ROBOMECH Journal

RESEARCH ARTICLE

Open Access

Actively listening twin robots for long-duration conversation with the elderly

Toshiaki Nishio[1*], Yuichiro Yoshikawa[1], Takamasa Iio[2], Mariko Chiba[3], Taichi Asami[4], Yoshinori Isoda[3] and Hiroshi Ishiguro[1]

Abstract

The number of isolated elderly people with few opportunities to talk to other people is currently increasing. Research is ongoing to develop talking robots for addressing the situation. The aim of the present study was to develop a talking robot that could converse with elderly people over an extended period. To enable long-duration conversation, we added a previously proposed active listening function for twining the robot dialogue system to prompt the user to say something. To verify the effectiveness of this function, a comparative experiment was performed using the proposed robot system and a control system with identical functions except the active listening function. The results showed that the conversation of the elderly subjects with the proposed robot system was significantly more than that with the control system. The capability of the developed robot system was further demonstrated in a nursing home for the elderly, where its conversation durations with different residents were measured. The results revealed that the robot could converse for more than 30 min with more than half of the elderly subjects. These results indicate that the additional function of the proposed talking robot system would enable elderly people to talk over longer periods of time.

Keywords: Human–robot interaction, Conversational robot, Elderly people, Multiple robots, Active listening

Introduction

Many nations around the world have an aging population, with Japan particularly observing an increase in the number of isolated elderly people who usually have few opportunities to talk to other people [1]. This has prompted research on robot technologies for supporting elderly people [2]. There are currently many commercial robotic devices that assist with health care for the elderly [3], as well as social robots that support the daily and social activities of elderly people, categorized as service and companion types [4]. Service-type robots include those that assist with mobility [5, 6], home care [7, 8], and telecommunication [9, 10]. The risk of dementia also seems to be lower in elderly people with significant opportunities for interaction with family and friends compared with those without such interactions [11]. Companion-type robots [12] are thus expected to help with managing the mental and social well-being of elderly people [13].

Many animal-type companion robots are currently used to provide elderly people with nonverbal communication opportunities [14], similar to the use of real animals for animal-assisted therapy [15]. To provide verbal communication, effort has been made to develop robots with human-like features [16]. Some researchers have demonstrated that the introduction of conversational robots to nursing homes for the elderly would be beneficial [17, 18]. For such robots to be useful through the day, they should be capable of continuous conversation without boring the user [19].

The aim of the present study was to develop a conversation strategy for a pair of robots that can cooperatively talk to and continue a conversation with an elderly person over an extended period. Although speech recognition

*Correspondence: nishio.toshiaki@irl.sys.es.osaka-u.ac.jp
[1] Graduate School of Engineering Science, Osaka University, Osaka, Japan
Full list of author information is available at the end of the article

and response generation technologies are rapidly becoming more sophisticated [20–22], it is unlikely that the error probability will be reduced to zero. It is for this reason that Arimoto et al. proposed the collaboration of a pair of robots for conversation with a person, to achieve robustness against errors in understanding human speech [23]. Iio et al. built a twin-robot system that utilized this conversational strategy, for inducing a positive attitude in an elderly person in a nursing home through alternate questioning of the subject [24]. To skip a question from one robot that might be difficult for the subject, the other robot answered the question when the subject did not reply within a fixed time period. However, the waiting time period is difficult to fix because some elderlies simply require more time to respond. Moreover, for simplicity, the robots only waited till the end of the subject's utterance before producing a verbal response. This makes it difficult to adapt to situations in which the elderly subject is actively talking on a specific topic over a long period with intermittent pauses.

In the present study, we developed a conversation strategy that includes an adaptive active listening mode for addressing the two problem situations mentioned above, namely, when the subject is talking a lot and when the subject requires some time to reply. To verify the effectiveness of the proposed method, a comparative experiment was performed in the laboratory. We found that robots using the proposed method could induce longer utterances from the participating elderly subjects. The potential of the method in a real-life environment was demonstrated through field tests conducted over 2 days in a nursing home for the elderly. Through measurement of the conversation durations, we confirmed that the proposed method enabled conversation for longer than 30 min with more than half of the elderly subjects.

System

Iio et al. proposed a system comprising multiple robots for communicating with visitors in an exhibition hall through YES/NO buttons [25]. This type of conversational robot system operates in three states, namely, asking questions, responding to answers, and changing the conversation topic. In a latter study [24], Iio et al. extended the conversational system to communicate with elderly people. To enable continuation of the conversation when the elderly subject does not respond to a question asked by the robot, two robots that could interact with each other were employed in the system. In the present study, we added a new active listening function to the architecture of the latter system, to enable conversation with an elderly subject for as long as 30 min. As shown in Fig. 1, the proposed system, referred to as CommU, consists of a pair of desktop humanoid robots

Fig. 1 Appearance of the robots

with different hair styles and colors and differing voice models indicative of their character and gender. The blue one is a boy while the red one is a girl. Child-like characters were used to dispel thoughts in the elderly subject about the robots having a harmful motive. To reduce the fabrication cost, the movements of the CommU robots were limited to three degrees of freedom (DoFs) for the neck, and one DoF for the mouth movements, respectively.

The human utterances are captured by a microphone array installed below the robots and connected to a cloud server for determination of the message and the appropriate robot responses. The responses are communicated to the robots, which then respond with synthesized voices through two speakers placed behind them. Motor commands are sent to the robots for neck or mouth movements corresponding to the responses.

Twin robot dialogue system [24]

To make the user feel attended to by the robots and acquire a sense of satisfaction and self-esteem, the system uses a strategy in which it asks a series of questions about the user on a given topic based on a database consisting of question–response sets. A question–response set includes (1) a question sentence, (2) a list of possible words expected from the user's answer, (3) specific responses of the system for each possible word of the user, and (4) ambiguous responses of the system when the answer of the user does not include any of the expected words. Table 1 presents an example of the question–response sets. The system robots wait for a while after asking a question. If the user answers with any of

Table 1 Example of a question–response set of the conversational robot system

(1) A question sentence	(2) Expected words	(3) An acknowledgement and comment	(4) An ambiguous response
Did you like sweets when you were a child?	"Yes,""yeah,""yep,""like"…	You liked sweets? Children usually like sweets, don't they?	–
	"No,""nope,""not like"…	You didn't like sweets? I think it's rare	–
	–	I see	I heard that children usually like sweets

the expected words that is successfully recognized, a response is selected and presented by the robots. Otherwise, an ambiguous comment is made to the user. The robots used in this study were designed to repeat questions pertaining to the same topic to avoid the potential risk of conversation breakdown. The topic of discussion was changed based on the satisfaction of any of the three conditions: (i) four questions pertaining to the same topic had been asked, (ii) the user provided a negative response to a question, or (iii) the system recognized the user's response based on keywords associated with other topics. It should be noted that an appropriate nonverbal behavior of both robots is associated with any utterance of the subject and is accordingly executed (e.g., the robots nod when they utter "*I see.*").

The system may make a comment that is more distantly related to the user's utterance when the system fails to recognize the words of, or intention behind, the utterance. To decrease the user's disappointment by this type of conversational breakdown, the speaking robot and the addressee robot can be changed when a specific comment is made [23]. We built a robust dialogue system with this strategy added to the basic repetitive questioning strategy based on the question–response database. This enabled alternation of the questioning between the two robots.

In the system, the two robots speak alternately. In questioning mode, when one robot queries the user, the other robot acknowledges the user's answer. Subsequently, the querying robot produces a comment or an ambiguous response. In the next query, the two robots exchange roles with each other. In this way, we intend to equalize the numbers of utterances made by the two robots.

Adaptive listening modes

We incorporated a function for actively switching the operation mode of the twin robot conversation system. In addition to the questioning mode, which is the normal operation mode of the system, two other modes were added, namely, listening and prompting modes. In the listening mode, the robots actively listen to the user on a topic instead of moving on to ask another question. Conversely, in the prompting mode, they focus on extracting small responses from the user when the latter is not forthcoming.

Figure 2 shows the switching between the different operation modes. When the utterance of the user lasts for more than 4 s, the system switches to the listening mode. In this mode, the robots first ambiguously ask the user to provide more information about the recent answer (e.g., "Please tell me the details of that story"). Then, the two robots alternately repeat a short reaction for every breath of the user, (e.g., "Then?" or "Hum") until the system judges that the user has finished talking. When the user does not utter something for 3 s after a short reaction of the robots, or when the user has repeated a short utterance that lasted less than 1 s for each of three short reactions of the robots, the system switches to the questioning mode. To avoid surprising the user by a potentially sudden topic change, one of the robots suggests asking the user another question, as if the first robot just remembered (e.g., "By the way, I have something to ask him (or her)"). Then, the second robot shows interest in the last question by the first robot (i.e., "Oh, what?"). In addition, when the user utters something for more than 1 s more than four times, the two robots praise the user in turn before switching back to the questioning mode (e.g., "You know a lot of things," "I can learn by talking with you," or "I want to talk with you more"). After one of the robots expresses praise, the other robot suggests asking the user another question.

The system switches to the prompting mode when the user has not answered the last three questions. In this mode, the two robots start to talk to each other to break the silence. They alternately utter in the conversation. The robots may ask the user a question as in the questioning mode, but the questions would be easier than those normally asked in the questioning mode. For

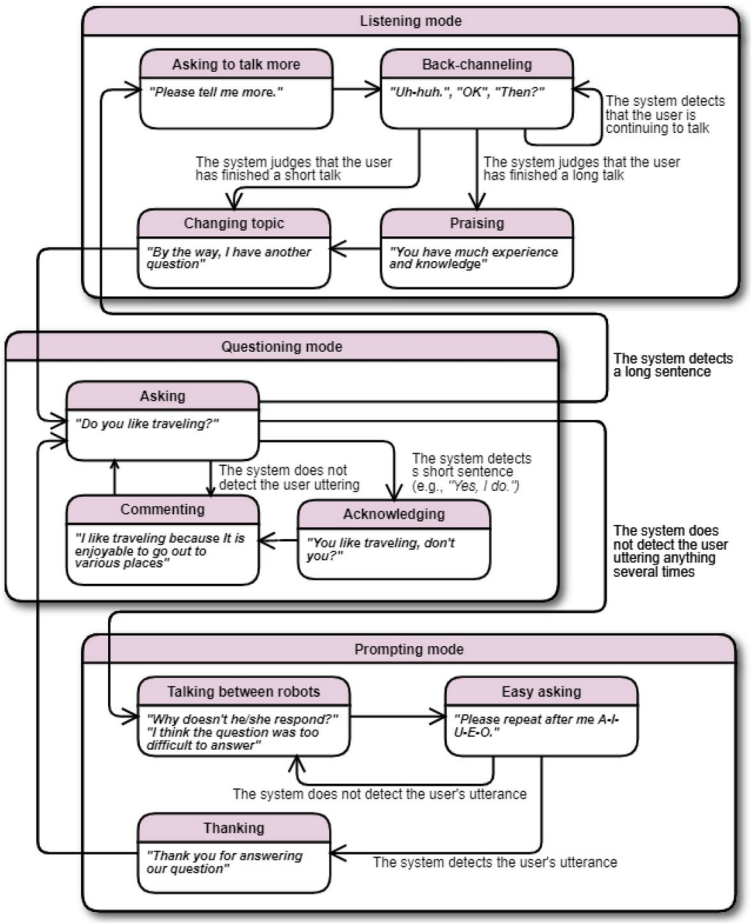

Fig. 2 Switching between the different modes of the robot conversation system

example, the robots may ask the user for their name. The robots may also utter something nonsensical such as the sequence of vowels (a-e-i-o-u), or move their hands, legs, or head. They also decrease the speed of their utterances to facilitate clear hearing by the user. If the user utters something to one of the robots, the system would switch back to the questioning mode with a comment of thanks by the other robot. If the user does not answer, the robots keep talking to each other. Each specific parameter of the switching system was determined by preliminary experiments to ensure smooth conversation between the robots and the elderly user.

Laboratory experiment

An experiment was conducted to examine the effect of the proposed twin robot system on the user's behavior and the user's impression of the system. The adaptive listening modes of the proposed twin robot conversation system was also comparatively evaluated by applying a control system without the modes. The healthy elderly participants of the experiment were required to talk to each system for a maximum of 15 min and the amount of utterances of the participants for the two systems were compared, as well as their impressions of the systems.

Participants

The study was approved by the ethics committee of the Graduate School of Engineering Science, Osaka University. The participants comprised 24 healthy elderly Japanese recruited by a public organization for human resourcing of the elderly. They included 15 males and nine females, ages 63–79 years (mean = 70.7 years, standard deviation (SD) = 4.55 years). The age range of the participants of the test of the proposed system with the adaptive listening function was 63–79 years (mean = 69.8 years, SD = 5.18 years old), whereas it was 67–75 years (mean = 70.9 years, SD = of 2.87 years) for the participants in the control test using the system without the innovative function. This reveals no significant difference between the two groups of participants

($t(17) = 0.64$ ns). The state of health of the participants allowed them to live in their house and visit the laboratory by themselves. None exhibited a recognizable dementia symptom.

Apparatus

The experiment was conducted in a room at Osaka University. Figure 3 shows an example scene of a participant talking to the robots. The experiment included three types of conversations: small talk at the beginning of the conversation; a question–response conversation; and another small talk at the end of the conversation. The entire conversation lasted for a maximum time length of 15 min. The time length was determined by a pilot experiment and ensured that each participant was allowed enough time to answer a number of questions for the collection of a sufficient amount of data. After every 5 min, the robots asked the participants whether they wanted to continue the conversation.

The question–response sets included 198 questions. Each question was assigned a topic such as childhood memories, travel experiences, health, and small talk. The considered conversation scenarios and the questions presented to the participants were designed in consultation with a scenario writer experienced in elderly care. The scenarios and questions were finalized after performing preliminary experiments with different elderly people. The script used in the present study was adapted to each participant to enable the robots to call the name of the participant. It should be noted that the difference between the control system and the present system is the incorporation of an adaptive active listening mode in the latter.

Procedure

In the experiment, the participant sat in front of the robots and the experimental procedure was explained to them. The participant was instructed to tell the robots about themselves as much as possible. Whenever the participant became tired of talking to the robots, he or she could stop the conversation. The experimenter then left the room and the robots were made to start talking. The participant conversed with either the proposed or the control system. After the conversation, they were required to complete a questionnaire about their impression of the system.

Fig. 3 Scene of a conversation between a participant and the robots during the laboratory experiment

Measurement

The proposed system was compared with the control system from both objective and subjective viewpoints. From an objective viewpoint, we evaluated the amount of utterances by measuring the total duration when the participant uttered something louder than a certain volume and calculated the average utterances relative to the total opportunities to answer. From a subjective viewpoint, a questionnaire was used to ask the participants about their feeling of being listened to. A questionnaire for nursing research in Japan was employed for this purpose [26]. The three requirements advocated for by Rogers [27], a pioneering researcher in the field of active listening, were covered in the questionnaire, namely, empathetic understanding, congruence, and unconditional positive regard.

Results

All the participants talked with the systems for more than 15 min. The average numbers of questions that the proposed system and the control system asked the participants in the questioning mode were 24.75 (SD = 9.23) and 33.25 (SD = 3.65), respectively. The average numbers of questions that the proposed system and the control system got responses from the participants were 24.75 (SD = 9.23) and 33.0 (SD = 3.52), respectively. The average number of questions that the listening mode of the proposed system was activated was 11.08 (SD = 5.21). No participant became silent in response to a question from the robots, and the prompting mode was thus not activated throughout the experiment.

Figure 4 shows box plots of the amount of utterances for the two considered systems. A Mann–Whitney U test was used to compare the results. The plot for the proposed system was observed to be significantly larger than that for the control system ($U = 32.0$, $p < 0.05$). As can also be seen from Fig. 5, there were no significant differences between the participants' feelings of being listened to for the two systems.

Field test

To demonstrate the ability of the proposed system to talk to elderly people for extended periods in real environments, a field test was conducted at a nursing home in Japan. Elderly residents of the nursing home talked to

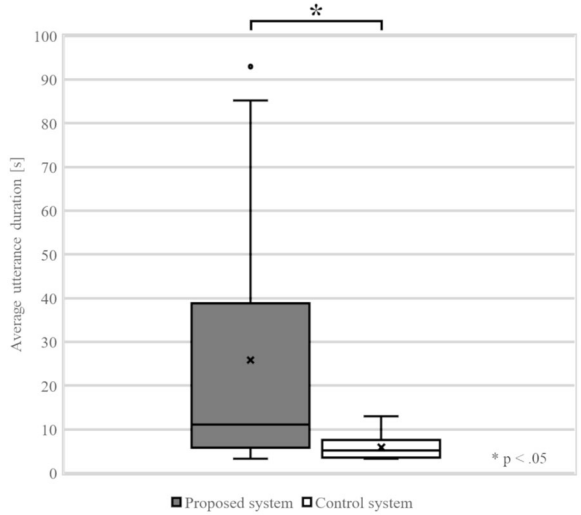

Fig. 4 Average utterance duration

the system, and the durations of the conversations were measured. To examine how the novelty of the system contributed to the conversation duration, the test was conducted over 2 days with an intervening day.

Participants

The nursing home field test was approved by the ethics committee of the Graduate School of Engineering Science, Osaka University. The participants were 12 females of ages 82–98 years (mean = 89.0 years, SD = 4.02 years). The data for two participants who could not hear the voice of the robots or human experimenter were excluded. Two other participants were not willing to join the experiment on the second day. According to information provided by the nursing home staff, four of the 10 participants whose data were considered exhibited mild dementia symptoms, four exhibited moderate symptoms, and two severe symptoms.

Apparatus

The robots were placed on a table in the conversation space at the corner of a corridor in the nursing home. Figure 6 shows an example scene of a participant talking to the robots. The question–response sets included 267 questions, which covered the same topics as in the laboratory experiment. In addition to the question–response sets, the system welcomed the participants and bid them farewell at the beginning and end of the conversation, respectively. While the welcome and farewell were used on both days of the test, the question–response sets

Fig. 6 Scene of a conversation during the field trial

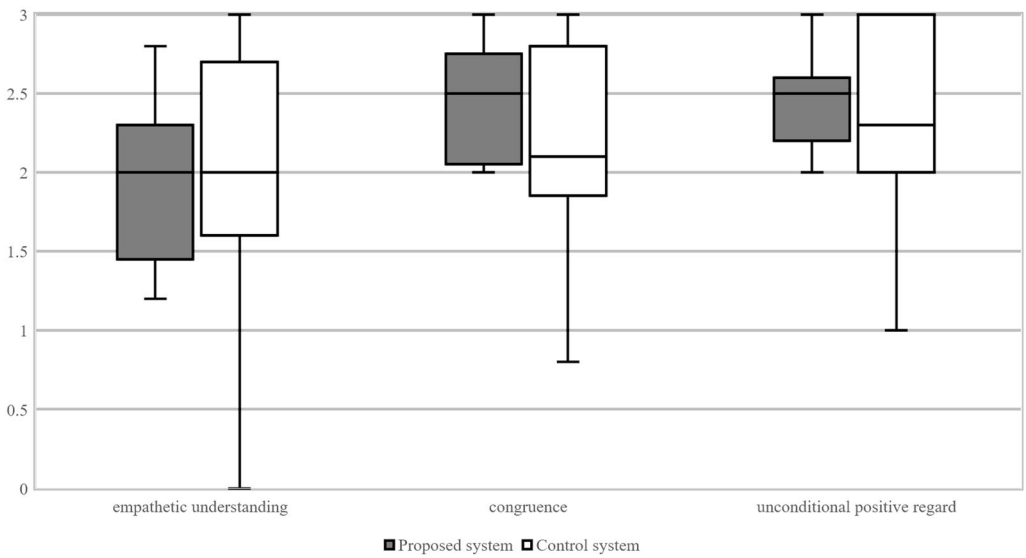

Fig. 5 Scores of feeling of being listened to

differed. The question–response script and the welcome and farewell utterances were adapted to each participant through incorporation of their names.

Procedure

The care giver in the nursing home brought a participant to the robot conversation space. The participant was seated in front of the robots and the experimenter sent a command to the system to begin the conversation. After 5 min, the robot asked the participant whether she wanted to continue the conversation. If she did not want to, the experimenter stopped the robots from talking. If the participant did not react to this question, the care giver was requested to determine her disposition. Every 5 min, the robots repeated the question of whether the participant wanted to continue the conversation. When the conversation duration exceeded 30 min, the farewell sequence was commenced instead of asking the next question. After the conversation, another experimenter interviewed the participant and the care giver about their impressions of the conversation.

Measurement

We measured the durations of the conversations between the system and the participants, namely the time between when the system produced the first word and when it produced the last one. The termination of a conversation duration was defined as when a participant clearly indicated her desire to stop the conversation. When a participant voluntarily decided to stop the conversation, the beginning of her utterance to stop the conversation was regarded as the termination time. If the care giver stopped the conversation, the beginning of the communication between the care giver and the participant was regarded as the termination time.

Results

Figure 7 shows a histogram of the conversation durations for the two days of the field test. On the first and second days, 50% and 62.5% of the participants respectively talked to the robots for longer than 30 min. The average conversation times were 24.03 min (SD = 9.67 min) and 25.88 min (SD = 9.07 min) on the first and second days, respectively.

The average numbers of questions that the participants were asked in the questioning mode was 26.33 (SD = 23.53). The average number of questions that the participants responded to was 20.28 (SD = 20.62). The average number of questions that the listening mode was activated was 8.17 (SD = 5.98). Five of ten participants experienced the prompting mode.

The feedback interviews considered all the test participants with the exception of the two who exhibited severe dementia. Three participants provided positive feedback comments such as "It was fun, "Please let us talk again," and "I remembered what I had forgotten, thank you." One participant provided an enthusiastic comment approximately 5 h after the experiment: "I talked with robots, and it was great." No participant provided a negative feedback such as a lack of comfort during their conversation with the robots. One care giver observed that some of the participants conversed more positively

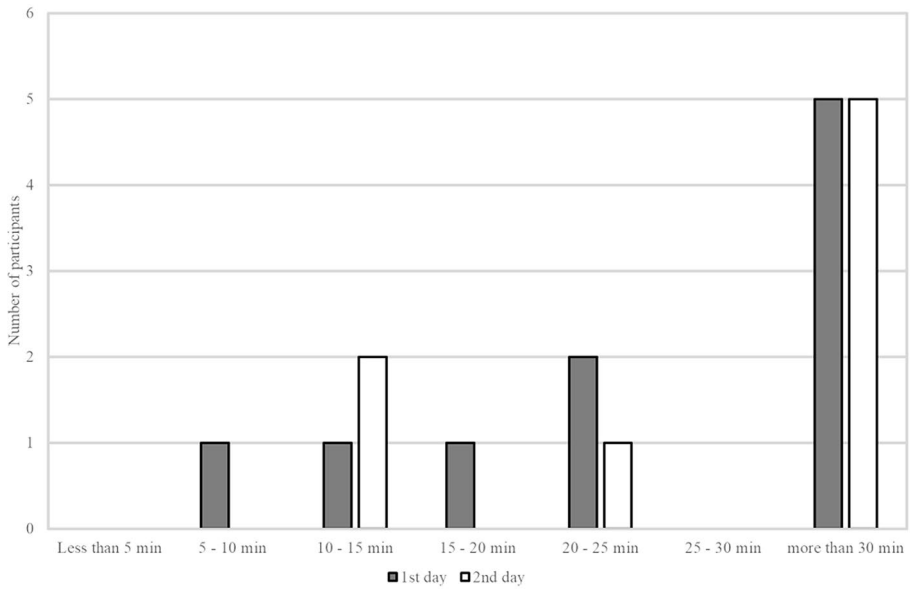

Fig. 7 Conversation durations for the field test participants

with the robots than they normally did with the caring giving staff of the nursing home. Some participants were also observed to talk longer to the robots and in a better mood than usual. These results suggest that the care givers would be disposed to using the proposed system to provide good conversation opportunities for their elderly wards. This is important for a broad application of the proposed system and other similar robotic systems.

Discussion

Implications

The laboratory experiment results indicated that the participants talked more in the listening mode, in which the robots requested the participants to provide more information about recent answers. We believe that the expected advantage of the prompting mode is that it encourages the elderly who are reluctant to respond to the robots' question to utter. However, because all of the participants in the laboratory experiment did not become silent in the conversation, the prompting mode was never activated. Thus, we could not statistically demonstrate the advantage of the prompting mode from the viewpoint of lengthening the duration of utterance. In contrast, video analysis of the field test revealed that the robots switched to the prompting mode for five participants who had remained silent. Three of them subsequently started or restarted to respond to simple and kind questions about their physical condition, such as "Are you okay?" and "Can you hear me?" While these cases show its potential, they could not do so statistically; thus, the prompting mode should be more carefully examined in the future in an experiment involving sufficient number of suitable participants.

Half of the participants continued with the conversation without quitting for more than 30 min on both days of the field test. Moreover, we received some positive feedback from the participants and care givers during the interviews conducted after the tests. The feedback suggested that the proposed system encouraged the elderly participants to talk more. The more frequent switch to the prompting mode in the nursing home field test was indicative of a lower engagement of the participants, attributable to their more degenerated health condition. The relationship between the participant engagement and their age and health condition is thus worthy of further investigation. From an applicative perspective, the 30-min conversation duration offers opportunities for a variety of potential uses of the proposed system. For example, the twin robots can be utilized for elderly cognitive rehabilitation through activities such as memory tests and/or quizzes. Further study to explore such applications is also worthy of pursuit.

Because the proposed system attempts to prompt a user to talk depending on their state, it is expected to provide a feeling of being listened to. However, no significant improvement in such feeling was observed in the present study. The maximum length of the conversation time during the pilot experiment performed using young participants was set to 15 min. This duration was considered sufficient for collecting adequate response samples for calculating the average amount of utterance. However, this duration may be too short for elderly participants to feel being listened to when questioned about personal topics, regardless of the adopted listening strategy. A probable alternative reason for this observation is that the proposed system did not deepen the conversation with regard to the prompted utterance. Because the proposed method complements existing methods for deepening conversation [28, 29], rather than being one for exclusive use, it should be integrated with such existing methods.

The proposed system utilizes a simple strategy for determining when the elderly user ends their speech by detecting silence over a prefixed duration. Although suitable parameters were determined through pilot experiments performed in this study, the voices of the participants and robots were observed to, on average, overlap approximately 6.5 times per session during the actual experiment. It would thus be worthwhile to update the system by incorporating relevant technologies for better detection of when a participant has finished talking, based on both semantic and prosodic information.

The practical use of the proposed robots in daily life would require them to maintain their conversational performance with elderly people over a long time. First, this would necessitate investigating whether elderly people would lose their motivation to talk to the robots after their first experience. The present field test did not reveal a significant negative tendency along this line over the two test days. This implies that the proposed system can maintain its performance for at least 2 days. CARESSES [30] is a related ambitious project aimed at providing the elderly with conversational opportunities over a span of a few weeks. However, it has been suggested that more than 2 months is required to investigate whether the novelty effect of talking robots would be eliminated when used over a long period [31]. Further study is thus needed to determine how long elderly people would accommodate the proposed robots as conversation partners.

Five of the 10 field test participants did not continue the conversation for up to 30 min on either of the two test days. Three of these spoke in voices that were too soft to be recognized by the system. They attempted to communicate with the robots by nodding or moving their hands, but the current system could also not recognize such

nonverbal responses. This contributed to the shortened conversation. To achieve a more robust system, it would be necessary to incorporate functions for recognizing nonverbal expressions, especially subtle ones [32]. This is another important avenue for further study.

Limitations

The present study had five major limitations. First, the number of participants in the field test was small, for which reason care must be taken in generalizing the findings. The small sample size was influenced by the cost of conducting a field test with elderly participants in deteriorating health states in a nursing home. The involvement of nursing home care givers without appropriate training also potentially impacted the quality of the field test.

Second, it is not clear in which mode, the listening or questioning mode, the participants were promoted to talk longer by the proposed method. To evaluate such a specificity of the influence, we need to know how long and from when the participants were promoted to do so; unfortunately, this was difficult to ascertain accurately in the current setup. For example, the dialogue system in the questioning mode of the current experiment was not completely flexible to the participant's motivation to talk. Namely, in the control condition, even though the participant wanted to keep talking, the system forcibly switched to the next question when it detected a breath group with a certain length, which could have underestimated the baseline of the talking duration. On the contrary, the participants in the experimental condition could have experience to talk longer for specific questions, which might promote them to talk longer even in the later questions. Due to the randomness of the questions and the varieties in the participant's preferences, the order of questions could not be controlled, which could have under- or over-estimated the specificity of the influence of the listening mode. Considering these problems in the setup, it was difficult to evaluate to what extent the participants were promoted to talk longer in either mode, only by analyzing the obtained data of uttering duration for each question. Thus, we need to run a different experiment where we give participants controlled opportunities to talk freely against the same questions after some turns to talk to the robots in various modes.

The third limitation is that it is not clear to what extent the proposed method is limited to multiple robots. There is a possibility that having both prompting and listening modes could improve the conversations of conventional dialogue systems with a single robot. In the dialogue system examined in this paper, however, making the dialogues in these modes sense or sound natural was accomplished not by utterances by a single robot but also by the coordination between the two robots. As future work, therefore, it is worth extending the dialogues in these modes to be accomplished only by a single robot and examining the merits and demerits of utilizing a single versus multiple robots with adaptive listening modes.

Fourth, this study adopted a root-driven conversation strategy. Mavridis has reported the potential of mixed initiative dialogues, wherein two robots and a user can take the initiative to break into a conversation [32]. However, the proposed system does not feature a mixed dialogue function, with its utility limited to the detection of key words in the responses of the user for the selection of the next discussion topic. The function for switching topics can therefore be further developed to facilitate mixed dialogue.

The fifth limitation of the study consists in its consideration of only Japanese culture in the development of the conversation scenario and questions presented to the participants. Bruno et al. [33, 34] presented a knowledge-based robot capable of adapting to the cultural background of the user for expanded utility. It has also been found that people with different cultural backgrounds exhibit different degrees of trust in robots [35, 36]. There is thus room for further study to examine the reproducibility of the present results for users with different cultural backgrounds.

Conclusions and future work

A twin robot dialogue system incorporating an adaptive active listening mode was developed for providing isolated elderly people with conversation opportunities over an extended time. A laboratory experiment showed that the elderly participants talked to the system significantly more than was observed for a previous system. A field test conducted over 2 days in a nursing home also revealed that half of the elderly participants conversed with the proposed system for more than 30 min. These results support the potential usefulness of the proposed system for enhancing elderly conversation in the real world. However, there is room for the further development of the system with integrated functions for broader understanding of the user, and for further evaluation of its long-term attractiveness to the elderly.

Acknowledgements

Not applicable.

Authors' contributions

TN and YY prepared the materials and collected and analyzed the data. TN wrote the first draft of the manuscript, which all the authors then commented on. The final manuscript was also written by TN and approved by all the authors. All authors read and approved the final manuscript.

Funding

This work was supported by the Japan Science and Technology Agency (JST) Exploratory Research for Advanced Technology (ERATO) (Grant No. JPMJER1401) and JSPS KAKENHI (Grant No. JP19H05691).

Availability of data and materials
The datasets used and/or analyzed in this study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
Ethical approval was granted by the ethics committee of the Graduate School of Engineering Science, Osaka University.

Consent for publication
Written consent was obtained from participants for publication of this report and any accompanying images.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Graduate School of Engineering Science, Osaka University, Osaka, Japan. [2]University of Tsukuba, Ibaraki, Japan. [3]NTT DOCOMO, Inc., Tokyo, Japan. [4]NTT Media Intelligence Laboratories, NTT Corporation, Kanagawa, Japan.

Received: 30 November 2020 Accepted: 14 June 2021
Published online: 24 June 2021

References

1. United Nations, Department of Economic and Social Affairs, Population Division (2017) World population prospects: key findings and advance tables. Working paper no. ESA/P/WP/248. Available via https://esa.un.org/unpd/wpp/publications/Files/WPP2017_KeyFindings.pdf. Accessed 11 Sep 2017.
2. Zafrani O, Nimrod G (2018) Towards a holistic approach to studying human–robot interaction in later life. Gerontologist 59(1):e26–e36
3. Neumann D (2016) Human assistant robotics in Japan – challenges and opportunities for European companies. EU–Japan Centre for Industrial Cooperation, Tokyo
4. Broekens J, Heerink M, Rosendal H (2009) Assistive social robots in elderly care: a review. Gerontechnology 8(2):94–103
5. Graf B, Hans M, Schraft RD (2004) Care-O-bot II—development of a next generation robotic home assistant. Auton Robot 16(2):193–205
6. Pollack ME, Brown L, Colbry D, et al. (2002) Pearl: a mobile robotic assistant for the elderly. In: AAAI workshop on automation as eldercare. AAAI, Palo Alto, pp 85–91.
7. Dario P, Guglielmelli E, Laschi C, Teti G (1999) MOVAID: a personal robot in everyday life of disabled and elderly people. Technol Disabil 10(2):77–93
8. Yamazaki K, Ueda R, Nozawa S et al (2012) Home-assistant robot for an aging society. Proc IEEE 100(8):2429–2441
9. van Osch M, Bera D, van Hee K, Koks Y, Zeegers H (2014) Tele-operated service robots: ROSE. Autom Constr 39:152–160
10. Camanio Care AB, Giraff. http://www.giraff.org, Accessed 1 Dec 2019.
11. Fratiglioni L, Wang H-X, Ericsson K, Maytan M, Winblad B (2000) Influence of social network on occurrence of dementia: a community-based longitudinal study. Lancet 355(9212):1315–1319
12. Robinson H, MacDonald B, Broadbent E (2014) The role of healthcare robots for older people at home: a review. Int J Soc Robot 6(4):575–591
13. Kachouie R, Sedighadeli S, Khosla R, Chu M-T (2014) Socially assistive robots in elderly care: a mixed-method systematic literature review. Int J Hum Comput Interact 30(5):369–393
14. Wada K, Shibata T, Saito T, Tanie K (2003) Effects of robot assisted activity to elderly people who stay at a health service facility for the aged. In: Proceedings 2003 IEEE/RSJ international conference on intelligent robots and systems. IEEE, Piscataway, pp 2847-2852
15. Fine AH (ed) (2010) Handbook on animal-assisted therapy: theoretical foundations and guidelines for practice. Academic Press, Cambridge
16. Hutson S, Lim SL, Bentley PJ, Bianchi-Berthouze N, Bowling A (2011) Investigating the suitability of social robots for the wellbeing of the elderly. Fla Dent J 44:578–587
17. Kanoh M, Oida Y, Nomura Y et al (2011) Examination of practicability of communication robot-assisted activity program for elderly people. J Robot Mechatron 23(1):3–12
18. Rouaix N, Retru-Chavastel L, Rigaud A-S, Monnet C, Lenoir H, Pino M (2017) Affective and engagement issues in the conception and assessment of a robot-assisted psychomotor therapy for persons with dementia. Front Psychol 8:950
19. Sasama R, Yamaguchi T, Yamada K (2011) An experiment for motivating elderly people with robot guided interaction. International conference on universal access in human–computer interaction. Springer, Berlin, Heidelberg, pp 214–223
20. Karpagavalli S, Chandra E (2016) A review on automatic speech recognition architecture and approaches. Int J Signal Process Image Process Pattern Recognit 9(4):393–404
21. Nassif AB, Shahin I, Attili I, Azzeh M, Shaalan K (2019) Speech recognition using deep neural networks: a systematic review. IEEE Access 7:19143–19165
22. Xiong W, Wu L, Alleva F, Droppo J, Huang X, Stolcke A (2018) The Microsoft 2017 conversational speech recognition system. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, Piscataway, pp 5934-5938
23. Arimoto T, Yoshikawa Y, Ishiguro H (2018) Multiple-robot conversational patterns for concealing incoherent responses. Int J Soc Robot 10(5):583–593
24. Iio T, Yoshikawa Y, Chiba M, Asami T, Isoda Y, Ishiguro H (2020) Twin-robot dialogue system with robustness against speech recognition failure in human-robot dialogue with elderly people. Appl Sci 10(4):1522
25. Iio T, Yoshikawa Y, Ishiguro H (2017) Retaining human-robots conversation: comparing single robot to multiple robots in a real event. J Adv Comput Intell Inform 21(4):675–685
26. Mishima N, Shinkoda H, Hukumoto H (1999) Active listening for nursing care. Medicus Shuppan Publishers, Osaka (**[In Japanese]**)
27. Rogers CR (1959) A theory of therapy, personality, and interpersonal relationships: as developed in the client-centered framework, vol 3. McGraw-Hill, New York, pp 184–256
28. Higashinaka R, Imamura K, Meguro T, et al. (2014) Towards an open-domain conversational system fully based on natural language processing. In: Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers. ACL, Stroudsburg, pp 928–939.
29. Sakai K, Higashinaka R, Yoshikawa Y, Ishiguro H, Tomita J (2018) Introduction method for argumentative dialogue using paired question-answering interchange about personality. In: Proceedings of the 19th annual SIGdial meeting on discourse and dialogue. ACL, Stroudsburg, pp 70–79.
30. Bruno B, Chong NY, Kamide H, Kanoria S, Lee J, Lim Y, Pandey AK, Papadopoulos C, Papadopoulos I, Pecora F, Saffiotti A (2017) The CARESSES EU-Japan project: making assistive robots culturally competent. Italian forum of ambient assisted living. Springer, Cham, pp 151–169
31. de Graaf MMA, Allouch SB, Klamer T (2015) Sharing a life with Harvey: exploring the acceptance of and relationship-building with a social robot. Comput Hum Behav 43:1–14
32. Mavridis N (2015) A review of verbal and non-verbal human–robot interactive communication. Robot Auton Syst 63:22–35
33. Bruno B, Chong NY, Kamide H, et al. (2017) Paving the way for culturally competent robots: a position paper. In: 26th IEEE international symposium on robot and human interactive communication (RO-MAN). Pp 553–560.
34. Bruno B, Recchiuto CT, Papadopoulos I et al (2019) Knowledge representation for culturally competent personal robots: requirements, design principles, implementation, and assessment. Int J Soc Robot 11(3):515–538
35. Andrist S, Ziadee M, Boukaram H, Mutlu B, Sakr M (2015) Effects of culture on the credibility of robot speech: a comparison between English and Arabic. In: Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction. pp 157–164.
36. Evers V, Maldonado H, Brodecki T, Hinds P (2008) Relational vs. group self-construal: untangling the role of national culture in HRI. In: 3rd ACM/IEEE international conference on human-robot interaction (HRI). pp 255–262.

Publisher's Note