Intuitive ways to instruct a social assistive robot: a study of natural interaction with elderly people

Shanee Honig, Vardit Sarne-Fleischmann, Yael Edan and Tal Oron-Gilad

Abstract— **Robots designed to support elderly people need to be able to recognize user intentions and operate accordingly. However, little is known about the intuitiveness of commands used to help older adults control a social assistive robot's movement. We examined which voice commands and/or hand gestures elderly people select to guide the motions of a robot. Twelve elders aged 69-78 participated in this 2-part study. The first part of the experiment implicitly elicited user commands in a simulated grocery store environment. In the second part, participants were explicitly asked to guide the robot to perform different navigational commands. The data was analyzed to consider age-related differences relative to previous experiments with younger adults, and differences between implicitly and explicitly selected communication forms. Results indicated that elderly people vary more in how they choose to instruct the robot than younger individuals. Implicitly used communication means were significantly different than explicitly selected ones, suggesting that elderly people may be unable to predict the hand gestures and voice commands that they are likely to use with a robot. Emerging patterns in how elderly participants guide a mobile social assistive robot could act as a baseline for future development of a natural, user-friendly aural and gesture vocabulary that is suitable for older users.**

I. INTRODUCTION

Aging populations have significantly enhanced interest in developing social assistive robots that help older adults maintain their independence [1]. For such robotic assistants to be effective, elderly people must be able to guide the robot in intuitive and comfortable ways [2]. Most robots that automatically detect and recognize natural gestures and speech are still based on predefined stimulus-response or stimulus-state-response systems [3], consequently, the success of robotic assistants depend in part on our ability to identify stimulus-response pairs that meet older users' expectations. Furthermore, to improve user experience and the robot's ability to decipher user input, it is important to better understand how elderly users would attempt to instruct robots in such situations. Yet, little is known about how they communicate with mobile robots and how their selected communication methods differ, if at all, from those of younger individuals and across different interaction scenarios.

This work was supported in part by the Ministry of Science, Technology and Space, Israel, (Follow me) under Grant 3-12060, in part by the Helmsley Charitable Trust through the Agricultural, Biological, and Cognitive Robotics Center, the Marcus Endowment fund through Ben-Gurion University of the Negev, and in part by the Rabbi W. Gunther Plaut Chair in Manufacturing Engineering through Ben-Gurion University of the Negev. The work of S. Honig is supported by Ben-Gurion University of the Negev through the High-tech, Bio-tech and Chemo-tech Scholarship.

All authors are from the Department of Industrial Engineering and Management, Ben-Gurion University of the Negev Beersheba 84105, Israel. Vardit Sarne-Fleishman is also with Afeka Tel Aviv Academic College of Engineering (e-mail: shaneeh@post.bgu.ac.il, sarne@post.bgu.ac.il, yael@bgu.ac.il, orontal@bgu.ac.il).

The present study seeks to identify which voice commands and/or hand gestures elderly people select to guide the movements of a social assistive mobile robot and evaluate factors that might impact their choices. The use of natural forms of communication such as speech and hand gestures for robot motion control is well established [4]. There is also a precedent of using user-defined interaction techniques to develop intuitive communication vocabularies for robots [5], [6]. Prior work implemented user-centered design methodologies to investigate natural navigation commands for drones [4], [7], however the preferred method of providing spatial instructions to ground robots may differ.

To the best of our knowledge, how elderly people prefer to guide the motion of ground robots has not yet been evaluated using formal user studies, even though guiding a robot to a new location is a commonly needed skill [8]. A first attempt at evaluating how people naturally guide the motion of a ground robot is described in our previous work [9]. Fourteen participants were asked to demonstrate different navigational commands while interacting with a robot. Results showed that certain commands (Stop, Follow me, Slow down, or Increase speed) had higher agreement scores and were more frequently communicated as voice commands. Commands that were more "direction-oriented" (e.g., Move right or left) received higher agreement being communicated via gestures. Some participants were observed to divide their instructions into sub-commands, or adjust their behavior in order to communicate their intentions to the robot.

The current study builds and expands upon [9] in two primary ways. First, [9] evaluated only explicit communication strategies of the participants (i.e. commands that participants consciously selected), whereas the current work investigates also implicit strategies used by participants during a natural human-robot interaction scenario (i.e. commands that participants use in practice without forethought). Adding the implicit condition provides insight as to people's ability to accurately predict and select in advance hand gestures and voice commands to be used while interacting with a robot. Second, whereas [9] evaluated young participants aged 23-28, the current study was performed on an elderly population (69 years and up). It is unclear from the literature whether elderly participants will instruct the robot differently than their younger counterparts. While age-related changes in

Fig. 1: A participant interacting with the robot.

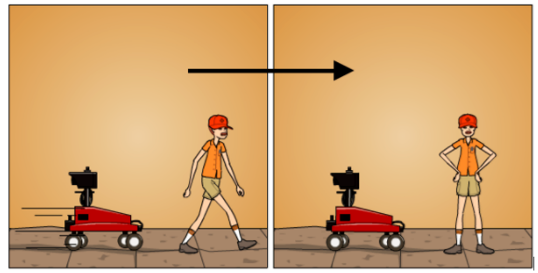

Fig. 2: A card demonstrating the Stop command.

sensory, cognitive and motor abilities have been shown to decrease the speed in which elderly can perform physical gestures in human-computer interactions [10], we are unaware of any work that evaluated how these changes influence choice of communication modality (voice, gesture or both) or mode (discrete, continuous, breaking command into sub-commands, behavioral) in human-robot interactions. Age-related differences and similarities are analyzed and presented for the explicit communication strategies evaluated in both studies. Emerging patterns in how elderly participants guide a mobile robot could act as a baseline for future development of a natural, user-friendly aural and gesture vocabulary that is suitable for older users

II. METHODS

A. Overview

In order to evaluate how elderly people choose to guide the motion of a robot, a two-part exploratory study was conducted. Twelve participants aged 69-78 (Avg=72.4, STD=4.4) participated in the study in exchange for monetary compensation. Five target navigational commands were evaluated (stop, follow me, move right, move left, move forward). The first part of the experiment elicited the use of each command implicitly using a simulated grocery store environment. Participants were told that the robot understood voice commands and gestures, and were asked to guide the robot to various predetermined locations. The second part of the experiment assessed the navigational commands explicitly. Participants were given seven cards, each describing a different command which they were required to direct the robot to perform. The Wizard of Oz (WoZ) technique was used to enable the robot to react to the commands that were communicated in both parts of the study. Completion of both parts of the study took about an hour.

B. Apparatus

The robot which participants interacted with throughout the study was a Pioneer LX robot, equipped with an integrated on-board computer with a 1.8 GHz Dual Core processor, and 2GB DDR3 RAM. The base of the robot is 50 cm wide, 70 cm long and 45 cm high. The robot has two drive wheels and four casters, and a maximum speed of 1.8 m/sec. A Microsoft Kinect V2 was mounted on a pan mechanism and connected by a 90 cm rod to the Pioneer LX in order to increase the robot's height and presence. A plastic basket was connected to the base of the robot using Velcro to enable the robot to carry small items (Fig. 1). Since we used the WoZ technique, the robot's autonomous navigation system was disabled and it was controlled by an experimenter via wireless joystick.

C. Procedure and Measurements

Participants were given a verbal and written description of the study and were instructed on how to use the tablet. They were then asked to fill informed consent forms and a demographics questionnaire. The robot was introduced to the participants as their personal shopping assistant who is capable of understanding voice commands and hand gestures. Participants were told to instruct the robot from their own point of reference (i.e. 'turn right' meant telling the robot to turn towards the participant's right, not the robot's right).

In the simulated grocery store task, they were required to guide the robot through three grocery-collecting stations, represented by a table and a marking on the floor beside it. The participants were instructed to guide the robot to the marking beside each station, and once the robot stood in place, take the grocery item from the table and place it in the robot's basket. Only once the item was in the basket could the participant lead the robot to the next station. The floor markings acted as a way to ensure that all the target navigational commands would be used: without the participant's knowledge, the experimenter controlling the robot would purposefully guide the robot to different positions relative to each marking: in the first station, the robot was taken to the right of the marking, to ensure participants had to guide the robot to the left; in the second station, the robot was taken to the left of the marking, to ensure the participants had to guide the robot to the right; in the third station the robot was stationed behind the marking to ensure participants had to guide the robot forward. In order to ensure participants used a 'Follow Me' command, the robot did not follow the participants to the first station unless they performed a deliberate action to make it walk after them. Participants completed post-session questionnaires after the third station.

In the second part of the study, participants received 7 pictorial cards, each including a command which the robot had to perform. The commands were identical to those elicited in the supermarket task. To prevent potential bias, for each command an initial and a final state were presented visually (Fig. 2). Participants were asked to instruct the robot to move from the initial state to the final state as depicted in the card. The order of the commands presented in the cards was counterbalanced between participants. Participants completed post-session questionnaires after each card.

All questionnaires were administered digitally using a tablet. The post session questionnaires used 5-point Likert scales with 5 representing ”Strongly agree” and 1 representing ”Strongly disagree”. At the end of each part of the study, a final questionnaire was given which used 5-point bipolar Likert scale with different ratings according to each question. Both parts of the experiment were filmed from two different angles for post-experiment analysis.

Two researchers were present during trials. One acted as a Wizard, controlling the robot's movements with an Xbox controller according to the participant's instructions. The wizard sat behind a table which blocked participants' view of the controller and took notes between sessions. Participants were told that this researcher was an observer and did not suspect that the robot was remotely controlled. The second researcher was responsible for providing participant instructions; collecting, preparing, and organizing the questionnaires administered.

D. Analysis

As in [9], each command was categorized on a two-dimensional axis: modality (voice, gesture or a combination of both) and mode (discrete, continuous, breaking the command into sub-commands, behavioral). Behavioral mode involved participants changing their own behavior in order to get the robot to do something (e.g. stop walking in order to indicate for the robot to stop). For voice, commands that used the same terminology were considered similar. For gestures, similarity was determined according to the movement on the XYZ axes. An agreement score based on [4], [11] was calculated for each command, for each of the modalities, to enable comparison (1). $\mathrm{P_r}$ represents the set of observed commands for each referent, and $\mathrm{P_i}$ represents a subset of similar commands observed for that referent. All categorizations were made by a single researcher. For more information, see [9].

$$A_r = \sum_{P_i} \left(\frac{P_i}{P_r} \right)^2 \tag{1}$$

III. RESULTS AND DISCUSSION

Table I presents the agreement scores calculated for each command in the grocery store and card tasks. The sum of the occurrences for each row may exceed the total number of participants (12) due to the use of mixed modalities.

Overall, agreement scores were low, except for the 'Follow me' voice command in the grocery store task (0.58 for 12 occurrences) and the 'Move right' gesture command in the cards tasks (0.72 for 6 occurrences). In contrast, in our former study on younger participants, high agreement scores were observed for four of the target commands. These results suggest a greater variance in how elderly people attempt to communicate with the robot as opposed to younger individuals. This variance may be a byproduct of the greater variance in physical and cognitive abilities among older adults or perhaps due to higher variance in mental models.

When comparing between the grocery store and card tasks, some commands had significantly different agreement scores. This may indicate that users have not formulated a consistent approach toward interacting with robots, possible due to the novelty of the technology. Furthermore, it appears that participants used gestures more frequently in the grocery store task than in the cards task. This may derive from the fact that most gestures in day-to-day communication are subconscious rather than purposefully selected [12]. In addition, most of the participants used a combination of voice and gestures when asked to perform direction-oriented commands (Move left / Move right) as opposed to more abstract commands (Follow me / Stop / Move forward). This is in line with findings among younger people [9].

The distributions of modality and mode according to commands and tasks are presented in Fig. 3 and Fig. 4, respectively. Participants mostly used voice commands or a combination of voice commands and gestures. Gestures alone were observed in only 3% of all cases in the cards task, and in 17% of all cases in the grocery store task. With regards to mode, differences between the two tasks were observed: while in the grocery store task the discrete and continuous modes were more common (57% and 35% respectively), in the card tasks using sub-commands was more dominant (55%), followed by discrete actions (38%). In the card tasks, continuous actions were observed only once (for the Stop command). This may be the result of different user goals: in the grocery store task, their goal was to get the robot to marked locations, pick up items and place them in the robot's basket. In contrast, the cards task geared participants to focus on specific robot actions and may have lead participants to overthink, resulting in a greater level of precision. These differences may indicate the general difficulty in implementing a pre-selected command vocabulary for elderly users.

Post sessions questionnaires indicated that in the grocery store task, 96% of the participants reported that the way they

TABLE I: Agreement scores calculated for each command in both parts of the study. The number of occurrences that were observed for each of the modalities appears in parentheses. Agreement scores higher than 0.5 with at least 6 occurrences are marked in bold.

	Grocery Store Task		Card Task	
Command	**Voice**	**Gesture**	**Voice**	**Gesture**
Stop	0.36 (10)	0.5 (6)	0.35 (7)	0.28 (5)
Follow me	**0.58 (12)**	0.2 (5)	0.39 (12)	N/A (1)
Move right	0.12 (11)	0.18 (9)	0.47 (12)	**0.72 (6)**
Move left	0.17 (12)	0.22 (8)	0.31 (12)	0.28 (8)
Move forward	0.22 (12)	0.44 (5)	0.33 (12)	0.56 (3)

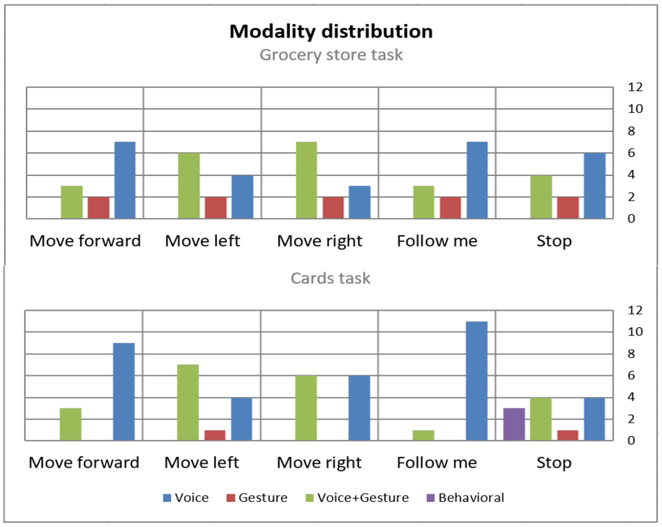

Fig. 3: Modality distribution (number of times observed) for the Grocery store task (top) and the cards task (bottom).

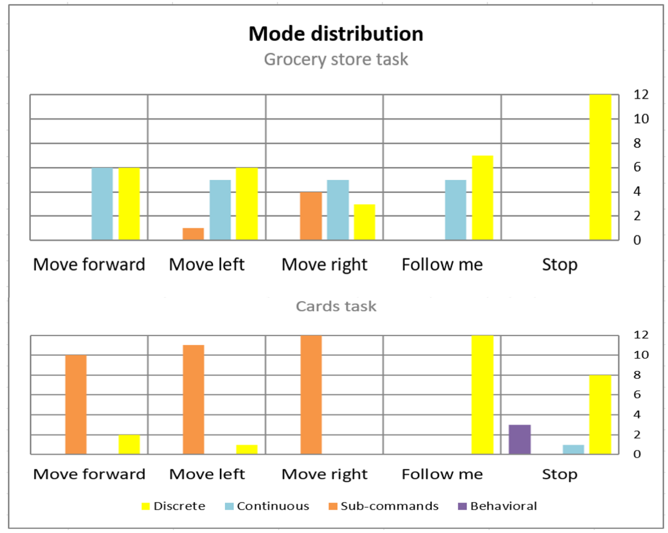

Fig. 4: Mode distribution (number of ties observed) for the grocery store task (top) and the card tasks (bottom).

chose to guide the robot was easy and did not require a lot of concentration. All participants reported that communicating with the robot was not physically tiring. In the card tasks, most participants said the way they chose to instruct the robot was intuitive (83%), not mentally or physically demanding (93% and 100% respectively), nor complicated (100%).

This study has several limitations. First, using visual representations of the target commands in the cards task led some participants to imitate the drawn character. Future experiments should find a better way to clarify how the cards should be used. Secondly, though participants were instructed to guide the robot according to their reference point, some participants still used the robot as the focal point, leading to inverse navigational commands. The data was matched during analysis to the correct command, however the point-of-reference may have influenced the communication methods used. Finally, a small sample size and the homogeneity of participants limit the generalizability of results. Future studies should include a larger number of elderly participants of different ages, with different backgrounds and different levels of technological competency.

In conclusion, it appears that elderly people varied more in how they chose to instruct a socially assistive robot than younger individuals. This might be in part the result of the greater variance in their physical abilities, but may also be ascribed to the limited presence of robots and technology in general in the daily life of this population. In addition, implicitly used communication means were significantly different than explicitly selected ones, suggesting that elderly people may be unable to predict the hand gestures and voice commands that they are likely to use with a robot. These behavioral patterns present interesting challenges that will need to be considered in order to create an aural and gesture vocabulary that enables older users to comfortably guide the motions of a socially assistive robot.

References

[1] J. M. Beer and L. Takayama, "Mobile remote presence systems for older adults: acceptance, benefits, and concerns," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on*. IEEE, 2011, pp. 19–26.

[2] W. Chen, "Gesture-based applications for elderly people," in *International Conference on Human-Computer Interaction*. Springer, 2013, pp. 186–195.

[3] N. Mavridis, "A review of verbal and non-verbal human–robot interactive communication," *Robotics and Autonomous Systems*, vol. 63, pp. 22–35, 2015.

[4] J. R. Cauchard, K. Y. Zhai, J. A. Landay *et al.*, "Drone & me: an exploration into natural human-drone interaction," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 2015, pp. 361–365.

[5] N. Otero, C. L. Nehaniv, D. S. Syrdal, and K. Dautenhahn, "Naturally occurring gestures in a human–robot teaching scenario," *Interaction Studies*, vol. 9, no. 3, pp. 519–550, 2008.

[6] M. Wongphati, Y. Kanai, H. Osawa, and M. Imai, "Give me a handhow users ask a robotic arm for help with gestures," in *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2012 IEEE International Conference on*. IEEE, 2012, pp. 64–68.

[7] E. Peshkova and M. Hitz, "Exploring user-defined gestures to control a group of four uavs," in *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*. IEEE, 2017, pp. 169–174.

[8] H. Zender, P. Jensfelt, and G.-J. M. Kruijff, "Human-and situation-aware people following," in *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*. IEEE, 2007, pp. 1131–1136.

[9] V. Sarne-Fleischmann, S. Honig, T. Oron-Gilad, and Y. Edan, "Multi-modal communication for guiding a person following robot," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2017, pp. 1018–1023.

[10] C. Stößel, H. Wandke, and L. Blessing, "Gestural interfaces for elderly users: help or hindrance?" in *International Gesture Workshop*. Springer, 2009, pp. 269–280.

[11] J. O. Wobbrock, H. H. Aung, B. Rothrock, and B. A. Myers, "Maximizing the guessability of symbolic input," in *CHI'05 extended abstracts on Human Factors in Computing Systems*. ACM, 2005, pp. 1869–1872.

[12] D. McNeill, "Where do most spontaneous representational gestures actually occur with respect to speech?" *Language and gesture*, vol. 2, pp. 186–198, 2000.