

Article

Twin-Robot Dialogue System with Robustness against Speech Recognition Failure in Human-Robot Dialogue with Elderly People

Takamasa Iio ^{1,2,*} , Yuichiro Yoshikawa ^{3,4}, Mariko Chiba ⁵, Taichi Asami ⁵, Yoshinori Isoda ⁵ and Hiroshi Ishiguro ^{3,4}

¹ Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan

² JST PRESTO, Kawaguchi-shi 332-0012, Japan

³ Graduate School of Engineering Science, Osaka University, Toyonaka 560-0043, Japan; yoshikawa@irl.sys.es.osaka-u.ac.jp (Y.Y.); ishiguro@irl.sys.es.osaka-u.ac.jp (H.I.)

⁴ JST ERATO, Kawaguchi-shi 332-0012, Japan

⁵ NTT DOCOMO, Inc., Tokyo 100-6150, Japan; mariko.chiba.eh@nttdocomo.com (M.C.); taichi.asami.mr@nttdocomo.com (T.A.); isoday@nttdocomo.com (Y.I.)

* Correspondence: iio@iit.tsukuba.ac.jp; Tel.: +81-29-853-5115

Received: 23 December 2019; Accepted: 17 February 2020; Published: 23 February 2020



Featured Application: Dialogue robot system for elderly people who have less opportunities to talk to other people.

Abstract: As agents, social robots are expected to increase opportunities for dialogue with the elderly. However, it is difficult to sustain a dialogue with an elderly user because speech recognition frequently fails during the dialogue. Here, to overcome this problem, regardless of speech recognition failure, we developed a question–answer–response dialogue model. In this model, a robot took initiative in the dialogue by asking the user various questions. Moreover, to improve user experience during dialogue, we extended the model such that two robots could participate in the dialogue. Implementing these features, we conducted a field trial in a nursing home to evaluate the twin-robot dialogue system. The average word error rate of speech recognition was 0.778. Despite the frequently high number of errors, participants talked for 14 min in a dialogue with two robots and felt slightly strange during the dialogue. Although we found no significant difference between a dialogue with one robot and that with two robots, the effect size of the difference in the dialogue time with one robot and that with two robots was medium (Cohen’s $d = -0.519$). The results suggested that the presence of two robots might likely encourage elderly people to sustain the talk. Our results will contribute to the design of social robots to engage in dialogues with the elderly.

Keywords: dialogue system; elderly people; two robots

1. Introduction

For elderly people, it is important to have opportunities to talk with someone daily. The act of talking with someone is a fundamental action for building social connectedness and reducing the feeling of social isolation. Social disconnectedness and perceived social isolation are likely to cause health risks [1–3], such as dementia [4,5], depression [6], and early death [7]. Therefore, it is important to increase opportunities for elderly people to have dialogues with other people.

However, increasing such opportunities are not as easy as they may seem. To see why, we considered a case in Japan, which is one of the countries with the highest rate of people aged 65 or over. Ninety percent of the elderly people who live with their families have dialogues every day in

some ways including telephone calls and e-mails. However, only 54.3% of those who live alone do so [8]. To increase opportunities for those living alone to have a dialogue, it was desirable for their family members, friends, neighbors, and hired caregivers to support them. Nonetheless, it was not easy for them to continue spending much time with the elderly people. In other words, there are limitations in human resources to support elderly people.

Social robots, including computer agents, are expected to increase opportunities for the elderly to engage in dialogues. Various applications of social robots for elderly people, such as schedule management, cognitive games, physical exercise suggestions, and information-provision [9,10], have been proposed so far. Although these applications can be useful in maintaining the health of elderly users, the aim of the applications is not to sustain a dialogue. Studies of communication for elderly people, for example Erber [11], Caris-Verhallen et al. [12], and Grainger [13], are premised on the idea that dialogue is important for elderly people. Furthermore, based on this idea, interventions that attempt to motivate residents of nursing care homes to interact with each other have been proposed; for example, staff training to raise awareness and to encourage caregivers to stimulate residents to interact [14–16]. Like these studies, under the assumption that much of the beneficial features of human social interaction is carried by dialogue, it does make sense to investigate whether these beneficial features of dialogue can also be realized by a non-human social agent. The first step towards producing and investigating such an application is to develop a system that can sustain a dialogue with elderly people for some time.

However, it is quite hard for robots to sustain a dialogue with the elderly. Kopp et al. [17] pointed out that “elderly users often have selectively impaired abilities, e.g., for auditory perception, articulation, adapting to a recommended interaction style, adhering to a clean turn-taking structure, or comprehending content of high information density [18,19]”. In particular, the difficulty of speech recognition in a dialogue with elderly people [18] is a critical issue in sustaining a coherent dialogue. In commonly used chat-bot systems, speech recognition failures basically because of nonsense responses and results in dialogue breakdown. It is unclear how a robot could sustain a coherent dialogue for a while under the situation where speech recognition would frequently fail.

Our goal is to develop a robot dialogue system that can sustain a coherent dialogue with elderly people for some time and also provide good user experience with the dialogues. To achieve this goal, we propose a question–answer–response dialogue model in which a robot takes the initiative in the dialogue by asking a user, various questions. Moreover, we propose an approach to extend the model such that two robots can participate in the dialogue. To evaluate how these features influence user’s dialogue time and user’s dialogue experience, we implement a dialogue system with two robots called the twin-robot dialogue system and conduct a field trial using the twin-robot dialogue system in a nursing home. We report the details of the trial and the results and finally, we discuss the implication of the results.

2. Related Work

Over the years, several categories of robot technologies have been proposed to support elderly people [9,10,20]. However, there are only two basic categories. One is robot technologies that physically support humans. These technologies include smart wheelchairs [21,22], prosthetic hands, and exoskeletons [23,24]. With these technologies, robots are considered as tools for extending the human body rather than as independent agents. The other is robot technologies that socially support humans. This category is further divided into two subcategories. One is robot technologies that support daily life tasks, such as medication and task schedule management [17,25,26], monitoring (fall detection) [27,28], household tasks [29,30], and shopping support [31]. The other subcategory is robot technologies that support health maintenance and psychological well-being improvement, such as pet-type robots (Paro) [32], conversational robots (including computer agents), AIBO [33,34], and NeCoRo [35]. Note that this categorization is formal. In fact, robot systems that belong to both

categories have also been proposed. For example, there have been wheelchairs that talk to elderly people who ride them [22] and robots that chat while shopping [31].

In this study, we review past studies of robot technologies for supporting the elderly through conversations. A mobile robotic assistant, Pearl [25], provided a reminder of the elderly daily activities, such as to visit the toilet every three hours and also to take medication. Regarding Pearl, a caregiver for an elderly client had input his/her daily activities in advance, and Pearl reminded the person based on the schedule. MonAMI Reminder [26] was also a schedule management assistant that allowed users to register their own schedule. Due to the difficulty in speech recognition, the input to the agent was provided with a digital pen and paper. Reminders were output, by voice, through embodied agents on a device. These robots and the agent have not been evaluated from the viewpoint of quality of a conversation. Ryan was a conversation robot for elderly people with dementia and depression [36]. Ryan has been developed by DreamFace Technologies, LLC. This robot has a head projection system that displays an animated avatar onto a mask and can show an emotive and expressive face. The robot has a touch-screen interface on its torso, which can be used as music player, narrated photo album, and video player. Furthermore, the robot can play cognitive games with a user using the screen and remind the user of daily activities with simple chats. Abodollahi et al. [36] installed this robot in a room for elderly participants and asked them to live together for 4 to 6 weeks. As a result, the average one-turn conversation between a participant and the robot was 198 times per day, which is relatively large. However, this paper does not show specific results, such as examples of dialogues, duration of a dialogue, and accuracy of speech recognition. Therefore, the quality of a dialogue was quite unclear.

A virtual assistant, Billy ('Billie'), to accompany and guide the elderly throughout the day was developed [17,37]. Billy basically performs schedule management like the MonAMI Reminder. However, with Billy, all inputs and outputs can be done through spoken-dialogue and natural confirmation signals like nodding or non-lexical cues. Furthermore, Billy can provide suggestions for leisure activities depending on the user. In their study, they took into account that elderly people seldom speak clearly, they cannot understand high-density dialogue, and they cannot perform turn-taking well. In view of this, they developed a robust and reliable interaction design called social cooperative behavior for the schedule management system. Although field trials started, the specific results have not yet been reported. Their system was sophisticated in terms of schedule management. However, with regards to achieving a conversation for a long while, it appeared to have low conversation support.

For robots specialized in assisting physical or cognitive activities, Ifbot could provide some activity programs such as Japanese language quizzes, singing songs, mouth exercise, and arithmetic. Although they used speech recognition in the programs, they said; "Almost all participants may have been dissatisfied with the robot's speaking and voice recognition functions." [38]. Matilda is a human-like (in appearance and attributes; e.g., voice, expressions, gestures, emotions) assistive communication robot (service and companion) in nursing homes in Australia [39]. Although users' impression of Matilda was assessed through a field trial in the nursing homes (Australia), the dialogue between elderly participants and the robot was not mentioned. Minami et al. [40] developed a dialogue robot system that chats with elderly people watching TV. This robot could provide responses by extracting social media comments related to TV programs. In addition to this function, the robot improves users' dialogue experience such as backchannel and repetition. However, the study did not evaluate the system with elderly people, and it did not consider dialogue breakdown due to speech recognition failures. Otaki et al. [41] developed a robot system that supported the co-imagination method for elderly people. With this robot system, elderly people talk to each other with a specified theme while looking at photos taken in their early lives. This method is designed to train the cognitive functions, which especially decline with aging, at an early stage of dementia [42]. The robot acts as a moderator of the conversation. Because the robot was remote-controlled by an operator, it is unclear how successful robot moderation between elderly participants is under the situation of frequent speech recognition failures. Sakakibara et al. [43] proposed a system that dynamically generated dialogue scenarios for

counseling patients with dementia. In the study, personalized conversations were generated using the history obtained in conversation and linked open data. However, they did not care about the situation of speech recognition failures and the system had not yet been evaluated. Jokinen proposed constructive dialogue models [44] and an architecture based on the model [45] for socially intelligent robots. A robot with this architecture can be aware of potentially interesting topics and of the user's attention, interest, and understanding through multimodal signals. Dialogue content is managed by topic tracking and anticipating possible continuations, calculated by coherence measures using the semantic distance between possible topics. With this architecture, Jokinen et al. [45] developed an application, using a robot assistant that instructed the human user on various task procedures related to elder care support services. However, the architecture is not applied to engage in dialogues with the elderly and is not concerned with speech recognition failures.

Through the above review, past studies of social assistive robots for elderly people appear to have paid little attention to speech recognition failure. They have not evaluated for how long dialogues with elderly participants had been continued. In contrast to the past studies, this paper focuses on developing and evaluating a robot dialogue system that can sustain a dialogue even in a situation where speech recognition frequently fails.

3. Question–Answer–Response Dialogue Model

We developed a question–answer–response dialogue model in which a robot takes the initiative in a dialogue by asking a user various questions in order to sustain the dialogue regardless of speech recognition failures. This model has two features:

- The model regards a dialogue as a transition process which comprises four states.
- Every time a state is transitioned to the next state, a system selects a suitable utterance (among a set of utterances defined in the next state) for the robot.

Here, we explain the details of the model.

3.1. States

The four states in a dialogue are as follows: question, answer, backchannel, and comment. Figure 1 shows part of a dialogue constructed from the four states. The question state means the robot asks the user a question. The answer state means the user answers the robot. The backchannel state means the robot provides a backchannel, such as “uh-huh”, “I see”, or “really?”. The comment state means that the robot comments on the question or the answer from the user.

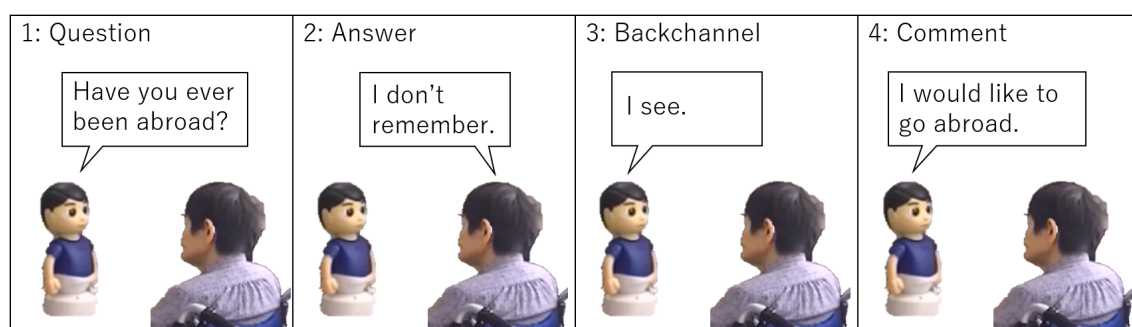


Figure 1. Dialogue constructed from the four states.

There are three reasons for adopting such a structure:

- The structure makes a dialogue robust against speech recognition failures. Although speech recognition may fail, a robot can sustain a coherent dialogue by providing an ambiguous response; independent of user answer. Consider the dialogue in Figure 1. The user might answer “yes”, “no”, or “so-so” instead of “I don't remember”. However, the response of the robot sounds

coherent in any of the three ways. It means that the robot can sustain a coherent dialogue even when the speech recognition result is doubtful.

- The structure makes it easy to control the direction of a dialogue. A dialogue with the elderly is basically non-task-oriented (i.e., chat). However, the dialogue often has some type of implicit purposes. For example, it would be expected to stimulate users' cognitive function abilities or to get users to pay attention to their daily lifestyle. To achieve such a purpose, a robot needs to mainly control the direction of the dialogue such that it can talk with the user about certain topics. Because the structure compels the robot to take the initiative in a dialogue, the robot would control the direction of the dialogue more easily than if it had been a free-style chat.
- The structure makes it easier to create a scenario of a dialogue. In the present natural language processing technology, it is still difficult to generate robot responses suitable for user utterances automatically. In particular, it is quite difficult to generate robot responses that facilitate dialogue with implicit purposes. In view of this, we had to create a scenario manually, consulting experts of such dialogues (i.e., caregivers). In general, creating a scenario from scratch is difficult. A scenario writer needs to consider many factors; for example, what kind of story the robot tells, how the story unfolds, whether each topic of the story remains coherent, when and how the robot asks, and for how long the robot should take the initiative. In contrast to the difficulty in considering such factors, using the proposed structure, a scenario writer needs to only create pairs of questions asked by a robot and responses to answers the user might give. This is relatively easier than creating scenarios from scratch.

3.2. Transition Rules

The state transition diagram is shown in Figure 2. Table 1 shows the transition rules for each state. The system transforms from one state to another when it receives some events. In this study, we defined three events: an event in which a robot completed an utterance (Robot Utterance Completion: RUC), an event in which the system recognized user utterance (User Utterance Recognition: UUR), and an event in which user utterance has not been recognized for a certain period of time (Timeout: TO).

The reason for separating the backchannel state from the comment state is to deal with a situation where a user remained mute. In general, when a user answered something, the robot should provide a backchannel like "I see" or "uh-huh" to show that the robot is listening to the user. However, when a user does not answer, the robot should not provide any backchannel because the backchannel would make the user feel strange. At that time, the system skips the backchannel state and transitions to the comment state as shown in the state transition diagram (Figure 2).

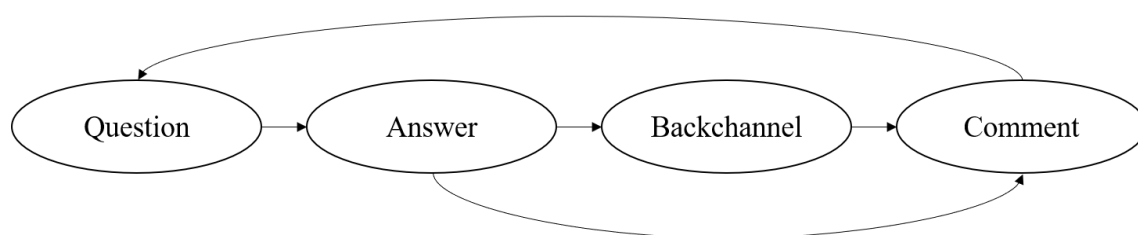


Figure 2. State transition diagram Of a dialogue with an elderly person.

Table 1. State transition table of the state transition diagram. The events are abbreviated as follows: RUC: Robot utterance completion, UUR: User utterance recognition, and TO: User utterance timeout.

State	RUC	UUR	TO
Question	Answer	-	-
Answer	-	Backchannel	Comment
Backchannel	Comment	-	-
Comment	Question	-	-

3.3. Utterance Selection in Each State

Every state (except for the answer state) has a set of objects for the robot utterances and an algorithm for selecting an object from the set.

The question state has a set of objects shown in the Listing 1. When the system transitions to the question state for the first time, it selects an object among the set randomly. After the second time, it selects an object with the same topic as the previous question among the set, except for objects that have already been selected. The reason for maintaining the same topic is that it would be unnatural to change topics with every question. However, to prevent the user from getting bored, the topic is changed every four times.

Listing 1: Objects of the question state.

```

1  [
2    {
3      Question: "Have you ever been abroad?",
4      Topic: "travel"
5    },
6    {
7      Question: "Where do you want to travel to?",
8      Topic: "travel"
9    },
10   {
11     Question: "What was your favorite food in childhood?",
12     Topic: "childhood"
13   },
14   ...
15 ]

```

The answer state has no set. The system selects no objects because it waits for the user to answer.

The backchannel state has a set of objects shown in the Listing 2. The question key is used to associate a backchannel with a question. The keyword, 'key', is a list of answers expected of the user in the answer state. The backchannel key is a sentence with a backchannel. The algorithm for selecting a backchannel is as follows: First, the system selects objects having the same question key as the previous question and it searches whether any of keywords of each object is in the speech recognition results. If a keyword is in the speech recognition results, the object having that keyword is selected. In contrast, if there is no keyword in the speech recognition results, the system selected the object without keywords (i.e., default backchannel). For example, as shown in Figure 1, when the user answered "I don't remember." to the question; "Have you ever been abroad?", the second object of the Listing 2 is selected.

Listing 2: Objects in the backchannel state.

```

1  [
2    {
3      Question: "Have you ever been abroad?",
4      Keyword: ["yes", "yeah"]
5      Backchannel: "Cool!"
6    },
7    {
8      Question: "Have you ever been abroad?",
9      Backchannel: "I see."
10   },
11   {
12     Question: "Where do you want to visit?",
13     Keyword: ["kyoto", "tokyo", "osaka"]
14     Backchannel: "Yeah."
15   },
16   ...
17 ]

```

The comment state has the same set of objects as in the backchannel state (Listing 3). The selection algorithm is also the same as that of the backchannel state.

Listing 3: Objects in the comment state.

```

1  [
2    {
3      Question: "Have you ever been abroad?",
4      Keyword: ["yes", "yeah"]
5      Comment: "You are a global person. I also want to go someday."
6    },
7    {
8      Question: "Have you ever been abroad?",
9      Comment: "I would like to go abroad someday."
10   },
11   {
12     Question: "Where do you want to visit?",
13     Keyword: ["kyoto", "tokyo", "osaka"]
14     Comment: "That's an exciting city. I want to go there too."
15   },
16   ...
17 ]

```

4. Participation of Multiple Robots

Although the question–answer–response dialogue model is expected to be robust against speech recognition failures, the dialogue generated by the model might be tedious for users. This is because the dialogue system is a one-way model; the robot asks a question, the user answers it, and the robot responds to the answer. If this continues for a while, the user might feel bored and stop the dialogue early.

We let two robots participate in the dialogue to decrease the tediousness of the model. There have been reports of advantages of using multiple robots in a dialogue. For example, when multiple robots participate in a conversation, a user seems to become insensitive to unnaturalness about consistency

in a dialogue [46,47]. In addition, the user tends to feel that he or she can talk easily with multiple robots [48,49] and the user is likely to experience eye contact with the robots [50]. Karatas et al. [51] developed a multi-agent system that interacts with a driver in a car and showed that using multiple agents reduces cognitive loads of the driver compared to using a single agent. Sakamoto et al. [52] conducted a field trial in which two robots provide information to passersby at the station and found that passersby were more likely to stop when two robots were talking than when a single robot was talking. Iio et al. [47] demonstrated that visitors in an exhibition hall tended to have a longer dialogue with multiple robots than a single one.

In this study, we defined the participation framework of a dialogue in which two robots and one user participated, according to Goffman [53]. The participation framework contains a speaker, an addressee, and a side-participant. The turn-taking system was implemented based on Sacks et al. [54]. The rules of turn-taking of each state in our system are described below.

In the question state, the rule of selecting a speaker differs between the first time and the second time or after. In the first time, either of the robots is selected as a speaker randomly. Another robot becomes a side-participant. After the first time, the speaker of the previous question is selected as a speaker. However, when the topic changes from the previous question, the side-participant of the previous question is selected as a speaker.

This approach is based on the concept of common ground [55] among the participants. When a speaker asks a question on a certain topic, the addressee and the side-participant would have common ground that the speaker appears to be interested in the topic. Such common ground would allow the speaker to continue asking questions on the same topic. Therefore, it is reasonable for the speaker of the previous question to continue questions. However, when the topic of a question changes, it is not easy to interpret the sudden topic shift on the common ground. Arimoto et al. [46] reported that the unnaturalness of the sudden topic shift would be alleviated by changing the speaker. Therefore, it is reasonable for the side-participant of the previous question to become a speaker when the topic changes from the previous question.

In the answer state, the user becomes a speaker. The speaker of the previous question would be regarded as an addressee from the viewpoint of the concept of adjacency pair [56].

In the backchannel state, the speaker of the previous question is selected as a speaker again. This is grounded on the concept of the sequence-closing third [57]. Since the backchannel state is regarded as post-expansion of the question–answer adjacency pair, it is reasonable that the speaker of the previous question, which were addressed in the previous answer state, become a speaker.

In the comment state, a speaker depends on a speech recognition result of the previous answer state. Although the speaker of the previous question is basically selected as a speaker, the side-participant of the previous question is selected when the result has a kind of negative expression (including “No”, “Nothing”, “Never”, etc.) or timeout. When the side-participant is selected, the side-participant speaks to the speaker; in other words, the speaker is selected as an addressee. In this manner, the speaker can easily continue asking a question in the next question state. A user’s answer with negative expressions might indicate low interest in the topic. Here, if the side-participant expresses interest in the topic by commenting on the previous question, it appears to be reasonable for the speaker to continue asking a question on the same topic from the viewpoint of common ground [55] because at least one participant shows the interest in the topic.

5. System

We developed a twin-robot dialogue system including the two features: the question–answer–response dialogue model and the participation of two robots in a dialogue. The hardware components of the system are shown in Figure 3 and the system architecture is shown in Figure 4.

A microphone array collects sounds. The sounds are integrated through noise reduction process by a microphone array. The integrated sound is then sent to the automatic speech recognition module, which recognizes the user utterance. We used a cloud speech recognition service provided by NTT

Docomo. The service receives a voice and returns the voice recognition results, which are sent to the utterance selection module. According to the selection rules (see Section 3.3), the utterance selection module selects an utterance from the database. The selected utterance is sent to the robot controllers. The voice recognition results are also sent to the nodding generation module as a signal of user speech. The nodding generation module sends a nodding motion to the robot controllers. Nodding is a motion for expressing that the robot is listening to the user in the nodding generation module. This motion is always executed in the answer state whenever the system received a speech recognition result. The robot controllers interpret the utterance with motions and execute them. After the execution is completed, the completion signal is sent to the utterance selection module. The utterance selection module selects a next utterance. As such, the system repeats selecting and executing an utterance according to a speech recognition result and its own behavior execution.

A social-conversational robot developed by VSTONE, CommU, was used as the dialogue partner in our system. This robot is desktop sized at 304 mm high, 180 mm wide, and 131 mm deep, weighing 938 g. CommU has three degrees of freedom (df) for its waist, 3 df for its neck, and 2 df for each eyes. The robot has two LEDs in its cheeks. The robot controller was a software server, which received a command, such as “speak” or “nod”. The robot controller controlled the robot according to a received command.

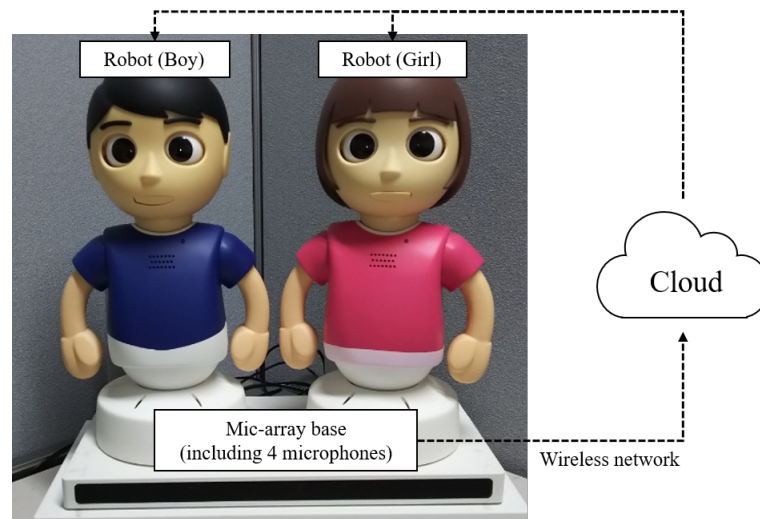


Figure 3. Hardware components of the system and the connections between them.

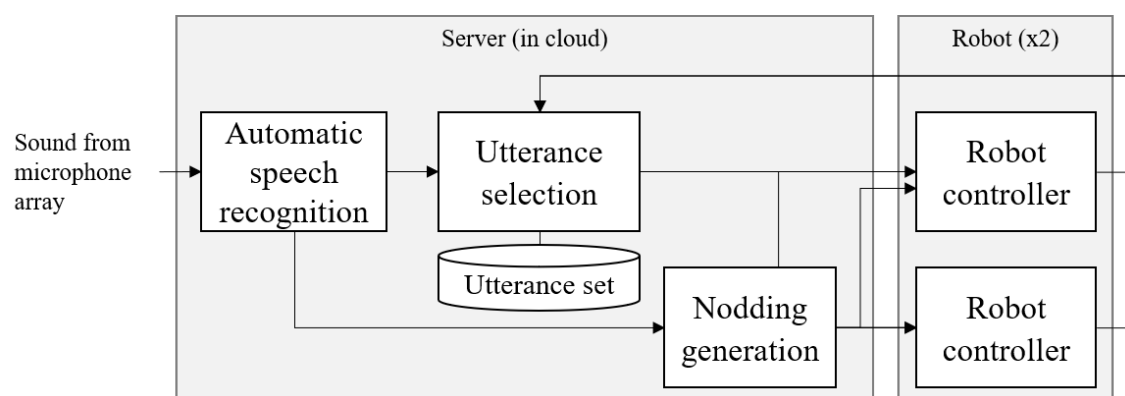


Figure 4. The system architecture.

6. Field Trial

6.1. Purpose

We conducted a field trial in a nursing home. There were two purposes for this field trial. The first was to investigate whether the twin-robot dialogue system can sustain a coherent dialogue with elderly people for a certain time and the second was to evaluate whether the system can provide good user experience of a dialogue.

6.2. Participants

Thirty elderly residents in a nursing home participated in the trial: 26 females and 4 males. They were native Japanese speakers. Their average age was 86.3 years (SD = 7.5). According to caregivers, 13 participants had no dementia, 4 had mild dementia, and the other 13 had advanced dementia.

The participants were recruited by caregivers of the nursing home. Before the trial, we explained the purpose and the procedure to the caregivers and asked them to recruit candidates who would like to participate in the trial. We sent an instruction document of the trial to their families and asked the families to fill out a consent form. The consent form had been approved by the ethical committee of Osaka University. The candidates whose families agreed that they should participate in the trial became the participants.

Furthermore, two caregivers participated in the trial to observe participants' behavior.

6.3. Scenarios

To achieve the purpose, it was desirable to design an experiment that could clarify two basic research questions: Whether or not the system has the question–answer–response dialogue model, and whether a single robot or two robots engage in a dialogue. However, without the question–answer–response dialogue model, it was obvious that the elderly could not continue a dialogue with the system. The reasons are as follows: the chatbot model commonly used in non-task-oriented chat systems generates responses based on the results of speech recognition. When speech recognition fails, the chatbot model generates a response that does not match the context of the dialogue. Since speech recognition frequently fails during a dialogue with elderly people, the system with the chatbot model would give unrelated responses in the dialogue in most cases. Therefore, we designed scenarios depending on whether only a single robot or two robots participate in a dialogue, which are as follows:

1. **One-robot scenario.** One robot participated in a dialogue. The robot performed tasks according to the question–answer–response dialogue model (see Section 3).
2. **Two-robot scenario.** Two robots participated in a dialogue. The robots performed tasks according to the question–answer–response dialogue model (see Section 3). They take turns according to the rules described in Section 4.

The field trial was a between-participant design. The participants were assigned to each scenario in such a way as to balance the dementia level of participants in of each scenario as shown in Table 2.

Table 2. Distribution of participants' cognitive capacities.

Cognitive Capacities	Scenario		Total
	One-Robot	Two-Robot	
Noting or mild	6	6	12
Severe	5	7	12
Total	11	13	24

6.4. Procedure

The procedure was as follows: A caregiver escorted a participant to a place of trial (Figure 5). The caregiver had the participant sit down on a chair in front of the robot. If the participant was using a wheelchair, the caregiver put the participant with the wheelchair in front of the robot. After escorting, the caregiver moved to a position behind the participant. Thus, the caregiver was invisible to the participant during the trial. Then, a controller greeted the participant and explained the task. The instruction was as follows: “This robot starts to talk to you in a little while. Please talk with it.” After the instruction, the controller started the system and the robot started a dialogue. As the participants were native Japanese speakers, the field trial was conducted in the Japanese language. The procedure that a caregiver takes an elderly person to the robot, encourages him or her to talk with the robot, and watches him or her from behind would be reasonable at least in the phase of introducing the robot system.

The dialogue continued following the flowchart of Figure 6. The robot said the introduction first. Next, the robot started a dialogue. In every 5 min during the dialogue, the robot asked the participant whether to continue the dialogue or not. When the participant gave a positive answer, the robot continued the dialogue. Otherwise, a negative answer ended the dialogue.

Here, we should note an inappropriate case caused by speech recognition errors. If the robot recognized that the participant answered positively even though the participant actually answered negatively, the robot would have continued the dialogue. Because such a situation must be avoided, an experimenter force-quit a program for a dialogue as soon as possible.

Considering the burdens of a participant, we limited the dialogue time to 15 min even if the participant would like to continue. The dialogue was recorded by video cameras.

The caregiver had observed the dialogue and filled out in a questionnaire about the participant behavior. When the dialogues ended or were force-quit, the caregiver took the participant to a place away from the robot. Then, an interviewer asked the participant a simple question. After that, the caregiver took the participant back to his or her room, and then escorted the next participant to the place of trial.



Figure 5. Twin-Robot dialogue system that talks with an elderly person.

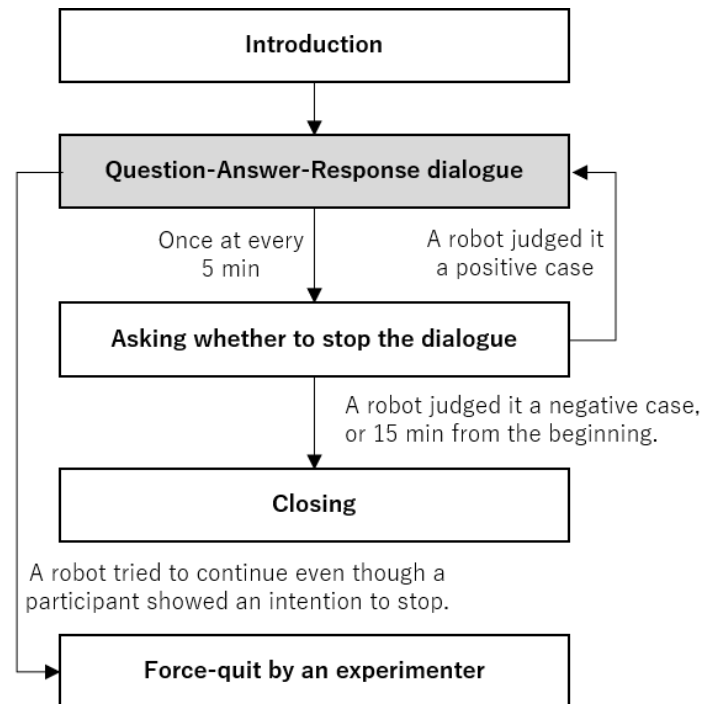


Figure 6. Flowchart of a dialogue in the trial.

6.5. Dialogue Contents

The number of questions we created was 55, which enabled a dialogue to run for approximately 27 min because it required approximately 30 s in the cycle for the robot to ask a question, receive an answer from the user, and respond to the answer. The details are shown in Table 3. The questions, backchannels, and comments were created by an expert in robot speech creation and elderly care.

Table 3. A part of the questions we prepared for the trial.

Topic	Number	Example
Childhood	16	"Where did you usually play?" "What toy did you want?" "Did you like to go to school?"
Travel	20	"Do you like travel?" "Where have you traveled so far?" "What was the best food you have had in travel?"
Health	19	"Have you had severe illness so far?" "Do you like walking?" "Do you have food you eat for health?"

6.6. Measurements

We measured the following values: word error rate (WER), dialogue time, user utterance time, and subjective impressions of the participant and caregiver.

6.6.1. Word Error Rate

WER is a typical metric of the accuracy of speech recognition [58]. In this experiment, the WER is for errors that occur when the robot recognizes the participant's speech. The WER was calculated as follows:

$$WER = (S + D + I) / N \quad (1)$$

where S , D , I , and N are the number of substitutions, deletions, insertions, and words in the reference, respectively. To compute the WER, we transcribed all participants' utterances in the dialogue. The WER was used to confirm the difficulty of speech recognition in a dialogue with elderly people.

6.6.2. Dialogue Time

Dialogue time is the time from when the robot starts a dialogue until the robot ends the dialogue or a participant leaves the seat. Dialogue time was used to evaluate how long the twin-robot dialogue system can sustain a dialogue with elderly people.

6.6.3. Participant Utterance Time

Participant utterance time is the time that a participant spent speaking during a dialogue. By watching the video of the dialogue, we recorded the time when the participant started speaking and when he or she ended speaking. Participant utterance time was used to investigate whether participants participated positively during a dialogue.

6.6.4. Participant Subjective Impression

We asked a participant the following question: "Did you feel something strange in that dialogue with the robot?" The question was asked to elicit their viewpoint of the naturalness of the dialogue. Indeed, we needed to use formal psychological measures to obtain accurate results; however, this was quite challenging to accomplish for the elderly with dementia.

6.6.5. Caregiver Subjective Impression

We asked the caregivers the following question: "Did the participant talk with the robot more positively, comparing to when he or she talked with staffs." The question used a 7-point Likert scale, in which one means 'strongly disagree', and seven means 'strongly agree'. The question was asked in order to find out if the participant had been in the same or different state as usual.

7. Results

We obtained the videos and the questionnaire results of 24 participants and analyzed the data. Although there had been 30 participants altogether, one participant could not continue a dialogue because he was not able to hear the robot's voice at all, while the other five participants had halted the dialogue owing to technical problems (e.g., network trouble, program bugs). The two-robot scenario had 13 participants, and one-robot scenario had 11 participants.

We used the Mann–Whitney U test to compare data between the scenarios, and the alpha-level set at 0.05. We used a computer software 'jamovi' [59] for this test.

7.1. Word Error Rate

Figure 7 shows the averages of the WERs. The total average of the WERs was 0.778 (SD = 0.144). The average of the WERs of the one-robot scenario was 0.789 (SD = 0.123), and that of the two-robot scenario was 0.768 (SD = 0.163). There was no significant difference between the scenarios ($U = 62.0$, $p = 0.608$, Cohen's $d = 0.148$).

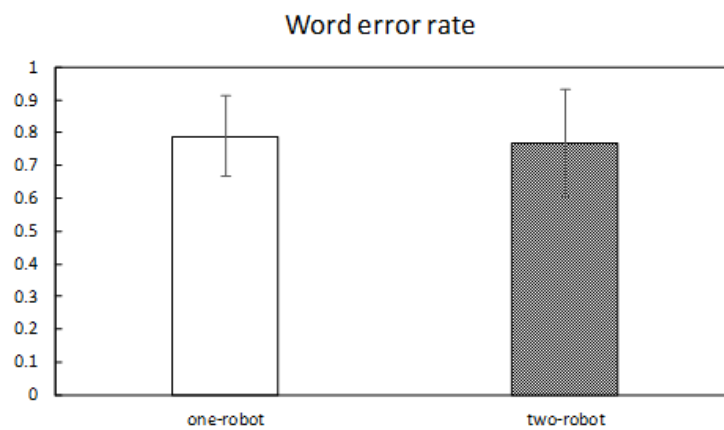


Figure 7. The averages of the word error rate (WER) in each scenario.

7.2. Dialogue Time

Figure 8 shows the averages of the dialogue time. The total average of the dialogue time was 12 min 51 s (SD = 4 min 52 s). The average of the dialogue time of one-robot scenario was 11 min 30 s (SD = 5 min 18 s), and that of two-robot scenario was 14 min (SD = 4 min 20 s). Here, milliseconds were round off. There was no significant difference between the scenarios ($U = 48.0$, $p = 0.188$, Cohen's $d = -0.519$).

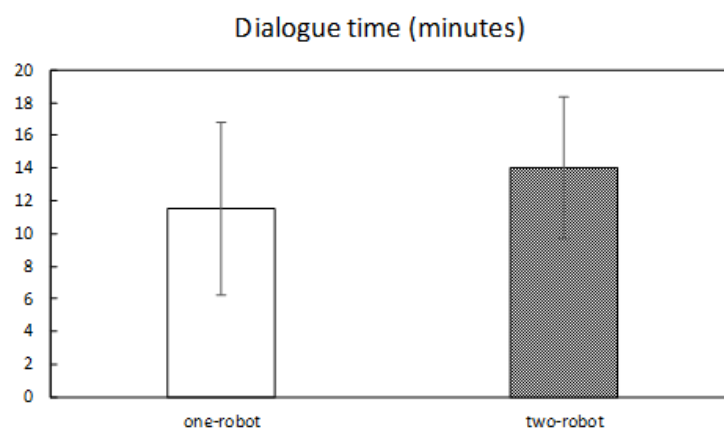


Figure 8. Averages of the dialogue time in each scenario.

7.3. Participant Utterance Time

Figure 9 shows the averages of the participant utterance time. The total average of the participant utterance time was 3 min 31 s (SD = 3 min 14 s). The average of the participant utterance time of one-robot scenario was 3 min 5 s (SD = 3 min 21 s), and that of two-robot scenario was 3 min 53 s (SD = 3 min 13 s). There was no significant difference between the scenarios ($U = 61.0$, $p = 0.569$, Cohen's $d = -0.242$).

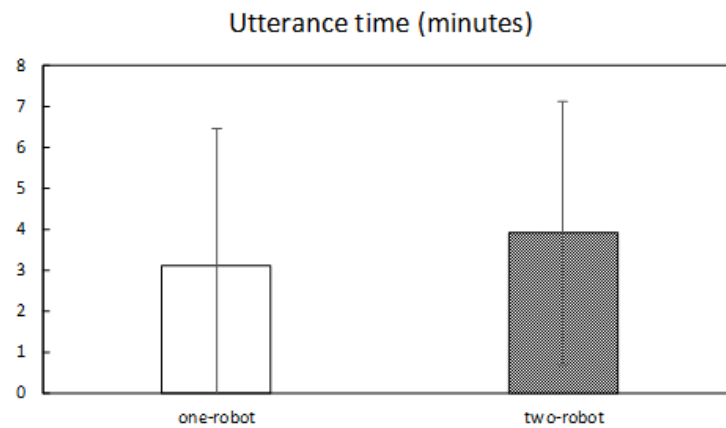


Figure 9. Averages of the user utterance time in each scenario.

7.4. Participant Subjective Impression

Figure 10 shows the results of the question to the participants, which was whether the participants have felt something strange in a dialogue. The numbers of the participants who answered “Yes”, “No”, and nothing were 3 (13%), 17 (71%), and 4 (17%), respectively in total. Those numbers were 1 (9%), 8 (73%), and 2 (18%) in one-robot scenario, and were 2 (15%), 9 (69%), and 2 (15%) in two-robot scenario, respectively.

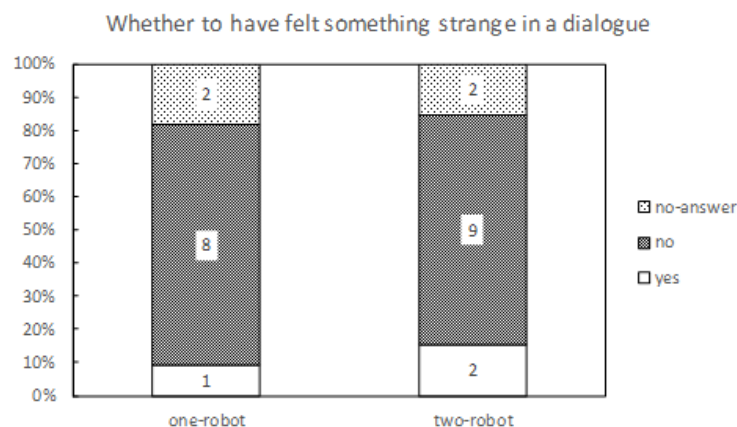


Figure 10. Results of the question to the participants in each scenario.

7.5. Caregiver Subjective Impression

Figure 11 shows the averages of the scores of the question to the caregivers, which is whether the participants had talked with the robot more positively than usual. The total average of the scores was 4.92 (SD = 1.89). The average of the scores of the one-robot scenario was 4.55 (SD = 1.86), and that of two-robot scenario was 5.23 (SD = 1.92). There was no significant difference between the scenarios ($U = 61.5$, $p = 0.568$, Cohen’s $d = -0.186$).

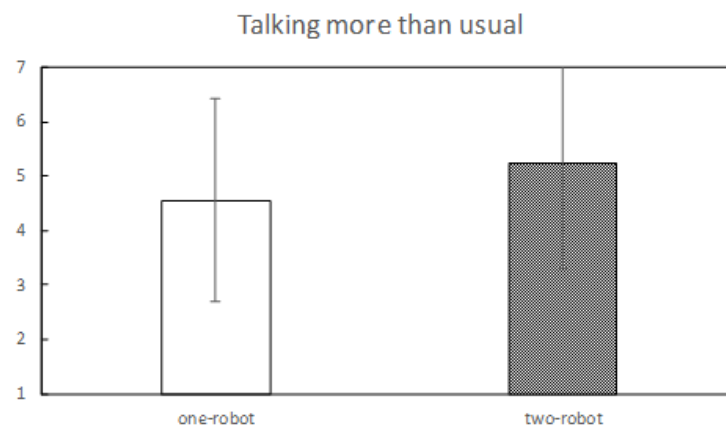


Figure 11. Averages of the scores of the question to the caregivers.

8. Discussion and Conclusions

8.1. Interpretation of the Results

The total average of the WERs was 0.778. This means that approximately 78% of the words in the utterances of the participants were mis-recognized. In general, it would be too difficult to continue a dialogue with this speech recognition accuracy. Despite the difficult situation, the system continued the dialogues for 12 min 51 s on average. This suggests that the twin-robot dialogue system could sustain a dialogue for a certain time regardless of speech recognition failures.

The average of the participant utterance time was 3 min 31 s, which was approximately 27% of the average dialogue time (cf. the average of the robot utterance times was approximately 5 min 51 s). In other words, the ratio of the participant utterance times to the robot utterance time was approximately 3:5. Because the gap between the utterance time of the participants and the robot was not so much, the participants can be considered to have positively participated in dialogue with the twin-robot dialogue system.

Regarding subjective impressions, 71% of participants answered that there was nothing strange in the dialogues with the robot. We believe that the system could have provided a dialogue without breakdown for many participants. In addition, the caregivers answered that they felt that the participants had been speaking more positively than usual. Because such positive participation might have involved a novelty effect that none of the participants has spoken to a robot before or an experimenter effect that the participants received special attention in the context of this experiment, we cannot justify whether the system was able to encourage some participants to participate more actively. To clarify the effect of the system on the positive participation, a long-term study is required.

In contrast, there was no significant difference between the one-robot scenario and two-robot scenario in each measurement. Therefore, it is still unclear if the use of two robots is effective in improving the user experience of dialogue. Nevertheless, regarding dialogue time, the effect size was medium (Cohen's $d = -0.519$). The results suggested that the presence of two robots might likely encourage elderly people to sustain the talk.

8.2. Effects on the Elderly with Dementia

The caregivers who observed the trial remarked, during an interview after the trial, that participants with dementia appeared to have really enjoyed the dialogue. To comprehend their opinion, we grouped the results of the following question posed to the caregivers, "Did the participant talk with the robot more positively, comparing to when he or she talk with staffs?" at the level of dementia (Figure 12). Although two-way ANOVA showed no interaction between groups and scenarios, the graph appears to suggest the trend that participants with severe dementia spoke more

actively than usual, in accordance with the observation caregivers made during the trial and confirmed during the interview after the trial.

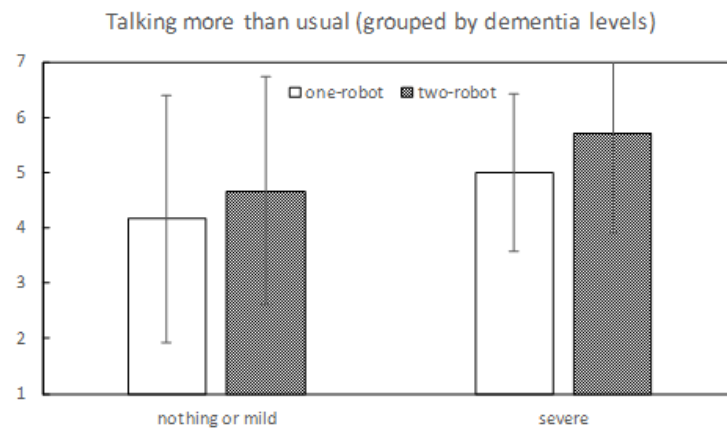


Figure 12. Averages of the scores of the question to the caregivers, grouped by dementia levels of the participants.

Furthermore, we grouped the participant utterance times precisely as we did for the caregiver impressions (Figure 13). This graph shows that some of the participants were talking for more than 5 min (i.e., Participants 2, 15, 18, 19, and 20). Especially, there were three participants with severe dementia in the two-robot scenario. Although we cannot conclude with certainty based on such small data, this result appears to provide a new hypothesis that some elderly people with severe dementia may actively speak when placed in a two-robot scenario.

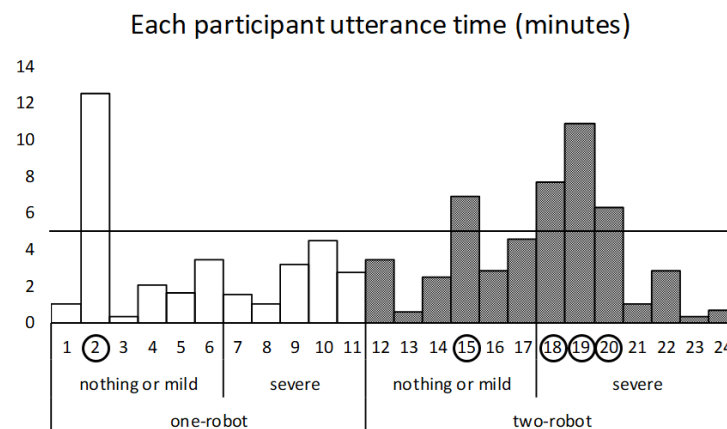


Figure 13. Utterance time of each participant grouped by dementia levels.

Further research are required to investigate whether using multiple robots is better for elderly people with severe dementia.

8.3. Influence of Topics on Participants' Utterance Time

In order to investigate whether the topic of the questions influence the verbal participation time of the participants, we calculated the mean of the utterance time of the participants in each topic. The results were as follows: In the one-robot scenario, the utterance time of the topic of travel, health, and childhood were 11.9 s (SD = 21.3), 6.8 s (SD = 5.1), and 14.7 s (SD = 27.4). In the two-robot scenario, the utterance time of the topic of travel, health, and childhood were 9.3 s (SD = 10.5), 9.2 s (SD = 11.5), and 12.0 s (SD = 12.7). We analyzed the results using two-way mixed ANOVA, which

has scenario factor and topic factor. The results showed that there was no main effect in the scenario factor ($F(1,20) = 0.01, p = 0.920$), no main effect in the topic factor ($F(2,40) = 2.094, p = 0.136$), and no interaction between the two factors ($F(2,40) = 0.933, p = 0.402$). Therefore, it is unlikely that differences in topics had a systematic effect on utterance time.

Because the variance of the utterance time is very large, the influence of the topic of the question on the utterance time appeared to be considerably dependent on the individual. As an interesting example, we found that a topic stimulated participants' memory and the participants began to talk about their life for a long time. Specifically, participant 2 spent about 2 min talking about her initial visit to the nursing home when the robot asked, "Have you ever had a honeymoon?". Moreover, in response to the question, "What class did you like in elementary school?" she had talked about her childhood struggles for about 3 min. Participant 18 spent about 3 min talking about their experiences during World War II when asked "Have you lived around here since you were a child?". Their utterance times were quite long considering that usual answers of other participants were only one or two sentences. More surprisingly, even the caregivers, who have been interacting with the participants in their daily lives, did not know the stories of the participants until that time. These examples are interesting from the aspect of robots being potentially able to elicit a much deeper story from the elderly if the robot chooses topics adjusted to the individual.

8.4. Pros and Cons of Our System

We found several pros and cons of the twin-robot dialogue system through the field trial.

- Leading a dialogue by a robot

Pros. Participants who have no topic to discuss might have easily taken part in a dialogue. In general, it is quite challenging for people to initiate a dialogue unless they have topics they would want to talk about. We found that many participants had no topic to discuss with the robot. By the robot initiating a dialogue, those participants could have participated in the dialogue without worrying about initiating it.

Cons. Dialogue initiation by the robot may have frustrated some participants in case they had something they would have preferred to talk about with the robot.

- Patterning a dialogue

Pros. Participants who are not good at communicating smoothly might have easily followed a dialogue because the user could have predicted the flow of the dialogue. This aspect should be important in dialogues for elderly people with declining cognitive ability.

Cons. Participants who have no communication problems might have felt bored earlier during a dialogue if the dialogue was monotonous.

- Choosing robot responses by using keyword match of user answers

Pros. This method was clearly robust against speech recognition failures. In our question-answer-response dialogue model, if the speech recognition result contains words of the keyword attribute, the backchannel (comment) associated with the keyword is selected. Otherwise, the backchannel (comment) associated with no keyword (i.e., the default backchannel (comment)) is selected. Therefore, when the speech recognition result is a broken sentence, the default backchannel (comment) is selected in most cases. Because the default backchannel (comment) is a sentence that is coherent for any answer, the dialogue was usually coherent even if the speech recognition fails.

Cons. There are two situations for a dialogue to break down. First, there is the case where a user asks a robot a question while the robot is in the "answer mode". Here, the sentence associated with the default attribute is selected unless it was time for the user to be posing a question to the robot. Because the sentence of the default attribute is not meaningful to the

question, the dialogue would become unnatural. Second, there is the case where a keyword is matched owing to speech recognition failures, although this will rarely happen. For example, let us consider the following situation: a robot asks “Which countries do you want to travel, France or England?”. Although a user answers France, the speech recognition result could be England. At this point, the dialogue would be strange because the robot would choose “England” as the response. To avoid this, we need more sophisticated algorithms.

8.5. Application of the System

The ability to talk with elderly people is becoming increasingly important for social robots because dialogues play an important role in building human–robot relationships. Social robots instruct elderly people to take medicines, exercise, undergo cognitive training, and suggest lifestyle improvements in order to sustain physical and mental wellbeing. Instructions in such situations may not work well if the relationship with elderly people—in other words, a sense of trust, security, and familiarity—is not built. Conversely, if the relationship between elderly people and social robots is well formed, the instruction will be more effective. Therefore, not only robots as companions but also a various other robots will need to have a certain level of dialogue with humans. The proposed twin-robot dialogue system would be useful from the viewpoint that they can sustain dialogues up to 12 min 51 s.

Author Contributions: Conceptualization, H.I. and Y.I.; methodology, Y.Y. and T.I.; software, M.C., T.A., and T.I.; validation, Y.Y. and T.A.; formal analysis, T.I. and M.C.; investigation, T.I., Y.Y., and M.C.; resources, Y.Y. and T.A.; data curation, T.I. and M.C.; writing—original draft preparation, T.I.; writing—review and editing, T.I., Y.Y., and T.A.; visualization, T.I.; supervision, H.I. and Y.I.; project administration, H.I. and Y.I.; funding acquisition, H.I. and Y.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NTT DOCOMO, INC., JST, PRESTO Grant Number JPMJPR1851, JST, ERATO Grant JPMJER1401 and JSPS KAKENHI Grant Number 19H05691.

Acknowledgments: We appreciate participants and caregivers who had participated in the field trial, and the facility manager that had admitted us to do the field trial.

Conflicts of Interest: Mariko Chiba, Taichi Asami and Yoshinori Isoda are working at NTT DOCOMO, INC., which is one of the funders of this study.

References

- House, J.S.; Landis, K.R.; Umberson, D. Social relationships and health. *Science* **1988**, *241*, 540–545. [[CrossRef](#)] [[PubMed](#)]
- Cornwell, E.Y.; Waite, L.J. Social disconnectedness, perceived isolation, and health among older adults. *J. Health Soc. Behav.* **2009**, *50*, 31–48. [[CrossRef](#)] [[PubMed](#)]
- Holt-Lunstad, J.; Smith, T.B.; Baker, M.; Harris, T.; Stephenson, D. Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspect. Psychol. Sci.* **2015**, *10*, 227–237. [[CrossRef](#)] [[PubMed](#)]
- Kotwal, A.A.; Kim, J.; Waite, L.; Dale, W. Social function and cognitive status: Results from a US nationally representative survey of older adults. *J. Gen. Intern. Med.* **2016**, *31*, 854–862. [[CrossRef](#)]
- Poey, J.L.; Burr, J.A.; Roberts, J.S. Social connectedness, perceived isolation, and dementia: Does the social environment moderate the relationship between genetic risk and cognitive well-being? *Gerontologist* **2017**, *57*, 1031–1040. [[CrossRef](#)]
- Heikkinen, R.L.; Kauppinen, M. Depressive symptoms in late life: A 10-year follow-up. *Arch. Gerontol. Geriatr.* **2004**, *38*, 239–250. [[CrossRef](#)]
- Berkman, L.F.; Syme, S.L. Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. *Am. J. Epidemiol.* **1979**, *109*, 186–204. [[CrossRef](#)]
- Japan Cabinet Office: Annual Report on the Ageing Society: 2018 (Summary), <Viewpoint 2> Establishing a New Paradigm of Health through Science and Technology in Ageing Society. Available online: https://www8.cao.go.jp/kourei/english/annualreport/2018/2018pdf_e.html (accessed on 20 February 2020).
- Broadbent, E.; Stafford, R.; MacDonald, B. Acceptance of healthcare robots for the older population: Review and future directions. *Int. J. Soc. Robot.* **2009**, *1*, 319. [[CrossRef](#)]

10. Kachouie, R.; Sedighadeli, S.; Khosla, R.; Chu, M.T. Socially assistive robots in elderly care: A mixed-method systematic literature review. *Int. J. Hum.-Comput. Interact.* **2014**, *30*, 369–393. [\[CrossRef\]](#)
11. Erber, N.P. Conversation as therapy for older adults in residential care: The case for intervention. *Int. J. Lang. Commun. Disord.* **1994**, *29*, 269–278. [\[CrossRef\]](#)
12. Caris-Verhallen, W.M.; Kerkstra, A.; Bensing, J.M. The role of communications in nursing care for elderly people: A review of the literature. *J. Adv. Nurs.* **1997**, *25*, 915–933. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Grainger, K. Communication and the institutionalized elderly. In *Handbook of Communication and Aging Research*; Routledge: Abingdon, UK, 2004; pp. 479–497.
14. Allen, C.I.; Turner, P.S. The effect of an intervention programme on interactions on a continuing care ward for older people. *J. Adv. Nurs.* **1991**, *16*, 1172–1177. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Allen-Burge, R.; Burgio, L.D.; Bourgeois, M.S.; Sims, R.; Nunnikhoven, J. Increasing communication among nursing home residents. *J. Clin. Geropsychol.* **2001**, *7*, 213–230. [\[CrossRef\]](#)
16. Nieuwenhuis, R. Breaking the speech barrier. *Nurs. Times* **1989**, *85*, 34. [\[PubMed\]](#)
17. Kopp, S.; Brandt, M.; Buschmeier, H.; Cyra, K.; Freigang, F.; Krämer, N.; Straßmann, C. Conversational Assistants for Elderly Users—The Importance of Socially Cooperative Dialogue. In Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications, Stockholm, Sweden, 15 July 2018.
18. Young, V.; Mihailidis, A. Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review. *Assist. Technol.* **2010**, *22*, 99–112. [\[CrossRef\]](#)
19. Yaghoubzadeh, R.; Kramer, M.; Pitsch, K.; Kopp, S. Virtual agents as daily assistants for elderly or cognitively impaired people. In Proceedings of the International Workshop on Intelligent Virtual Agents, Edinburgh, UK, 29–31 August 2013; pp. 79–91.
20. Feil-Seifer, D.; Mataric, M.J. Defining socially assistive robotics. In Proceedings of the 9th International Conference on Rehabilitation Robotics—ICORR 2005, Chicago, IL, USA, 28 June–1 July 2005; pp. 465–468.
21. Gomi, T.; Griffith, A. Developing intelligent wheelchairs for the handicapped. In *Assistive Technology and Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 1998; pp. 150–178.
22. Shiomi, M.; Iio, T.; Kamei, K.; Sharma, C.; Hagita, N. Effectiveness of social behaviors for autonomous wheelchair robot to support elderly people in Japan. *PLoS ONE* **2015**, *10*, e0128031. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Kazerooni, H.; Steger, R.; Huang, L. Hybrid control of the Berkeley lower extremity exoskeleton (BLEEX). *Int. J. Robot. Res.* **2006**, *25*, 561–573. [\[CrossRef\]](#)
24. Kiguchi, K.; Rahman, M.H.; Sasaki, M.; Teramoto, K. Development of a 3DOF mobile exoskeleton robot for human upper-limb motion assist. *Robot. Auton. Syst.* **2008**, *56*, 678–691. [\[CrossRef\]](#)
25. Pollack, M.E.; Brown, L.; Colbry, D.; Orosz, C.; Peintner, B.; Ramakrishnan, S.; Thrun, S. Pearl: A mobile robotic assistant for the elderly. In Proceedings of the AAAI Workshop on Automation as Eldercare, Edmonton, AB, Canada, 28–29 July 2002; Volume 2002, pp. 85–91.
26. Beskow, J.; Edlund, J.; Granström, B.; Gustafson, J.; Skantze, G.; Tobiasson, H. The MonAMI Reminder: A spoken dialogue system for face-to-face interaction. In Proceedings of the Tenth Annual Conference of the International Speech Communication Association, Brighton, UK, 6–10 September 2009.
27. Noury, N.; Rumeau, P.; Bourke, A.K.; ÓLaighin, G.; Lundy, J.E. A proposal for the classification and evaluation of fall detectors. *IRBM* **2008**, *29*, 340–349. [\[CrossRef\]](#)
28. Iio, T.; Shiomi, M.; Kamei, K.; Sharma, C.; Hagita, N. Social acceptance by senior citizens and caregivers of a fall detection system using range sensors in a nursing home. *Adv. Robot.* **2016**, *30*, 190–205. [\[CrossRef\]](#)
29. Dario, P.; Guglielmelli, E.; Laschi, C.; Teti, G. MOVAID: A personal robot in everyday life of disabled and elderly people. *Technol. Disabil.* **1999**, *10*, 77–93. [\[CrossRef\]](#)
30. Graf, B.; Hans, M.; Schraft, R.D. Mobile robot assistants. *IEEE Robot. Autom. Mag.* **2004**, *11*, 67–77. [\[CrossRef\]](#)
31. Iwamura, Y.; Shiomi, M.; Kanda, T.; Ishiguro, H.; Hagita, N. Do elderly people prefer a conversational humanoid as a shopping assistant partner in supermarkets? In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, 6–9 March 2011; pp. 449–456.
32. Shibata, T.; Wada, K. Robot therapy: A new approach for mental healthcare of the elderly—a mini-review. *Gerontology* **2011**, *57*, 378–386. [\[CrossRef\]](#) [\[PubMed\]](#)

33. Kanamori, M.; Suzuki, M.; Oshiro, H.; Tanaka, M.; Inoguchi, T.; Takasugi, H.; Yokoyama, T. Pilot study on improvement of quality of life among elderly using a pet-type robot. In Proceedings of the 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation. Computational Intelligence in Robotics and Automation for the New Millennium (Cat. No. 03EX694), Kobe, Japan, 16–20 July 2003; Volume 1, pp. 107–112.
34. Tamura, T.; Yonemitsu, S.; Itoh, A.; Oikawa, D.; Kawakami, A.; Higashi, Y.; Fujimoto, T.; Nakajima, K. Is an entertainment robot useful in the care of elderly people with severe dementia? *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2004**, *59*, M83–M85. [[CrossRef](#)] [[PubMed](#)]
35. Libin, A.; Cohen-Mansfield, J. Therapeutic robocat for nursing home residents with dementia: Preliminary inquiry. *Am. J. Alzheimer's Dis. Other Dement.* **2004**, *19*, 111–116. [[CrossRef](#)] [[PubMed](#)]
36. Abdollahi, H.; Mollahosseini, A.; Lane, J.T.; Mahoor, M.H. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. In Proceedings of the 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids), Birmingham, UK, 15–17 November 2017; pp. 541–546.
37. Yaghoubzadeh, R.; Buschmeier, H.; Kopp, S. Socially cooperative behavior for artificial companions for elderly and cognitively impaired people. In Proceedings of the 1st International Symposium on Companion-Technology, Ulm, Germany, 23–25 September 2015.
38. Kanoh, M.; Oida, Y.; Nomura, Y.; Araki, A.; Konagaya, Y.; Ihara, K.; Kimura, K. Examination of practicability of communication robot-assisted activity program for elderly people. *J. Robot. Mechatron.* **2011**, *23*, 3. [[CrossRef](#)]
39. Khosla, R.; Chu, M.T.; Kachouie, R.; Yamada, K.; Yamaguchi, T. Embodying care in Matilda: An affective communication robot for the elderly in Australia. In Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, Miami, FL, USA, 28–30 January 2012; pp. 295–304.
40. Minami, H.; Kawanami, H.; Kanbara, M.; Hagita, N. Chat robot coupling machine responses and social media comments for continuous conversation. In Proceedings of the 2016 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Seattle, WA, USA, 11–15 July 2016; pp. 1–6.
41. Otaki, H.; Otake, M. Interactive Robotic System Assisting Image Based Dialogue for the Purpose of Cognitive Training of Older Adults. In Proceedings of the 2017 AAAI Spring Symposium Series, Stanford, CA, USA, 27–29 March 2017.
42. Otake, M.; Kato, M.; Takagi, T.; Asama, H. The coimagination method and its evaluation via the conversation interactivity measuring method. In *Early Detection and Rehabilitation Technologies for Dementia: Neuroscience and Biomedical Applications*; IGI Global: Hershey, PA, USA, 2011; pp. 356–364.
43. Sakakibara, S.; Saiki, S.; Nakamura, M.; Yasuda, K. Generating personalized dialogue towards daily counseling system for home dementia care. In Proceedings of the International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management, Vancouver, BC, Canada, 9–14 July 2017; pp. 161–172.
44. Jokinen, K. Dialogue models for socially intelligent robots. In Proceedings of the International Conference on Social Robotics, Qingdao, China, 28–30 November 2018; pp. 127–138.
45. Jokinen, K.; Nishimura, S.; Watanabe, K.; Nishimura, T. Human-robot dialogues for explaining activities. In Proceedings of the 9th International Workshop on Spoken Dialogue System Technology, Singapore, 18–20 April 2019; pp. 239–251.
46. Arimoto, T.; Yoshikawa, Y.; Ishiguro, H. Multiple-robot conversational patterns for concealing incoherent responses. *Int. J. Soc. Robot.* **2018**, *10*, 583–593. [[CrossRef](#)]
47. Iio, T.; Yoshikawa, Y.; Ishiguro, H. Retaining Human-Robots Conversation: Comparing Single Robot to Multiple Robots in a Real Event. *J. Adv. Comput. Intell. Inform.* **2017**, *21*, 675–685. [[CrossRef](#)]
48. Todo, Y.; Nishimura, R.; Yamamoto, K.; Nakagawa, S. Development and evaluation of spoken dialog systems with one or two agents through two domains. In Proceedings of the International Conference on Text, Speech and Dialogue, Pilsen, Czech Republic, 1–5 September 2013; pp. 185–192.
49. Shibahara, Y.; Yamamoto, K.; Nakagawa, S. Effect of sympathetic relation and unsympathetic relation in multi-agent spoken dialogue system. In Proceedings of the 2016 International Conference On Advanced Informatics: Concepts, Theory and Application (ICAICTA), George Town, Malaysia, 16–19 August 2016; pp. 1–6.

50. Arimoto, T.; Yoshikawa, Y.; Ishiguro, H. Nodding responses by collective proxy robots for enhancing social telepresence. In Proceedings of the Second International Conference on Human-Agent Interaction, Tsukuba, Japan, 28–30 October 2014; pp. 97–102.
51. Karatas, N.; Yoshikawa, S.; Okada, M. Namida: Sociable driving agents with multiparty conversation. In Proceedings of the Fourth International Conference on Human Agent Interaction, Singapore, 4–7 October 2016; pp. 35–42.
52. Sakamoto, D.; Hayashi, K.; Kanda, T.; Shiomi, M.; Koizumi, S.; Ishiguro, H.; Ogasawara, t.; Hagita, N. Humanoid robots as a broadcasting communication medium in open public spaces. *Int. J. Soc. Robot.* **2009**, *1*, 157–169. [[CrossRef](#)]
53. Goffman, E. *Forms of Talk*; University of Pennsylvania Press: Philadelphia, PA, USA, 1981.
54. Sacks, H.; Schegloff, E.A.; Jefferson, G. A simplest systematics for the organization of turn taking for conversation. In *Studies in the Organization of Conversational Interaction*; Academic Press: Cambridge, MA, USA, 1978; pp. 7–55.
55. Clark, H.H.; Carlson, T.B. Hearers and speech acts. *Language* **1982**, *58*, 332–373. [[CrossRef](#)]
56. Schegloff, E.A.; Sacks, H. Opening up closings. *Semiotica* **1973**, *8*, 289–327. [[CrossRef](#)]
57. Schegloff, E.A. *Sequence Organization in Interaction: A Primer in Conversation Analysis I (Vol. 1)*; Cambridge University Press: Cambridge, UK, 2007.
58. McCowan, I.A.; Moore, D.; Dines, J.; Gatica-Perez, D.; Flynn, M.; Wellner, P.; Bourlard, H. *On the Use of Information Retrieval Measures for Speech Recognition Evaluation (No. REP_WORK)*; IDIAP: Martigny, Switzerland, 2004.
59. The Jamovi Project. Jamovi (Version 0.9) [Computer Software]. 2019. Available online: <https://www.jamovi.org> (accessed on 20 February 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).