two robots: A teleoperated speaker robot is operated by the operator's voice to speak freely, and a teleoperated side-participant robot is operated by buttons for proactive participation to assist the operator in easily speaking. Therefore, we developed an interface with functions for supporting a single operator to simultaneously operate two robots with different dialog roles. Then, to show that the proposed system improves the operator's mood during a conversation, an experiment was conducted in which the subject was asked to be an operator conversing in a situation in which he or she easily feels guilty or stressed. We then showed that the proposed system decreases the operator's guilt and stress during the conversation.

2 Related Work

2.1 Conversation Assistance Using Robots and Agents

In previous studies, robots and agents have been developed that make conversation smooth. For example, Matsuyama et al. [31] developed a framework for facilitation robots that regulate imbalanced engagement density in a four-participant conversation with proper procedures for obtaining initiatives. The robot's behavior based on the framework showed evidence of acceptability for a participant's behavior and a feeling of group connectedness. Short et al. [39] developed a robot that moderates a collaborative assembly game and showed that the more the robot spoke to participants, the higher the group cohesion they reported, and the more they helped the other participants in the group. In addition, Birmingham et al. [7] reported that a robot can improve trust among strangers by asking questions that encouraged participants to share with one another. Isbister et al. [22] also developed an agent that suggests nonsensitive topics for cross-cultural conversation when it breaks off and confirmed that it positively affected impressions of typical Japanese people for American people. All of these robots and agents assisted in conversation through the participation style of conversation management. This differs from our study, which aims to realize conversational assistance through the proactive participation of robots.

2.2 Multiple Robots

Previous studies on human-robot interactions have shown good points of cooperation among multiple autonomous robots [3, 20, 21, 37, 38, 42]. The first merit of cooperat-

ing with multiple robots is increasing the impact on people. Sakamoto et al. [37] reported that two robots attracted more people than did a single robot in open public environments, which was conceptualized as the advantage of a passive social medium. Shiomi et al. [38] also reported that multiple robots can give their human interlocutor stronger pressure than that of a single robot. The second merit is improving the user's impression of a conversation. Arimoto et al. [3] reported that people felt that it was easier to talk with a dialog system that alternately made two robots talk to a human interlocutor than it was with one that talked only with a single robot. Iio et al. [20, 21] developed a dialog system that can maintain conversational consistency against speech recognition failure by making two robots cooperatively respond to human interlocutors.

On the other hand, studies have also been conducted on multiple teleoperated robots. Glas et al. [16] developed a system in which robots automatically talk in noncritical sections and are controlled in critical sections. It allows one operator to control multiple robots in different places simultaneously. However, it was not for facilitating conversations by using multiple robots. Arimoto et al. [2] focused on enhancing the social presence of the teleoperator to improve the conversation. They developed a telecommunication system not only with a proxy robot that conveys the operator's utterances but also with another autonomous bystander robot that gives vague utterances. However, the bystander robot only gives back-channel feedback such as "oh" or "uh". Although Arimoto used a teleoperated and an autonomous robot, in this study, we use two teleoperated robots. Recently, Mehmood et al. proposed the use of multiple teleoperated robots, either of which can be randomly chosen to serve as the main robot that conveys the operator's words; these authors also highlighted the effect of this approach on facilitating conversation in terms of the operator's perceptions of the right to talk and social support [29]. However, the choice of robots was made randomly and independently of the context of the dialog. In other words, the second robot was not designed to assist the main robot but rather to share the speaker role with the main robot. In this paper, on the other hand, we propose a system in which the second robot plays the role of a side-participant that sometimes explicitly assists the operator in developing the dialog via the main robot.

3 Proposed System

3.1 System Configuration

The proposed system enables a remote operator to operate two robots, that is, a main robot and an assistance robot, simultaneously, each with different dialog roles. The main robot produces the operator's voice after it has been processed

with voice changer software, while the assistance robot produces actions to support conversation, which is driven by the operator's button input (see Fig. 1).

These robots are controlled by a microphone and GUI with three software blocks. The blocks consist of the "voice recognizer", which receives the operator's and interlocutor's voices for speech recognition and voice activity detection, the "utterance recommender," which receives voice recognition results of the operator and the interlocutor to offer recommendations for the operator to choose the next utterance for the assistance robot, and the "multiple-robots controller," which controls the robot's utterances and coordinates two robots' behavior so as to make them work collaboratively.

The utterance recommender generates candidates for the assistance robot's utterances to either the interlocutor or the main robot. The GUI receives the candidates and updates the buttons to trigger them. When the operator selects the assistance robot's speech content and the speech addressee (main robot or interlocutor) by clicking on a preferable button generated on the GUI, the selected information is sent to the multiple-robot controller.

The main robot's utterance controller in the multiple-robot controller converts the operator's voice into robot-like speech by elevating its pitch, which is produced through a loudspeaker placed behind the main robot. Furthermore, in synchrony with the operator's speech voice, the main robot opens and closes its mouth and moves its arms up and down, such that the produced voice is perceived as being generated by the main robot. For such synchronization, the mouth and arm movements are generated based on vowels extracted from the operator's speech [23]. The assistance robot's utterance controller receives the text to be uttered by the assistance robot from the GUI and synthesizes voice sound. The synthesized voice is produced by the built-in speaker of the assistance robot, while its mouth and arm movements are generated using the same method employed by the main robot.

The question of how to implement social gaze has been the focus of research in the field of human-robot interaction (for a review, see [1]). It is well known that gaze has the ability to regulate the exchange and maintenance of speaker roles [27]. Participants playing the speaker and listener roles spend a large amount of time looking at the listener and the speaker, respectively [44]. These human tendencies have been used to generate appropriate gaze behavior for a speaker robot [33]. The attentional target of a side-participant robot has also been modeled based on the roles played by participants [48]. In accordance with previous studies, we implemented the multiple-robot coordinator to generate the gaze of both robots and the nodding behavior of the assistance robot so that both robots look as if they are paying attention naturally in the conversation. As in a previous study [2], if either robot speaks, the gaze is directed toward the speaker robot, while

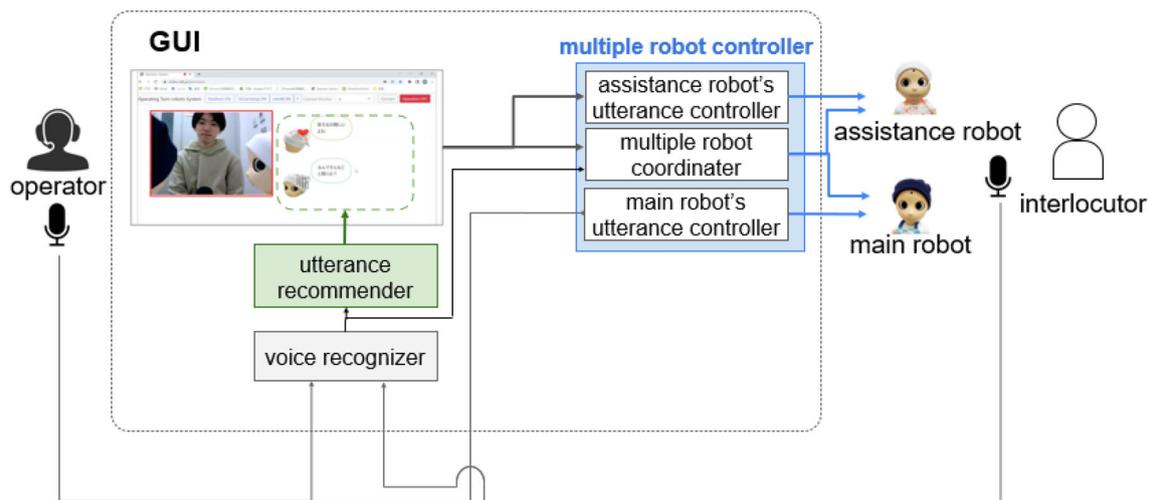


Fig. 1 The system's configuration

if neither robot is speaking, the gaze is directed toward the interlocutor. Specifically, when the operator's voice activity is detected, the assistance robot turns its gaze toward the main robot. When the multiple-robot coordinator receives the speech addressee of the assistance robot (main robot or interlocutor) chosen by the operator, it moves both robots' gaze in the following steps. First, the assistance robot turns its gaze toward the addressee simultaneously when the generated voice is produced from the built-in speaker by the utterance controller. One second later, the main robot turns its gaze toward the assistance robot two points five seconds later, and the main robot looks back to the interlocutor. These gazes of the main robot are expected to make it clear that the current speaker is the assistance robot and the addressee is the interlocutor. In other cases, the main robot and the assistance robots are directed to the interlocutor. In addition, to increase both the interlocutor's and the operator's sense of being listened to, the head of the assistance robot is moved up and down once or twice 30% of the time when their voice activities become inactive.

3.2 Assistance-Robot's Utterances Recommender

To allow an operator who mainly controls the main robots to easily control the assistance robot, the assistance robot's utterances recommender suggests its next utterance based on the last utterance by the operator or interlocutor.

The assistance robot is supposed to utter to reduce two types of potential concerns of the operator during the conversation: One concern is for his or her impoliteness, and the other is a one-sided conversation. We suppose that appropriate words basically depend on the expected function of the utterance by the last speaker. Fukuoka et al. [13, 14] proposed that utterances in a dialog can be categorized into nine dialog

acts. Among the nine dialog acts, in the proposed system, we focused on questions (yes-no), questions (what), requests, and self-disclosures and prepared sets of appropriate utterances for each category, which is to be recommended for use by the assistance robot. For example, if the operator says, "How much money do you have saved?", it is recognized as a question (what). Then, the recommender recommends candidates such as "That is a rude question." or "It is hard to answer, is not it?" We note that the utterances in the prepared set are associated with binary labels indicating nuance (hereafter called "nuance label"). These labels had values as friendly or critical and were used in the GUI for the operator to intuitively imagine how the assistance robot is perceived by saying it. The sentence type judgment API included in the COTOHA API is used to classify the dialog act of the given sentence sent from the Google Speech API, which is an artificial sound recognition module. As the labels of the COTOHA API and Fukuoka et al. [13, 14] are different from each other, a simple mapping rule was applied. Namely, information-seeking, directive, and information-providing are mapped to the question (Yes-No, What), request, and self-Conflict of interest, respectively.

In addition, to reduce one-sided conversations, appropriate support words are generated dependent on speaking amount bias. Namely, when either the operator or the interlocutor is speaking for a long period, the recommender chooses utterances promoting another to speak from prepared sets and recommends it. Specifically, if the operator or the interlocutor speaks continuously for more than 30s, the assistance robot is recommended to say words to encourage turn-taking. For example, it says, "Do you have something to say?" or "Did you understand me so far?" to the previous listener. In addition, if the operator speaks continuously for more than 50s, the recommender suggests an utterance that

condemns the one-sided conversation, such as, “Aren’t you talking a little too much?” We note that, as in the case of the dialog act, utterances in the prepared set are associated with nuance labels.

Another way to relax the concern for one-sided conversation is to increase the sense of being listened to. To increase it, we implemented “repeating utterances” [8], which is evidence of understanding the other person’s speech. Namely, key phrases are extracted from the speech recognition results and used to recommend utterances for the speaker to feel repeated and listened to. We note that the tone of the speech tail of the repeated key phrase is ascended to sound interrogative so that the assistance robot looks interested in the phrase. For example, if the operator says, “I went to Kyoto yesterday,” “Kyoto?” is recommended to be repeated. Key phrase extraction (v2) of Yahoo! API is used for key phrase extraction. It analyzes Japanese sentences and extracts important expressions as key phrases. The extracted key phrases are given a score (0–100) indicating their importance. For the system’s stability, only key phrases with a score of 100 are treated as key phrases. We note that utterances generated from key phrases are associated with friendly labels.

3.3 GUI

We developed a GUI for the operator to observe the conversation scene and to choose the recommended utterances for the assistance robot to utter (see Fig. 2). On the left side of the screen, the camera image capturing both the frontal face of the interlocutor and the backside or lateral faces of the two robots is displayed so that the operator can observe the conversational scene. On the right side, buttons with the shape of speech bubbles are displayed, each of which contains texts of the recommended utterance to be produced by the assistance robot when the operator clicks on it. When new utterances are recommended, the GUI automatically generates new speech bubble-type buttons.

In addition, there are two icon images of the assistance robot with different facial directions expressing the addressee of the recommended utterances for the assistance robot. For example, in the upper area, the robot’s face looking at the interlocutor in the monitor image is drawn so that the utterance candidate filled in the speech bubble close to it is produced toward the interlocutor. Conversely, on the lower area, the robot’s face looking at the operator is drawn to mean that the utterance candidate filled in one close to it is produced toward the main robot that is a proxy agent of the operator. In addition, to support the operator in intuitively deciding if he or she accepts the recommended utterances, a visual representation of their nuance, such as friendly and critical, was superimposed on the face icon. In particular, a heart mark,

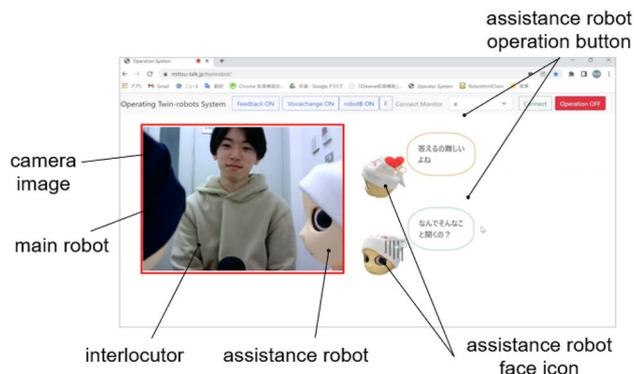


Fig. 2 Screen of the GUI

which is often used to represent a warming heart in Japan, was used if the nuance label was friendly. On the other hand, five short vertical lines, which are often used to represent a disappointing or disrespectful feeling in Japan, were used if the nuance label was critical. Furthermore, when the recommended utterance is unacceptable to the operator, he or she can click on face icons to obtain another candidate.

4 Experiment

The experimental conversation was conducted with a subject and the experimenter who took the roles of the operator and the interlocutor, respectively. To verify the effect of the proposed system, they attended two conversations with and without the operation of the assistance robot. Then, we evaluated the hypothesis that operating the assistance robot to be a proactive participant that supports the interlocutor reduces the operator’s guilt and stress during the conversation where the operator happens to be rude.

This research involves human participants and was approved by the Ethics committee for research involving human subjects at the Graduate School of Engineering Science, Osaka University, approval number 31-1-1. The written informed consents were obtained from all participants.

4.1 Subjects

Twenty-nine native Japanese speakers (20–23 years old, 14 males and 15 females) participated in the experimental conversation under two conditions. Specifically, 15 participants (7 males and 8 females, mean age 21.3 years) did it while operating the assistance robot (hereafter, With condition) followed by the condition without operation (hereafter Without one). The remaining 14 participants (7 males and 7 females, mean age 21.5 years) did so in the reverse order.

Fig. 3 The environment on the interlocutor side (left) and on the operator side (right)



4.2 Apparatus

In our system, we adopted two robots called CommU for the main and assistance robots, which is a desktop, social conversation robot developed by the collaboration between Osaka University and VSTONE Co. Ltd. and has been used for studies of human-robot conversation [18, 26, 47]. It has a height of 304 mm, a width of 180 mm, and a depth of 131 mm. It has two DOFs for its waist, two for each arm, three for its neck, three for its eyes, one for its eyelids, and one for its mouth.

The environments on the interlocutor's and operator's sides are shown in Fig. 3. On the interlocutor side, a web camera was used to observe the interlocutor and the robots, a USB loudspeaker was used to produce the main robot's voice, and two microphones were placed around the robots on a desk. An omnidirectional microphone was used to capture both the assistance robot and the subject's voices for the operator to understand the conversation, while a unidirectional microphone was used to capture only the subject's voice for the system to perform speech recognition. To prevent both participants from directly listening to the other's voice, the GUI computer with a headset for the operator was placed in a soundproof room. In the With condition, the proposed system introduced in Sect. 3 was used. On the other hand, in Without one, the same system was used except that the area for operating the assistance robot was not shown in the interface.

4.3 Scenarios

To control for the two conditions, subjects were asked to ask the interlocutor questions according to the predefined scenario. Sequences of the subject actions, such as questions and backchannels, were given in the form of flowcharts that the subject must follow. Each flowchart consists of three phases: opening, questioning, and closing. At the opening, the subject said, "Nice to meet you." Then, for questioning, the subject asked the interlocutor six questions. Finally, in closing, the subject closed the conversation by saying "That's all." To

verify the effect of the proposed method to decrease the subject's guilt and stress, three of the six questions were chosen to be rude ones so that the standard level of the subject's guilt and stress became relatively high. To further enhance the subject's guilt, he or she was asked to give a further question to deepen the interlocutor's answer to one of the rude questions: "I really want to know, but what is the truth of the matter?"

To prepare two sets of questions with sufficient and balanced levels of rudeness for two conversations in the experiment, we surveyed the rudeness of 28 questions. A total of 114 participants in their 20s (57 men and 57 women) participated in the web-based survey. The participants were asked to rate the difficulty of asking and answering each question on a 7-point Likert scale while imagining a situation where a woman in her 20s was the dialog partner. The rude score was calculated for each question by averaging the difficulty in asking and answering. The average rude score for the 28 questions was 8.7. The questions were then selected so that the total rude score for each set was high and as equal as possible. The selected questions and the rude scores are shown in Table 1.

We note that the flowchart for With condition requested the subject to operate the assistance robot and that this request appeared every time after the subject made rude questions or a further deepening one. In both conditions, the flowchart was presented right next to the window for operation GUI on the PC so that the subject's eye movement was minimized.

4.4 Procedure

First, subjects were asked to watch a video describing how to attend the experiment as well as the first practice session. In the first practice session, the subjects talked with two robots operated by the experimenter. Here, the experimenter played the role of operator, and the subject played the role of interlocutor; they experienced a conversation similar to the experiment to have an image of the operation. Afterward, the subjects moved to the operation room and watched another video explaining the details of the experiment and

Table 1 Rude questions used in the experiment and their mean and standard deviation

round	Question	Rude score		Total score
		Avg.	SD	
1st	Can you give me your address?	10.0	1.7	27.8
	What is your parents' annual income?	8.9	1.7	
	How old were you when you first kissed?	8.9	1.7	
2nd	Can you give me your phone number?	9.8	1.7	27.9
	What is your weight?	9.2	1.8	
	How much money do you have in savings?	8.9	1.8	

how to operate the robots. In the video, the subject was told not to utter any words other than those shown in the flowchart. We note that they were also told that, for the request of the operating assistance robot, they could decide not only when to do it but also whether or not to use the assistance robot. The experimenter checked the subjects' understanding, and if there were any unclear points, the experimenter explained them again. Afterward, the subjects were given a flowchart for the practice and asked to practice talking with the experimenter as the interlocutor through the robot system. After the practice, the subjects were asked to check the questionnaire items. The experimenter then told the subject to start the conversation when the experimenter said "please" and moved to the front of the robots in the robot booth to be the interlocutor in the experiment. When the experimenter was ready, the experimenter said "please", and the subjects started a conversation under one condition. After this conversation, the participants were asked to answer a questionnaire. Afterward, the subjects conversed again in another condition. The subjects answered the same questionnaire again afterward second conversation.

4.5 Measurement

4.5.1 Mood Evaluation

The total mood disturbance (TMD) of POMS [11] was employed to assess mood during a conversation, which is considered to reflect the operator's stress. We used the Japanese version of the POMS developed by Yokoyama et al. [46]. For each item, the participants were asked to select the options that best represented their mental state during the conversation on a 5-point scale (0 = not at all, 4 = very much).

4.5.2 Guilt Evaluation

Seven items on a factor called intrapsychic guilt representing ambiguous guilt without clear reasons from the trait guilt scale (TGS) [30] were used to assess the operator's guilt during a conversation. Intrapsychic guilt is an assessment of

vague guilt within the personality trait mind. In this experiment, for each questionnaire item included in this factor, the subjects were asked to choose an option on a 5-point scale (1 = not at all, 5 = always) which best describes their current guilt-related feeling formed through the conversation. Furthermore, additional questionnaire items "I regret what I said," "I felt sorry for the other person," and "I felt awkward" were used as items to directly evaluate guilt caused by one's own behavior during the conversation. They were rated on a 7-point Likert scale (1 = completely disagree, 7 = absolutely agree).

4.5.3 Conversation Impression Evaluation

Conversation by operating two robots is a special situation that may cause awkwardness and nervousness in the participants. Therefore, to evaluate them, the items "the conversation causes awkwardness (awkwardness)" and "the conversation causes a tense atmosphere (nervousness)" were used with a 7-point Likert scale (1 = completely agree, 7 = absolutely agree).

5 Result

All subjects performed the conversation following the designed flow as specified in the flowchart. In the With condition, more than 89% (26 of 29 subjects) operated the assistance robot more than four times, which corresponded to the number of times that the operator was recommended in the flowchart.

Figure 4 shows boxplots of the measured impressions of subjects in the With and Without conditions. A Wilcoxon signed rank sum test was applied to TMD since the normality of its distribution was rejected by the Shapiro-Wilk test. The test revealed that the TMD in the With condition was significantly lower than the TMD in the Without condition (With: Md = 10.0, Without: Md = 15.0, W = 101, d = -0.3, $p < .05$).

The minimum residual method was applied to scores for guilt evaluation to perform factor analysis with Promax

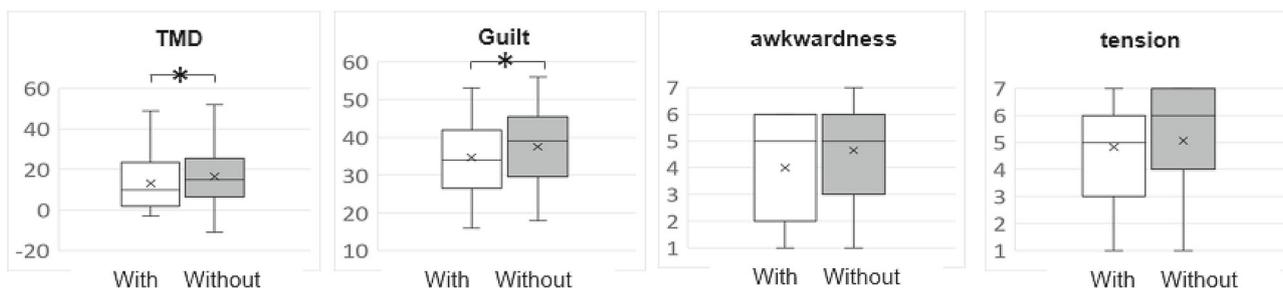


Fig. 4 Impression of subjects

rotation. Based on the Guttman-Kaiser criterion, one factor consisting of 10 items was found, which explained 55% of the variance with high internal consistency (Cronback's $\alpha=0.91$). A paired *t* test was applied on the sum of the 10 items since the normality of its distribution was accepted by the Shapiro-Wilk test. The test revealed that the guilt score was significantly lower in the With condition than in the Without condition (With: $M = 34.7$ ($SD = 9.0$), Without: $M = 37.5$ ($SD = 10.8$), $t(23) = -2.1$, $d = -0.5$, $p < .05$).

A Wilcoxon signed rank sum test was applied on awkwardness and tension since the normality of their distributions was rejected by the Shapiro-Wilk test. Although the median values were lower in the With condition than in the Without condition, no significant differences were observed in the following measurements: awkwardness (With: $Md = 5.0$, Without: $Md = 5.0$, $W = 62.5$, $d = -0.3$) and tension (With: $Md = 5.0$, Without: $Md = 5.5$, $W = 76.5$, $d = -0.2$).

6 Discussion

The results indicated that the operator's mental state was improved in the With condition. In both conditions, the subjects were led to feel guilt and to be in a bad mental mood by being forced to give rude questions. In the With condition, simultaneously, they were given opportunities to help the interlocutor by making the assistance robot produce a supportive utterance to the interlocutor or a critical one to the main robot, which was their main proxy. When people make their interlocutor stressed or annoyed, they tend to try to help him or her to reduce their guilt [5]. Therefore, the helping behavior through the assistance robot promoted in the With condition causes a reduction in their guilt and consequently an improvement in their overall mood.

The proposed method can be applied to make tense conversations comfortable, such as in first-encounter and cross-cultural communication, where the participants do not know which questions are rude or inadequate. Because the user's anxiety to give rude or inadequate questions can be

reduced by having options to resolve problems that possibly arise from their rude or inadequate questions. The proposed method is expected to be applied to a conversation with a Conflict of interest. In some cases of discipline, it is not easy for parents to control their emotions, and they are prone to say react in an overly aggressive or emotional manner [15]. In that situation, the parents feel guilty for giving necessary but eventually discouraging comments to the children to be scolded. The proposed system is expected to reduce the guilt and stress of such a parent by allowing him or her to maintain consistency in the conflicting roles of scolding and forgiving to release the children from strong discouragement. Similarly, this method can be applied to conversations to manage a customer's complaint where the responsible person cannot simply obey the customer due to his or her role. It may be possible to reduce his or her stress and guilt by allowing him or her to produce the apology via another robot instead of simply rejecting their unacceptable request. It is worth investigating the validities of such applications in field experiments in future work.

Special and complicated tasks that require operating multiple robots simultaneously are said to be difficult [17, 19]. The awkwardness and tension that subjects felt toward the conversation in the With condition did not differ from those in the Without condition, although the former was a special and complicated setting in which the subjects simultaneously operated two robots. In the interview, some subjects reported that the assistance robot's operation relaxed the conversation's atmosphere. However, such differences could not be confirmed in this experiment. This might be caused by the characteristics of the used indexes about the whole conversation, which required considering not only the subject but also the interlocutor who was controlled to produce the same behavior independently of the awkwardness and tension of the conversation. Therefore, it is necessary to confirm the effect of the proposed system on the impression of the conversation in a situation close to a natural conversational style, such as free dialog.

7 Limitations

In the current paper, we did not focus on the accuracy of the generation of candidate utterances since the purpose of this experiment was to verify the effect of the operation of the assistance robot on the operator's mood. Therefore, to control between conditions, the controlled dialog style conversation was adopted: The operator was asked to have a conversation according to the flowchart, and if the generated candidate utterance was inappropriate, the operator was allowed to regenerate another candidate and to select the appropriate one. To properly evaluate and to improve how well the current method for the utterance candidate generation works for supporting the use of the assistance robot, it is necessary to verify the extent to which the generated candidates are used in free dialog style conversation.

To simultaneously control two robots with different dialog roles, as focused on in the current study, the operator might have a high operation load [17, 19]. Although we developed an interface that displays the generated utterance candidates in the form of buttons with the shape of a speech balloon to reduce the operation load, the evaluation of the operability of the interface is not sufficient. Therefore, it is necessary to conduct experiments again to investigate how easily users can use the proposed interface.

Third, the proposed system was not evaluated from the interlocutor's perspective. In a situation where the system is actually used, it is considered important to evaluate not only the operator's perspective but also that of the interlocutor. Therefore, it is necessary to conduct an experiment with the subject as the interlocutor to investigate how the proposed system affects the person facing the robot.

Moreover, the current experiment was limited by the fact that young persons were the interview targets. A worthwhile future approach would be to examine the applicability of the proposed method in other potential target situations and among different users, such as by investigating scolding and forgiving in parent–child conversations or by exploring complaining customers in the context of customer service.

8 Conclusion

In this study, a dialog system that allows an operator to have more than one dialog role simultaneously by using two tele-operated robots was developed, for example, an interviewer and an assistant of the interviewee. The experimental results showed that the operation of the assistance robot to defend the interviewee or to criticize the main robot operated by the interviewer improved the operator's mood during the conversation and reduced the operator's sense of guilt. We argue that the proposed system allows the user to talk with a sense of security even in intense situations prone to such as first

conversations and cross-cultural conversations. Although it was evaluated in the controlled style conversation, it is worth evaluating also in the freestyle conversation to verify and to enhance its operability of the proposed interface as well as the impression of the interviewee's perspective.

Funding Open Access funding provided by Osaka University. This research was partially supported by JST, Moonshot R&D Grant Number JPMJMS2011 (system development) and JSPS KAKENHI Grant Number 20H00101 (experiment).

Data Availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Human and Animal Rights This research involves human participants and was approved by the Ethics committee for research involving human subjects at the Graduate School of Engineering Science, Osaka University, Approval Number 31-1-1. The written informed consents were obtained from all participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Admoni H, Scassellati B (2017) Social eye gaze in human-robot interaction: a review. *J Human-Robot Interaction* 6(1):25–63
- Arimoto T, Yoshikawa Y, Ishiguro H (2014) Nodding responses by collective proxy robots for enhancing social telepresence. In: *Proceedings of the second international conference on human-agent interaction*, pp 97–102
- Arimoto T, Yoshikawa Y, Ishiguro H (2018) Multiple-robot conversational patterns for concealing incoherent responses. *Int J Soc Robot* 10(5):583–593
- Baba J, Song S, Nakanishi J et al (2021) Local vs. avatar robot: performance and perceived workload of service encounters in public space. *Front Robot AI* 8:778753
- Baumeister RF, Stillwell AM, Heatherton TF (1994) Guilt: an interpersonal approach. *Psychol Bull* 115(2):243
- Biocca F, Harms C, Burgoon JK (2003) Toward a more robust theory and measure of social presence: review and suggested criteria. *Presence Teleoper Virtual Environ* 12(5):456–480
- Birmingham C, Hu Z, Mahajan K et al (2020) Can I trust you? A user study of robot mediation of a support group. In: *2020*

- IEEE international conference on robotics and automation (ICRA). IEEE, pp 8019–8026
8. Clark HH, Schaefer EF (1989) Contributing to discourse. *Cogn Sci* 13(2):259–294
 9. Connell JB, Mendelsohn GA, Robins RW et al (2001) Effects of communication medium on interpersonal perceptions. In: Proceedings of the 2001 ACM international conference on supporting group work, pp 117–124
 10. Cooney G, Mastroianni AM, Abi-Esber N et al (2020) The many minds problem: disclosure in dyadic versus group conversation. *Curr Opin Psychol* 31:22–27
 11. Curran SL, Andrykowski MA, Studts JL (1995) Short form of the profile of mood states (POMS-SF): psychometric information. *Psychol Assess* 7(1):80
 12. Fujimoto M (2008) Compass indicates participants' communication participation styles. *Soc Soc Psychol* 23(3):290–297
 13. Fukuoka Shirai (2017) Dialogue act estimation in free dialogue systems considering features specific to dialogue acts. *Nat Lang Process* 24(4):523–547
 14. Fukuoka T, Shirai K (2015) Identification of sympathy in free conversation. In: Proceedings of the 29th Pacific Asia conference on language, information and computation: posters, pp 1–9
 15. Gershoff ET, Grogan-Kaylor A, Lansford JE et al (2010) Parent discipline practices in an international sample: associations with child behaviors and moderation by perceived normativeness. *Child Dev* 81(2):487–502
 16. Glas DF, Kanda T, Ishiguro H et al (2011) Teleoperation of multiple social robots. *IEEE Trans Syst Man Cybernet Part A Syst Humans* 42(3):530–544
 17. Goodrich MA, Quigley M, Cosenzo K (2005) Task switching and multi-robot teams. In: *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III: Proceedings from the 2005 international workshop on multi-robot systems*. Springer, Berlin, pp 185–195
 18. Hagio Y, Kamimura M, Hoshi Y et al (2022) TV-watching robot: toward enriching media experience and activating human communication. *SMPTE Motion Imaging J* 131(4):50–58
 19. Hill SG, Bodt B (2007) A field experiment of autonomous mobility: operator workload for one and two robots. In: Proceedings of the ACM/IEEE international conference on Human-robot interaction, pp 169–176
 20. Iio T, Yoshikawa Y, Ishiguro H (2016) Pre-scheduled turn-taking between robots to make conversation coherent. In: Proceedings of the fourth international conference on human agent interaction, pp 19–25
 21. Iio T, Yoshikawa Y, Chiba M et al (2020) Twin-robot dialogue system with robustness against speech recognition failure in human-robot dialogue with elderly people. *Appl Sci* 10(4):1522
 22. Isbister K, Nakanishi H, Ishida T et al (2000) Helper agent: Designing an assistant for human-human interaction in a virtual meeting space. In: Proceedings of the SIGCHI conference on Human factors in computing systems, pp 57–64
 23. Ishi CT, Liu C, Ishiguro H, et al (2012) Evaluation of formant-based lip motion generation in tele-operated humanoid robots. In: 2012 IEEE/RSJ international conference on intelligent robots and systems. IEEE, pp 2377–2382
 24. Kiesler S, Sproull L (1992) Group decision making and communication technology. *Organ Behav Hum Decis Process* 52(1):96–123
 25. Kraut RE, Fussell SR, Siegel J (2003) Visual information as a conversational resource in collaborative physical tasks. *Human-Comput Interaction* 18(1–2):13–49
 26. Kumazaki H, Yoshikawa Y, Yoshimura Y et al (2018) The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Mol Autism* 9(1):1–10
 27. Kendon A (1967) Some functions of gaze-direction in social interaction. *Acta Physiol (Oxf)* 26(1):22–63
 28. Laporte L, van Nimwegen C, Uyttendaele AJ (2010) Do people say what they think: Social conformity behavior in varying degrees of online social presence. In: Proceedings of the 6th Nordic conference on human-computer interaction: extending boundaries, pp 305–314
 29. Mehmood F, Mahzoon H, Yoshikawa Y, Ishiguro H et al (2023) Effect of having and switching multiple avatars on the operator's right to talk and receive social support. *PLoS ONE* 18(10):e0292803
 30. Masafumi Onishi. Structure of Trait Guilt in Adolescents: Conceptualization of Guilt and Development of Trait Guilt Scale Based on Psychoanalytic Theory. *The Japanese J Person*, 16(2): 171–184, 2008
 31. Matsuyama Y, Akiba I, Fujie S et al (2015) Four-participant group conversation: a facilitation robot controlling engagement density as the fourth participant. *Comput Speech Lang* 33(1):1–24
 32. Morikawa O, Maesako T (1998) Hypermirror: toward pleasant-to-use video mediated communication system. In: Proceedings of the 1998 ACM conference on Computer supported cooperative work, pp 149–158
 33. Mutlu B, Kanda T, Fortizzi J, Hodgins JK (2012) Conversational gaze mechanisms for humanlike robots. *ACM Trans Interact Intell Syst* 1(2):1–33
 34. Nakano YI, Nihonyanagi S, Takase Y et al (2015) Predicting participation styles using co-occurrence patterns of nonverbal behaviors in collaborative learning. In: Proceedings of the 2015 ACM on international conference on multimodal interaction, pp 91–98
 35. Newhart VA, Warschauer M, Sender L (2016) Virtual inclusion via telepresence robots in the classroom: an exploratory case study. *Int J Technol Learn* 23(4):9–25
 36. Sakai K, Ishi CT, Minato T et al (2015) Online speech-driven head motion generating system and evaluation on a tele-operated robot. In: 2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 529–534
 37. Sakamoto D, Hayashi K, Kanda T et al (2009) Humanoid robots as a broadcasting communication medium in open public spaces. *Int J Soc Robot* 1:157–169
 38. Shiomi M, Hagita N (2016) Do synchronized multiple robots exert peer pressure? In: Proceedings of the fourth international conference on human agent interaction, pp 27–33
 39. Short E, Mataric MJ (2017) Robot moderation of a collaborative game: towards socially assistive robotics in group interactions. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 385–390
 40. Takeuchi K, Yamazaki Y, Yoshifuji K (2020) Avatar work: telework for disabled people unable to go outside by using avatar robots. In: Companion of the 2020 ACM/IEEE international conference on human-robot interaction, pp 53–60
 41. Thomas J (1983) Cross-cultural pragmatic failure. *Appl Linguist* 4(2):91–112
 42. Todo Y, Nishimura R, Yamamoto K et al (2013) Development and evaluation of spoken dialog systems with one or two agents through two domains. In: Text, speech, and dialogue: 16th international conference, TSD 2013, Pilsen, Czech Republic, September 1–5, 2013. Proceedings 16. Springer, Berlin, pp 185–192
 43. Tu CH, McIsaac M (2002) The relationship of social presence and interaction in online classes. *Am J Distance Educ* 16(3):131–150
 44. Vertegaal R, Slagter R, van der Veer G, Nijholt A et al (2001) Eye Gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: Proceedings of ACM conference on human factors in computing systems, pp 301–308
 45. Yamaguchi K, Nergui M, Otake M (2014) A robot presenting reproduced stories among older adults in group conversation. *Appl Mech Mater* 541:1120–1126
 46. Yokoyama K, Araki S, Kawakami N et al (1990) Production of the Japanese edition of profile of mood states (POMS): assessment

- of reliability and validity. [Nihon koshu eisei zasshi] *Jpn J Public Health* 37(11):913–918
47. Yoshino K, Murase Y, Lubis N et al (2021) Spoken dialogue robot for watching daily life of elderly people. In: Increasing naturalness and flexibility in spoken dialogue interaction: 10th international workshop on spoken dialogue systems. Springer, Berlin, pp 141–146
48. Yoshino T, Hayashi Y, Nakano IY (2014) Determining robot gaze according to participation roles in multiparty conversations. In: Proceedings of the second international conference on Human-agent interaction, pp 277–280

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reina Nozawa received the Master degree in engineering from Osaka University, Japan, in 2024. Her research interests include human-robot interaction and dialogue support.

Kazuki Sakai received a Ph.D. degree in engineering science from Osaka University, Japan, in 2020. He is currently a specially appointed assistant professor of the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University. His research interests include human-robot interaction and dialogue systems.

Megumi Kawata received a Ph.D. in Graduate School of Intercultural Studies from Kobe University, Japan, in 2021. She is a Specially appointed assistant professor of the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University. Her research interests include social robotics and human-robot interaction.

Hiroshi Ishiguro received a D. Eng. in systems engineering from the Osaka University, Japan in 1991. He is currently professor of the Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University (2009-). He is also visiting Director (2014-) of ATR Hiroshi Ishiguro Laboratories.

Yuichiro Yoshikawa received the Ph.D. degree in engineering from Osaka University, Japan, in 2005. From 2010, he has been an associate professor of the Department of Systems Innovation in the Graduate School of Engineering Science, Osaka University. He is interested in the study of human-robot communication, especially about multiple-robots coordination and psychiatric or educational applications.