Temporal Patterns in Multi-modal Social Interaction between Elderly Users and Service Robot

WANG, Ning, DI NUOVO, Alessandro <http://orcid.org/0000-0001-5308-7961>, CANGELOSI, Angelo and JONES, Ray

Temporal Patterns in Multi-modal Social Interaction between Elderly Users and Service Robot

Ning Wang, Alessandro Di Nuovo, Angelo Cangelosi, and Ray Jones

***Abstract*—Social interaction, especially for older people living alone is a challenge currently facing human-robot interaction (HRI). User interfaces to manage service robots in home environments need to be tailored for older people. Multi-modal interfaces providing users with more than one communication option seem promising. There has been little research on user preference towards HRI interfaces; most studies have focused on utility and functionality of the interface. In this paper, we took both objective observations and participants' opinions into account in studying older users with a robot partner. Our study was under the framework of the EU FP7 Robot-Era Project. The developed dual-modal robot interface offered older users options of speech or touch screen to perform tasks. Fifteen people aged from 70 to 89 years old, participated. We analyzed the spontaneous actions of the participants, including their attentional activities (eye contacts) and conversational activities, the temporal characteristics (timestamps, duration of events, event transitions) of these social behaviours, as well as questionnaires. This combination of data distinguishes it from other studies that focused on questionnaire ratings only. There were three main findings. First, the design of the Robot-Era interface was very acceptable for older users. Secondly, most older people used both speech and tablet to perform the food delivery service, with no difference in their preferences towards either. Thirdly, these older people had frequent and long-duration eye contact with the robot during their conversations, showing patience when expecting the robot to respond. They enjoyed the service. Overall, social engagement with the robot demonstrated by older people was no different from what might be expected towards a human partner. This study is an early attempt to reveal the social connections between human beings and a personal robot in real life. Our observations and findings should inspire new insights in HRI research and eventually contribute to next-generation intelligent robot development.**

***Index Terms*—Human-robot interaction, ageing population, multi-modal user interface, social behaviours, service robot.**

This work was partially supported by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant No. 288899 and the National Natural Science Foundation of China under Grant No. 61803039.

Ning Wang is with the School of Information Engineering, Chang'an University, Middle-section of Nan'er Huan Road, Xi'an, Shaanxi Province, 710064, China, School of Computing, Electronics and Mathematics, Plymouth University, Drake Circus, Plymouth PL4 8AA, United Kingdom and the Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong nwang@cse.cuhk.edu.hk

Alessandro Di Nuovo is now with the Centre for Automation and Robotics Research, Sheffield Hallam University, United Kingdom, he was with the School of Computing, Electronics and Mathematics, Faculty of Science and Engineering, Plymouth University, United Kingdom a.dinuovo@shu.ac.uk

Angelo Cangelosi is a Professor of Machine Learning and Robotics with School of Computer Science, The University of Manchester, United Kingdom, and Visiting Professor with the School of Computing, Electronics and Mathematics, Faculty of Science and Engineering, Plymouth University, United Kingdom angelo.cangelosi@manchester.ac.uk

Ray Jones is a Professor of Health Informatics with School of Nursing and Midwifery, Faculty of Health and Human Sciences, Plymouth University, United Kingdom ray.jones@plymouth.ac.uk

I. Introduction

A. Elderly People-Robots Interaction

The ageing society is a difficult challenge facing most developed countries. As many older people would like to live independently in their own homes [1], there is a high demand for staff in housekeeping and domestic service. According to the United Nations, the birth rate is well below replacement level [2] and these workforce shortages are getting worse. To address the challenge there have been many attempts to develop service robots that can be easily used by older people in their own homes. As many older people experience gradual cognitive and perception decline, new designs in future retirement apartments include not only service robots, but also smart homes, and assistive facilities like mobility support [3]. For example, Robot *Robear* developed by RIKEN and Sumitomo Riko is a nursing care robot capable of lifting patients or disabled people from bed to wheelchair and providing stand-up support [4]. Nowadays, emotional & social intelligence has become a desired character of personal robots. According to MIT scholars Breazeal *et al.* [5] robots are expected to behave with proper social manners to fulfil the cognitive needs of interacting with humans. The target customers for emotional and social assistive robotic platforms range from pre-school children to independent living elders [6]. Typical products like *Nao* and *Pepper* by Aldebaran and Softbank have gained great success in the edutainment robot market[1]. Elderly users have high expectations from a robot as they tend to treat it like a human being [7], and it is evident that the user interface is key to fulfilling these social commitments.

B. User Interfaces

Currently user interfaces may include keyboard, touch screen, body posture, hand gesture, and speech as modes of interaction. Multi-modal interaction is more likely to be accepted by elderly robot users [8]. Older people usually prefer natural and straightforward ways of communication like speech, buttons and touch screen to other complicated ones when coping with machines like robots [9], [10], [11], or assistive equipment, like mobility aids [12]. The EU FP7 Project Robot-Era [2] [13] was a large-scaled integration project that aimed to implement robotic systems like service robots, smart homes and assistive living facilities, to offer older people safe home environments and quality daily life. Two individual but fully inter-changeable interfaces were specifically designed to suit the needs of older people: a speech user interface (SUI) and a graphic user interface (GUI) via a touch screen.

[1]https://www.softbankrobotics.com/emea/en/robots
[2]http://www.robot-era.eu/

Apart from verbal human language, robots are expected to undertake conversations with a number of people with non-verbal cues. For example, gaze cues have been studied as a way for robots to establish the roles of people involved in human-robot conversations [14]. Other studies have examined joint attention between human user and robot during spoken interaction [15] [16] [17]. In [18], an audio-visual social interaction corpus was collected and annotated to investigate behaviour patterns of robot users and their attention while interacting with the robot. More recently, the timing of interaction has been explored to design robots that can embody these behavioural resources. Yu *et al.* investigated the temporal patterns of human joint attentional process in a multi-modal human-robot interaction word-learning task [15]. The precise timing in the course of human-robot interaction has been studied to coordinate the head movement and utterance of the robot [19]. In some studies, human-robot communication is uni-modal, that is human voice as the sole interactive media. In other studies, investigations with HRI modalities have been carried out with young adults [15] [16] [20], which limit the generalizability of the findings to older people.

C. Research Highlights

In this study, we investigated the temporal patterns of older people interacting with service robot by means of speech and/or graphical interfaces in real-world scenarios. Users were free to choose the modalities they preferred under our experimental settings, their spontaneous social behaviours (including where they were looking, their conversation with the robot or others) in performing a specific task were captured and analyzed. The elderly users' time course of integrating the HRI modalities were observed. The purpose was to investigate how older people act when completing a task with a robot given two interfaces in real time, and how can they coordinate their "talking" (by speech) and "reading" (on touch screen) actions during the process. This is to reveal complementary effects between dual-modalities interfaces, which in turn will guide the development of future multi-modal interfaces. The highlights of this research work were:

- *Multi-modal interaction*: Integrating spoken and graphical interfaces to complete a single task. The participant manipulated the robot through speech and/or touch screen.
- *Real-time manipulation*: All interaction between participants and the robot took place in real time. Timestamps of the interactive data (utterance, eye direction and touch screen activity) were recorded.
- *Real-world conditions*: The experiment was conducted in real environment, no simulated conditions were employed.
- *Elderly-users engagement*: Fifteen older people (average age 80.5) participated in the experiment. They were 12 females and 3 males and all were British English speakers.

II. User Platform for Robot-Era

The domain-specific SUI was developed to meet older people's needs in HRI. As the Robot-Era robots were targeted at a European market, the SUI was supported in multilingual scenarios, including English, Italian, and Swedish. The Robot-Era SUI developed was based on the framework of a Nuance speech recogniser & parser[3], a Ravenclaw based dialogue manager provided by Olympus[4] , and a speech synthesizer by Acapela[5] . The speech recognition program performed grammar-based recognition to detect keywords and phrases that corresponded to user commands. The dialogue manager controlled the flow of the conversation with the robot. The Olympus dialogue manager was an open-source resource, supporting multi-modal infrastructure [21]. Speech recognition was conducted in the context of service-tailored grammars. For each language, the grammars were configured in a user-centred manner, and continued to be refined during the course of pilot study. To manage the balance between recognition rate and computational costs to a satisfactory level, in addition to developing context-aware grammars for reduced complexity [22], we managed to switch the dialogue flow among all available services without rebooting the module. It was noted that before each dialogue movement, a confirmation with the user was made to avoid wrong actions caused by speech recognition failures. The Olympus dialogue manager, employing advanced policy to handle errors from the Nuance recognizer, also contributed to a better dialogue flow. The SUI was activated by predefined "waking up" phrases. Grammars were loaded upon request dynamically. Users could use either phrases or sentences to navigate from one service to another within the SUI "menu". Details about the Robot-Era SUI can be found in [23].

The GUI was built with web technology in order to be accessible by all devices that have a connection with the system, from smartphones to tablets, from personal computers to laptops. The user could store the address of the home page and of the various services and, then, access them using the preferred device. The pages making up the interface included a main menu presenting all available services from the robot, and individual pages for respective services, a confirmation page to confirm the user option. Sounds were played for certain events, e.g., text notifications, interactions or warnings.

Both Robot-Era interfaces were fully employable for all services, namely:

- *communication*: via mobile phone or skype.
- *shopping*.
- *cleaning*.
- *food delivery*.
- *indoor escort*.
- *object manipulation*.
- *garbage collection*.
- *laundry*.
- *reminding*.
- *surveillance*.
- *mobility support*.

Under the umbrella of Robot-Era, the functionality, utility, acceptability and efficiency of the developed robotic services

[3]http://www.nuance.com/
[4]http://wiki.speech.cs.cmu.edu/olympus/index.php/Olympus/
[5]http://www.acapela-group.com/

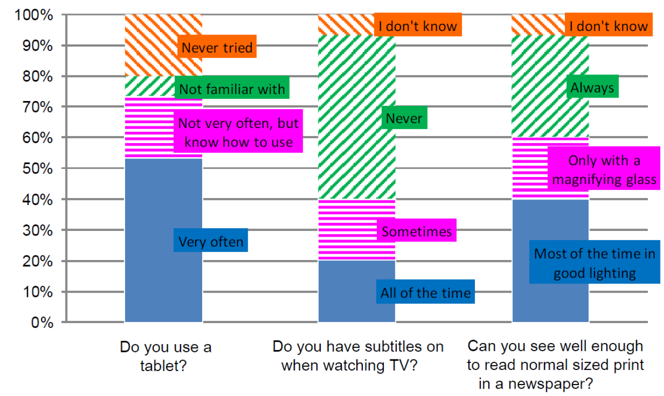

Fig. 1. Participant background demographic. Three questions were asked about their use of tablet, their eyesight and hearing condition.

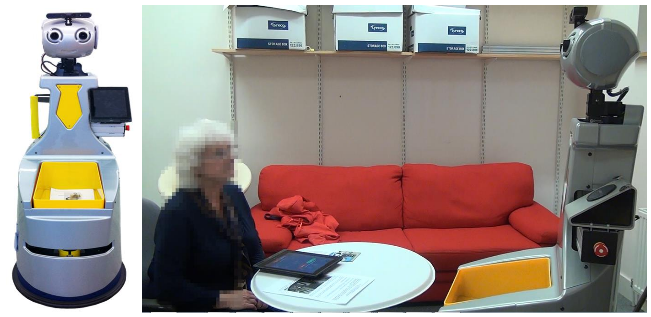

Fig. 2. The SCITOS G5 mobile robot platform (left) and a social interaction scenario captured between an older participant and the robot (right).

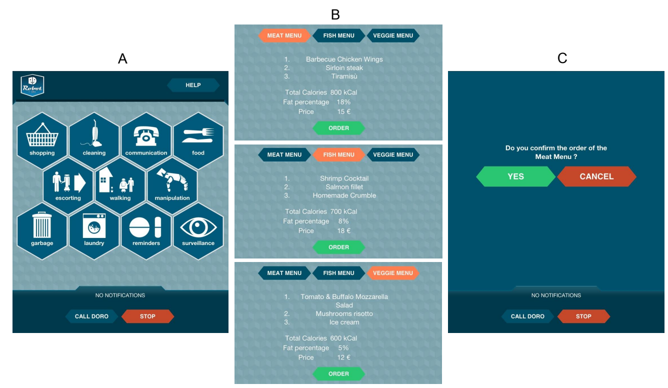

Fig. 3. The graphical user interface (GUI) of Robot-Era. A: All available service options including the food delivery; B: Food menus offered: meat menu, fish menu, and veggie menu; C: Confirmation before making the order.

and the intelligent living environments were assessed in real-world circumstances by the older people. A full evaluation of the Robot-Era services acceptance and functionality satisfaction can be found in [24], where it is shown that the Robot-Era system has the potential to be developed as a socially acceptable and believable provider of robotic services to facilitate older people to live independently in their homes.

III. Experiments and Data

A. Participants

The participants of this experiment were 15 residents (12 females and 3 males) from retirement apartment block in Plymouth, United Kingdom. They were aged from 70 to 89, with an average age of 80.5. They were all native British English speakers. Their computer use, hearing and eyesight status are shown in Figure 1.

B. Experimental Setup

The robot platform employed in Robot-Era was SCITOS G5 by MetraLabs, which was 4.5 feet tall wheeled mobile robot with a surface plate for carrying objects and a mount for a touch-screen tablet. Through the graphical and/or speech interface, an elderly person could control the robot to carry out a series of indoor and outdoor jobs like housekeeping, shopping and surveillance. In this experiment, the HRI task to undertake was food delivery. Participants were free to use either the GUI, SUI, or both, to complete the task. These two interfaces were fully synchronised and interchangeable, i.e., participants could choose one interface to undertake the first part of the task, then give it up and use the other one to continue. The collected data captured the spontaneous interaction that took place between the older people and the robot, i.e., conversation and eye direction with all time-stamps annotated. Figure 2 shows the SCITOS G5 robot platform and the image captured from the human-robot social interaction. Figure 3 demonstrates the GUI pages for food delivery service, one of the tailored Robot-Era services for older people. A detailed description about the SUI has been recently reported in [23].

C. Procedure and Task

Prior to an experiment, the experimenter met and greeted participants, answered any questions to ensure they were comfortable and willing to continue. The participant was then given the information sheet, completed the informed consent form and a short pre-experiment questionnaire to collect demographic information. The experimenter then gave a brief introduction about the robot and the food delivery task. Participants were asked to imagine that they had reduced mobility and to order the robot to have a meal delivered for them. Participants could choose to interact with the robot using speech and/or via the tablet. There was no significant physical contact with the robot and no risk to participants. The experimenter was present with participants at all times in case assistance was needed. There was also a technician present but hidden and separate from the participant, who supervised the robot software and hardware. As shown in Figure 2, in the experiment, the participant was seated in a simulated living room environment, with the Robot SCITOS G5 (Nickname: Johnny) standing opposite. The tablet interface was placed on the table in between the robot and the participant. Each test lasted for around 45 minutes.

To initiate the conversation, the participant said "Hello/Hi Robot." or "Hello/Hi Johnny." After the robot had woken up, it asked a few questions to the participant about the service they were interested in, and the follow-up details about their

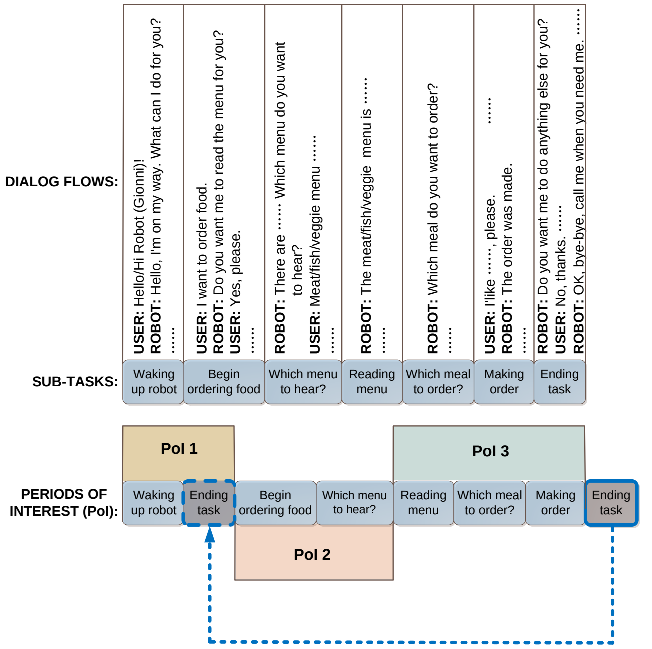

Fig. 4. The three Periods of Interest (PoIs) defined in the food delivery task.

requests. These questions were simple and easy to understand. Meanwhile, the participant could get the same information from the tablet pages. If no answer from the participant was heard for a certain time, or the answer was irrelevant, the robot repeated the question one more time. An instruction sheet was placed on the table for a quick help about what to say if needed. Figure 4 elaborates the overall task flow, the sub-tasks defined in it, as well as the example dialogue excerpts.

D. Data Characteristics and Annotation

For each participant, we collected questionnaire user opinions, open-question interview data, and frontal & profile video streams. The HRI food delivery task included 7 sub-tasks: *Waking up (robot)*, *Begin ordering food*, *Which menu to hear? Reading menu*, *Which meal to order? Making order*, and *Ending the task*. The audio-visual data was segmented on an individual-specific basis, as the timing of segments varied from one participant to another. To facilitate data analysis, we further merged these data segments sequentially into longer sections, which we called the Periods of Interest (PoIs). Figure 4 illustrates how we defined a PoI in the food delivery task.

We annotated all PoI data with free software ANVIL, which specializes in annotating video clips [25]. It supports multi-layered annotation in transcribing a video stream, in a user-configured coding style. Figure 5 demonstrates the workspace of ANVIL in annotating a video file. Table I indicates the four-layered decoding scheme employed in this study. Four layers were identified – speech waveform, participant gaze direction, dialog state, and tablet activity. The participant gaze direction could be towards the *Robot*, the *Tablet*, or *Others* including the experimenter. The dialogue state refers to the stage of conversation currently going on between the participant and robot, for example, *Robot Speaking*, *User Speaking*, etc. Intermittent pauses in conversations were identified either as *Robot Waiting* (for participant's response) or *User Waiting* (for the robot to respond). The last track – tablet activity, indicates whether *Button pressed* or *Screen refresh* events were taking place. Any single mode transition leads to a new event in a concerned PoI. Using Figure 6 as an example, the robot was speaking and the participant was looking at the robot at the beginning, then suddenly, the participant turned to look at the tablet instead, this marks the ending of Event 1 and the beginning of Event 2. Likewise, in Event 4, the participant's gaze was towards the tablet when the robot was speaking. The subsequent tablet activity, ”screen refresh”, ends Event 4, and leads to a new Event 5. In this process, the timestamps of all decoded events, i.e., onset and offset timings, were recorded in the annotation files.

TABLE I
ANVIL DATA ANNOTATION SCHEME.

Track	Value set
Waveform	*N/A*
Gaze	*Robot*, *Tablet*, *Others* (including the experimenter)
Dialog state	*User speaking*, *User waiting* (for robot's response), *Robot speaking*, *Robot waiting* (for user's response)
Tablet activity	*Screen refresh*, *Button pressed*, *None*

From the video we found some instances when participants turned to look at the experimenter and talked with her. These usually happened when participants were waiting for the response from the robot, or were not certain what to say.

IV. Results

The purpose of this investigation was to explore real-world interactions when there are more than one communication channel. In this study, the two modalities were ”talking” and ”reading”. Previously, researchers have compared these two modalities only from questionnaire ratings [11]. In this study, we also assessed temporal interactive patterns. The robot and participant took turns in a conversation and participants might have eye contact with the robot as they liked, just as we do in our daily person-person communication. Apart from direction of the eye gaze, the time allocation on each direction was also taken as an interaction pattern. To summarize, the social attributes of the multi-modal HRI data were derived from the following perspectives:

- Speech *vs*. tablet: which one to choose and at which time.
- Time distribution:
  - Proportion of time the participant was looking at the robot, tablet, or other places (e.g., the experimenter).
  - Proportion of time the robot/participant was speaking or waiting for the other to respond.

A. Attentional behaviours – Where to look at?

Figure 7 shows the time distribution of different participants towards the robot, tablet and/or other places during the HRI task. Most participants spent most time (69.6%) looking at

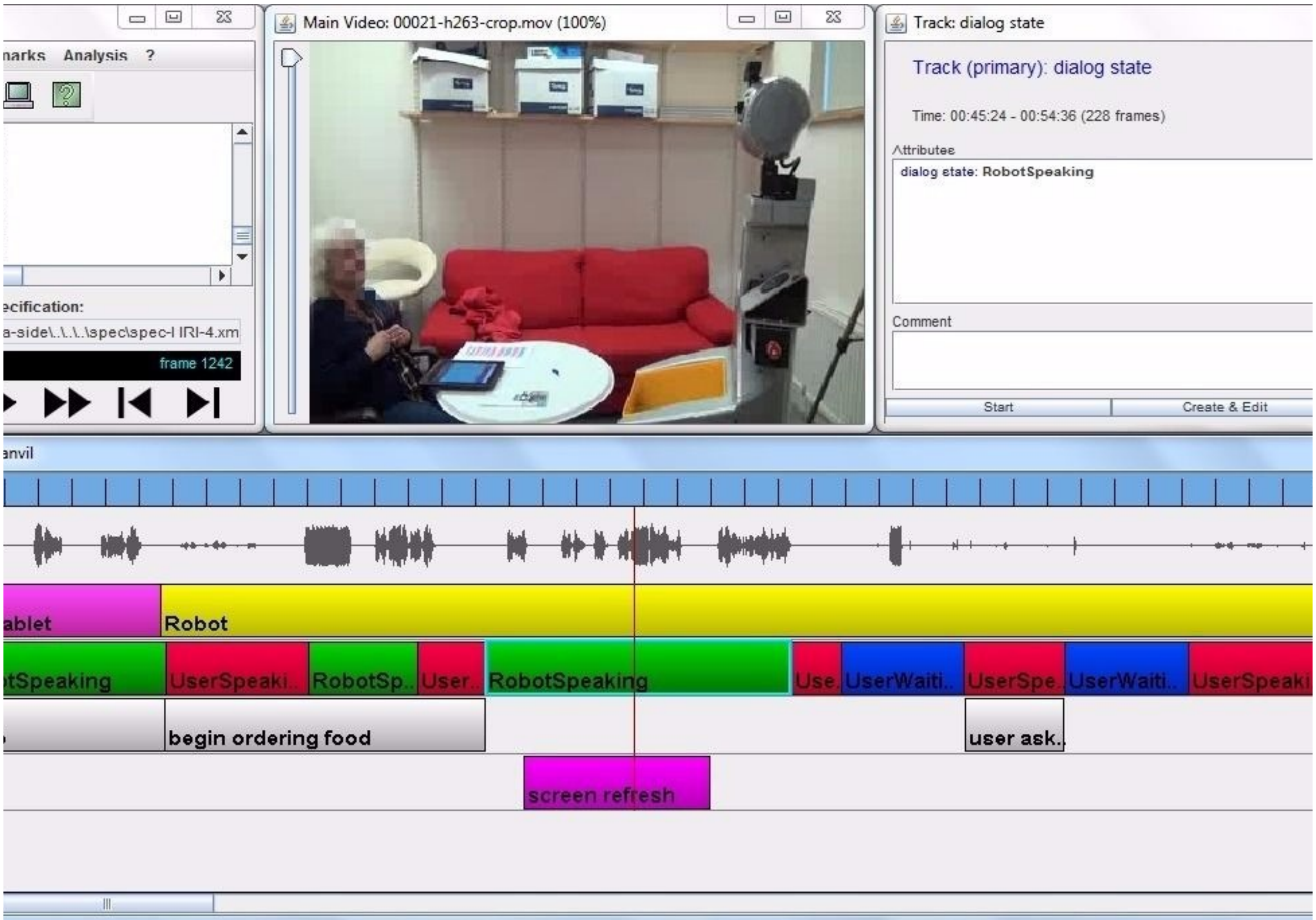

Fig. 5. ANVIL workspace of decoding a video file.

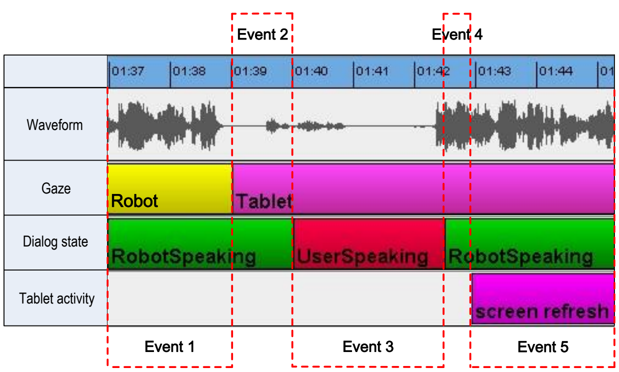

Fig. 6. Decoded events in an interactive period.

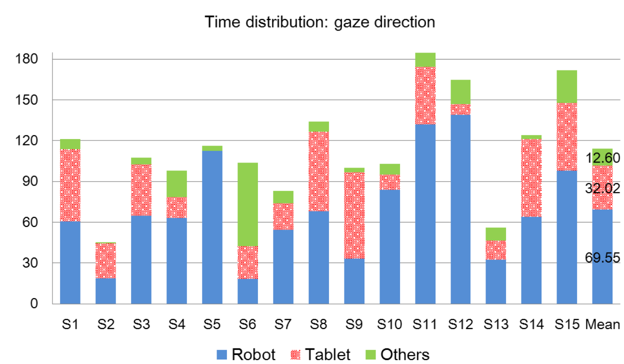

Fig. 7. Time distribution of participant gaze towards Robot, Tablet, or Others during the HRI task (in second).

the robot. Eye contact has long been identified as a kind of non-verbal social interaction between human beings [26], or even with an animal [27], so this finding is consistent with human-human interactions.

B. Conversational behaviours – Speak or not?

Figure 8 shows the time distribution of dialogue states in the conversations that occurred between the participant and the robot. The dialogue excerpts have been divided into four states – robot speaking, user speaking, robot waiting (for the

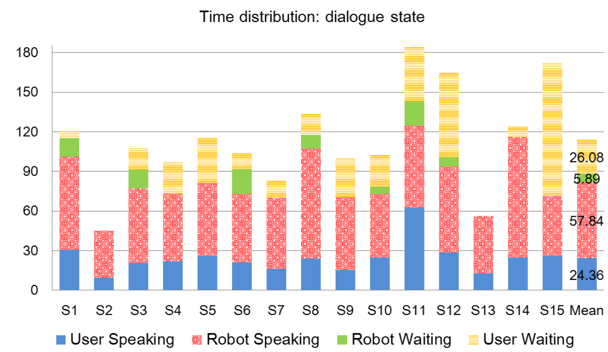

Fig. 8. Time distribution of the dialogue states in participant-robot conversations during the HRI task (in second).

participant to respond), and user waiting (for the robot to respond), respectively. On average, robot speech counted for more than half (57.8%) of the conversation time (Figure 8). The second longest portion is *User Waiting*, which is 26.1%, slightly higher than the portion of User Speaking at 24.4%. The robot only spent 5.9% of time on waiting for participants to respond, which reveals that in a real-world HRI scenario like this, human users usually take the robot as a real partner and attempt to respond on most occasions.

C. Preference towards dual-modalities: "Talking" or "reading"?

Participants had two options for HRI communication: they could read all necessary information from the touch screen and press buttons to send commands, or talk with the robot about their preferences. Either option was independent and interchangeable with the other, participants could switch from one interface to the other at any time without discontinuing the task. Reading messages from a touch screen may be easier for an older person with hearing problems, while for blind people or those with difficulty in reading, the speech interface offers an alternative.

We investigated how participants coordinated their "talking" and "reading" actions. Table II shows the time spent on gazing at the robot, tablet, or the experimenter by participants during a dialogue. Participants spent most of the time looking at the robot except during the *Robot waiting* state, when they tended to turn to the experimenter for assistance or clarification. Participants mostly looked at the robot (82.5%) rather than the tablet (14.5%) when they were speaking to the robot. However, participants' gaze time towards the robot dropped to 52.7% and increased towards the tablet (38.5%) when the robot was speaking. This was because most participants chose to read the food menu details from the tablet while listening to the robot. On the other hand, when the dialogue flow broke, there were two different situations. First, when a participant was sure that the robot was waiting (which only happened for 6 participants), he/she might feel puzzled, and so turned either to the experimenter for help (40.8%), or towards the tablet for clues (36.5%). Otherwise, when the users were expecting

TABLE II
TIME ALLOCATION OF HUMAN PARTICIPANT'S GAZE DIRECTION DURING THE HRI DIALOGUE.

User speaking	**Robot**	**82.52%** $\pm$ 17.25%
	Tablet	14.52% $\pm$ 14.15%
	Others	2.96% $\pm$ 7.51%
Robot speaking	**Robot**	**52.66%** $\pm$ 25.01%
	Tablet	38.45% $\pm$ 21.54%
	Others	8.89% $\pm$ 16.23%
Robot waiting	Robot	22.80% $\pm$ 22.13%
	Tablet	36.45% $\pm$ 20.58%
	Others	40.75% $\pm$ 37.54%
User waiting	**Robot**	**67.84%** $\pm$ 23.82%
	Tablet	13.77% $\pm$ 20.15%
	Others	18.39% $\pm$ 20.46%

responses from the robot, they still tended to look at the robot (67.8%).

Overall, our observations showed that the older people's eye contact to communicate with the robot was similar to human-human contacts.

D. PoIs

In this part, we further took the HRI task activities into account, to see whether they affect the social manners of participants. Based on the attributes of the food delivery task, we identified three PoIs. Figure 9 shows the details of eye direction and dialogue state transitions of the participants in our test, with regard to the three PoI slots.

1) PoI 1: Brief greeting: This part consisted of a brief conversation when waking the robot up at the beginning and ending the task at the end of the HRI job. There were only greeting and farewell conversations taking place here, which were quite short. Both parties took part in the dialogue. The average duration for participant and robot speaking times were 4.4 and 8.9 seconds, respectively. In terms of gaze direction, participants spent 15.5 seconds on greeting with their social partner – the robot, which accounted for 74.9% of the entire PoI time. Participants spent only 2.1 seconds gazing at the tablet.

2) PoI 2: Question & answer: This PoI covered the main activities that could happen when ordering a meal at a restaurant. There was continuous conversation between participant and robot. The average speaking time of the participant and robot were 12.8 and 20.0 seconds, respectively. For gaze direction, on average, each participant spent 31.1 seconds on eye contact with the robot, which was much longer than on the tablet (8.8 seconds).

3) PoI 3: Reading & listening: This PoI contained many speech excerpts from the robot, and few from the participant. This PoI was a reading & listening focused one, where the tablet displayed information that was fully synchronized with the robot's speech. In this PoI, participants spent similar amounts of time looking at the robot (22.9 seconds) and the tablet (21.2 seconds). Meanwhile, the robot talked more than participants (28.9 versus 7.2 seconds).

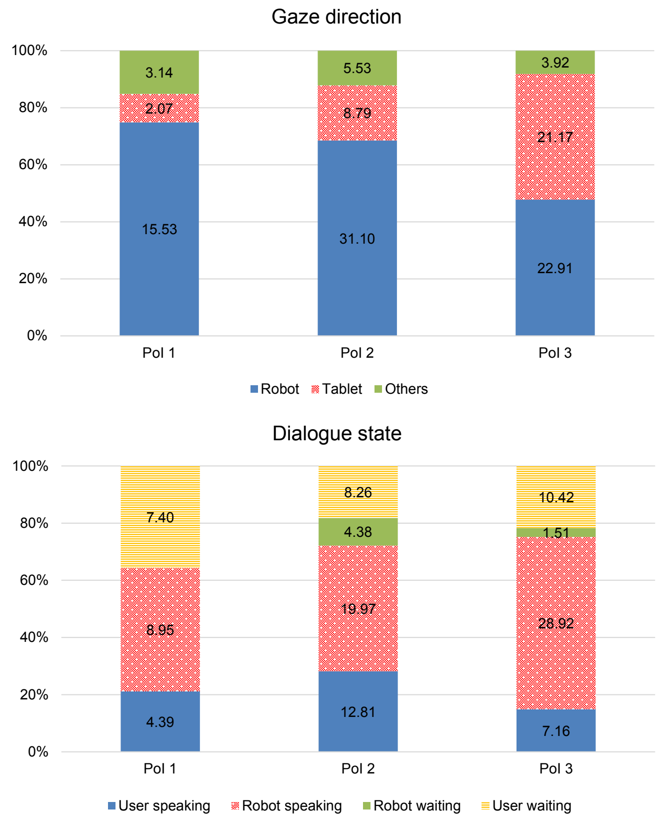

Fig. 9. The gaze direction and dialogue state transitions of participants in each of the period of interest (PoI) area. The data labels on the bars show the exact time spent (in seconds) on this category.

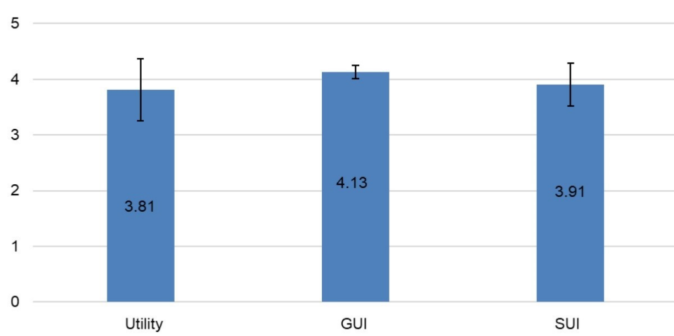

Fig. 10. Statistics of questionnaire ratings regarding system utility, GUI and SUI from all participants.

E. Questionnaire opinions

Questionnaire data was collected on three aspects: system utility, preference towards the SUI, and that for the GUI. There were 27 questions in the survey. Participants rated their attitudes towards each topic on five-point Likert scales (1 – Strongly disagree, 2 – Disagree, 3 – Neither agree nor disagree, 4 – Agree, 5 – Strongly agree). Some example questions are listed in Table III. Questions were not tagged with categories, and were mixed randomly when they were shown to participants.

1) Overall performance: Generally speaking, the user interface system (SUI + GUI) was acceptable for participants.

TABLE III
SOME QUESTIONNAIRE TOPICS IN THE SURVEY.

Utility	*- I thought the food delivery service was easy to use.* *- I felt confident about using the food delivery service.* *- I would use the robot for doing the food delivery service, in case of need.* *- I enjoyed using the robot for doing the food delivery.* *- I felt the robot understood what I wanted to do.*
SUI	*- I found it easy to speak to the robot to perform the food delivery service.* *- I understood what I could say to the robot to perform the food delivery service.* *- I thought I got sufficient information from the dialogue with the robot to finish the food delivery service.* *- I am satisfied with the conversation with the robot about the food delivery service.* *- I enjoyed using the speech to conduct the food delivery service.*
GUI	*- I found the tablet easy to use to perform the food delivery service.* *- I could clearly read the messages on the tablet.* *- I understood what buttons I needed to press to perform the food delivery service.* *- I thought the response of the tablet is quick enough for me to use the food delivery service.* *- I enjoyed using the tablet to conduct the food delivery service.*
SUI *vs.* GUI	*- I prefer to use speech rather than tablet for the food delivery service.* *- I prefer to use the tablet rather than speech for the food delivery service.* *- I mainly used speech for the food delivery service, only used the tablet when I felt needed.* *- I mainly used the tablet for the food delivery service, only spoke to the robot when I felt needed.* *- I'm confident that I can use only speech (no tablet) to complete the food delivery service.* *- I'm confident that I can use only the tablet (no speech) to complete the food delivery service.*

Figure 10 shows questionnaire opinions of participants on the general utility, GUI and SUI, respectively. The average score for the utility of the user interface as a whole is 3.8, while GUI and SUI scored higher as individuals.

2) GUI or SUI?: There was no preference between GUI and SUI; participants scored 3.8 ± 0.4 for SUI and 3.8 ± 0.2 for GUI. All participants, except two with vision problems, used both modalities for HRI. They enjoyed having two options for the service, and being able to switch from one to the other.

Before ending the experiment, participants were asked a few open questions such as what did you like/dislike about the system you used, and what are your suggestions to improve the system. In this part of the session, participants were free to discuss anything relevant about the robot and the experiment. Participants involved in this study were all very enthusiastic about the idea of having a robot at home, and were keen to support our research by participating in future studies.

V. Conclusions

The populations of all developed countries are ageing. Most older people prefer to live at home as long as possible and to deal with this challenge, service / caring robots are being developed to provide domestic and nursing services. These intelligent robots communicate, interact and collaborate with human beings through user interfaces. Besides utility and functionality, social intelligence is also expected from robots. To evaluate what is preferable in a social interface, we observed and studied older peoples' social behaviours in performing a food delivery task with a service robot, and the ways they manage to achieve this. We carried out an experiment with 15 older participants in the context of the EU FP7 Robot-Era Project. Two ways were provided to complete the task, by speech or by using a touch screen tablet. We analyzed the spontaneous actions of the participants, including their attentional activities (eye contacts) and conversational activities, questionnaires, and temporal characteristics (timestamps, duration of events, event transitions) of these social behaviours. This combination of data distinguishes it from most of other studies that have focused on questionnaire ratings only.

There are three main findings. First, the design of the Robot-Era interface was very acceptable for older users. Secondly, most older people used both SUI and GUI to perform the food delivery service, with no difference in their preferences towards either SUI or GUI. Thirdly, the older users had frequent and long-time eye contact with the robot during their conversations, and showed patience when expecting responses from the robot. They enjoyed the service. Overall, the level of social engagement with the robot demonstrated by elderly users was no different from what might be expected towards a human partner. This study was an early attempt to reveal the social connections between human beings and a personal robot in real life. Our observations and findings should inspire new insights in HRI research and eventually contribute to next-generation intelligent robot development.

VI. ACKNOWLEDGMENTS

This work was partially supported by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant No. 288899 and the National Natural Science Foundation of China under Grant No. 61803039. We thank our participants from Wesley Court Plymouth.

REFERENCES

[1] M. A. Blythe, A. F. Monk, and K. Doughty, "Socially dependable design: The challenge of ageing populations for hci," *Interacting with Computers*, vol. 17, no. 6, pp. 672–689, 2005.

[2] United Nations, "Population facts," October 2017.

[3] P. Mayer, C. Beck, and P. Panek, "Examples of multimodal user interfaces for socially assistive robots in ambient assisted environments," in *Proc. IEEE International Conference on Cognitive Infocommunications*, 2012, pp. 401–406.

[4] (2015). [Online]. Available: https://www.theguardian.com/technology/2015/feb/27/robear-bear-shaped-nursing-care-robot

[5] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 708—713.

[6] J. Linert and P. Kopacek, "Robots for education (edutainment)," in *IFAC-PapersOnLine*, Elsevier, Ed., vol. 49, no. 29, 2016, pp. 24–29.

[7] C. Smarr, A. Prakash, J. Beer, T. Mitzner, C. Kemp, and W. Rogers, "Older adults preferences for and acceptance of robot assistance for everyday living tasks," in *Proc. Human Factors and Ergonomics Society Annual Meeting*, 2012, pp. 153–157.

[8] D. Tang, B. Yusuf, J. Botzheim, N. Kubota, and C. Chan, "A novel multimodal communication framework using robot partner for aging population," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4540–4555, 2015.

[9] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects," *Personal and Ubiquitous Computing*, vol. 17, no. 1, pp. 127–144, 2013.

[10] M. Al-Razgan, H. Al-Khalifa, M. Al-Shahrani, and H. Al-Ajmi, *Touch-based mobile phone interface guidelines and design recommendations for elderly people: A survey of the literature*, ser. Lecture Notes in Computer Science. Berlin Heidelberg: Springer, 2012, vol. 7666, pp. 568–574.

[11] A. Di Nuovo, F. Broz, T. Belpaeme, A. Cangelosi, F. Cavallo, R. Esposito, and P. Dario, "A web based multi-modal interface for elderly users of the Robot-Era multi-robot services," in *Proc. IEEE International Conference on Systems, Man and Cybernetics*, 2014, pp. 2186–2191.

[12] J. Schneider, S. Irgenfried, W. Stork, and H. Wörn, "A multimodal human machine interface for a robotic mobility aid," in *Proc. IEEE International Conference on Automation, Robotics and Applications*, 2015, pp. 289–294.

[13] "Robot-Era project: Implementation and integration of advanced robotic systems and intelligent environments in real scenarios for the ageing population," FP7-ICT-Challenge 5: ICT for Health, Ageing Well, Inclusion and Governance. Grant agreement number 288899.

[14] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita, "Footing in human-robot conversations: How robots might shape participant roles using gaze cues," in *Proc. ACM/IEEE International Conference on Human Robot Interaction*, 2009, pp. 61–68.

[15] C. Yu, M. Scheutz, and P. Schermerhorn, "Investigating multimodal real-time patterns of joint attention in an HRI word learning task," in *Proc. ACM/IEEE International Conference on Human-Robot Interaction*, 2010, pp. 309–316.

[16] L.-P. Morency and T. Darrell, "From conversational tooltips to grounded discourse: Head post tracking in interactive dialog systems," in *ACM International Conference on Multimodal Interaction*, 2004, pp. 32–37.

[17] S. Sheikhi and J.-M. Odobez, "Combining dynamic head pose-gaze mapping with the robot conversational state for attention recognition in human-robot interactions," *Pattern Recognition Letters*, 2014.

[18] M. A. Sehili, F. Yang, and L. Devillers, "Attention detection in elderly people-robot spoken interaction," in *Proc. ICMI Workshop on Multi-modal, Multi-Party, Real-World Human-Robot Interaction*, 2014, pp. 7–12.

[19] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka, "Precision timing in human-robot interaction: Coordination of head movement and utterance," in *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 131–140.

[20] M. Staudte and M. Crocker, "Visual attention in spoken human-robot interaction," in *Proc. ACM/IEEE International Conference on Human-Robot Interaction*, 2009, pp. 77–84.

[21] D. Bohus and A. Rudnicky, "The ravenclaw dialog management framework: Architecture and systems," *Computer Speech and Language*, vol. 23, no. 3, pp. 332–361, 2009.

[22] O. Lemon, "Context-sensitive speech recognition in Information-State Update dialogue systems: results for the grammar switching approach," in *Proc. Eighth Workshop on the Semantics and Pragmatics of Dialogue*, 2004, pp. 49–55.

[23] N. Wang, F. Broz, A. Di Nuovo, T. Belpaeme, and A. Cangelosi, *Recent Advances in Nonlinear Speech Processing*, ser. 2190–3018. Springer International Publishing, 2016, vol. 48, ch. A user-centric design of service robots speech interface for the elderly, pp. 275–283.

[24] F. Cavallo, R. Esposito, R. Limosani, A. Manzi, R. Bevilacqua, E. Felici, A. D. Nuovo, A. Cangelosi, F. Lattanzio, and P. Dario, “Robotic services acceptance in smart environments with older adults: User satisfaction and acceptability study,” *J Med Internet Res*, vol. 20, no. 9, 2018.

[25] M. Kipp, *Handbook of Corpus Phonology*. Oxford University Press, 2014, ch. ANVIL: A Universal Video Research Tool, pp. 420–436.

[26] M. L. Knapp, J. A. Hall, and T. G. Horgan, *Nonverbal communication in human interaction*, 8th ed. Boston, USA: Wadsworth, 2013.

[27] B. Hare and M. Tomasello, “Human-like social skills in dogs?” *Trends in Cognitive Sciences*, vol. 9, no. 9, pp. 439–444, 2005.