https://doi.org/10.20965/jaciii.2017.p0675



**Paper:**

# Retaining Human-Robots Conversation: Comparing Single Robot to Multiple Robots in a Real Event

**Takamasa Iio, Yuichiro Yoshikawa, and Hiroshi Ishiguro**

Osaka University / JST ERATO
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
E-mail: {iio, yoshikawa, ishiguro}@irl.sys.es.osaka-u.ac.jp
[Received November 20, 2016; accepted May 2, 2017]

**In human-robot conversation in a real environment, low speech recognition and unnatural response generation are critical issues. Most autonomous conversational robotic systems avoid these issues by restricting user input and robot responses. However, such restrictions often render the interaction boring because the conversation becomes predictable. In this study, we propose the use of multiple robots as a solution for this problem. To explore the effect of multiple robots on a conversation, we developed an autonomous conversational robotic system and conducted a field trial in a real event. Our system adopted a button interface, which restricted user input within positive or negative intention, and maintained a conversation by choosing the most suitable of the prepared static scenarios. Through the field trial, we found that visitors who conversed with multiple robots continued their conversation for a more prolonged period, and their experience improved their impression on the conversation, in contrast to the visitors who conversed with a single robot.**

**Keywords:** human-robot interaction, multiple robots, social robot, field trial

## 1. Introduction

Robots that are able to talk to people have been used as an event companion to provide information or for promotional purposes at various places, such as museums [1, 2], expositions [3], receptions [4], shops [5], stations [6], and shopping malls [7]. Such robots are expected to attract visitors and create a positive impression regarding the topic that the robots would discuss about. To realize this purpose, one of the most important factors is the ability to maintain a conversation with visitors, which is defined as to communicate information between a speaker and an addressee via natural language. This is because robots that are unable to maintain a conversation cannot achieve the purpose, which is to provide information and to fulfill promotional purposes.

However, robots have presented difficulty in maintaining a conversation with people in a real environment because of technological issues, low speech-recognition accuracy, and unnatural response generation. These issues cause conversational collapse; eventually, people cannot concentrate on a conversation. Most human-robot interaction (HRI) studies in real environments have avoided these issues by two ways. The first is a remote operation, known as Wizard of Oz (WOZ) method. In this method, human operators play the role of speech recognition and response generation. Through this method, a natural conversation can be held; however, it has a cost on operator labor. Therefore, the use of the WOZ method in real applications is difficult. The second method is more reasonable than the WOZ method in real applications; it restricts user input and robot response. The autonomous conversational robotic systems of this method restrict the topic of a conversation, and accept specific user input that is related to the topic. The systems lead a conversation by choosing one of the scenarios that has been prepared in advance based on specific user input. Therefore, strong restrictions of user input and response options decrease conversational collapse. However, such strong restrictions are often boring to users because the conversation becomes predictable. Therefore, the users often lose the sense of having a conversation, which may feel as if they are engaging in a conversation, and finally abandon their efforts to maintain the conversation.

Our objective is to provide a method to improve the sense of conversation of the user and maintain a conversation for a longer period, even under restrictions in user input and response options in a real environment. For this purpose, we propose the use of multiple robots in a conversation. In the present work, multiple robots are defined as two or more robots. There are studies in which the potential of multiple-robot conversation has been suggested [8–10]. The aforementioned related works coupled with our central idea are introduced in Section 2. Furthermore, we challenged ourselves in conducting a field trial in a public interactive art event, in which visitors could converse with robots as they wished; in other words, they were free on how long they continued to a conversation.

As previously mentioned, autonomous conversational robotic systems realistically need to restrict user input and response options. Regarding this point, we decided to develop a system that would be quite substantially restricted because we wished to demonstrate the effects of using

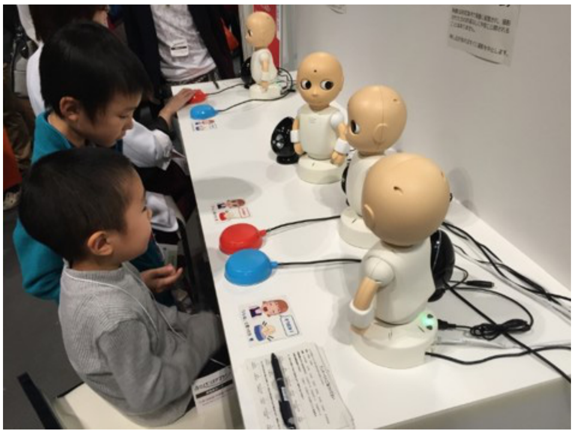

**Fig. 1.** Multiple robots system via button interface.

multiple robots in a realistic system. Our system had a static conversational model and prepared scenarios designed for the event as shown in Section 3, so that robots did not give nonsense responses. Instead of using speech recognition, we restricted user input by using a button interface, as shown in **Fig. 1**, to allow visitors to represent positive or negative intention (see Section 4). Although the use of buttons to express user intentions is different from the usual conversation style, we regard such communication as a conversation styles in the sense that a user and a robot communicate information each other. In Section 5, we will explain our field trial in a public exhibition event to evaluate the effect of using multiple robots in a conversation; the results will be presented in Section 5 and will be discussed in Section 6. Finally, we will conclude our study in Section 7.

## 2. Multiple Robots Conversation

Takahashi et al. developed mediation robots that determined utterances by searching comments in existing social media platforms and promoted conversation opportunity for humans [9]. Takahashi conducted a user study, where a user watched TV together with the mediation robots. The results showed that people did not increase the conversation opportunity; however, they tended to more enjoy watching TV with multiple robots than with a single robot. Moreover, Hayashi et al. developed robots as a passive social medium in which multiple robots conversed with one another [10]. They conducted a field experiment at a station to investigate the effects of the robots. As a result, they found that people were more likely to stop to listen to a conversation of two robots than to listen to one robot. The aforementioned works suggested positive effects of using multiple robots as a social medium; however, the researchers did not examine the effect that the use of multiple robots would have on a conversation.

There were a few studies that investigated the effects that the use of multiple robots would have on a conversation between humans and robots. Arimoto et al. reported that a person who conversed with two robots felt less ignored from the robots than a person who conversed with a single robot [8]. These results suggested that using multiple robots may positively affect certain nonverbal aspects of a conversation; nevertheless, it was not clear whether humans felt as if they held a conversation, which is highly influenced by verbal aspects. In other words, it is unknown whether people feel that they can hold a good conversation by using multiple robots.

In this work, we aim to explicitly evaluate human-robot conversation from the viewpoint of conversation quality.

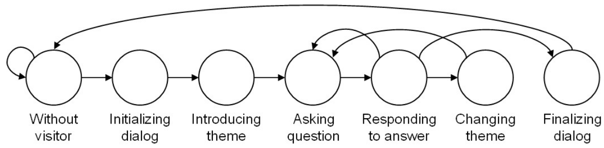


**Fig. 2.** Dialog-state model.

## 3. Conversation Design

We intended to design a conversation between a person visiting a social event and robots that are exhibited in the event. The robots do not use artificial speech recognition to avoid conversational collapse. This means that the robots cannot answer a question that has originated from a person. Therefore, we modeled the conversation as follows: the robots would chat with each other before a visitor would come (without-visitor state). The robots would greet, and would then start talking to the visitor when he/she would approach (initializing-dialog state). The robots would introduce a theme (introducing-theme state), and would continue to discuss certain topics of the theme (asking-question state and responding-to-answer state). If time permits, the robots may change the theme (changing-theme state); otherwise, they close the conversation (finalizing-theme state). This conversation model is illustrated in **Fig. 2**. We will provide more details on each state in the sections below.

### 3.1. Without-Visitor State

The without-visitor state is a state in which no visitor is in front of robots. In this state, if the robots do not do anything, a passing visitor would not pay any particular attention to the robots. This causes the loss of an opportunity for conversation. Therefore, the robots should do something, even if it is minute. In this study, until a visitor comes, multiple robots would have a chat among each other; a single robot would talk to itself at intervals of a few seconds because it is unnatural to continue talking without a pause. It should be noted that multiple robots present an advantage in creating a natural scene at the without-visitor state. The without-visitor state would continue until a visitor would approach. The arrival of a visitor was detected via buttons; in other words, the without-visitor state transitioned to the next state, namely the initializing-dialog state, when a visitor pushes either buttons.

### 3.2. Initializing-Dialog State

The initializing-dialog state is a preparatory state for a conversation. When a visitor approaches, he/she is not ready for a conversation with robots. The robots should greet and instruct how to converse so that the visitor may naturally enter a conversation. Below, we provide an example ($R_a$ and $R_b$ represent the robots, and H represents the visitor. $R_a \rightarrow$ H represents that $R_a$ is speaking to H.):

$R_a \rightarrow$H: Hello.
$R_b \rightarrow$H: If you agree with my idea, push that blue button. Otherwise, push that red button.

### 3.3. Introducing-Theme State

In the introducing-theme state, a theme of topics is provided, on which the robots plan to converse. Typically, people tend to set a theme for their talk prior to becoming familiar with the details of each topic. An example of scripts in this state is provided below:

$R_a \rightarrow$H: Hey, will you listen to my problem?
$R_b \rightarrow R_a$: Do you think what happened all of a sudden?
$R_a \rightarrow R_b$: I just wanted to say that no matter what.

### 3.4. Asking-Question State

The asking-question state is a state in which robots ask a question about a topic of a theme that has been selected in the previous state (either the introducing-theme state or the changing-theme state) to a visitor. The question would be posed in such a manner such that it could be answered positively (yes) or negatively (no). The following script is an example of this state:

$R_a \rightarrow$H: Sometimes, our doctors move our joints.
$R_a \rightarrow R_b$: Isn't that painful?
$R_b \rightarrow R_a$: Yeah, very.
$R_a \rightarrow R_b$: Humans need to handle robots more gently.
$R_a \rightarrow$H: Do you think so?

### 3.5. Responding-To-Answer State

The responding-to-answer state is a state in which robots respond to a visitor answer. A visitor answer can be classified as in the following: positive, negative, and timeout. In fact, robots wait for visitor input; however, they resume a previous conversation (e.g., timeout), if a visitor indicates that he/she is neither positive nor negative for long time because it would be unnatural to keep the visitor waiting. Below, we describe examples that correspond to positive, negative and timeout, respectively:

[Positive]

$R_a \rightarrow$H: That's so you! You are kind.
$R_b \rightarrow R_a$: If all the people are like you, it will be easier for us robots to live.

[Negative]

$R_a \rightarrow$H: That's horrible.
$R_b \rightarrow R_a$: I think I'm not motivated to help humans.

[Timeout]

$R_a \rightarrow$H: Well, if something gets broken, it can be fixed though.
$R_b \rightarrow R_a$: Just replace the broken part, and that's it.

After the responding-to-answer state has been completed, it transitions to the asking-question state; however, it transitions to the changing-theme state or the finalizing-dialog state, according to the conditions. The conditions depend on the number of questions asked. We can set a condition, for example, the responding-to-answer state transitions to the asking-question state until the count reaches to three times.

### 3.6. Changing-Theme State

The changing-theme state is a state in which a new theme is provided. Robots should maintain the same theme for a while in a conversation because of its consistency. However, it is unnatural to continue conversing on only one theme. Therefore, the robots need to change the current theme and replace it with a new one. The timing of changing depended on the number of topics on the current theme told by robots. The following script is an example of a changing-theme state.

$R_a \rightarrow R_b$: Well, I've been thinking about my problem.
$R_b \rightarrow R_a$: What happened all of a sudden?
$R_a \rightarrow R_b$: I want you to know how hard it is to be a robot.
$R_a \rightarrow$H: It's no big deal. Please listen to me, just a little.

### 3.7. Finalizing State

The finalizing-dialog state is a state for ending the conversation with a visitor. Robots working at a social event are expected to communicate with several visitors. Therefore, they should end a conversation following an appropriate timing, so that they may hold a conversation with the next visitor. For example, the timing of ending the conversation may be after the completion of conversing about two themes. An example of such a scenario is presented below:

$R_a \rightarrow R_b$: Hey, time is almost up, right?
$R_b \rightarrow R_a$: Yes.
$R_b \rightarrow$H: It's time to switch to the next human.
$R_a \rightarrow$H: Great talking to you. See you. Bye.

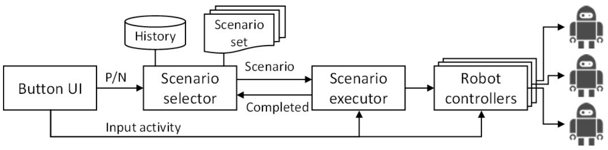


**Fig. 3.** System architecture.

It should be noted that there is a problem in this dialog state model, namely that robots continue talking even when a visitor leaves during a conversation because the model does not consider whether a visitor is in front of the robots or not. Actually, every state should be able to transition to the without-visitor state when a visitor moves far away. However, we did not implement this concept because we did not employ any sensors for detecting visitors who would step away from the robots. In future works, we aim to solve this problem by using human detection sensors such as Kinect.

## 4. Autonomous Conversational System

**Figure 3** illustrates the architecture of the developed system. When a visitor pushes either the positive or the negative button, his/her input is sent to the scenario-selection module. The scenario-selection module chooses one scenario out of the scenario set, and then sends the scenario to the scenario executor module. The scenario executor module parses the scenario, and sends the sequence of commands of speech, gesture, and gaze to the robot controllers, according to the scenario. After the scenario has been executed, the scenario executor module sends back a completion signal to the scenario selector module. Then, the scenario selector module chooses the following scenario. This means that the system can hold a conversation through a sequence of scenarios. Visitor input is used as positive or negative information, as well as a trigger that resumes the process of the scenario executor module that awaits the visitor input. Furthermore, it is used as a trigger via which robots can pay attention to a visitor. In the following sections, we will explain in more details.

### 4.1. Scenario

The scenario is a data file which contains a description of how robots behave and the attributes that are used for scenario selection. It is divided into two elements: the scenario attributes and the behavior definition. **Fig. 4** illustrates a sample scenario.

#### 4.1.1. Scenario Attributes

A scenario attribute contains four variables: the dialog state, the user answer, the theme, and the topic. These variables are used as conditions for choosing a scenario. The dialog state, as described in Section 3, has seven values. The user answer is the visitor input to which the scenario corresponds. The user answer has three values:

```
DialogState:  "RespondingAnswer",
UserAnswer:   "Positive",
Theme:        "RobotFrustration"
Topic:        "SpeechRecognition"
Behavior:
 0, H, "That's so you! You are kind."
 1, H, "If all the people are like
       you, it will be easier for us
       robots to live."
 0, 1, "Yeah!"
```

**Fig. 4.** Scenario sample.

positive, negative, and timeout. The theme and the topic are similar to a label that features the content of a scenario. The theme is a more general concept than the topic. For instance, the sample scenario that is shown in **Fig. 4** is chosen if the current dialog state is the responding-to-answer state, the visitor input is positive, and the robots talks about speech recognition in the context of robotic frustration.

#### 4.1.2. Behavior Definition

The behavior definition includes the timing of a robot to converse with another robot. As shown **Fig. 4**, we describe the identity of the speaker-robot (the first item) and the identity of the target robot or that of the target human to which the speaker-robot talks to (the second item), and how many seconds later an utterance begins from the completion of the last utterance (the third item), as well as the content of the utterance (the forth item). When a robot asks a question to a visitor, a wait-for-input tag can be added at an utterance in a scenario to indicate that the robots are waiting for visitor input. For example, the first behavior of the sample scenario means that robot $R_a$ says "That's so you! You are kind" to a visitor after 1.0 s from the last utterance.

This scenario defines only the behavior of the speaker. The behaviors of the listeners are generated automatically based on a predefined rule. While a speaker talks, the other robots (listeners) look at the speaker, and nod with a random time lag (the range is from 0.0 to 1.0 s) at the end of its utterance.

### 4.2. Scenario Selector

In our architecture, there is always one "scenario" under execution. A scenario is a data file that controls the utterance, the gesture, and the gaze. In the system, each scenario corresponds to a theme, topic, and visitor input in every dialog state. The scenario selector is a dialog state transition-based system, in which the current scenario is periodically matched with the pre-implemented dialog state network (**Fig. 2**). The scenario selector chooses one scenario out of the scenario set, according to the current dialog state, visitor input, theme, topic, and history. The scenario set is a storage that contains all scenarios. The history is the place where the executed scenario is stored. There are two reasons for using history. One is that a robot does not execute a scenario that has already been


**Fig. 5.** Button interface.

executed. It would be unnatural to repeat the same speech in a conversation. The other reason is to check the changing theme condition and the closing dialog condition.

### 4.3. Scenario Executor

The scenario executor receives a scenario from the scenario selector, and then parses the scenario. As a result of parsing, the scheduled sequence of commands—which are Say, Gesture, and Look—to be sent to the robot controllers is generated. The Say command makes a robot speak a sentence. The Gesture command makes a robot perform gestures, such as nodding or bowing. The Look command makes a robot look toward a predefined place. In this study, because we fixed the human position, the position of each robot, such as the static label approach, did not significantly matter.

The scenario executor sends back a signal after a scenario is completed. The signal is used as a trigger for the scenario selector to choose the next scenario.

### 4.4. Robot Controller

The robot controller is a module for the operation of a robot based on a command received from the scenario executor. The robot controller extracts a sequence of commands that move the motors of each axis according to the gaze or the gesture command. In addition, it extracts a sequence of commands to play an audio file, and moves a motor for the opening and the closing of the mouth as per the Say command.

### 4.5. Button User Interface

The button user interface has three functions. Each function is executed depending on the following dialog states.

(1) To represent user intention: We employed two physical buttons as user input devices, each of which represents the positive and the negative input (**Fig. 5**). When either of buttons is pushed after the question posed by the robot to a user (the asking-question state shown in **Fig. 2**), the dialog state transitions to the responding-to-answer state; then, a scenario that is suitable for the button input is selected and executed. Next to the buttons, we placed pictures that correspond to a positive or negative input in order for a visitor to intuitively understand the meaning of buttons.

(2) To detect a user who is motivated to talk to robots: When either of buttons is pushed during the robot–robot conversation (without-visitor state, as shown in **Fig. 2**), the robots stop talking and look at the user; then, they start a conversation with the user. In other words, the state transitioned to the initializing-dialog state. This means that the buttons are used as a detecting device.

(3) To pay attention to a user who pushes a button: When either one of the buttons is pushed at any other state other than the abovementioned states, the robots look at the user if this gaze moving do not conflict with gaze moving prescheduled in the scenario executed at that time. The robots return their gaze 2 s after the first gaze moving, if the second gaze moving does not conflict with the prescheduled gaze moving. Thanks to this function, a user is expected feel that the robots are aware of him/her.

It should be noted that we may use different interfaces, such as a touch panel or a keyboard, if the devices accept two types of input.

#### 4.5.1. Positive, Negative, and Timeout

The button user interface module sends the visitor input to the scenario selector module. The input is of two types: positive and negative. Although we used physical buttons in this study, we may use any device that can accept two or more types of input, such as a keyboard or a touch panel.

#### 4.5.2. Input Activity as a Trigger

Furthermore, the button user interface module sends the visitor input activity to the scenario executor and robot controllers. In the scenario executor, the input activity is used as a trigger to resume a scenario that awaits visitor input. In robot controllers, the activity is used as a trigger for robots to pay attention to user input. As it is not friendly for robots not to respond to visitor input, we configured robots to look toward the visitor, when the visitor provides input.

### 4.6. Robot

In our system, we employed CommU (**Fig. 6**), which is a social conversation robot developed by VSTONE. This robot is a desktop size, which is with a height of 304 mm, a width of 180 mm and a depth of 131 mm, and a weight of 938 g. CommU has three degrees of freedom (DOFs) at its waist, three DOFs at its neck, and two DOFs at each eye, thus being able to flexibly control its gaze, as shown **Fig. 6**. This flexible gaze control is important for human-like social behaviors, such as turn taking in a conversation, establishing engagement via eye contact, and expressing its attention. The robot has a camera at the center of the forehead, a microphone on the left side of the chest and a speaker at the center of the chest; however, these devices were not used in this experiment.

## 5. Field Trial

### 5.1. Hypothesis

Based on past research [8–10], we hypothesized that multiple robots can continue a conversation with visitors

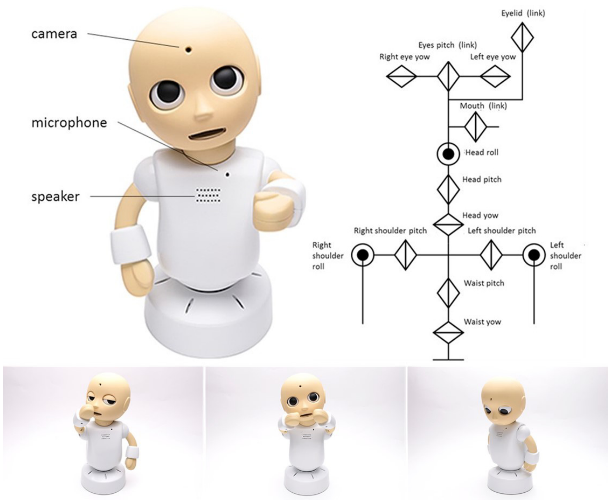


**Fig. 6.** The robot used in this experiment, called CommU (above) and examples of its nonverbal expressions (below).

for a longer period than a single robot. Moreover, regarding the conversational impression, multiple robots will improve the following, compared to the single robot;

(1) Sense that a conversation is established with robots.

(2) Sense that their opinions are conveyed to robots.

(3) Sense that their thoughts on the contents that have been provided by the robots were meaningful.

To investigate the validity of the hypothesis, we conducted a field trial, where multiple robots and a single robot conversed with visitors in an interactive art event.

### 5.2. Scenario

We had to select a theme for the scenarios considering the two following options: (1) the theme has both a positive perspective and a negative perspective, and is an open question, and (2) the theme is related to robots. The first standard is considered because each robot is attributed one different perspective. We considered that such a division of roles would help multiple robots to produce a powerful effect. The second standard was considered because it would not be natural that robots would converse on topics that would not be related to robots, in a public event.

Based on the above two standards, we created the following three themes; (a) a robot agrees with nursing care for the elderly by robots, (b) a robot disagrees with nursing care for the elderly by robots, and (c) a robot becomes frustrated with being robot. Regarding the first and second themes, both the positive and the negative perspective are considered, and the themes are related to the robot itself. The difference between the first theme and the second theme is whether the topics of each theme are positive or negative. The third theme contains a positive perspective and a negative perspective as well; for example, a robot can continue working without a recess, as long as there is electricity; however, its body becomes too hot owing to the electricity.

**Fig. 7.** Field trial at the Tokyo Design Week event.

We created five topics for each theme. After starting a conversation with a user, the robots randomly selected one theme and three topics within the theme. The time allocated to one topic was approximately from 1 to 1.5 min. The total time of a conversation, including the initialization and the finalizing, was approximately five minutes if the robots conversed on three topics.

### 5.3. Procedure

We conducted a field trial during the Tokyo Design Week, which was an interactive art event, for three days. Our developed system operated from 11:00 to 20:00 (from 11:00 to 16:00 on the final day). The system was installed at the site of the interactive robot section (**Fig. 7**).

Visitors of the event could freely experience the system without the permission of the exhibitor. Visitors who participated in the trial sat on a chair in front of the center of one of the robots. When a visitor group—such as a family—approached, one of its members would sit on the chair and the others would sit on extra chairs or stood next to or behind the seated one. The member sitting on the center chair almost always pushed the buttons; however, other members rarely did.

Visitors were instructed to push a blue button if they agreed with the opinion of the robot, or to push a red button if he/she did not agree. The instruction was offered by an exhibitor, and was illustrated through picture cards as well (**Fig. 5**), to ensure that visitors would be able to easily understand. They could freely behave and exit from a conversation anytime. To provide an opportunity for other visitors to experience the conversation, the number of scenarios initiated by the robots was coordinated in a manner that the conversation time would be limited to approximately 5 min. The settings were limited because of the field trial, and do not restrict our multiple-robots conversation design. We randomly chose visitors who experienced the system, and we asked them to answer a questionnaire.

### 5.4. Conditions

To verify our hypothesis, we made two conditions, namely a single-robot and a multiple-robots condition. In the single-robot condition, one robot conversed with one visitor. When there was no visitor in front of the robot, it talked to itself. In the multiple-robots condition, three robots conversed with a visitor. The reason for using three robots instead of two robots is because it would be easy to maintain a consistent position for each robot during a conversation. By using three robots, we could assign different positions in a topic to each of them. For example, in the topic of nursing care by robots, the first robot agreed that robots care for seniors; however, the second robot disagreed, and the third robot maintained a neutral position. Of course, a single robot can well explain claims of the aforementioned positions; nevertheless, it is easier for three separate robots to continue to claim their respective positions than it is for a single robot. When no visitor was in front of the robots, they talked to each other.

To compare these conditions fairly, we created almost the same scenarios for each condition. Their differences were in terms of the manner of expressing their claims or their responses to another robot. In the multiple-robots condition, the robots directly stated their claim; for example "I think that the nursing care by robots is a very good thing." or "Robots should not care for seniors, I think." Contrary to this direct claim, a single robot indirectly stated its claims; for example, "Someone thinks that the nursing care by robots is a very good thing" or "There are opinions that robots should not care for seniors." The contents of the scenarios as well as the total times of the scenarios were almost same for both conditions.

### 5.5. Measurement

#### 5.5.1. Conversation Time

The impression that a conversation makes on the visitor, as well as whether the behavior of a visitor changed or not, are both interesting points. Social robots are required to prolong the conversation time, so that robots can provide more information or promote an advertisement. From this standpoint, it is interesting to compare the conversation time of each condition. To measure the conversation time, we recorded conversations with video cameras, and we manually annotated the video offline. The conversation times were coded as the duration from the time when the robot said "hello" to the time that the visitor withdrew, or the robot said "goodbye." It should be noted that even if a visitor withdrew from a conversation, the robots could not stop the conversation because they did not have a means to detect the visitor position. Therefore, there were cases in which visitors came in the middle of a conversation with the previous visitor. We did not include these visitors because the results would not be correct.

#### 5.5.2. Questionnaire

We asked the following three questions based on our hypothesis:

*Q1. Did you feel that you had a conversation with the robot(s)?*

*Q2. Did you feel that your opinion was conveyed to the robot(s)?*

*Q3. Did you feel that you deepened your thoughts on the topics presented by the robot(s)?*

All questionnaire items were evaluated on a one to seven point scale, where seven represented the most positive opinion. The first question aimed to verify whether the sense of conversation of the visitor changed or not. The second question aimed to compare the two conditions in terms of how well were the opinions–positions communicated across. The third question was asked with the purpose to reveal whether the sense of contemplation of the visitor was different between the two conditions.

## 6. Results

### 6.1. Conversation Time

Through video coding, we obtained the data of 139 participants for the single-robot condition and the data of 148 participants for the multiple-robot condition. The means of their conversation times were 160.5 s (SD = 91.9) for the single-robot condition and 200.7 s (SD = 90.9) for the multiple-robot condition, as shown in **Fig. 8(a)**. As the data do not indicate a normal distribution, as shown **Fig. 8(b)**, we analyzed them via the Mann–Whitney U test. The test revealed a significant difference between these conditions ($z = 3.33$, $p = .0013$). These analysis results indicate that participants who conversed with multiple robots joined the conversation for a longer period than those who conversed with a single robot.

Furthermore, as shown in **Fig. 8(b)**, approximately half of the visitors who conversed with multiple robots continued a conversation for an additional 4 min. The scenarios that we used in the field trial lasted for approximately 5 min; more specifically, the main topic of the conversation would last for 4 min. This means that the visitors maintained a conversation until the robots finished talking on the main topic.

### 6.2. Questionnaire About Conversational Impression

Through this questionnaire survey, we obtained data from a total of 176 participants. **Table 1** lists the breakdown of participants, categorized by age and sex. Most of our listeners were in the 10–50 demographic. The ratio of males to females for the two conditions was quite different because the participants in this field trial were free to choose either one of the two conditions. We will consider the effect of gender distinction in the Discussion section, after presenting the results of questionnaire survey.

#### 6.2.1. Sense of Conversation

The mean scores of the first question that enquires upon the sense of conversation of the participant were 4.55 (SD = 1.71) for the single-robot condition and 5.06 (SD = 1.41) for the multiple-robots condition. The boxplot of the scores is shown in **Fig. 9** (the left graph). We analyzed these data via the Mann–Whitney U test, and the results indicated a significant difference between these conditions ($z = 2.045$, $p = .041$). The analysis results indicate

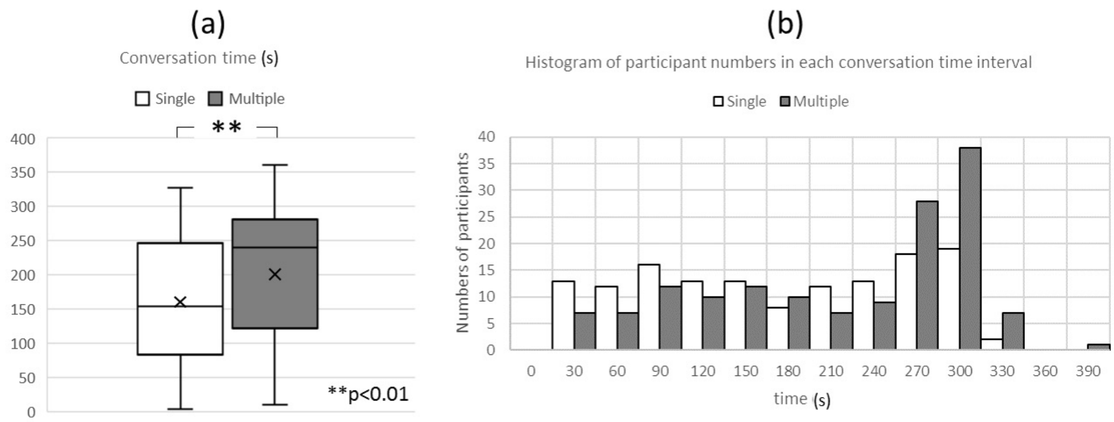


**Fig. 8.** Boxplot (a) and histogram (b) of conversation times in each condition.

**Table 1.** Demographic of participants.

| | Single | | Multiple | |
|---|---|---|---|---|
| | Male | Female | Male | Female |
| 0–9 | 4 | 5 | 5 | 4 |
| 10–19 | 3 | 6 | 4 | 12 |
| 20–29 | 7 | 28 | 9 | 15 |
| 30–39 | 3 | 10 | 12 | 10 |
| 40–49 | 2 | 13 | 6 | 5 |
| 50–59 | 6 | 1 | 3 | 1 |
| 60–69 | 0 | 0 | 1 | 0 |
| 70– | 0 | 0 | 0 | 1 |
| Total | 25 | 63 | 40 | 48 |

that participants who conversed with multiple robots felt more strongly that they could converse with these robots than those who conversed with a single robot.

#### 6.2.2. Sense of Conveying Opinion

The mean scores of the second question that enquires upon the sense of conveying opinion were 4.33 (SD = 1.54) for the single-robot condition and 5.06 (SD = 1.41) for the multiple-robots condition. The boxplot of the scores is shown in **Fig. 9** (the middle graph). The Mann–Whitney U test revealed a significant difference between these conditions ($z = 3.33$, $p = .0013$). The analysis results indicate that participants who conversed with multiple robots felt more strongly that they could convey their opinion to these robots than those who conversed with single robot.

#### 6.2.3. Sense of Deepening Thought

Owing to two participants not completing the third question, the data were obtained for 87 visitors, for each condition.

The mean scores of the third question that enquires upon the sense of deepening thought of the participant were 4.53 (SD = 1.31) for the single-robot condition, and 5.06 (SD = 1.24) for the multiple-robots condition. The boxplot of the scores is shown in **Fig. 9** (the right graph). The Mann–Whitney U test revealed a significant difference between these conditions ($z = 3.33$, $p = .0013$). The analysis results indicate that participants who conversed with multiple robots felt more strongly that they could deepen their thoughts on the topic presented by the multiple robots than those who conversed with single robot.

## 7. Discussion

### 7.1. Gender Distinction

This field trial was held in a large interactive art event. As visitors were free to look at and explore any exhibit, we could not control them in our trial. As a result, our data contained certain gender distinctions. Therefore, we analyzed whether the gender distinction biased our results. In this analysis, we did not focus on the video data; we only focused on the questionnaire results because there were indistinctive cases owing to mixed groups of male and female visitors, such as couples or families.

The gender-segregated results of questionnaire are listed in **Table 2**. A two-way analysis of variance (ANOVA) is commonly used to analyze such data; however, we used the Mann–Whitney U test for each condition because the variances of each group were different. The analysis results revealed that there was a marginally significant difference between male and female visitors for the multiple-robot condition in Q2 ($z = 1.76$, $p = .078$). Although significant differences were not observed between male and female visitors, in **Table 2**, we may observe the tendency that males rated a conversation with multiple robots higher than females. These data are interesting because they suggest a new hypothesis, in which males prefer to converse with multiple robots more than females do.

Past studies on human-robot interaction have reported that there are gender differences in user impressions of robots. For example, Kuo et al. found that elderly men had a more positive attitude toward robots in healthcare than women [11]. Schermerhorn et al. reported that men tended to think of a robot to be more human-like, and were

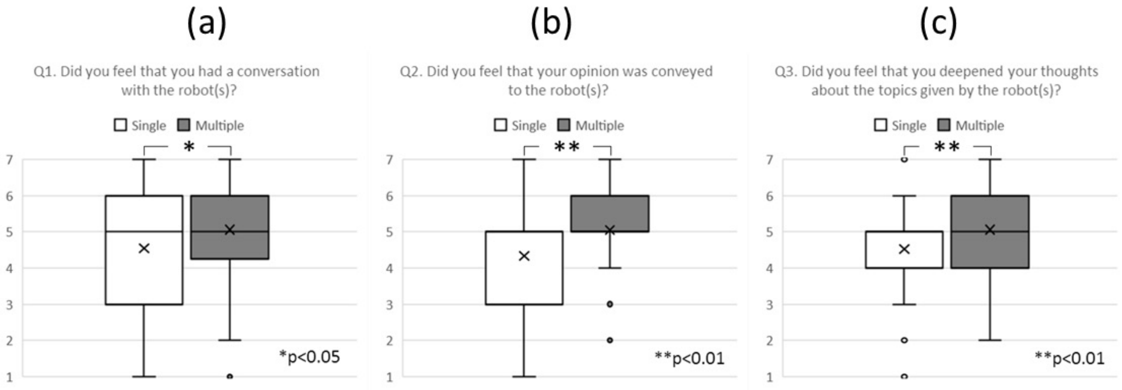


**Fig. 9.** Results from the questionnaire.

**Table 2.** Gender-separated results from the questionnaire.

| | Q1 | | Q2 | | Q3 | |
|---|---|---|---|---|---|---|
| | Single | Multiple | Single | Multiple | Single | Multiple |
| Male | 4.32 | 5.25 | 4.36 | 5.33 | 4.56 | 5.2 |
| Female | 4.63 | 4.90 | 4.32 | 4.81 | 4.52 | 4.94 |

more socially facilitated in a task than women [12]. These studies suggest that men tend to regard robots as being human-like and social beings more than women do. The results that multiple robots were evaluated higher by men than women in our trial may be attributed to the fact that men tended to consider robots as social beings.

Moreover, nonverbal features, such as a sense of the gaze and presence of the robot, may be related to the gender differences of our results as well. It has been known that men prefer to be approached from the side than from the front, and women prefer to be approached from the front than from the side [13]; and in human-robot interaction, a similar phenomenon has been observed [14]. Multu et al. rated a robot more positively in a conversation when they were looked at more; in contrast, women rated it more positively when they were looked at less [15]. Considering these past works, multiple-robot conversation might be preferred because in total, multiple robots would look at user face more than a single robot would; moreover, they spoke to the user from his/her front, as well as from his/her side.

In fact, there were not enough data to conclude gender effects on multiple-robots conversation because we did not focus on gender difference in this field trial. We would need to conduct more experiments for the investigation of the relation between gender difference and the use of multiple robots in a conversation.

### 7.2. Effect of Multiple Robots

The conversation times and questionnaire results showed that multiple robots improved the quality of conversations more than a single robot. We speculate that the reason may be the dynamics of interaction generated by multiple robots.

From video observations, we found that visitors and robots in multiple-robots conversations seemed more dynamic than those in a single-robot conversation. Visitors conversing with a single robot fixed their gaze on the single robot. The single robot fixed its gaze on the visitors, as well. Contrary to these static behaviors, visitors conversing with multiple robots moved their gaze to look at a speaking robot. In a similar manner, the robots moved their gaze at every turn taking. These nonverbal behaviors have the potential of creating an impression that a conversation is natural [16, 17]. As a result, such dynamic interaction might help the visitors engage in a conversation.

Furthermore, as well as nonverbal behaviors, verbal communication in a multiple-robots conversation sounded more dynamic than that in a single-robot conversation. Good speakers can provide an interesting topic, as well as various aspects of the topic, including positive or negative opinions. In our field trial, a single robot would express opinions from the positive, negative, and neutral viewpoints of a topic; for example, "Some people agree with nursing care by robots, whereas others are concerned for their safety." This seemed to be similar to a lecture by a teacher; therefore, the conversation might be boring for visitors. Contrary to such a boring conversation, multiple robots expressed their own position by providing consistent opinions regarding a theme; for example, one robot agreed with robotic nursing care, another disagreed, and the other robot exhibited a neutral attitude toward it. The form of this talk seemed similar to a debate. The form could develop an atmosphere that would make visitors deeply think about the opinions that the robots expressed.

In conclusion, we consider that the dynamics of nonverbal behaviors and verbal communication in multiple-robots conversation would contribute to improve the quality of a conversation.

### 7.3. Effect of Quantity of Robots

Regarding the effect of the quantity of robots that participate in a conversation, we consider that it is important to either use a single robot or two or more robots because two or more robots should make the expression of robot's attention through its gaze richer than a single robot. During a conversation, a single robot continues to pay attention to only one user; however, two or more robots can shift their attention to either a user or another robot. Such a shifting attention from two or more robots is expected to improve the feelings of the user toward the fact that robots are making efforts to listen to his/her speech more than a single robot with a fixed attention.

Regarding the comparison between two and three robots, we consider that three robots are superior to two robots in the case of a debate-like conversation, which would enable a user to contemplate on a topic. If three robots participate in a conversation on a certain topic, each can assume one of the three basic roles, namely the positive, the negative, and the neutral role. A user would develop empathy for any one of these three roles. A user who is feels empathy for one of the robots may become more immersed in a conversation.

The use of four or more robots would complicate a conversation; this would make it difficult for a user to understand which robots are in which position regarding a topic. In addition, such a predominantly robot situation would give a user a formidable impression. Therefore, we consider that increasing the number of robots may not always improve the sense of conversation of the user.

### 7.4. Implication

The contribution of our study is that holding a conversation with multiple robots resulted in a prolonged conversation and improved human impression of the conversation more than holding a conversation with a single robot. Furthermore, our findings were based on human-robot interaction in a real exhibit event, where participants were free to look and explore the exhibits. Using multiple robots will become one of the most efficient manners to fulfill promotional or advertising purposes in an event.

Moreover, our findings will be useful in the field of educational robots [18, 19]. For such robots, it is important to continue toward conversations with children, and to make them consider a topic mentioned by the robots. According to our findings, multiple robots are considered appropriate for this requirement. For example, discussions with multiple robots that have different perspectives among each other will be more exciting and attractive than a lecture from single robot.

### 7.5. Limitation

In this field trial, we did not use speech recognition and response generation technologies because practical performance was of high priority. However, using multiple robots has the potential of such technologies being applied in conversations among robots. Even if the speech recognition accuracy and the naturalness of the response generation reach the human level, robots will not avoid failures of speech recognition or response generation. If these failures occur, a conversation would become incoherent. According to Iio [20], although a single robot does not perform well in recovering from broken coherence, multiple robots are able to maintain a coherent conversation by using a special turn-taking pattern between the robots, even if the robots would misinterpret human intention. Therefore, we consider that using multiple robots might be more efficient in a conversation in which speech recognition or response generation would be used, rather than in our scenario-based conversation with a button interface.

As we did not use any sensors, our system could not detect participants that withdrew from the conversation. The robots could not stop a conversation when participants disengaged. Therefore, in certain instances, when participants approached, the robots would be discussing the scenario of the previous participant. Although this was an issue as a practical system, it did not bias our findings because our video analysis focused on participants who engaged from the beginning of a conversation.

## 8. Conclusion

We developed an autonomous conversational robotic system that controls multiple robots to lead a conversation based on human response through a button interface. Our field trial in a real event revealed that multiple robots can maintain a conversation with visitors for a longer period than a single robot. Approximately half of the visitors who conversed with multiple robots maintained a conversation for an additional 4 min; this means that they were engaged until the end of the conversation. Furthermore, visitors who conversed with multiple robots had better impressions of their conversation than those who conversed with a single robot. The findings will contribute to practical applications, such as advertising companion robots or educational robots in schools.

**Acknowledgements**

This work was supported by the JST ERATO, grant number JPM-JER1401, Japan.

**References:**

[1] W. Burgard, et al., "The Interactive Museum Tour-Guide Robot," National Conf. on Artificial Intelligence (AAAI1998), pp. 11-18, 1998.

[2] S. Thrun, et al., "Minerva: A Second-Generation Museum Tour-Guide Robot," IEEE Int. Conf. on Robotics and Automation (ICRA1999), pp. 1999-2005, 1999.

[3] R. Siegwart, et al., "Robox at Expo.02: A Large Scale Installation of Personal Robots," Robotics and Autonomous Systems, Vol.42, pp. 203-222, 2003.

[4] R. Gockley, et al., "Designing Robots for Long-Term Social Interaction," IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2005), pp. 1338-1343, 2005.

[5] H. M. Gross, et al., "Shopbot: Progress in Developing an Interactive Mobile Shopping Assistant for Everyday Use," IEEE Int. Conf. on Systems, Man, and Cybernetics (SMC2008), pp. 3471-3478, 2008.

[6] M. Shiomi, et al., "A Semi-Autonomous Communication Robot – a Field Trial at a Train Station –," ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI2008), pp. 303-310, 2008.
[7] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "A Communication Robot in a Shopping Mall," IEEE Trans. on Robotics, Vol.26, pp. 897-913, 2010.
[8] T. Arimoto, Y. Yoshikawa, and H. Ishiguro, "Cooperative Use of Multiple Robots for Enhancing Sense of Conversation without Voice Recognition," SIG-SLUD, B5(2), pp. 76-77, 2015 (in Japanese).
[9] T. Takahashi, M. Kanbara, and N. Hagita, "A Social Media Mediation Robot to Increase an Opportunity of Conversation for Elderly: Mediation Experiments Using Single or Multiple Robots," Technical Committee on Cloud Network robotics (CNR), Vol.113, No.84, pp. 31-36, 2013 (in Japanese).
[10] K. Hayashi, D. Sakamoto, T. Kanda, M. Shiomi, S. Koizumi H. Ishiguro, T. Ogasawara, and N. Hagita, "Humanoid robots as a passive-social medium – a field experiment at a train station," ACM/IEEE 2nd Annual Conf. on Human-Robot Interaction, pp. 137-144, 2007.
[11] H. I. Kuo, et al., "Age and gender factors in user acceptance of healthcare robots," The 18th IEEE Int. Symp. on Robot and Human Interactive Communication, RO-MAN 2009, pp. 214-219, 2009.
[12] P. Schermerhorn, M. Scheutz, and R. C. Crowell, "Robot social presence and gender: Do females view robots differently than males?" Proc. of the 3rd ACM/IEEE Int. Conf. on Human robot interaction, pp. 263-270, 2008.
[13] D. J. Fisher and D. Byrne, "Too close for comfort: Sex differences in response to invasions of personal space," J. of Personality and Social Psychology, Vol.32, No.1, pp. 15-21, 1975.
[14] S. D. Syrdal, K. Dautenhahn, S. Woods, L. M. Walters, and L. K. Koay, "'Doing the right thing wrong' – Personality and tolerance to uncomfortable robot approaches," The 15th IEEE Int. Symp. on Robot and Human Interactive Communication, ROMAN 2006, pp. 183-188, 2006.
[15] B. Mutlu, J. Forlizzi, and J. Hodgins, "A storytelling robot: Modeling and evaluation of human-like gaze behavior," 2006 6th IEEE-RAS Int. Conf. on Humanoid robots, pp. 518-523, 2006.
[16] M. Imai, T. Ono, and H. Ishiguro, "Physical relation and expression: joint attention for human-robot interaction," IEEE Trans. on Industrial Electronics, Vol.50, pp. 636-643, 2003.
[17] M. Shimada, Y. Yoshikawa, M. Asada, N. Saiwaki, and H. Ishiguro, "Effects of observing eye contact between a robot and another person," Int. J. of Social Robotics, Vol.3, No.2, pp. 143-154, 2011.
[18] M. Shiomi, T. Kanda, I. Howley, K. Hayashi, and N. Hagita, "Can a Social Robot Stimulate Science Curiosity in Classrooms?" Int. J. of Social Robotics, pp. 1-12, 2015.
[19] T. Kanda, R. Sato, N. Saiwaki, and H. Ishiguro, "A two-month Field Trial in an Elementary School for Long-term Human-robot Interaction," IEEE Trans. on Robotics, Vol.23, No.5, pp. 962-971, 2007.
[20] T. Iio, Y. Yoshikawa, and H. Ishiguro, "Pre-scheduled Turn-Taking between Robots to Make Conversation Coherent," Proc. of the 4th Int. Conf. on Human Agent Interaction, pp. 19-25, 2016.

**Name:**
Takamasa Iio

**Affiliation:**
Specially Appointed Assistant Professor, Graduate School of Engineering Science, Osaka University

**Address:**
1-3 Machikaneyamacho, Toyonaka, Osaka, Japan

**Brief Biographical History:**
2012- Researcher, Intelligent Robotics Laboratories, ATR
2015- Specially Appointed Assistant Professor, Graduate School of Engineering Science, Osaka University

**Main Works:**
• T. Iio, Y. Yoshikawa, and H. Ishiguro, "Pre-scheduled Turn-Taking between Robots to Make Conversation Coherent," Proc. of the 4th Int. Conf. on Human Agent Interaction, pp. 19-25, 2016.

**Membership in Academic Societies:**
• The Robotics Society of Japan

**Name:**
Yuichiro Yoshikawa

**Affiliation:**
Associate Professor, Graduate School of Engineering Science, Osaka University

**Address:**
1-3 Machikaneyamacho, Toyonaka, Osaka, Japan

**Brief Biographical History:**
2005- Researcher, Intelligent Robotics Laboratories, Advanced Telecommunication Research Institute International
2006- Researcher, JST ERATO Asada Synergistic Intelligence Project
2010- Associate Professor, Graduate School of Engineering Science, Osaka University

**Main Works:**
• "Social Skill Acquisition Model through Face-to-Face Interaction: Local Contingency for Open-Ended Development," frontiers in Robotics and AI, Vol.24, March 2016.

**Membership in Academic Societies:**
• The Robotics Society of Japan
• Japanese Cognitive Science Society
• The Virtual Reality Society of Japan

**Name:**
Hiroshi Ishiguro

**Affiliation:**
Distinguished Professor, Graduate School of Engineering Science, Osaka University

**Address:**
1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan

**Brief Biographical History:**
2000-2001 Associate Professor in the Department of Computer and Communication Sciences at Wakayama University
2001-2002 Professor in the Department of Computer and Communication Sciences at Wakayama University
2009- Professor of Department of Systems Innovation in the Graduate School of Engineering Science at Osaka University
2014- Visiting Director of ATR Hiroshi Ishiguro Laboratories

**Main Works:**
• His research interests include distributed sensor systems, interactive robotics, and android science. Especially, his android studies are well-known very much in the world.

**Membership in Academic Societies:**
• IEEE Robotics and Automation Society (RAS)
• Association for Computing Machinery (ACM)
• Robotics Society of Japan (RSJ)