Simultaneous Dialogue Services Using Multiple Semiautonomous Robots in Multiple Locations by a Single Operator: A Field Trial on Souvenir Recommendation

Kazuki Sakai, Megumi Kawata, Alexis Meneses, Hiroshi Ishiguro, and Yuichiro Yoshikawa

***Abstract*—Recently, teleoperation systems have been developed enabling a single operator to engage with users across multiple locations simultaneously. However, under such systems, a potential challenge exists where the operator, upon switching locations, may need to join ongoing conversations without a complete understanding of their history. Consequently, a seamless transition and the development of high-quality conversations may be impeded. This study directs its attention to the utilization of multiple robots, aiming to create a semiautonomous teleoperation system. This system enables an operator to switch between twin robots at each location as needed, thereby facilitating the provision of higher-quality dialogue services simultaneously. As an initial phase, a field experiment was conducted to assess user satisfaction with recommendations made by the operator using twin robots. Results collected from 391 participants over 13 days revealed heightened user satisfaction when the operator intervened and provided recommendations through multiple robots compared with autonomous recommendations by the robots. These findings contribute to the formulation of a teleoperation system that allows a single operator to deliver multipoint conversational services.**

***Index Terms*—Social HRI, telerobotics and teleoperation, natural dialog for HRI, multiple robots, field experiment.**

I. Introduction

THE rapid development of large language models has considerably improved the interactive capabilities of autonomous conversational robots. However, complex conversations such as solving problems [1], [2] and exhibiting empathy [3], [4] are in the nascent stage of development. Numerous studies have focused on conversational robots that incorporate human intervention, that is, teleoperation [5], [6]. Typically, these robots manage individual conversations under the guidance of a single operator. However, due to the scarcity of human resources, the expectation for a single operator to handle multiple conversations simultaneously has emerged.

In response to this demand, researchers have introduced semiautonomous teleoperation systems that empower a single operator to engage with multiple virtual agents concurrently, facilitating the simultaneous provision of spoken conversations [7], [8]. In these systems, conversations are primarily handled autonomously, with the operator monitoring the conversation history. However, when challenges arise that the autonomous agent cannot resolve or when interruptions are needed, such as deepening the conversation, the system transitions from autonomous to teleoperated mode. This flexibility enables the operator to deliver high-quality conversations requiring both autonomous agent follow-up and human intervention at multiple locations simultaneously. However, the potential challenge arises in instances where the operator may need to join a conversation without a full grasp of its history, hindering the smooth transition and development of high-quality interactions.

To address this problem, we propose the coordination of multiple robots to buy time for an operator to respond in a conversation. A conversation between multiple robots makes a user more likely to accept information [9]. This phenomenon can be used not only to buy time but also to develop an easily understandable takeover mechanism. Furthermore, we consider supporting the development of conversations for an operator during operation. Operating another robot when conversing buys time to develop the conversation [10]. Therefore, we investigated the effect of multiple robots on dialogue facilitation.

First, this study examined whether a conversation between multiple robots is accepted by the user. In particular, we evaluated whether such systematic interactions between robots do not cancel the effects of operator intervention. In the proposed semiautonomous teleoperation system, two robots participate in the same conversation in each location and send the information of the conversation of each booth to a single operator. The system enables the operator to switch between locations as required, enhancing the quality of dialogue services. The system incorporates two key functions: a takeover function allows the

Manuscript received 10 January 2024; accepted 30 April 2024. Date of publication 23 May 2024; date of current version 30 May 2024. This letter was recommended for publication by Associate Editor B. Bruno and Editor G. Venture upon evaluation of the reviewers' comments. This work was supported in part by JSPS KAKENHI under Grant JP19H05691 (development for dialogue system) and Grant 20H00101 (experiment), in part by JST, Moonshot R&D under Grant JPMJPS2011 (development for base system), and in part by Innovation Platform for Society 5.0 under Grant JPMXP0518071489 (for data collection), Japan. *(Corresponding author: Kazuki Sakai.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the ethical committee of the Graduate School of Engineering Science at Osaka University, Japan.

The authors are with the Graduate School of Engineering Science, Osaka University, Toyonaka 5608531, Japan (e-mail: sakai.kazuki@irl.sys.es.osaka-u.ac.jp; kawata@irl.sys.es.osaka-u.ac.jp; alexis.meneses@irl.sys.es.osaka-u.ac.jp; ishiguro@irl.sys.es.osaka-u.ac.jp; yoshikawa@irl.sys.es.osaka-u.ac.jp).

This letter has supplementary downloadable material available at https://doi.org/10.1109/LRA.2024.3404752, provided by the authors.

Digital Object Identifier 10.1109/LRA.2024.3404752

robots to autonomously compose a conversation for the operator to hear in a takeover situation from the robots to the operator, and guides the operator in producing their first words after joining a conversation without requiring a comprehensive understanding of the dialogue history. Additionally, a dialogue support function allows the operator to easily proceed their conversation by having another robot speak by pressing buttons, and assists the operator in smoothly transitioning topics to provide necessary information for the dialogue service. To evaluate this system, we conducted a field experiment at Nifrel, an aquamuseum in Osaka, Japan. The scenario involved customers visiting booths operated by two twin robots near the museum exit, where the robots inquired about the customer's experiences and preferences related to the aquamuseum, ultimately recommending souvenir items. The recommendation task was selected because the proposed system can assist an operator to engage in a complex conversation in which autonomous dialogue systems are not feasible, and we considered the recommendation to be a typical example of a complex conversation. The study aims to assess whether remote conversation by an operator enhances satisfaction with recommendations compared to those provided autonomously by the robots.

The contributions of this study are as follows: (i) we developed a dialogue system that utilizes two supporting functions (takeover and dialogue support functions) for two robots, (ii) we conducted a field experiment involving numerous participants, and (iii) we confirmed that the two supporting functions give a positive impression for the users.

II. Related Work

In human-robot interaction (HRI), the development of dialogue systems has attracted considerable attention. A recent survey summarized dialogue management in HRI [11]. The survey classified methods of dialogue management, such as rule-based, probabilistic, and hybrid methods. In addition, multimodal systems are being developed, particulary because robots are capable of expressing nonverbal behavior. For example, Wolfert et al. conducted a survey on robot gesture generation [12], and Inoue et al. developed a method for generating laughter in android robots to provide empathic dialogue [13]. There have been studies on operator intervention in conversational tasks with autonomous conversational robots [5], [6]. Shiomi et al. developed a mechanism for a robot to ask an operator for help when a problem cannot be solved autonomously [5]. Extending this mechanism, research is ongoing to develop a system that allows a small number of operators to intervene in multiple robots.

Research has explored teleoperating systems designed for simultaneous conversations managed by a single operator. Zheng et al. introduced a semiautonomous operation system where the operator adjusts intervention timing in navigation and conversation tasks simultaneously, guided by a proactive timing model [14], [15]. Kawahara et al. proposed a framework for a semiautonomous teleoperation system wherein one operator engages in simultaneous conversations with multiple participants in a pseudo manner [7]. Kawai et al. introduced a simultaneous job interview system, enabling one operator to conduct three interviews simultaneously [8]. Their pilot experiment indicated enhanced robotic ability to appropriately respond to what a human says compared with a system without human intervention. Notably, these studies did not incorporate support mechanisms for the operator.

However, as large-scale language models advance, the conversational capabilities of dialogue systems are rapidly improving, raising the question of whether operator intervention remains necessary. In call center applications, certain studies [1], [2] argue that autonomous agents are beneficial for easy and simple cases, while human intervention is desirable for more advanced and complex cases to solve problems swiftly. Other studies [3], [4] report that while the performance of basic functions, such as backchannel communication, by autonomous dialogue agents is comparable to that of human operators, the performance of advanced functions, such as showing empathy, is insufficient. Consequently, semiautonomous teleoperation systems, where an operator switches in conversations from autonomous robots, become essential. However, a key challenge is the smooth transition for the operator during conversation switching. This study addresses this challenge by focusing on assistance functions for the operator through the utilization of multiple robots.

Employing multiple robots proves advantageous for both autonomous robots and operators. Studies indicate that coordinated interaction between multiple robots can motivate conversations, even when speech recognition faces challenges [16], [17]. Other research highlights the enhancement of emotional aspects, such as cuteness [18] or apology [19], through the use of multiple robots. Conversely, for the operator, a study suggests that using multiple robots can improve telepresence [20]. However, this study focused solely on the nodding behavior of accompanying robots and did not investigate conversational support. In our study, we implement conversational support functions for the operator.

III. Semiautonomous Teleoperation Systems Using Multiple Robots

We have engineered a semiautonomous teleoperation system employing two robots. Initially, both robots operate autonomously, individually gathering a user's experiences and preferences in their respective locations. While the conversation is ongoing, a single operator intervenes, utilizing either robot to provide recommendations. Henceforth, we designate the robot operated by the operator as the proxy robot, and the other robot, supporting the operator, as the assistant robot.

A. System Architecture

Fig. 1 shows the system architecture of the overall system. The system consisted of an autonomous dialogue system (local side) and a teleoperation system (remote side).

1) Autonomous Dialogue System: User utterances undergo recognition through speech recognition, with the results transmitted to the dialogue manager. A SpeechRecognition interface

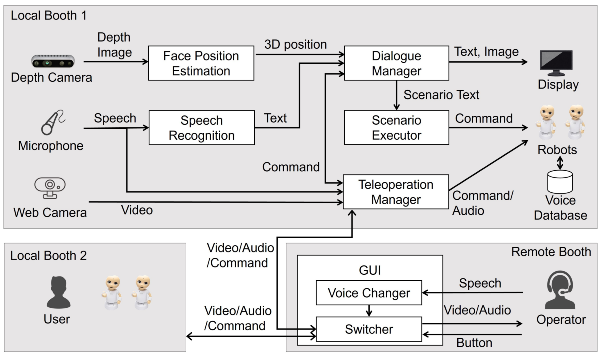

Fig. 1. System architecture. The system consists of autonomous dialogue systems and a teleoperation system. The autonomous system consists of five modules. The system uses the user's voice and video as input, and the robot's speech and display as output. The operator monitors multiple booths with a single GUI. The operator communicates with each autonomous dialogue system using voice and button operations as interface inputs.

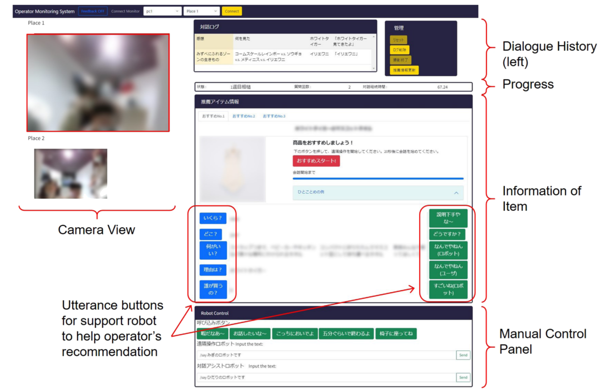

Fig. 2. Graphical user interface for teleoperation. The left side of the screen shows videos from the web camera and microphone in each booth. The right side shows the conversation progress of each booth, the information of items to be recommended, and buttons for robots and systems.

of the WebSpeechAPI[1] implemented in Chrome browser on Windows was used for speech recognition. The dialogue manager conducts morphological analysis on the utterance texts to identify keywords, encompassing agreement/disagreement and comments, employing the word match method. A predefined list of possible words signifying agreement or disagreement, such as "yes" and "prefer" for agreement, and "no" and "dislike" for disagreement, is utilized for keyword recognition. If the morphological analysis yields these prepared words, the module outputs the corresponding category (i.e., agreement/disagreement). For comment recognition, content words (nouns, adjectives, and verbs) are extracted from the morphological analysis results.

Subsequently, the dialogue manager formulates next dialogue scenarios following a predefined dialogue flow. A template featuring variable word slots, filled with appropriate words, such as item names, is employed for scenario generation. The template encompasses not only the robots' utterances but also their gaze targets and gestures. The dialogue scenario is transmitted to the scenario executor, which schedules and dispatches the commands to the robots as per the schedule. Notably, to maintain consistency in the proxy robot's voice, all wave files containing laboratory members reading out all possible utterance texts in the scenario using a voice changer were pre-stored in the voice database. A voice changer with pitch-shifting capabilities is applied when the proxy robot speaks. The assistant robot's voice is generated using text-to-speech software. The voice of the robots was changed to emphasize multiparty dialogue, that is, to give the impression that the robots have different personalities even though a single operator controls multiple robots. Another reason is to easily distinguish which robot is speaking in a high noise environment.

To enhance user comprehension of questions and recommendations, the dialogue manager displays the words related to the queried items and images of the recommended goods. Establishing eye contact between the user and robots is facilitated by face position estimation, which determines the 3D position of the user's face using a depth camera and updates gaze targets accordingly.

2) Teleoperation System: In Fig. 2, the graphical user interface (GUI) designed for the operator is depicted. The left side of the screen shows videos from the web camera and microphone via Web Real-Time Communication (WebRTC). Clicking on a video enlarges it and makes the audio audible. Selecting a video results in switching the operator's monitored booth. The operator's monitored conversation progress is displayed on the right side. A tabular summary of the conversation history, sent by the dialogue manager, is presented at the top right of the screen. Notably, the summary serves as an indicator of progress. The central region on the right side exhibits recommendation information and operation buttons. Multiple recommendation candidates are organized into tabs, facilitating the selection of a recommended item by switching between tabs. A request button (highlighted in red at the center) allows the operator to request an operation from the dialogue manager. On pressing the button, the command is sent to the dialogue manager and the operator waits to start the takeover conversation. After completing the takeover conversation, the microphone of the operator automatically unmutes and the operator can start speaking. When the operator speaks, their voice is altered by using a voice changer with the same pitch-shifting capabilities. Simultaneously, the original voice undergoes processing for mouse shape estimation [21], ensuring synchronization of the operator and robot's mouse movements. When either robot speaks, the other robot looks at the speaking robot automatically. If the operator cancels the (waiting for) operation, they press a halting button that is automatically transformed by the request button. Operation buttons (highlighted in blue on the left) are associated with questions regarding the recommended item. Activating an operation button prompts the assistant robot to pose the question to the proxy robot. The operator responds to the question, enabling easy introduction of new topics. Additional buttons (highlighted in green on the right) are linked to supportive utterances by the assistant robot. By responding to the robot's utterance personally, the operator can smoothly navigate the conversation, even in situations where it may not be progressing well or the user's

[1]https://developer.mozilla.org/en-US/docs/Web/API/SpeechRecognition

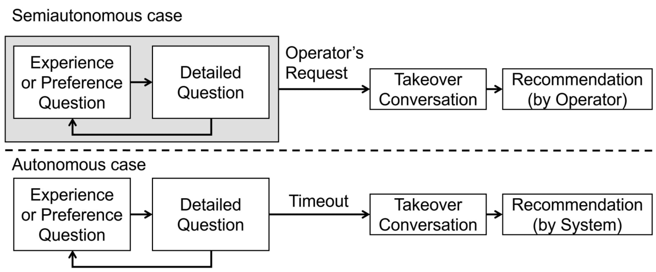

Fig. 3. Dialogue flow. The structure of the conversation is the same in both cases, but the transition to recommendation differs depending on the operator's behavior. In the semiautonomous case, recommendation is triggered by the operator's request even in the middle of questions, whereas in the autonomous case, recommendation occurs automatically when all questions are asked.

response is less than satisfactory. The lower part of the GUI provides a space for free speech for both robots.

B. Dialogue Flow

Fig. 3 shows two dialogue flows: semiautonomous and autonomous. Both flows consist of the preference collection part (experience or preference question and detailed question) and recommendation part (takeover conversation and recommendation).

In the preference collection part, the assistant robot initiates by asking questions about the user's experiences and preferences. The proxy robot then follows up with more detailed questions to elaborate on the initial inquiries. Following each round of preference collection, the robots select three items for recommendation and send them to the operator's GUI. The operator, within a defined time limit, chooses one of the items, and upon making the selection, presses the request button to initiate the operation. Upon receiving the request, the system proceeds to the recommendation part, either triggered by the operator or automatically when the time limit elapses.

Before the recommendation in both scenarios, a takeover conversation occurs where the two robots engage in dialogue to provide the operator with a smooth transition, summarizing the ongoing conversation. For instance, the assistant robot might express, "*I'm glad that you like white tigers as I do. Hey, could you introduce some goods that are suitable for the user who prefers white tigers?*" Subsequently, the operator commences the recommendation through the proxy robot. During the recommendation, the operator can facilitate their interaction by controlling the assistant robot. For instance, when the operator intends to discuss the price, they press the corresponding button on the GUI. The assistant robot then queries the proxy robot with, "*How much is it?*" allowing the operator to effortlessly mention the price. Upon completing the recommendation, the operator presses the halting button, and the system continues the conversation autonomously.

Alternatively, when the system autonomously handles the recommendation, the robots present the recommended goods using a predefined template. An example dialogue is detailed in Table I.

TABLE I
EXAMPLE DIALOGUE WITH THE RECOMMENDATION BY THE AUTONOMOUS SYSTEM. AR, PR, AND U REFER TO THE ASSISTANT ROBOT, PROXY ROBOT AND USER, RESPECTIVELY

AR	Hello, what did you see in Nifrel?
U	Hmm, I saw a hippo, a pygmy hippo.
AR	You saw the pygmy hippo! It is good for a humorous face.
PR	What do you think?
U	It is so cute!
PR	I see.
AR	I think seeing the pygmy hippo means that you went to the area of Waterside. For the four living things in the area of Waterside, what impressed you?
U	A white tiger.
AR	I see, it is good for a great shape and cool face!
PR	What do you think?
U	The cool face!
PR	I see.
AR	By the way, there are many living things in Nifrel. Which do you prefer, crocodiles or pigeons?
U	I prefer crocodiles.
AR	You like crocodiles!
	[*Repeat similar preference and deepen question*]
AR	I am gradually getting to know you. I'm glad that you like white tigers as I do. Hey (AR looks to PR), could you introduce some goods that are suitable for the user that prefers white tigers?
PR	Hmm, in the museum shop, there are playing cards with Karuta.
AR	What is it?
PR	It is not only playing cards but also Karuta. Many living things are drawn in it.
AR	That sounds good! Do you know it?
U	I do not know it.
PR	I see. It is created because we can be interested in the living things with joy.
AR	Great! I think we can enjoy by using it.
PR	Yes, I suggest you to see it which is in the museum shop.
AR	Oh, it is time to finish the conversation.

C. Recommendation Method

In this study, we employed a subjective similarity-based preference estimation method [22], [23] for generating recommendations. The subjective similarity model serves as a relational representation of items, reflecting a specific person's criteria. This method allows the robot to estimate preferences in a humanlike manner, whereby the next queried item is somewhat related to previous items.

To construct the subjective similarity model, we solicited a person's input on similarity data between items. The adaptation involved 96 living things and eleven places within the museum. For preference inquiries, we compiled a list of 55 items, including 28 living things associated with the museum, 10 adjectives related to feelings, and 17 items related to souvenirs. Additionally, for recommendation purposes, we selected 19 goods available in the museum shop. A laboratory person repeated a survey (about 1 minute per survey) to classify eight randomly selected items 200 times, and the data was subjected to analysis using an infinite relational model [24] to generate the subjective similarity model. Utilizing this model within the estimation method allows the system to gauge the degree of preference for the items recommended.

IV. FIELD EXPERIMENT

To examine the impact of our semiautonomous teleoperation system, a field experiment was conducted at Nifrel from 10 am to 6 pm on weekdays and from 9:30 am to 7 pm on weekends

Fig. 4. Experimental setup: (a) overall and (b) detailed scene.

and holidays, spanning the period of November 18 to November 30, 2021. The ethical committee of the Graduate School of Engineering Science at Osaka University, Japan, granted approval for all procedures involved in this experiment.

A. Apparatus

Fig. 4 illustrates the configurations of the event booths. Two booths existed and each booth was separated by partitions, and the wall was covered with green cloth for chroma key compositing. Within each booth, three desks were arranged. The first desk housed two CommU robots[2], a display, a web camera, a microphone, a depth camera, a pair of stereo speakers, and a red button device. The CommU robots featured synchronized mouth movements with their voices, possessing 14 degrees of freedom for various nonverbal gestures such as nodding. A laptop PC and recording camera were positioned on the second desk behind the first one. The third desk, located behind the subject for phototaking, included another display and web camera. Adjacent to the booth, between two setups, an additional desk held another laptop PC and a printer for generating photo stickers, which served as gifts.

B. Subject

A total of 708 visitors participated in the experiment. We excluded data from subjects who (1) left in the middle of the conversation due to visitors' availability or system trouble, (2) did not answer a questionnaire, and (3) answered a questionnaire two times or more in one conversation. Consequently, data from 213 participants (80 men and 133 women) in the autonomous condition and 178 participants (65 men and 113 women) in the semiautonomous condition were included in the analysis.

C. Stimuli

In this study, we conducted a comparison between two conditions: Semiautonomous and Autonomous. In both conditions, preference acquisition was conducted by autonomous robots. In the semiautonomous condition, the recommendation process was undertaken by an operator. Two female operators, who had no prior experience working in the museum, were hired. Each operator was assigned responsibility for two booths simultaneously. Before the experiment, the operators underwent training on using the GUI and making recommendations. They were instructed to select one of the goods displayed in the GUI based on the conversation history and subsequently recommend the chosen item through the robots. Additionally, operators were guided on simultaneously managing conversations in two booths.

Under the autonomous condition, the robots autonomously recommended goods following the predetermined dialogue flow. The assignment of conditions was randomized, considering both weekdays and weekends. However, if the operator encountered difficulties in managing the conversation during the semiautonomous condition, the condition was switched to autonomous. Notably, a conversation provided in the autonomous mode and a conversation provided in the semiautonomous mode are the same when the operator cannot intervene.

To enhance visitor participation in the experiment, we incorporated a feature allowing the system to capture a simulated photograph in which the background is replaced with an image related to the aquamuseum. We used chroma key compositing to create the simulated photograph. We recruited staff to print the photo stickers.

To determine the goods to be recommended, we asked a staff member at the aquamuseum to select nineteen items from the museum shop. This staff member also filled out a questionnaire providing detailed information about each item, including price, target age, selling points, photos, and characteristics. Based on this information, the experimenter manually created the utterance texts for each item.

D. Procedure

The experiment was publicly announced to passersby and visitors through notification posters, providing information about the ongoing experiment. The posters included details about the purpose of the experiment, precautions, the procedure, how to withdraw from the experiment, and the handling of collected data. This study employed an opt-out approach, wherein participants who were unwilling to have their data included could request its removal, even if they had already engaged in the conversation.

Participants, upon reading the instructions on the poster, could voluntarily decide to partake in the experiment by pushing a designated button, initiating the conversation. Upon completion of the conversation, the robots prompted participants to fill out questionnaires using a web form. A 2D code was displayed for participants to scan using their smartphones. After completing the questionnaire, participants received a printed photo sticker of the composite photo as a memento before leaving the area.

E. Evaluation

The self-developed questionnaire comprised two parts: four items about participants and eleven items about their impressions of the dialogue. The first part included inquiries about gender, age, active participation status, and whether the participant had previous experience with the experiment. The remaining eleven items addressed participants' impressions consisting of five aspects listed in Table II. A seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) was used, with

[2] https://www.vstone.co.jp/products/commu/index.html

TABLE II
QUESTIONNAIRE USED IN THE EXPERIMENT

Satisfaction	Did you enjoy the conversation with the robots? Do you wish to talk to the robots again? Were you satisfied with the conversation?
Robots' understanding	Do you think these robots can have a capability to understand the conversation? Did you feel that the robots understood your preferences?
Presentation of goods	Did you feel that the presentation of the goods by the robots was natural? Did you feel that the presentation of the goods by the robots was humanlike?
Intrusiveness	Did you feel that the presentation of the goods by the robots was intrusive?
Desire to purchase	Were you interested in the goods introduced by the robots? Do you want the goods introduced by the robots?

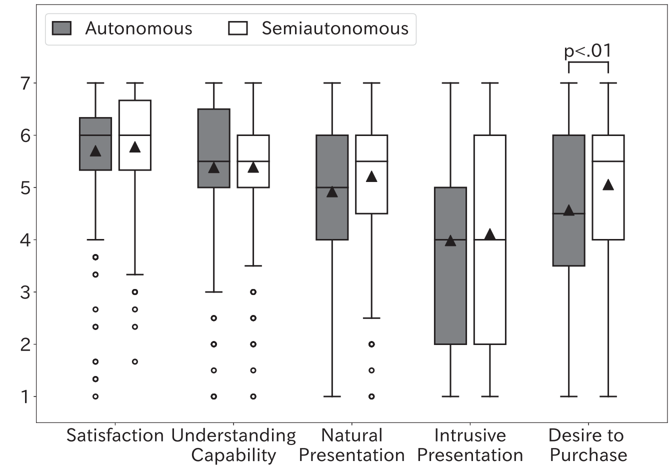

Fig. 5. Boxplots of the average scores. Multiple regression analysis was conducted to find the significance between the conditions.

TABLE III
RESULTS OF THE MULTIPLE REGRESSION AFTER THE STEPWISE METHOD

		Estimate	Std. Error	t-val	p-val
Satisfaction	(Intercept)	5.494	0.124	44.427	<.001
	Age	−0.007	0.003	−2.084	.038
	Gender	0.200	0.120	1.666	.097
	Active participation	0.506	0.116	4.345	<.001
Understanding capability	(Intercept)	4.409	0.385	11.453	<.001
	Gender	0.228	0.142	1.611	.108
	Active participation	0.622	0.138	4.510	<.001
	Number of recognition failure	0.067	0.040	1.696	.091
	Time of preference acquisition	−0.005	0.003	−1.570	.117
Natural presentation	(Intercept)	5.127	0.258	19.849	<.001
	Gender	0.312	0.156	1.993	.047
	Active participation	0.381	0.152	2.507	.013
	autotime	−0.003	0.001	−2.066	.040
Intrusive presentation	(Intercept)	4.503	0.146	30.823	<.001
	Gender	−0.720	0.185	−3.899	<.001
Desire to purchase	(Intercept)	4.250	0.130	32.602	<.001
	Condition	0.459	0.157	2.927	.004
	Active participation	0.684	0.156	4.379	<.001

the midpoint value of four corresponding to "undecided." The average scores for each measurement were calculated.

A multiple regression analysis was conducted to explore how experimental factors, system performance, and participant backgrounds influenced the scores. The dependent variable was the average score on each questionnaire item. Independent variables included condition (1: Semiautonomous, 0: Autonomous), age (numeric), gender (1: female, 0: male), active participation status (1: main participant, 0: bystander), the number of participations (1: repeater, 0: no experience), the number of recognition failures (numeric), duration of conversation before the recommendation (numeric), and duration of conversation during the recommendation (numeric). A stepwise method was employed to select effective factors.

The net promoter score (NPS) [25] was used to assess satisfaction. NPS is a single scale to measure customers' loyalty and has been widely used in business. Participants were asked, "How likely are you to recommend these robots to a close friend or colleague on a 0 to 10 scale?" Ratings of 0 to 6 were categorized as "detractors," 7 or 8 as "passives," and 9 or 10 as "promoters." NPS was calculated by subtracting the percentage of detractors from the percentage of promoters, resulting in a range from −100 (poor) to 100 (excellent).

F. Result

First, we report the statistics of the conversation. The average duration for users who completed the conversation was 359.1 s (SD = 65.9). Under the semiautonomous condition, the average duration before the recommendation (including preference questions and takeover) was 104.0 s (SD = 34.2), the operator's conversation duration was 45.3 s (SD = 18.2), and the duration between pushing the request and halting buttons was 81.3 s (SD = 19.0). Under the Autonomous condition, the duration before the recommendation was 202.8 s (SD = 39.6), and the duration of the recommendation was 59.3 s (SD = 6.6).

On days designated as the semiautonomous condition, 366 conversations reached completion, with the operator handling the conversation 286 times. Specifically, when both booths were occupied, the operator completed the earlier recommendation 194 times, of which 148 times (148/194=76.3%) the operator successfully initiated the recommendation for the other booth. The remaining instances were hindered by system issues, such as unstable network connections or the inability to commence the next recommendation promptly after completing the first.

Fig. 5 presents boxplots of the average scores for each condition in the questionnaire. Table III provides the results of the multiple regression analysis using the stepwise method. For satisfaction, no significant effect was found for the conditions; however, active participation and age emerged as significant factors. This suggests that individuals of younger age may score higher satisfaction than their older counterparts, and participants who directly engaged in conversation with the robot tended to express higher satisfaction. Regarding the perception of the robots' understanding capabilities, no significant effect was observed

TABLE IV
RESULT OF NPS. THE NUMBERS IN THE PARENTHESES ARE RESIDUAL, AND THE STAR MARK MEANS THAT THE RESIDUAL IS SIGNIFICANTLY HIGH

	Promoters	Passives	Detractors	NPS
Autonomous	57 (0.079)	46 (−2.799∗)	110 (2.433∗)	−24.9
Semiautonomous	47 (−0.079)	61 (2.799∗)	70 (−2.433∗)	−12.9

for the conditions, with only active participation influencing the score. Participants who directly conversed with the robot perceived a higher level of understanding compared to those who did not. Concerning the naturalness of the presentation, no significant effect was found for the conditions. However, gender, active participation, and the duration of preference acquisition were identified as significant factors. Female participants and those who directly interacted with the robot perceived a greater naturalness in the presentation, while longer durations in preference acquisition were associated with lower scores. For the perceived intrusiveness of the introduction, no significant effect was found for the conditions, with gender being the sole factor influencing the score; specifically, female participants reported a lower sense of intrusiveness. Regarding the desire to purchase, the condition factor significantly affected the score, indicating that participants in the Semiautonomous condition expressed a greater desire to purchase compared to those in the autonomous condition. Additionally, active participation contributed to an increase in the score.

Table IV presents the NPS results. The score for the Semiautonomous condition is higher than that for the Autonomous condition, indicating greater participant satisfaction with the operator's recommendation. For a more in-depth analysis, we conducted a χ^2 test to examine the degree of bias in cells. The results showed statistical significance ($\chi^2(2) = 8.892, p = .012$). A residual analysis following the χ^2 test suggests that more participants in the Semiautonomous condition belong to the Passives category compared with those in the Autonomous condition. Additionally, fewer participants in the Semiautonomous condition belong to the Detractors category compared with those in the Autonomous condition.

V. DISCUSSION

The NPS and desire-to-purchase results suggest that the operator's conversation could mitigate the limitations of autonomous recommendation effectiveness. This is attributed to the operator's ability to tailor presentations to users rather than simply showcasing goods. These findings support the effectiveness of combining teleoperated and autonomous systems for spoken conversation robots, as proposed in previous work [7], [8]. The results also offer insights for designing systems that facilitate operators' interactions in such situations.

However, there were no significant differences observed in terms of understanding ability or the perceived naturalness of the presentation. For understanding ability, both operator and autonomous robots exhibited attentive listening during the preference acquisition phase, potentially influencing the results. Regarding the naturalness of goods presentation, the limited information available to both the operator and autonomous robots may have restricted opportunities for enhancing naturalness. The multiple regression results indicated that naturalness was negatively impacted by a longer duration before the recommendation. Extended preference acquisition durations may result in the acquisition of more preferences, potentially diminishing the relevance of recommended items, even when the assistant robot mentions one of the preferences in the takeover conversation. The conditions did not affect satisfaction levels because the enjoyment of talking to a robot and taking pictures strongly influenced satisfaction regardless of recommendation. We assume that the absence of differences in intrusiveness was because no specific instructions were given to the operators. Furthermore, the participants who directly participated in the conversation presented higher ratings to the questionnaire items other than intrusiveness. In cases of families with children, the parents answered the questionnaire. Therefore, the standard of the scores could be lower for adults. Additionally, factors such as subjects' age and gender influenced the results. In future experiments, results should be analyzed considering user attributes. Future experiments should carefully analyze results, taking into account user attributes.

During the busiest time of the day, a total of 17 conversations occurred in the two booths in 1 h. Considering that each conversation lasted approximately 6 min, the two booths were almost always occupied. The minimum time between the operator ending the previous conversation and starting the next, that is, from pressing the halting button to pressing the next request button, was 9 s. This could be attributed to the difficulty to understand the conversation history in this amount of time. Therefore, a summary of the dialogue is necessary. However, we did not evaluate this phenomenon. The degree of fatigue was not measured, but should be evaluated in the future. We analyzed 1 h of video during the busy time of the day and found that the operator used the buttons 2.1 times per conversation. Because the interaction per button lasted approximately 10 s, we can estimate that half of the conversation duration was occupied by this interaction. However, we did not evaluate system operability. Therefore, this should be considered in the future, including the decision on whether to make the system autonomous.

One limitation is the lack of investigation into the ease of conversation for the operator in a multiple-robot setting. While the operator in this experiment possessed reasonable conversational abilities, they were a novice when it came to presenting goods. However, in some instances, the conversation was smoothly initiated after the takeover conversation. Additionally, scenes were noted where the operator controlled the assistant robot by pushing buttons during the operation and responded to the assistant robot's utterances. These observations confirm that with minimal practice, the operator can generate reasonable recommendations. However, because the operator was instructed to speak freely, the quality and quantity of each subject's experience could have varied. Further experiments are needed to investigate the effects of using multiple robots and evaluate the contents of the operator's recommendation. Another limitation is that the results were obtained in a scenario with two booths for parallel conversations. Moreover, the subjective similarity model was based on only one person, which affected recommendation performance. As the number of booths increases, the operator's capacity to handle conversations and the time allocated per conversation may decrease. Future developments should focus on implementing a booth selection mechanism and an effective remote conversation strategy to enhance overall satisfaction. The upper number of booths can be predicted using the Erlang c formula [26], which allows us to calculate the maximum number of requests sent by the systems that an operator can cope with in one hour. Investigating the relationship between the number of booths and robots and cognitive load is crucial.

In this study, a simple natural language processing technique was used. Conventional language processing techniques such as intent classification using tools such as DialogFlow, RASA, and large language models may improve the natural language understanding capabilities and the overall user experiences.

VI. Conclusion

In this study, we introduced a multirobot dialogue system designed for parallel conversations facilitated by a single operator. A field experiment was performed to assess the operator's recommendation performance using the proposed dialogue system, achieving positive outcomes in terms of satisfaction and purchase intention, although the understanding capability and natural and intrusive presentations were not confirmed. These findings offer valuable insights for designing dialogue systems with a limited number of operators capable of handling simultaneous parallel conversations. Future work will involve a comprehensive evaluation of the proposed dialogue system's functionality with multiple robots, along with the development of strategies to enhance multirobot interaction and increase the number of feasible simultaneous conversations. We aim to improve the proposed dialogue system by applying large language models for language processing and effective takeover conversations.

Acknowledgment

The authors would like to thank NIFREL for use of facilities.

References

[1] M. Adam, M. Wessel, and A. Benlian, "AI-based chatbots in customer service and their effects on user compliance," *Electron. Markets*, vol. 31, no. 2, pp. 427–445, 2021.

[2] S. Kraus et al., "Customer service combining human operators and virtual agents: A call for multidisciplinary ai research," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 15393–15401.

[3] K. Inoue, D. Lala, K. Yamamoto, S. Nakamura, K. Takanashi, and T. Kawahara, "An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions," in *Proc. 21th Annu. Meeting Special Int. Group Discourse Dialogue*, 2020, pp. 118–127.

[4] Y. Muraki, H. Kawai, K. Yamamoto, K. Inoue, D. Lala, and T. Kawahara, "Semi-autonomous guide agents with simultaneous handling of multiple users," in *Proc. Int. Workshop Spoken Dialogue Syst. Technol.*, 2023, pp. 1–12.

[5] M. Shiomi, D. Sakamoto, T. Kanda, C. T. Ishi, H. Ishiguro, and N. Hagita, "A semi-autonomous communication robot: A field trial at a train station," in *Proc. 3rd ACM/IEEE Int. Conf. Hum. Robot Interact.*, 2008, pp. 303–310.

[6] K. Okuoka, Y. Takimoto, M. Osawa, and M. Imai, "Semi-autonomous telepresence robot for adaptively switching operation using inhibition and disinhibition mechanism," in *Proc. 6th In. Conf. Hum.-Agent Interact.*, 2018, pp. 167–175.

[7] T. Kawahara, N. Muramatsu, K. Yamamoto, D. Lala, and K. Inoue, "Semi-autonomous avatar enabling unconstrained parallel conversations–seamless hybrid of WOZ and autonomous dialogue systems," *Adv. Robot.*, vol. 35, no. 11, pp. 657–663, 2021.

[8] H. Kawai, Y. Muraki, K. Yamamoto, D. Lala, K. Inoue, and T. Kawahara, "Simultaneous job interview system using multiple semi-autonomous agents," in *Proc. 23rd Annu. Meeting Special Int. Group Discourse Dialogue*, 2022, pp. 107–110. [Online]. Available: https://aclanthology.org/2022.sigdial-1.12

[9] K. Hayashi et al., "Humanoid robots as a passive-social medium: A field experiment at a train station," in *Proc. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2007, pp. 137–144.

[10] T. Kubota, T. Isowa, K. Ogawa, and H. Ishiguro, "Development and verification of an onsite-operated android robot working cooperatively with humans in a real store," (in japanese), *The Trans. Hum. Interface Soc.*, vol. 22, no. 3, pp. 275–290, 2020.

[11] M. M. Reimann, F. A. Kunneman, C. Oertel, and K. V. Hindriks, "A survey on dialogue management in human-robot interaction," *J. Hum.-Robot Interact.*, 2024, just Accepted.

[12] P. Wolfert, N. Robinson, and T. Belpaeme, "A review of evaluation practices of gesture generation in embodied conversational agents," *IEEE Trans. Human-Mach. Syst.*, vol. 52, no. 3, pp. 379–389, Jun. 2022.

[13] K. Inoue, D. Lala, and T. Kawahara, "Can a robot laugh with you?: Shared laughter generation for empathetic spoken dialogue," *Front. Robot. AI*, vol. 9, 2022, Art. no. 933261.

[14] K. Zheng, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Supervisory control of multiple social robots for navigation," in *Proc. 8th ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2013, pp. 17–24.

[15] K. Zheng, D. F. Glas, T. Kanda, H. Ishiguro, and N. Hagita, "Supervisory control of multiple social robots for conversation and navigation," *Trans. Control Mech. Syst.*, vol. 3, no. 10, pp. 76–92, 2014.

[16] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Multiple-robot conversational patterns for concealing incoherent responses," *Int. J. Social Robot.*, vol. 10, no. 5, pp. 583–593, 2018.

[17] H. Sugiyama, T. Meguro, Y. Yoshikawa, and J. Yamato, "Avoiding breakdown of conversational dialogue through inter-robot coordination," in *Proc. 17th Int. Conf. Auton. Agents MultiAgent Syst.*, 2018, pp. 2256–2258.

[18] M. Shiomi, R. Hayashi, and H. Nittono, "Is two cuter than one? Number and relationship effects on the feeling of kawaii toward social robots," *PLoS One*, vol. 18, no. 10, 2023, Art. no. e0290433.

[19] Y. Okada, M. Kimoto, T. Iio, K. Shimohara, and M. Shiomi, "Two is better than one: Apologies from two robots are preferred," *PLoS One*, vol. 18, no. 2, 2023, Art. no. e0281604.

[20] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Nodding responses by collective proxy robots for enhancing social telepresence," in *Proc. 2nd Int. Conf. Hum.-agent Interact.*, 2014, pp. 97–102.

[21] C. T. Ishi, C. Liu, H. Ishiguro, and N. Hagita, "Evaluation of formant-based lip motion generation in tele-operated humanoid robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2012, pp. 2377–2382.

[22] K. Sakai, Y. Nakamura, Y. Yoshikawa, and H. Ishiguro, "Effect of robot embodiment on satisfaction with recommendations in shopping malls," *IEEE Robot. Automat. Lett.*, vol. 7, no. 1, pp. 366–372, Jan. 2022.

[23] K. Sakai, Y. Nakamura, Y. Yoshikawa, S. Kano, and H. Ishiguro, "Expressing robot's understanding of human preference based on successive estimations during dialog," *Int. J. Hum.–Computer Interact.*, pp. 1–22, 2023.

[24] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinite relational model," in *Proc. Conf. Assoc. Advance. Artif. Intell.*, 2016, pp. 381–388.

[25] F. F. Reichheld, "The one number you need to grow," *Harvard Bus. Rev.*, vol. 81, pp. 46–54, 2003.

[26] A. K. Erlang, "Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges," *Post Office Elect. Engineer's J.*, vol. 10, pp. 189–197, 1917.