



Field Testing of Ro-Tri, a Robot-Mediated Triadic Interaction for Older Adults

Jing Fan¹ · Akshith Ullal¹ · Linda Beuscher⁵ · Lorraine C. Mion² · Paul Newhouse^{3,4} · Nilanjan Sarkar⁶

Accepted: 3 February 2021 / Published online: 21 February 2021

© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract

Older adults residing in long term care (LTC) settings commonly experience apathy, a neuropsychiatric condition with adverse consequences of increased morbidity and mortality. Activities that combine social, physical and cognitive stimuli are most effective in engaging older adults with apathy but are time consuming and require significant staff resources. We present the results from an initial pilot field study of our socially assistive robotic (SAR) system, Ro-Tri, capable of multi-modal interventions to foster social interaction between pairs of older adults. Seven paired participants attended two sessions a week for three weeks. Sessions consisted of robot-mediated triadic interactions with three types of activities repeated once over the 3 weeks. Ro-Tri gathered quantitative interaction data, head pose, vocal sound, and physiological signals to automatically evaluate older adults' activity and social engagement. Ro-Tri functioned smoothly without any technical issues. Older adults had > 90% attendance and 100% completion rate and remained engaged with the system throughout the study duration. Participants' visual attention toward the SAR system and their partners increased 7.2% and 4.7%, respectively, with their interaction effort showing an increase of 2.9%. Older adults and LTC staff had positive perceptions with the system. These initial results demonstrate Ro-Tri's ability to engage older adults, encourage social human-to-human interaction, and assess the changes using quantitative metrics. Future studies will determine SAR's impact on apathy in LTC older adults.

Keywords Socially assistive robotics · Multi-user human–robot interaction · Virtual reality · Field study · Older adults · Cognitive impairment · Long term care · Assisted living

1 Introduction

In the US older adults are projected to represent nearly 21 percent of the total population by 2030 [1]. With aging, older adults experience chronic health conditions, functional limitations, and dementia [1–3]. The prevalence of dementia, including Alzheimer's disease and other related disorders,

increases with age, from 16% of those 65–74 years to 36% of those 85 and older [2]. Dementia impacts communication, interaction ability, judgement, and memory as well as ability to perform basic activities of daily living. These deficits often result in the need for long-term care (LTC), either in assisted living or in nursing home settings [2, 4].

✉ Akshith Ullal
akshith.ullal@vanderbilt.edu

Jing Fan
jing.fan@vanderbilt.edu

Linda Beuscher
linda.m.beuscher@vanderbilt.edu

Lorraine C. Mion
mion.3@osu.edu

Paul Newhouse
paul.newhouse@vanderbilt.edu

Nilanjan Sarkar
nilanjan.sarkar@vanderbilt.edu

¹ Department of EECS, Vanderbilt University, Nashville, TN, USA

² College of Nursing, The Ohio State University, Columbus, OH, USA

³ Center for Cognitive Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

⁴ Geriatric Research, Education and Clinical Center, Tennessee Valley Veterans Affairs Medical Center, Nashville, TN, USA

⁵ School of Nursing, Vanderbilt University, Nashville, TN, USA

⁶ Department of Mechanical Engineering, Vanderbilt University, Nashville, TN, USA

Up to 72% of persons with Alzheimer Disease and related dementias suffer from apathy; it is associated with further cognitive decline, functional deficits, reduced quality of life, social isolation, and increased mortality [5–10]. Moreover, it produces increased stress, burden and frustration for formal and informal caregivers [11–13]. Since few pharmacologic options exist [14], a major strategy is to foster older adults' engagement in activities [14, 15]. Indeed, the US Centers for Medicare & Medicaid mandates LTC settings include activities as part of the individual's plan of care.

Although neither pharmacologic or non-pharmacologic therapies can treat dementia or slow or stop the progression at present, reviews and meta-analyses indicate that cognitive intervention, exercise and physical activity intervention are beneficial to people with Alzheimer's disease and have positive effects on cognitive function [13–15]. The treatment of apathy in these individuals also remains primarily non-pharmacological. Multimodal strategies that combine social, cognitive and physical domains, tailored to the individual, appear most successful; those activities that highlight social engagement in the multimodal intervention appear most effective [16–18]. Unfortunately, delivering these activities requires significant personnel resources. Many LTC facilities have inadequate staffing, either in labor quantity and/or skill mix [19]. As a result, leaders in health care and nursing have called for the use of advanced technology, including robotics, as a strategy to deal with nursing shortages [20, 21].

SAR systems have been developed to provide interventions focused mainly on either physical or cognitive domains. Recently, a few studies, mainly tested in the laboratory, have used SARs to administer physical and cognitive based interventions. [22–24]. We added a social component to these interventions and tested our SAR system in the laboratory setting; the SAR system provided multimodal intervention and received positive feedback from participants. [25, 26] Although these multimodal SAR systems show promising results in laboratory settings, there is a need to determine how older adults would react to these types of interventions in LTC settings and over multiple exposures. The real-world environment is more unstructured compared to a laboratory setting, where the researchers have control over many important factors that impact the smooth functioning of the SAR. For example, lighting, noise level, and room setup are less controllable in the field. In addition, a laboratory-based study cannot accurately gauge participants' enthusiasm and engagement as participants who already have spent time travelling to the laboratory are less likely to drop out. The laboratory studies [22–26] often included relatively healthy older adults. However, in LTC settings, participants will have multiple chronic conditions and commonly need ambulatory (such as wheelchairs) or medical devices (such as oxygen supplement). These factors make SAR-based interventions in LTC settings more challenging.

Before evaluating the efficacy of multimodal SAR interventions on apathy among older adults with dementia, we needed to determine whether the SAR system would operate successfully in the natural setting, whether older adults would continue to use the system past the first one or two sessions, whether they would enjoy interacting with the robotic system, and whether they would enjoy working with their partners over time. These issues are critical to ensure successful and sustainable robotic intervention with older adults in LTC settings. The main objective of the current study was to explore the feasibility, acceptability, and effect of the Ro-Tri SAR system on engagement and social interaction of older adults residing in LTC settings and whether Ro-Tri SAR was capable of delivering multimodal stimuli involving physical, cognitive and social stimuli to older adults in LTC settings over a longer duration.

Over a 3-month period, we recruited 7 pairs of older adults at two local LTC facilities to participate in a 3 week pilot field study to examine (1) feasibility and acceptance, (2) whether older adults remain engaged over time with SAR-based activities; (3) the robot's ability to encourage communication and social engagement between two older adults; and (4) the feasibility of gathering quantitative data on older adults' task performance, physiological indicators of stress, and engagement in activities. The primary contributions of this work are: (1) demonstration of the capability of the Ro-Tri SAR system in LTC settings where a pair of older adults participated in many-to-one human–robot interaction (HRI) over 6 sessions; and (2) data analysis of the results to show how the HRI influenced human–human interaction (HHI) between the paired participants based on Cohen-Mansfield's engagement model for group activities [27]. Although preliminary, findings from this study can provide insight for design, implementation and testing of SAR system interventions that are feasible and effective in LTC systems. The remainder of the paper is organized as follows. Section II presents literature relevant to our work. Section III describes the model of engagement used in our work, rationale for the data collection and robot behaviors, and the protocol of the field study. Finally, in Section IV and V, we present the results of the field study and discuss the implications and future directions of this research, respectively.

2 Literature survey

We summarize the published literature in two broad areas that are relevant for our current work – the state of HRI focusing on older adults and the deployment of robotic technologies in LTC facilities. Therapeutic robotic systems can be broadly categorized as animal robots to provide companionship [28], telepresence robots to facilitate social connections with families and caregivers [29], and socially

assistive robotic (SAR) systems to provide activity-oriented therapies, such as physical exercise and memory games [30]. SAR systems, including animal robots, are designed specifically for social interactions with the intention of detecting and meaningfully responding to older adults' attention and behavior, and thus have significant potential for addressing physical, cognitive, and social conditions. However, early studies of SAR either used the Wizard of Oz (WoZ) experimental paradigm [31, 32] that required a human operator to control the robot or used open-loop robotic platforms [33–37] with pre-programmed robotic behaviors. WoZ design places interaction burden on a human operator whereas open-loop robotic platforms are limited in their capacity for HRI and lack real-time dynamic adaptation based on interaction. More advanced closed-loop robotic systems allow the robot to dynamically alter its interaction based on real-time human interaction. A number of closed-loop SAR platforms have been designed in recent years to engage older adults in eating [30, 38], cognitive stimulation [30, 38–40], and chair exercises [41]. Commercially available robots NAO, Robo-Philo, and Manoi-PF01 have been programmed to instruct older adults and correct their gestures during physical exercise routines [41–45]. The majority of these closed-loop SAR systems were developed for one-to-one HRI to engage older adults in physical and cognitive activities.

Group activities are essential to promote an enriched social environment with opportunities for older adults to have social contact and reduce the risk of social isolation [16, 18, 46]. By adding other human participants in the SAR-based activities, many-to-one (i.e., multiple humans interacting with one robot) interaction will likely lead to human-to-human interaction among the HRI participants. Cruz et al. found that activities within a group setting engaged older adults not only in the stimuli provided by the activity, but also by the simultaneous social contact [46]. In [47], a social robot was used to provide multi-sensory behavioral therapy to a group of 2–3 dementia patients. It was found that in addition to the increase in engagement between the social robot and the primary interactor, there was a significant increase in interactions among participants not directly interacting with the social robot. Similarly in [48], a SAR system intervention was used to understand social interactions between intergenerational groups where each group consisted of a child, an adult, and an older adult. It was observed that the robot in combination with the child increased older adult's interactions. A similar observation was noticed in a larger study [49], which included a single-blind, randomized controlled trial (RCT) of 40 older adults receiving an intervention from a socially assistive pet robot PARO. Other studies have also made similar inferences [50, 51], including our own previous laboratory-based triadic HRI study [25], where we observed an increase in both

verbal and non-verbal communication between the two older adults when the robot provided prompts to elicit HHI.

Technology assisted social interaction among older adults can help alleviate their social isolation and/or loneliness and in turn increase their motivation and engagement in activity-oriented therapies [16]. In this context, multimodal therapies that combine cognitive, physical, and social activities are more effective than a single modality intervention [16]. Realizing the importance of many-to-one HRI, several studies have recently emerged to facilitate group cognitive stimulation, chair exercises, and conversation [35, 37, 39, 52, 53] with promising results in engaging older adults. The majority of these SAR systems were not designed with an intent to promote interpersonal social interaction among older adults; instead they focused on interacting with the group as a whole. In addition, these systems were developed based on a single class of activity—either physical, cognitive, or conversation.

We believe that the next generation of SAR-based activity-oriented therapies should not only engage older adults in physical and cognitive exercises, but also foster interpersonal social interaction with multimodal stimuli embedded into the system. Collectively, laboratory-based user studies [22–26] indicated the potential for SAR-based interaction to involve more than one older adult, to administer multimodal activities with the aid of the robot, and to quantitatively measure older adults' social interaction and activity engagement.

Several clinical trials have been conducted in LTC settings involving residents with dementia that examined the effect of a SAR on engagement and various neuropsychiatric symptoms [51, 54]. The most frequently used SAR has been the seal robot PARO, an animal SAR designed specifically for those with dementia [55]. Clinical trials of PARO have been conducted in LTC settings in Japan, Australia, New Zealand, Norway, and the US [56–61]. Studies varied in sample size (10–415), research design (pre-post, cross-over, nonrandomized and randomized clinical trials, cluster randomized trials), intervention design (individual versus group, facilitated versus non-facilitated sessions), length of sessions (10–45 min), frequency of sessions (1–3 sessions/week), duration of intervention (1–12 weeks), and outcomes (depression, apathy, quality of life, sleep, agitation, and psychoactive medications). Studies yielded mixed results, but with enough evidence of efficacy for an animal SAR to aid some older adults with various neuropsychiatric symptoms. However, animal robots are limited in their ability to actively engage older adults in cognitive and physical activities since its sole intent is to provide social or emotional connectedness.

A limited number of studies have evaluated the performance and user acceptance of more advanced SAR systems in the field. Robot Brian 2.1 was placed at a LTC facility

for two days and interacted with 40 older adults to play a memory card game or monitor a meal-eating activity [30]. Robot Tangy scheduled and played Bingo games with seven residents at a LTC facility [62]. Each resident participated in at least two sessions out of the six total group sessions. Field trial of robot Matilda, an assistive companion robot designed to improve emotional wellbeing of older adults, was conducted with 70 residents from three residential care facilities over a three-day period [52]. Field studies using a social robot for a duration of 3–4 weeks [63] and a life sized robotic platform for a duration of 7–8 weeks [51] as dyadic companions have also been conducted to increase the level of user engagement by providing contextual interactions with a responsive augmented reality environment. In another study [64], a robotic exercise tutor was tested with 6 residents in a nursing home for a single-session HRI and tested with 12 visitors at a day care center for multi-session HRI (1–5 sessions, mean: 2.58 sessions). Overall, the majority of the participants in these studies were engaged and participated with the SAR systems. However, participants had limited exposure to the SAR systems; most only interacted with the system once. In addition, few of these systems were developed to facilitate human–human interaction (HHI) and social interaction among older adults were either not observed or not discussed in field studies conducted in group settings.

3 Method

3.1 Engagement Model and Types of Data Collection

The effect of a dementia-related intervention activity for older adults can be measured by their interest or involvement in the activity. This is most commonly measured by engagement, defined as the act of being involved or occupied with an external stimulus [27, 64]. One of the most well-known models of engagement for older adults with dementia is the Comprehensive Process Model of Engagement (CPME) [64], which focuses on the attributes that influence engagement and the factors that are influenced by engagement. The model was expanded to the Comprehensive Process Model of Group Engagement (CPMGE) to involve additional factors that affect engagement in a group setting. The dimensions used in the CPMGE model to assess engagement is measured with Group Observational Measurement of Engagement (GOME) instrument [27]. Other metrics such as Revised Index for Social Engagement (ISE) [65], Social Observations Behaviors Residents Index (SOBRI) [66], Ethnographic-Labian-Inspired Coding system of Engagement (ELICSE) and Evidence-based Model of Engagement-related Behavior (EMODEB) [67] have been used to

measure engagement in field studies. Although these metrics have detailed components, they require a human experimenter monitoring continuously to obtain them. The field studies in which these engagement metrics were used were also of a relatively longer time frame (6 months–2 years) compared to the GOME model which can be used for short group activities. We used (GOME) as a guide to measure engagement in our study since the Ro-Tri SAR system is also designed as a multi-participant system to promote social interaction in a group setting. The CPME has been used in other studies [25, 26], including as the basis of the group engagement model in [68] that specifically focused on the physiological substrate of engagement measured through the electrodermal activity (EDA) of the participants.

GOME describes mainly two dimensions in measuring engagement: individual and group related attributes. The individual attributes consist of 4 components, i.e., activity attendance, level of engagement to the activity, attitude towards the activity, and active participation in the group activity. The group attributes consist of 3 components, which are group size, the negative and the positive reactions among the group members.

In our study, we collected data from a variety of sources: SAR system, staff ratings, individual's behavioral and conversational data, and older adults' attitudes and perceptions. Measures included activity attendance, game performance score, physical effort across sessions, activity and social engagement measured by head pose angles, the duration measure of each head turn towards partner and the number of times one is speaking. These measures correspond to various dimensions of GOME. For example, the activity attendance maps to participant attendance, game performance maps to active participation, physical effort maps to level of engagement, head pose towards the system maps to attention level and to a smaller extent active participation, head pose towards partner and number of times one is speaking maps to active participation regarding collaborative aspects of the game. With regard to participants' attitudes towards the activity, the SAR system, and towards other group members we used 3 investigator-developed instruments: Robot Acceptance Survey (RAS) [69], a staff-completed visual analog scale (VAS), and a post experiment feedback survey. RAS gathers the attitude and perception of the participants towards the interaction with the SAR system. Their responses indicate positive and negative reactions. For example, "I enjoyed playing the game with the robot" or "I found the robotic interaction similar to a real person" can be considered as positive reactions, whereas "I found the robot boring" or "I found the robot intimidating" can be considered as negative reactions. The 5-item VAS survey was completed by the staff caregivers based on their observations of the participants' interactions with each other and interest in the robot activities. Hence it contributes to the

determination of positive and negative interactions that the group participants have with each other from the perspective of the staff.

In our earlier laboratory study, the emotive EEG set (www.emotiv.com) was used in conjunction with an EEG index developed in [25] to estimate older adults' engagement with the activity. In the field setting the use of the EEG was not considered due to practical considerations. Rather, we measured physiological responses of the participants by using the wearable physiological sensor, E4 (www.empatica.com). This provided an implicit mechanism to monitor the participants' stress during the SAR multimodal intervention. E4 sensors have been used in previous studies as a measure of engagement [68]. E4 sensor measures one's skin conductance and heart rate and from these two signals stress can be inferred using affective computing [70]. Besides assessing stress response, we were also interested in the accuracy of the E4 sensor when used in activities that involved physical motion.

3.2 Robotic System

The Ro-Tri architecture used for the field study was similar to the architecture presented in our previous publications [25, 26] with the following changes: (1) removal of the EEG acquisition module and use of E4 wristbands to collect physiological signals to infer older adults' implicit mental states, stress in particular; (2) extension of the interaction duration of the Simon Says activity to match with the interaction duration of the Book Sorting task (tasks are described in the next section); and (3) modification of the quantitative data acquisition modules for the Simon Says activity to match with that of the Book Sorting task.

3.2.1 Multimodal Activities

Ro-Tri, as shown in Fig. 1, was capable of administering four activities: Simon Says, Book Sorting taking turns (take turns mode), Book Sorting together (simultaneous mode), and Book Sorting with additional rules (post-test). Our systems engaged two older adults simultaneously in physical, cognitive, and social activities with the robot. Simon Says was based on an imitation game where each older adult and the robot took turns to direct a gesture and expected that others would follow only if the gesture was introduced with the utterance “Simon says”. For this activity, we used a Kinect v1 for gesture recognition and the computer monitor was turned off. There were five rounds of group interaction. The first round was introduction, during which the robot and older adults took turns to introduce their names and greet each other. The second to fourth rounds were “Simon says” play. In each session, the robot first acted as a leader to demonstrate an arm movement, initially vocalizing “Simon says” that required older adults to copy the movement and the second time without vocalizing “Simon says” in the instruction. Then the robot asked one older adult to be the leader and the robot and the other older adult would be the followers. This round ended with the other older adult also having the chance to play as the leader. Algorithms were designed for the robot to demonstrate and recognize three gestures, which were wave, raise arms up, and extend arms to the side. When older adults demonstrated a gesture, the robot had the ability to mirror their upper arm movements. The last round was when the robot thanked the older adults and asked them to wave goodbye to each other. The Simon Says activity has an embedded physical component through arm movements, a cognitive component through registering whether to follow

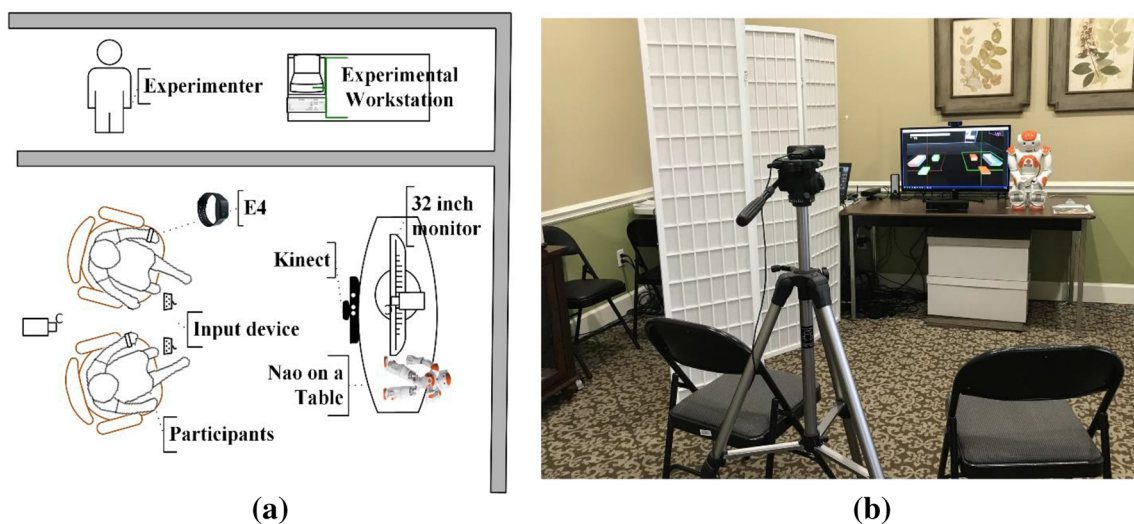


Fig. 1 **a** Ro-Tri System configuration used in the study; **b** Ro-Tri Setup at a Retirement Community

the leader's command, and a social component in requiring older adults to take turns as leader.

The virtual Book Sorting task was an activity with two triadic interaction modes, 'take turns' interactions and 'simultaneous' interactions. Here the triadic interaction refers to the situation where one robot interacts with two older adults. This task allows older adults to collect books into color matched bins in a virtual environment with natural grasp and arm movement. In take turns mode, only one older adult interacts with the virtual environment at a time, the other older adult cannot move virtual books but is free to help his/her partner. In simultaneous mode, both older adults can interact with the virtual environment at the same time. The Book Sorting activity used a Kinect v2 for motion-based interaction with a virtual reality (VR)-based book sorting game displayed on the computer monitor. The VR-based book sorting game consisted of different colored books and color matched bins to deposit the books (Fig. 2 left). Each older adult had a color-coded hand cursor displayed on the monitor that they could manipulate through large range arm movements and open/close hand gestures. For example, when older adults moved their arms to the left, the hand cursor would move to the left of the monitor until it reached the left boundary of the workspace. When older adults' hand cursors overlapped with books, they could grab the books by closing their hands. We defined the rules in the game to reward collaborative behaviors that occurred for the purpose of fostering HHI.

In take turns and simultaneous interactions, collaboration occurred when older adults helped each other by moving books closer to each other's bins, whereas in the additional rules post-test interaction, collaboration occurred when older adults moved the same book in the same direction. In the take turns interaction, there was only one hand cursor displayed on the monitor, and older adults were required to wait for their partners to finish before they could control the hand cursor. In the simultaneous interaction, two hand cursors were allowed and older adults could play at the same time. The robot facilitated the older adults both in take turns and simultaneous interactions with the purpose of maintaining and enhancing task engagement and HHI. This was realized by continuously evaluating older adults' interactions and providing feedback to engage them in motion-based

interaction, encouraging them to help one another, and celebrating their accomplishments in the game. In the post-test, older adults were not told to move the same book together. Yellow books were used in this task and the goal of the task was for the older adults to figure out the unknown collaborative rule through social interaction (Fig. 2 right). We expected them to explore different ways to interact with the system and gradually figure out that they needed to collaborate to move books. The robot would provide a hint half way through the session if older adults were not able to move books at all. These activities have an embedded physical component through arm and hand movements, a cognitive component through a matching and sorting exercise, and a social component through the need to communicate and collaborate in order to successfully complete the task. The SAR systems provided feedback to engage older adults in the activity as well as in HHI.

3.2.2 Robot Behaviors

The activities for triadic interaction were carefully designed to encourage communication the older adults. The Simon Says activity required older adults to copy each other's gestures, whereas the Book Sorting activities had collaborative rules embedded specifically for HHI through team work. Most importantly, the robot behaviors were designed to engage older adults in the activity as well as to elicit HHI.

The robot behaviors were governed by a hierarchical state machine (HSM) with different robot states for HRI and interaction events to trigger the transitions from one state to another. Since system development is not the focus of this work and given the complexity of the HSM, we provide a brief description of robot behaviors for social HRI and to elicit HHI. Details of the HSM and the design rationale are in [25, 26].

The robot interacts with older adults socially through social gaze, gestures, and utterance. Example robot gestures are clap, celebrate, wave, point towards human, and interact with the VR activity. There is no physical contact between the robot and the older adults. Based on older adults' real time interaction with the activities, Ro-Tri system automatically computes performance metrics and these metrics trigger robot behaviors to facilitate HRI and HHI. For example, in the Simon Says activity, if only one older adult correctly followed the robot's demonstrated gesture, the robot encouraged them to pay attention to each other's gestures by saying "Did everyone get this right?" For the Book Sorting activity, if both older adults struggled to move virtual books, the robot helped them by pointing out how to complete the task. If only one older adult had difficulty, the robot would direct the other older adult to offer help by saying "[Name 0], can you help [Name 1] with how to [area to improve]?". We designed the robot behaviors to induce

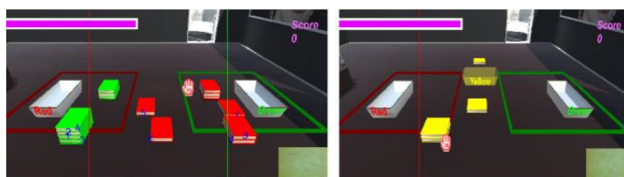


Fig. 2 Book sorting activity interface (left); Post test—yellow book task interface (right). (Color figure online)

social communication between older adults as much as possible, either by referring to their names such as “[Name 0], remember that [Name 1] cannot move the hand cursor pass the green vertical line. If you move the book further away, [Name 1] cannot grab the book”, by directly asking one older adult to help another, or by prompting them to collaborate in the task.

3.3 Field Study

3.3.1 Experimental Setup

With approval from the Vanderbilt University Institutional Review Board, the Ro-Tri system was sequentially placed at two local retirement communities and used by older residents. Eligibility criteria for participants included: (1) age 70 years or older; (2) ability to hear as screened by the Whisper Test [71, 72] with or without hearing aids; iii) ability to see as screened by ability to read newspaper print with or without eyeglasses; iv) ability to move arms as screened by the ability to raise arms up, forward and to the side; and v) ability to cognitively participate in the various robotic activities. The experimental setup and materials are shown in Fig. 1. Participants sat in front of and facing the system. NAO was positioned to the side of the computer monitor. The Kinect was placed on the edge of the table facing the two participants. An administrator operated the experimental workstation in a separated space. The primary robot-mediated activities for the paired older adults were the Book Sorting activity alternated with the Simon Says activity (Table 1). Each pair interacted with Ro-Tri twice per week for three weeks within a month. Two types of activities were presented each week, and the activities were arranged in such a way that the participants did not take part in the same activity in two successive sessions. This was done to reduce habituation effect, to keep the older adults interested, and also to reflect activities in LTC settings where, in general, activity options vary each week [18, 73]. Before the

triadic interaction, each participant went through an orientation to become familiar with the virtual Book Sorting activity as well as the robot movements and speech. The estimated interaction duration, which was 10 minutes, only included the time needed to interact with Ro-Tri. The whole session also involved putting on E4 sensors, calibration and baseline data recording, adjusting the robot’s speaking volume, adjusting software parameters, as well as collecting data from older adults; thus the complete session lasted approximately 40 min each.

3.3.2 Procedure and Participants

We conducted the field study first at Sycamores Terrace Retirement Community (www.sycamoresterrace.com) with 9 older adults and then at Elmcroft Senior Living (www.elmcroft.com/community/elmcroft-of-brentwood-tennessee) with 6 older adults. At Sycamores Terrace Retirement Community, Ro-Tri was set up in a vacant apartment. Participants interacted with Ro-Tri in the living room and the administrator operated the experimental workstation in the bedroom. At Elmcroft Senior Living, Ro-Tri was set up in the corner of a library with a room divider to separate the experimental workstation from participants (Fig. 1). A total of 14 older adults (7 pairs, mean age: 82.7, 3 had normal cognition, 10 had mild cognitive impairment (MCI), and 1 had Alzheimer’s dementia) completed the field study. One older adult dropped out after the second session due to her hearing aid issue and her peer was paired with another older adult and restarted from session one. At the start of each session, we placed E4 sensors on participants’ non-dominant wrists and recorded three minutes of baseline physiological responses while the participants were asked to sit quietly. We then reminded them how to interact with the system.

In Simon Says sessions, we told them only arm movements were recognized by the robot and reminded them to pull the trigger button of the Razer Hydra controller after they answered robot’s questions. In Book Sorting sessions,

Table 1 Experimental protocol and timeline used for each pair of participants

Session (Week)	Activity description	Estimated duration	Comment
Orientation	One-to-one HRI & task orientation	9 min	Older adults get familiar with interacting with the system and the robot
Session 1 (Week 1)	Simon Says	9 min	–
Session 2 (Week 1)	Book Sorting (take turns)	9 min	Allow older adults to practice in the virtual environment for 3–5 min before HRI
Session 3 (Week 2)	Book Sorting (simultaneous) & Book Sorting (post-test)	12 min	Allow older adults to practice in the virtual environment for 3–5 min before HRI
Session 4 (Week 2)	Simon Says	9 min	Repeat Session 1
Session 5 (Week 3)	Book Sorting (take turns)	9 min	Repeat Session 2
Session 6 (Week 3)	Book Sorting (simultaneous) & Book Sorting (post-test)	12 min	Repeat Session 3

we asked them to practice moving their hand cursors and grabbing books for a few minutes. Practice was followed by a short calibration that recorded Kinect's head pose angles when we asked older adults to look at the robot, the computer monitor, and their partners as well as Kinect's sound source angles when we asked each older adult to read a sentence. The experimenters then started the robotic interaction and stayed out of sight of the participants during the interaction. Finally, participants filled out a post experiment evaluation questionnaire at the end of each week, i.e., after sessions 2, 4, and 6.

3.4 Data Collection and Analysis

Two types of data were collected during the study, which were data logged automatically by Ro-Tri and data forms filled out by the participants and the caregivers. Human provided data included surveys for participants' acceptance of the system and a Visual Analog Scale (VAS) for caregivers' opinion about the participants engagement behaviors. Prior to implementation and conclusion of the study, the participants completed the Robot Acceptance Scale (RAS, 7-point scale, 1 most positive to 7 most negative response) that we developed previously [69]. The staff completed a VAS (0–10 continuous scale, 0 most negative to 10 most positive response) for assessment of the extent to which participants interacted with others and were interested in the robot sessions. At the end of each week, participants completed a post experiment evaluation questionnaire (7- point scale, 1 most negative to 7 most positive response) that elicited opinions about the activities and robot sessions.

Data collected by Ro-Tri were participants' interaction data and activity states, participants' head pose angles, Kinect's sound source angles as an indicator of sound source direction, participants' physiological responses from the E4 sensor, and the robot's behaviors. Interaction data logged participants' interaction with the Book Sorting task as well as their upper body skeleton position data. From interaction data and activity states, we computed an *effort metric* representing the amount of effort exerted by the participants during HRI. For Book Sorting tasks, the effort was the amount of book movements to collect one's own book or to help others. For Simon Says activity, the effort was the accumulated elbow and wrist movements. Participants' head pose yaw angles served as a coarse estimation of their gaze directions; a measure used previously where approximate gaze direction is adequate for the task [74]. Head pose yaw angles were zero when participants looked straight ahead, decreased when they looked to the right, and increased when they looked to the left. From the calibration data, which logged participants' head poses when they looked at the computer monitor, the robot, and their partners, we calculated head pose yaw angle ranges for head towards the robot,

head towards the computer monitor, and head towards the other person (Fig. 3). These ranges allowed us to compute automatically the amount of times older adults' paid visual attention to the computer monitor or the robot, as well as the amount of times and the number of times older adults moved their heads towards their partners. We classified visual attention to the system as *activity engagement* and visual attention to the other older adult as *social engagement*.

To compute activity engagement based on head pose yaw angles, the ranges of head pose towards the computer monitor and robot and the thresholds for head pose yaw angles towards human were used first to segment raw head pose data into intervals of data that belonged to activity engagement (i.e., head posed towards the computer monitor or robot), social engagement (i.e., head posed towards partners), or neither. Intervals belonging to activity engagement were summed together to calculate the total activity engagement duration. For social engagement, we first generated candidates for start timestamps when older adults potentially initiated a looking behavior. These candidates were selected from the intervals belonging to social engagement. In order to reduce accidental count of head turns due to noisy data, we set a 1 s threshold so that the start time of the next head

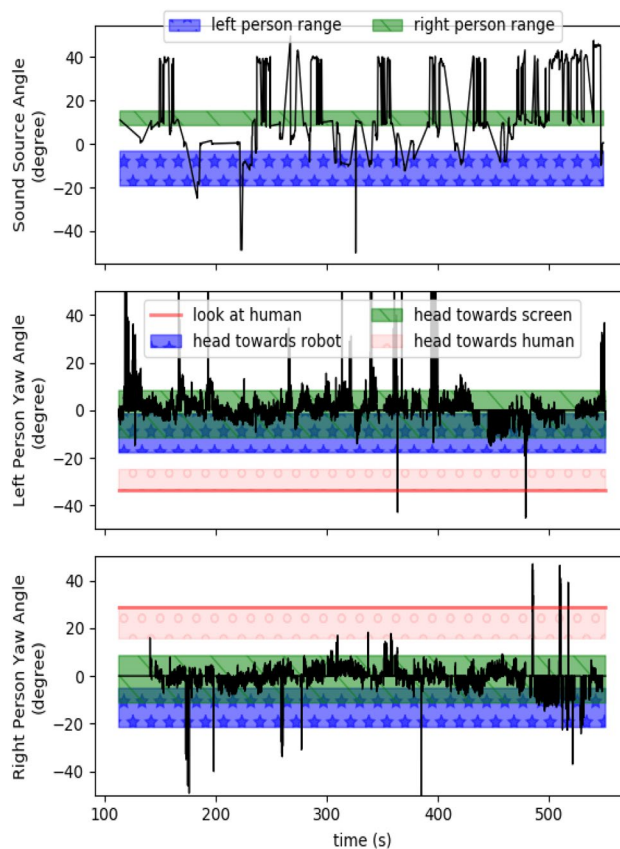


Fig. 3 Raw Sound Source Angle and Head Pose Yaw Angle Data for One Session of the Field Study

turn must be at least 1 s later than the end time of the previous head turn. The end timestamp for a selected candidate was calculated by merging the intervals associated with the candidate and outputting the end time of the merged interval. Each candidate represented a potential head turn. All the candidates were passed through three thresholds to filter out artifacts, such as very short durations, older adults' hands in front of their faces, or head pose data interpolation. From the remaining candidates, we calculated the social engagement duration and the number of times older adults looked towards their partners.

Sound source angles data were used to estimate the start and end of vocal sounds made by the older adults. From calibration data, we were able to compute ranges of sound source angles that captured each older adult's vocal sounds. In Fig. 3, the green band indicates when the right person was talking and the blue band indicates when the left person was talking in one HRI session. To compute automatically the amount of time older adults were talking and the number of times they spoke, we first segmented the raw sound source angles into intervals of data that belonged to the left speaker, right speaker, or neither based on the ranges and confidence levels of the detection algorithm. Second, the start times and the end times of these intervals were mapped to their closest integers in seconds, respectively, by applying the floor and ceiling functions. After this mapping, some intervals might overlap. We then merged all the overlapped intervals and finally summed the duration of these intervals to calculate the total amount of time older adults made vocal sounds during the triadic HRI. We also computed the number of times they were speaking as the count of these intervals after merging.

E4 sensor recorded peripheral physiological data including photoplethysmogram (PPG) and electrodermal activities (EDA). The sampling rates for PPG and EDA were 64 Hz and 4 Hz, respectively. The data were examined and three classes of features were used. PPG related features included heart rate and heart rate variability. EDA related features included mean skin conductance level (SCL), standard deviation of SCL, mean amplitude of skin conductance response (SCR), standard deviation of SCR, maximum amplitude of SCR, and rate of SCR. Finally, the temperature related features included mean and standard deviation of skin temperature. Heart rate (HR), which was used as a proxy measure of positive or negative emotions, was computed by detecting peaks in the PPG signal. Heart rate variability (HRV) measures the specific changes in time (or variability) between successive heart beats; it refers to the oscillation of the interval between consecutive heartbeats. HRV has been used as an indication of mental effort and stress [75]. EDA provides a measure of the resistance of the skin. This resistance decreases due to an increase of sudation, which usually occurs when one is experiencing emotions such as stress or

surprise. Tonic and phasic components of EDA were decomposed separately from the original signal [76]. The tonic component is the baseline level of EDA and is referred to as skin conductance level (SCL). The phasic component is the part of the signal that changes when stimuli are presented and is known as skin conductance response (SCR). Lang et al. discovered that the mean value of the SCR is related to the level of arousal [77]. EDA is a strong indicator of affective arousal in general [76]. Gjoreski et al. has used skin temperature data from a E4 wristband to predict stress level [70]. From the physiological data, we were interested in seeing whether it was possible to detect occurrences when the participants were stressed and when they were relatively at ease during HRI and HHI. The three-minute baseline data were used to remove feature variations due to time and individual differences. Specifically, the heart rate, heart rate variability, mean SCL, mean amplitude of SCR, and mean skin temperature features were subtracted by their respective baseline values and divided by their respective baseline standard deviation values. For the remaining features, the baseline values were subtracted from the features.

4 Results

4.1 Ro-Tri Algorithm Validation

The abovementioned algorithms for automatically computing the amount of times and the number of times older adults talked to and looked towards the other person were validated using data recorded during previous laboratory tests. In our previous laboratory experiments, paired older adults performed Book Sorting tasks (take turns and simultaneous) under the guidance of a robot. A trained research assistant manually analyzed the video and audio recordings and logged the start and end timestamps for each talking and looking behavior as the ground truth. The start and end timestamps were automatically generated by the algorithms and were validated against the ground truth. We validated the algorithms based on data from 8 older adults. The validation results are shown in Table 2. In general, the head pose analysis algorithm could detect with high accuracy the amount of time and the number of times older adults looked towards their partners. The start time deviation for correctly detected looks had a mean value of 0.25 s and a standard deviation of 0.14 s. The end time deviation for correctly detected looks had a mean value of 0.30 s and a standard deviation of 0.21 s. The sound source angle analysis algorithm could detect with high accuracy the number of times older adults spoke. For the duration of speaking, the algorithm had high precision but many missed detections. Therefore, the speaking duration was excluded in our field study data analysis.

Table 2 Validation results of Ro-Tri automatic evaluation algorithms

Data Type	Measure	Precision (%)	Recall (%)
Head Pose	Amount of times looking towards partners	98.30	91.52
	Number of times looking towards partners	95.65	86.27
Sound Source Angle	Amount of times talking	99.40	65.41
	Number of times talking	97.92	87.04

4.2 Ro-Tri Logged Data Analysis Results

The system worked as designed. Fourteen participants completed all 6 sessions. One participant dropped out after completion of session 2 due to issues with her hearing aids. Her peer was paired with another newly recruited participant and the pair restarted from session 1. For the post-test task, older adults were able to figure out the unknown collaborative rule and move yellow books together through communication with their partners for 10 out of 14 sessions. The robot provided hints to help them for the other 4 sessions. Table 3 lists the participants' engagement across 6 sessions as measured by their interaction effort, head pose, and sound source angle. On average, participants spent 77.7% of the time looking at Ro-Tri and 2.3% of the time looking towards their partners. The total average engagement including both system and social was 83.5% for the participants. The number of times looking towards their partners and the number of times talking across 6 sessions were 0.41 times per minute and 3.72 times per minute, respectively. The duration of each looking behavior towards their partners had an average value of 2.94 s. Since participants' interaction effort, visual attention and communication varied for different activities, we normalized the engagement results in order to compare results and demonstrate changes over 6 sessions.

For each activity, Simon Says, Book Sorting take turns, and Book Sorting simultaneous, we computed the best engagement values by taking the average of the top three values for that activity. The worst engagement values were derived based on the nature of the engagement measure and the activity. For effort, visual attention towards

partners, and verbal communication, the worst values were zero. Whereas for visual attention towards the system, the worst value was the head pose range towards the system divided by 180 assuming that participants looked at different directions randomly. Because the data were not normally distributed, we used min–max normalization where the engagement results were first subtracted by the worst engagement value, and then divided by the absolute difference between the best and worst engagement values. After normalization, the higher the value, the greater the engagement. Figure 4 shows changes in interaction effort, visual attention, and verbal communication over 6 sessions. The group results were the mean values of the 14 participants. In addition to group results, we plotted changes of engagement for some individual older adults as examples. Older adults' interaction efforts were maintained throughout the HRI sessions with slight improvement towards the end, 2.9% at session 6. Eight out of 14 participants' efforts increased from session 1 to session 6. For head pose data, participants' activity engagement represented by percentage of time they looked at the system increased slightly, by 7.2% at session 6. The change of visual attention over 6 sessions varied among individuals, as illustrated by two participants' results (S211 and S305). From session 1 to session 6, nine participants paid more visual attention to Ro-Tri and five of them paid less attention. Seven out of 8 participants who paid more attention to the system also had increased interaction effort. We combined HHI and HRI engaging metric from head pose data and the overall engagement (Fig. 4f) showed an increase of 5.1% from session 1 to session 6.

Table 3 Participants' engagement across six sessions (M—Mean, SD—standard deviation)

Data Type		Session1 ^a M (SD)	Session2 M (SD)	Session3 M (SD)	Session4 ^a M (SD)	Session5 M (SD)	Session6 M (SD)
Interaction Data	Effort/min	1.08 (0.77)	13.19 (5.45)	14.64 (8.18)	1.08 (0.73)	13.33 (6.34)	13.74 (5.12)
Head pose activity engagement	Duration—% of total session time	71.0% (14.6%)	87.6% (6.9%)	76.1% (15.5%)	70.7% (19.2%)	84.1% (9.8%)	76.4% (14.1%)
Head pose social engagement	Duration—% of total session time	1.3% (1.5%)	1.7% (3.0%)	3.1% (6.0%)	3.2% (4.7%)	2.0% (3.7%)	2.4% (3.1%)
	Count/min	0.29 (0.32)	0.48 (0.51)	0.50 (0.56)	0.35 (0.41)	0.37 (0.41)	0.47 (0.54)
	Duration (sec) /count	2.82 (1.75)	1.71 (0.87)	2.57 (2.46)	4.89 (3.45)	2.61 (1.88)	3.04 (1.33)
Sound source angle	Count/min	2.48 (2.29)	3.67 (4.17)	5.35 (5.64)	3.41 (1.71)	3.80 (4.52)	3.59 (3.59)

^aSession 1 and Session 4 are Simon Says, hence the Effort/min values are not directly comparable to other sessions

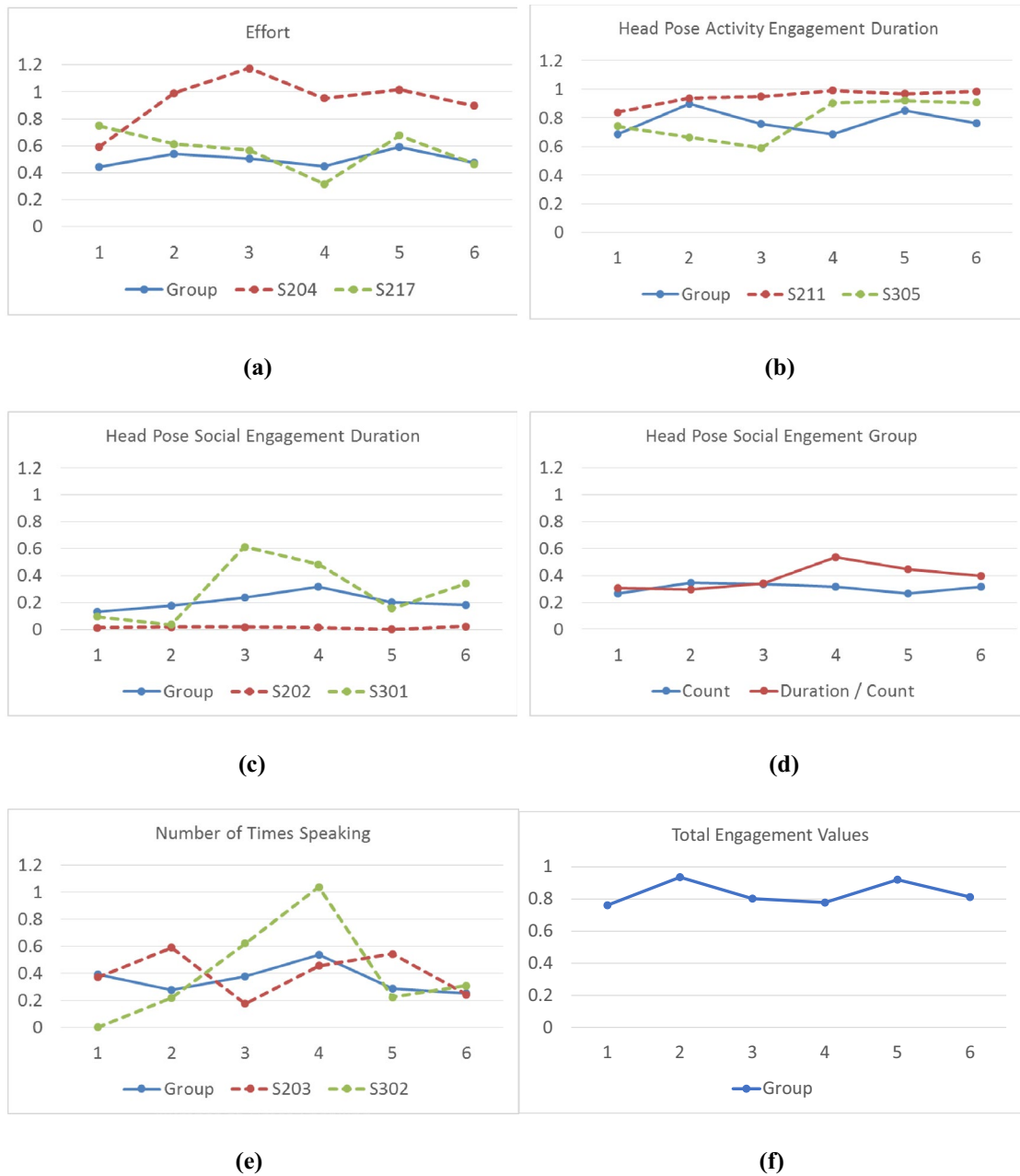


Fig. 4 Changes of various engagement metrics across sessions. X axis is session id. Y axis is the normalized engagement metric. **a** physical effort (Book Sorting—amount of book movements to collect one’s own book or to help others. Simon Says—accumulated elbow

and wrist movements); **b** head pose towards the Ro-Tri system; **c** head pose towards the other older adults in terms of duration; **d** head pose towards the other older adults in terms of count and duration per count; **e** speaking frequency; and **f** total engagement value ((**b**) + (**c**))

The percentage of time the participants looked at their partners continued increasing from session 1 to session 4. Eventually, participants’ visual attention in terms of duration towards their partners increased by 4.7%, which was slightly less than the increase of their visual attention towards the system. Seven participants paid more visual attention to their partners from session 1 to session 6. Only 2 participants’ visual attention to both Ro-Tri and their partners decreased

from session 1 to session 6. The rest of the participants either paid more visual attention to both HRI and HHI (4 out of 12) or paid more attention to the system and less attention to their partners or vice versa. The number of times participants looked towards their partners also increased slightly, by 4.7% at session 6. Overall, participants looked towards their partners at a frequency similar to session 1 during the experiment. However, they spent longer duration for each

looking behavior, increased by 8.8%. Finally, for verbal communication results, participants talked more during week 2 as compared to week 1. During week 3, their verbal communication results reduced, falling below that of week 1.

By observing the video recordings of the experiment, we selected 20 instances of data where the participants were stressed and 19 instances where the participants were relatively calm. Each instance was one and half minute in duration and labeled by a research assistant, either as “calm” or “stressed”. Waikato Environment for Knowledge Analysis (WEKA) was used for feature selection and model training. The wrapper subset evaluation method using the best attribute technique (forward direction) was used to select the best features and four machine learning algorithms were used to predict the stress level. The machine learning algorithms were evaluated with five-fold cross validation. The best performing features were found to be the mean SCL, mean SCR, mean heart rate variability, and mean skin temperature. The machine learning algorithms applied as well as the corresponding classification results are shown in Table 4. When compared with the baseline, we found that when the participants were stressed, three features showed noticeable changes; mean SCL when stressed was found to be 1.38 times higher than the baseline mean SCL. Mean SCR when stressed was found to be 5.36 times higher than the baseline mean SCR. Maximum SCR when stressed was found to be 3.11 times higher than the baseline maximum SCR value. Mean SCL and maximum SCR results were found to be statistically significant at 0.05 level with p -values of 0.010 and 0.042, respectively. Mean SCL was found not to be significant with p -value of 0.065. A total of 39 data points for each of 1.5 min were used.

4.3 Human Provided Data Analysis Results

The RAS survey was conducted to determine participants’ acceptance and anticipated use of the robotic system based on performance expectancy, effort expectancy, and attitude towards using the system. All participants completed the pre RAS and 13 completed the post RAS. Participants’ perceptions became more positive for all the subscales and RAS after the experiment (Table 5). Wilcoxon signed-rank test

Table 4 Classification results between stress/stress-free instances from the E4 sensor

Machine learning algorithm	Accuracy (%)	Precision	Recall	F-1 score
Random forest	75.0	0.75	0.75	0.75
Decision stump	71.4	0.76	0.71	0.70
Decision tree (J48)	71.4	0.71	0.71	0.71
Logistic regression	67.8	0.74	0.68	0.66

Table 5 Older adults’ attitudes and perceptions toward robots (RAS)

	Pre ^a M (SD)	Post ^a M (SD)	Z	p	r
Performance Expectancy	2.84 (0.57)	2.69 (0.78)	0.36	0.749	0.07
Effort Expectancy	2.83 (0.77)	2.52 (0.70)	1.55	0.133	0.30
Attitude	2.57 (0.55)	2.12 (0.66)	1.81	0.075	0.35
RAS	2.70 (0.57)	2.36 (0.67)	1.75	0.084	0.34

^aLower values are more positive, 7-point scale

results are shown in the table, including the standard score of the Wilcoxon signed ranks, p value, and effect size. Effort expectancy subscale, attitude subscale, and RAS were more positive with a medium effect size. VAS was completed by caregivers or staff who were familiar with the participants. The five questions used for VAS and their results are shown in Table 6. After six HRI sessions, staff rated participants social interaction during daily activities improved by 6.2%. Participants were observed to be more interested with Ro-Tri, 8.2% improvement on anticipation of robot session and 2.3% decrease on complaints about robot session. Participants’ engagement on daily activities decreased by 1.6%.

Post experiment evaluation gathered participants’ interests and acceptability on robot sessions after each week of HRI. Examples of questions were “Enjoyed attending the robot sessions” (interest on robot session), “Looked forward to interacting with another resident for the robot sessions” (interest on triadic interaction), “The robot was able to keep your attention” (acceptability of robot), “Doing the Book Sorting activity for future studies” (acceptability of activity), and “How interesting or boring were the Simon Says activity” (interest on activity). In general, participants’ interests and acceptability on the robot, triadic interaction with partners, and the activity were positive and maintained over the 3 weeks (Table 7).

5 Discussion and Conclusion

In this small field study, we found the Ro-Tri SAR system to be feasible and acceptable to older adults residing in LTC. Importantly, the system was able to deliver multimodal stimuli involving physical, cognitive and social domains to more than one person at a time. The study provided valuable lessons with regard to using SAR systems in LTC. First, the completed field study took 6 months. We needed to find and recruit a second site to provide sufficient pairs of participants. The physical set up at these settings posed challenges that took some additional time (e.g., lack of internet). Fifteen participants were recruited to take part in the study. Eight participants (4 pairs) completed all 6 sessions at the

Table 6 Staff's ratings of older adults' social engagement (VAS results)

Question	Pre ^a M (SD)	Post ^a M (SD)	Change (%)	Z	p	r
How social would you rate the participant?	7.97 (2.05)	7.97 (1.69)	0	–	–	–
To what extent would you say the participant likes to come out and do activities?	7.21 (3.05)	7.05 (2.91)	– 1.6	0.71	0.518	0.14
To what extent would you say the participant likes to talk to other residents, staffs, or family?	8.08 (1.85)	8.71 (1.44)	6.2	1.54	0.132	0.29
To what extent would you say the participant looks forward to attend the robot sessions?	8.62 (1.42)	9.44 (0.71)	8.2	1.68	0.102	0.36
To what extent would you say you observed the participant complained about the robot sessions?	9.41 (1.33)	9.64 (0.65)	2.3	0.31	0.844	0.07

^aHigher values are more positive, 0–10 continuous scale

Table 7 Older adults' post experiment evaluations

	Week 1 ^a M (SD)	Week 2 ^a M (SD)	Week 3 ^a M (SD)
Interest on robot session	6.33 (0.94)	6.33 (1.26)	6.52 (0.61)
Interest on triadic interaction	6.19 (1.03)	5.88 (1.70)	6.33 (0.91)
Acceptability of robot	6.21 (1.01)	6.33 (0.81)	6.54 (0.55)
Acceptability of activity	6.15 (0.95)	6.40 (0.83)	5.80 (1.10)
Interest on activity	6.19 (1.01)	6.30 (0.78)	5.96 (1.13)

^aHigher values are more positive, 7-point scale

first retirement community and 6 participants (3 pairs) completed the study at the second retirement community. We had 91.7% attendance over a 100% activity completion rate per session, demonstrating acceptability. Additional data from older adults and staff confirmed acceptability, tolerance, and interest in Ro-Tri system. Participants' perceptions on Ro-Tri were more positive after the three-week experiment and their interest and acceptability were high for both the robot and the activities. Importantly, they enjoyed interacting with another resident for the robot sessions. Some of the outcome measures had a medium effect size, which suggests more pairs of participants or longer study might lead to significant results. There was a decrease in the acceptability of and the interest in activity components of the post experiment survey at week 3. We speculate this may have been a result of the similar types of activities offered during Week 3 (both Book Sorting activities). There may have not been enough variability in the activities that resulted in a slight decrease in interest. Alternatively, this was the final week of the study that may have resulted in some habituation.

Ro-Tri logged several types of data during HRI to evaluate older adults' engagement in terms of interaction effort, visual attention to the system and to another older adult and verbal communication during HRI. In general, participants' engagement levels were maintained throughout the study. Quantitative data analysis results indicated an increase of 7.2% and 4.7% between session 1 and 6 in visual

attention of the participants towards the SAR system and their partners, respectively. Similarly, observational rating scale data as measured by the VAS scale showed an increase of 8.2% and 6.2% (for the questions "To what extent would you say the participant looks forward to attend the robot sessions?" and "To what extent would you say the participant likes to talk to other residents, staffs, or family?" of the VAS scale), respectively. Results indicated that the participants became more collaborative as the sessions progressed. They also demonstrate agreement between the SAR measured and human rated metrics. A slight decrease in social engagement and number of times speaking can be observed after session 4, which we believe is caused due to the change in activity.

Simon Says activity (session 1 and 4) differs from the Book Sorting activity (sessions 2–3 and 5–6) in that it encourages both HRI and HHI through direct participation of the robot throughout the task. The robot encourages HRI with the participants (in addition to conducting the game) by asking their names and engaging in small conversations such as "I am glad to meet you participant name" and "I guess this is the first time you have seen a robot". In this activity, the robot is directly involved in the task, either by following a leader participant or acting as a leader itself. It also encourages HHI by asking participants to meet and greet each other and discuss with each other when either participant makes an error during the game. The robot had an average collaboration prompt of 5 per session and the participants had an average head pose social engagement of 2.2% per session. In comparison, in the "Book Sorting" activity, the robot is mainly focused on HHI and acts more like a coach; the robot reminds the participants to engage with each other during the activity. Here the interaction with the robot is not as a copartner as the robot is not directly involved in the game. The robot only intervenes if the participants are not already collaborating compared to "Simon Says" activity where the robot is engaged throughout. If there is no need for the robot to engage older adults in the book sorting activity, the robot will fade into the background. This can be seen from the collected data on robot behaviors during the Book Sorting

task; when the average head pose social engagement % of the participants had been low, 1.7% and 2.3%, the robot's average collaboration prompts had been higher, 3.28 and 3.42 for session 2 and session 6 respectively. Similarly, when the average head pose social engagement % of the participants had been comparatively higher, 1.9% and 3.1%, the robot's average collaboration prompts had been lower, 2.57 and 2.42 for session 5 and session 3 respectively. Thus, the robot responded to the decrease in social engagement from the participants by increasing its collaborative prompts. This does create variation for the metrics, but the variation of activity is one factor we considered when designing multimodal activities for older adults (e.g., in the case for Book Sorting, the robot was encouraging collaboration and hence the higher social engagement).

We have also noted that as the sessions progressed the quality of the interactions increased, i.e., the participants gave appreciative feedbacks, made funny comments, etc. This can be seen in Fig. 4d where the duration/count increases for the head pose. The number of times speaking metric also decreased due to the same reason, that in the initial sessions, the participants used to only query each other for game relevant information. There were slight improvements for visual attention towards the system and partners from session 1 to session 6. The learning effect is an issue, which guided our experiment design and protocol in terms of the number and the manner in which the sessions were conducted during the field study. Increasing the number of sessions without introducing more variety either in terms of different games or partners may lead to boredom and lack of engagement by the participants. Solutions to offset this problem include having more than 2 players participate in a single activity session (our system architecture has the capability) or introducing randomization in partner selection after a fixed number of sessions.

E4 sensors were used to collect physiological signals from older adults. It was easy to apply the sensor and none of the participants complained about wearing a wristband. We extracted 10 features from the physiological data to estimate the stress level of participants during HRI. The ability to monitor the stress level will enable a future stress-sensitive robotic system. This way Ro-Tri could provide more personalized feedback to engage older adults in activity-oriented therapies. We obtained the highest accuracy using Random forest classifier (75%—Table 4). In general, higher classification rates were obtained using tree-based classification methods. It is important to note that the activities in this study were not specifically designed to induce stress or have stress free intervals during the interaction, which we believe is the reason for not obtaining higher accuracies. From the observations in the field study, participants were stressed in mainly two cases. First, a participant was stressed when the participant was having difficulty during the activity but

the robot, without recognizing this, kept giving instructions about the following steps. For example, initially many older adults had difficulty in moving their hands in 3D, which was mapped to a 2D screen for the Book Sorting task. However, the robot did not recognize this difficulty and it kept on reminding them to collaborate. Second, when one of the participants performed well and the other person was having difficulty, the poorly performing participant received advice from both the robot and his/her partner, which caused confusion and stress. Participants were also stressed during the yellow book task if they could not figure out the rule and the robot purposefully did not offer help.

To the best of our knowledge, only a small number of studies have shown the effectiveness of multimodal interactive intervention in a real world setting and measured activity and social interaction of older adults over time. The Ro-Tri system worked flawlessly in the field. The key findings of this study are: (1) the Ro-Tri based intervention was well tolerated by the older adults; (2) they stayed engaged over several weeks of robotic intervention with no drop out; (3) the LTC management was excited to test the robotic intervention in their facilities, which is essential for its adoption; and (4) most importantly, we observed HHI during the intervention, which has the potential to address apathy and social engagement in older adults. Our measures are sensitive to measuring change over time and will be useful in future studies of older adults residing in retirement settings.

Although the current study was successful in establishing the feasibility of SAR-based intervention in LTC settings and demonstrating potentially useful results in terms of engagement and HHI, it had several limitations. First, the sample size and session duration were not large enough to provide evidence on the efficacy or impact of Ro-Tri on older adults' activity and/or social engagement in daily life. Although 6 sessions were relatively longer exposure to the SAR system as compared to several other works in this field, more exposure to the SAR will likely provide important insight regarding how older adults' engagement changes over time. Randomization of older adults to each session was considered but determined to be not practical to implement for this small pilot study. Larger clinical trials are planned with this additional methodology. The current results focus on the changes of group engagement. Each individual older adult's interaction effort, visual attention, and verbal communication over 6 sessions changed differently; thus future studies will need to determine those activities and individual characteristics that are most effective in engaging older adults. We observed situations where one older adult performed very well whereas the other performed poorly. Some older adults were sensitive to their performance as compared to their partners and this might change their response to the system. In the future, we will add more robot behaviors to help reduce the gap between two older adults' task performance.

We will also conduct more in-depth analysis of the data, including analysis of experimental videos and looking into each individual's change of engagement and physiological response. We did not examine apathy, our long-term goal for this system. The staff-completed scale on older adults' social interactions and responses to the robot experiments was investigator-developed. In future studies, a validated and reliable tool to examine the effect on apathy will be utilized. Finally, we will continue the development of more robot-mediated activities and testing of SAR systems based on results and knowledge gained from this field study.

Overall, notwithstanding the above-mentioned limitations, the Ro-Tri system was able to provide older adults with technology assisted intervention in a safe and user-friendly manner. It could quantitatively measure the participants' engagement and was effective in engaging older adults in activities and increasing their collaborative social behavior. Seven pairs of participants were involved in the 6 session long study, and their impressions of the system were positive. The modular architecture of the system allows it to be expanded to include more participants with a more complex layered interaction module with various combinations of multimodal physical, cognitive, and social strategies. The current system interaction module uses a Hierarchical State Machine (HSM), but it can be easily changed to incorporate newer machine learning based algorithms with minimal changes to the other components of the system. The physiological indicator of stress could potentially be added to make the system stress-sensitive. We are encouraged by the acceptability and tolerability results, flawless functioning of the Ro-Tri system in real world, and the enthusiasm from the LTC management and staff to include robotic interaction in their workflow. We believe that this work validates the robustness of the Ro-Tri system and provides important field testing results on the potential of robotic intervention of older adults in LTC settings.

Acknowledgements The authors would like to thank all the participants and the staff of Sycamore Terrace and Belmont Village, Nashville TN for their help in conducting the study.

Funding This work was supported by the National Institute of Health Grant 1R21AG050483-01A1.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Federal Interagency Forum on Aging-Related Statistics (2016) Older Americans 2016: key indicators of well-being. Fed Interag Forum Aging-Related Stat
2. Alzheimer's Association (2018) Alzheimer's disease facts and figures. *Alzheimer's Dement* 14:367–429
3. US Department of Health, Human Services (2010) Multiple chronic conditions—a strategic framework: optimum health and quality of life for individuals with multiple chronic conditions. DC US Dep Heal Hum Serv, Washington, p 2
4. Spillman BC (2014) Why do elders receiving informal home care transition to long stay nursing home residency. Rep to Dep Heal Hum Serv Assist Secr Plan Eval Off Disabil Aging Long-Term Care Policy, Contract HHSP23337033T
5. Tierney SM, Woods SP, Weinborn M, Bucks RS (2018) Real-world implications of apathy among older adults: Independent associations with activities of daily living and quality of life. *J Clin Exp Neuropsychol* 40:895–903. <https://doi.org/10.1080/13803395.2018.1444736>
6. Lansdall CJ, Coyle-Gilchrist ITS, Vázquez Rodríguez P et al (2019) Prognostic importance of apathy in syndromes associated with frontotemporal lobar degeneration. *Neurology* 92:E1547–E1557. <https://doi.org/10.1212/WNL.0000000000007249>
7. Volicer L (2018) Behavioral problems and dementia. *Clin Geriatr Med* 34:637–651. <https://doi.org/10.1016/j.cger.2018.06.009>
8. D'Iorio A, Vitale C, Piscopo F et al (2017) Impact of anxiety, apathy and reduced functional autonomy on perceived quality of life in Parkinson's disease. *Parkinsonism Relat Disord* 43:114–117. <https://doi.org/10.1016/j.parkreldis.2017.08.003>
9. Nijsten JMH, Leontjevas R, Smalbrugge M et al (2019) Apathy and health-related quality of life in nursing home residents. *Qual Life Res* 28:751–759. <https://doi.org/10.1007/s11136-018-2041-y>
10. Merrilees J, Dowling GA, Hubbard E et al (2013) Characterization of apathy in persons with frontotemporal dementia and the impact on family caregivers. *Alzheimer Dis Assoc Disord* 27:62–67. <https://doi.org/10.1097/WAD.0b013e3182471c54>
11. Chodzko-Zajko WJ, Proctor DN, Singh MAF et al (2009) Exercise and physical activity for older adults. *Med Sci Sport Exerc* 41:1510–1530
12. Cattan M, White M, Bond J, Learmouth A (2005) Preventing social isolation and loneliness among older people: a systematic review of health promotion interventions. *Age Soc* 25:41–67
13. Aguirre E, Woods RT, Spector A, Orrell M (2013) Cognitive stimulation for dementia: a systematic review of the evidence of effectiveness from randomised controlled trials. *Age Res Rev* 12:253–262
14. Farina N, Rusted J, Tabet N (2014) The effect of exercise interventions on cognitive outcome in Alzheimer's disease: a systematic review. *Int Psychogeriatr* 26:9–18
15. Groot C, Hooghiemstra AM, Raijmakers P et al (2016) The effect of physical activity on cognitive function in patients with dementia: a meta-analysis of randomized control trials. *Age Res Rev* 25:13–23
16. Livingston G, Sommerlad A, Orgeta V et al (2017) Dementia prevention, intervention, and care. *Lancet* 390:2673–2734
17. Anderson-Hanley C, Arciero PJ, Brickman AM et al (2012) Exergaming and older adult cognition: a cluster randomized clinical trial. *Am J Prev Med* 42:109–119
18. Cohen-Mansfield J, Marx MS, Dakheel-Ali M, Thein K (2015) The use and utility of specific nonpharmacological interventions for behavioral symptoms in dementia: an exploratory study. *Am J Geriatr Psych* 23:160–170
19. Long Term Care Coalition Nursing Home Staffing 2018 Q4. <https://nursinghome411.org/nursing-home-staffing-2018-q4/>
20. Hussain A, Rivers PA, Glover SH, Fottler MD (2012) Strategies for dealing with future shortages in the nursing workforce: a review. *Heal Serv Manag Res* 25:41–47
21. Broadbent E, Stafford R, MacDonald B (2009) Acceptance of healthcare robots for the older population: review and future directions. *Int J Soc Robot* 1:319

22. Mois G, Collete BA, Renzi-Hammond LM et al (2020) Understanding robots' potential to facilitate piano cognitive training in older adults with mild cognitive impairment. *ACM/IEEE Int Conf Human-Robot Interact*. <https://doi.org/10.1145/3371382.3378299>
23. Martinez-Martin E, Costa A, Cazorla M (2019) PHAROS 2.0—A PPhysical assistant ROBot system improved. *Sensors (Switzerland)* 19:1–18. <https://doi.org/10.3390/s19204531>
24. Costa A, Martinez-Martin E, Cazorla M, Julian V (2018) PHAROS—PPhysical assistant ROBot system. *Sensors (Switzerland)* 18:1–19. <https://doi.org/10.3390/s18082633>
25. Fan J, Bian D, Zheng Z et al (2017) A robotic coach architecture for elder care (ROCARE) based on multi-user engagement models. *IEEE Trans Neural Syst Rehabil Eng* 25:1153–1163. <https://doi.org/10.1109/TNSRE.2016.2608791>
26. Fan J, Mion LC, Beuscher L, et al (2020) SAR-connect : a socially assistive robotic system to support activity and social engagement of older adults. *IEEE Trans Robot Under Rev (Conditionally accepted : Minor revision)*
27. Cohen-Mansfield J, Hai T, Comishen M (2017) Group engagement in persons with dementia: the concept and its measurement. *Psych Res* 251:237–243. <https://doi.org/10.1016/j.psychres.2017.02.013>
28. Shibata T, Wada K (2011) Robot therapy: a new approach for mental healthcare of the elderly—a mini-review. *Gerontology* 57:378–386
29. Coradeschi S, Cesta A, Cortellessa G, et al (2013) Giraffplus: combining social interaction and long term monitoring for promoting independent living. In: 2013 the 6th international conference on human system interaction (HSI). IEEE, pp 578–585
30. McColl D, Louie WYG, Nejat G (2013) Brian 2.1: a socially assistive robot for the elderly and cognitively impaired. *Robot Autom Mag IEEE* 20:74–83. <https://doi.org/10.1109/mra.2012.2229939>
31. Looije R, Neerinx MA, Cnossen F (2010) Persuasive robotic assistant for health self-management of older adults: design and evaluation of social behaviors. *Int J Hum Comput Stud* 68:386–397
32. Le Maitre J, Chetouani M (2013) Self-talk discrimination in human–robot interaction situations for supporting social awareness. *Int J Soc Robot* 5:277–289
33. Klamer T, Allouch SB (2010) Acceptance and use of a social robot by elderly users in a domestic environment. In: 2010 4th international conference on pervasive computing technologies for healthcare (PervasiveHealth). IEEE, pp 1–8
34. Shibano T, Ho Y, Kono Y, et al (2010) Daily support system for care prevention by using interaction monitoring robot. In: 2010 IEEE/RSJ international conference on ntelligent robots and systems (IROS). pp 3477–3482
35. Kanoh M, Oida Y, Nomura Y et al (2011) Examination of practicability of communication robot-assisted activity program for elderly people. *J Robot Mechatronics* 23:3
36. Matsusaka Y, Fujii H, Okano T, Hara I (2009) Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration. In: The 18th IEEE international symposium on robot and human interactive communication, 2009. RO-MAN 2009. IEEE, pp 472–477
37. Bäck I, Makela K, Kallio J (2013) Robot-guided exercise program for the rehabilitation of older nursing home residents. *Ann Long-Term Care Clin Care Aging* 21:38–41
38. McColl D, Nejat G (2013) Meal-time with a socially assistive robot and older adults at a long-term care facility. *J Hum-Robot Interact* 2:152–171
39. Louie WYG, Vaquero T, Nejat G, Beck JC (2014) An autonomous assistive robot for planning, scheduling and facilitating multi-user activities. In: 2014 IEEE international conference on robotics and automation (ICRA). pp 5292–5298
40. Tapus A, Tapus C, Matarić M (2010) Long term learning and online robot behavior adaptation for individuals with physical and cognitive impairments. In: *Field and service robotics*. Springer, Berlin, Heidelberg, pp 389–398. https://doi.org/10.1007/978-3-642-13408-1_35
41. Fasola J, Matarić MJ (2013) A socially assistive robot exercise coach for the elderly. *J Hum-Robot Interact* 2:3–32
42. Yoshino K, Kouda M, Shanjun Z (2012) Correct motion advice on rehabilitation instruction robot by superimposing instructor CG model. In: 2012 fifth international conference on intelligent networks and intelligent systems (ICINIS). pp 333–336
43. Simonov M, Delconte G (2015) Humanoid assessing rehabilitative exercises. *Methods Inf Med* 54:114–121
44. Görer B, Salah AA, Akin HL (2013) A robotic fitness coach for the elderly. In: *International joint conference on ambient intelligence*. Springer, pp 124–139
45. Gadde P, Kharrazi H, Patel H, MacDorman KF (2011) Toward monitoring and increasing exercise adherence in older adults by robotic intervention: a proof of concept study. *J Robot*. <https://doi.org/10.1155/2011/438514>
46. Cruz J, Marques A, Barbosa A et al (2013) Making sense (s) in dementia: a multisensory and motor-based group activity program. *Am J Alzheimer's Dis Other Dementias* 28:137–146
47. Šabanovic S, Bennett CC, Chang WL, Huber L (2013) PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *IEEE Int Conf Rehabil Robot*. <https://doi.org/10.1109/ICORR.2013.6650427>
48. Short ES, Swift-Spong K, Shim H, et al (2017) Understanding social interactions with socially assistive robotics in intergenerational family groups. In: *RO-MAN 2017—26th IEEE Int Symp Robot Hum Interact Commun* 2017-Janua:236–241. <https://doi.org/https://doi.org/10.1109/ROMAN.2017.8172308>
49. Chang WL, Šabanović S, Huber L (2013) Use of seal-like robot PARO in sensory group therapy for older adults with dementia. *ACM/IEEE Int Conf Hum-Robot Interact*. <https://doi.org/10.1109/HRI.2013.6483521>
50. Amirabdollahian F, Op Den Akker R, Bedaf S, et al (2013) Accompany: Acceptable robotiCs COMPanions for AgeiNG years—multidimensional aspects of human-system interactions. In: 2013 6th Int Conf Hum Syst Interact HSI 2013, pp 570–577. <https://doi.org/https://doi.org/10.1109/HSI.2013.6577882>
51. Hanheide M, Hebesberger D, Krajník T (2017) The when, where, and how: an adaptive robotic info-terminal for care home residents. *ACM/IEEE Int Conf Hum-Robot Interact Part F1271:341–349*. <https://doi.org/10.1145/2909824.3020228>
52. Khosla R, Chu M-T (2013) Embodying care in Matilda: an affective communication robot for emotional wellbeing of older people in Australian residential care facilities. *ACM Trans Manag Inf Syst* 4:18
53. Matsuyama Y, Taniyama H, Fujie S, Kobayashi T (2010) Framework of communication activation robot participating in multiparty conversation. In: *AAAI fall symposium: dialog with robots*
54. Bemelmans R, Gelderblom GJ, Jonker P, De Witte L (2012) Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *J Am Med Dir Assoc* 13(114–120):e1
55. Yu R, Hui E, Lee J et al (2015) Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: study protocol for a randomized controlled trial. *JMIR Res Prot* 4:e4189
56. Takayanagi K, Kirita T, Shibata T (2014) Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot. *PARO Front Aging Neurosci* 6:257

57. Moyle W, Cooke M, Beattie E et al (2013) Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial. *J Gerontol Nurs* 39:46–53
58. Robinson H, Broadbent E, MacDonald B (2015) Group sessions with Paro in a nursing home: structure, observations and interviews. *Australas J Age* 35:106–112
59. Robinson H, MacDonald BA, Kerse N, Broadbent E (2013) Suitability of healthcare robots for a dementia unit and suggested improvements. *J Am Med Dir Assoc* 14:34–40
60. Petersen S, Houston S, Qin H et al (2017) The utilization of robotic pets in dementia care. *J Alzheimer's Dis* 55:569–574
61. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C (2015) Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. *J Am Med Dir Assoc* 16:867–873
62. Li J, Louie W-YG, Mohamed S, et al (2016) A user-study with tangy the bingo facilitating robot and long-term care residents. In: 2016 IEEE international symposium on robotics and intelligent sensors (IRIS). IEEE, pp 109–115
63. Feng Y, Barakova EI, Yu S et al (2020) Effects of the level of interactivity of a social robot and the response of the augmented reality display in contextual interactions of people with dementia. *Sensors (Switzerland)* 20:1–12. <https://doi.org/10.3390/s20133771>
64. Cohen-Mansfield J, Dakheel-Ali M, Marx MS (2009) Engagement in persons with dementia: the concept and its measurement. *Am J Geriatr Psych* 17:299–307. <https://doi.org/10.1097/JGP.0b013e31818f3a52>
65. Gerritsen DL, Steverink N, Frijters DHM et al (2008) A revised index for social engagement for long-term care. *J Gerontol Nurs* 34:40–48. <https://doi.org/10.3928/00989134-20080401-04>
66. Mabire JB, Gay MC, Vrignaud P et al (2016) Social interactions between people with dementia: pilot evaluation of an observational instrument in a nursing home. *Int Psychogeriatr* 28:1005–1015. <https://doi.org/10.1017/S1041610215002483>
67. Perugia G, van Berkel R, Díaz-Boladeras M et al (2018) Understanding engagement in dementia through behavior. The ethographic and Laban-inspired coding system of engagement (ELICSE) and the evidence-based model of engagement-related behavior (EMODEB). *Front Psychol* 9:1–18. <https://doi.org/10.3389/fpsyg.2018.00690>
68. Perugia G, Diaz-Boladeras M, Catala-Mallofre A et al (2020) ENGAGE-DEM: a model of engagement of people with dementia. *IEEE Trans Affect Comput*. <https://doi.org/10.1109/TAFFC.2020.2980275>
69. Beuscher LM, Fan J, Sarkar N et al (2017) Socially assistive robots: measuring older adults' perceptions. *J Gerontol Nurs* 43:35–43
70. Gjoreski M, Luštrek M, Gams M, Gjoreski H (2017) Monitoring stress with a wrist device using context. *J Biomed Inform* 73:159–170
71. MacPhee GJA, Crowther JA, McAlpine CH (1988) A simple screening test for hearing impairment in elderly patients. *Age Age* 17:347–351
72. Strawbridge WJ, Wallhagen MI (2017) Simple tests compare well with a hand-held audiometer for hearing loss screening in primary care. *J Am Geriatr Soc* 65:2282–2284
73. Mansbach WE, Mace RA, Clark KM, Firth IM (2017) Meaningful activity for long-term care residents with dementia: a comparison of activities and raters. *Gerontologist* 57:461–468. <https://doi.org/10.1093/geront/gnv694>
74. Zheng Z, Fu Q, Zhao H et al (2017) Design of an autonomous social orienting training system (ASOTS) for young children with autism. *IEEE Trans Neural Syst Rehabil Eng* 25:668–678. <https://doi.org/10.1109/TNSRE.2016.2598727>
75. van Ravenswaaij-Arts CMA, Kollee LAA, Hopman JCW et al (1993) Heart rate variability. *Ann Intern Med* 118:436–447
76. Cacioppo JT, Tassinary LG, Berntson G (2007) *Handbook of psychophysiology*. Cambridge University Press, Cambridge
77. Lang PJ, Greenwald MK, Bradley MM, Hamm AO (1993) Looking at pictures: affective, facial, visceral, and behavioral reactions. *Psychophysiology* 30:261–273

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.