Socially Assistive Robot Exercise Coach: Motivating Older Adults to Engage in Physical Exercise

Juan Fasola and Maja J. Matarić

Abstract. We present the design, implementation, and user study evaluation of a socially assistive robot (SAR) system designed to engage elderly users in physical exercise aimed at achieving health benefits and improving quality of life. We discuss our design methodology, which incorporates insights from psychology research in the area of intrinsic motivation, and focuses on maintaining engagement through personalized social interaction. We describe two user studies conducted to test our design principles in practice with our system. The first study investigated the role of praise and relational discourse in the exercise system by comparing a relational robot coach to a non-relational robot coach. The second study compared physical vs. virtual embodiment in the task scenario. The results of both studies demonstrate the feasibility and overall effectiveness of the robot exercise system.

Keywords: socially assistive robotics, human-robot interaction, exercise therapy, intrinsic motivation, embodiment, older adults.

1 Introduction

The growing aging population is increasing the demand for healthcare services worldwide. By the year 2050, the number of people over the age of 85 will increase five-fold [1], while the shortfall of nurses and caregivers is already an issue [2]. Regular physical exercise has been shown to be effective at maintaining and improving the overall health of elderly individuals [3]. Social interaction, and specifically high perceived interpersonal social support, has also been shown to have a positive impact on general mental and physical

Juan Fasola · Maja J. Matarić
Interaction Lab, Computer Science Department,
Viterbi School of Engineering, University of Southern California,
Los Angeles, CA, USA
e-mail: {fasola,mataric}@usc.edu

J.P. Desai et al. (Eds.): *Experimental Robotics*, STAR 88, pp. 463–479.
DOI: 10.1007/978-3-319-00065-7_32

wellbeing [4], in addition to reducing the likelihood of depression [5]. Thus, the availability of physical exercise therapy, social interaction, and companionship will be critical for the growing elderly population; socially assistive robotics (SAR) has the potential to help to address this need.

In this paper, we present the approach, design methodology, implementation details, and user study evaluation of a novel socially assistive robot system that aims to motivate and engage elderly users in simple physical exercise. Our SAR system approach incorporates insights from psychology research into intrinsic motivation and contributes clear design principles for SAR-based therapeutic interventions. For system evaluation, we conducted two user studies with older adults with the following aims: 1) to validate the system approach and its effectiveness in gaining user acceptance and motivating physical exercise in older adults; 2) to study the effect of praise and relational discourse in the system towards increasing user motivation; and 3) to investigate the effect of embodiment in the system by comparing user evaluations of similar physically and virtually embodied SAR exercise coaches.

2 Related Work

The literature that addresses assistive robotics intended for and evaluated by the elderly is limited but growing. Representative work includes robots that focus on providing assistance for functional needs, such as mobility and health monitoring [6], navigation and schedule reminders [7], as well as social and emotional needs, such as reducing depression [8] and increasing social interaction [9]. Matsusaka et al. developed an exercise demonstrator robot to aid lead human demonstrators during simple arm exercises to a training group [10]. While similar to our work, this robot was not autonomous and did not have any sensors for which to perceive the users; hence, it did not provide any real-time feedback, active guidance, or personalized training, all of which are employed by our system.

Social agent coaches have previously been developed to autonomously assist individuals in tasks such as physical exercise [11, 12], but have largely been conversational and minimally interactive, without actively monitoring the activity itself. Our system provides a clear distinction in that the agent, a robot in our case, not only provides active feedback and task monitoring, but is also directly responsible for instructing and steering the task as well. Hence, our agent is both an administrator and active participant in the health-related activity.

While previous studies have investigated the positive effect of physical embodiment within the context of human-agent interaction (e.g., [12, 13, 14]), most have recruited a participant pool consisting primarily of young adults. Embodiment studies that have targeted the elderly population include the work of Heerink et al. [15], which investigated the acceptance of assistive social agents by older adults. While similar to our work, the robot used in their evaluation was a table-top robot (the iCat), and was either controlled via a human operator during interaction with elderly users (Wizard of Oz study), or, like their screen agent, interacted with users through a touch-screen interface. Furthermore, the interaction consisted primarily of short, informational or utility interactions (e.g., medication/agenda reminders, weather forecast, companionship), lasting about 5 minutes and often for a single session. In contrast, our SAR system was designed to engage elderly users in fluid, highly interactive exercise sessions, completely autonomously, while providing active feedback, motivation, and guidance on the task, across multiple sessions of interaction, each lasting 10-20 minutes in duration.

3 Robot Exercise Coach

Our approach to designing our SAR system to help address the physical exercise needs of the elderly population was motivated by two basic axioms regarding what a SAR agent must possess, namely: 1) the ability to influence the user's intrinsic motivation to perform the task, and 2) the ability to personalize the social interaction to maintain user engagement and build trust in the task-based human-robot relationship

3.1 Design Principles

In following the above basic axioms, we developed five design principles for the SAR system; all are general and can be applied to any SAR-based therapeutic intervention. The design principles stated that the robot coach should be:

1) *Motivating.* The coaching style and interaction methodology of our SAR exercise system was guided by psychology research in the area of intrinsic motivation, which has been shown to be more effective than extrinsic motivation in achieving long-term user task compliance and behavior change [16]. The motivational techniques utilized by our system were primarily derived from Csikszentmihalyi's theory of flow [17], which asserts that people are intrinsically motivated under conditions of optimal challenge. Towards this end, we focused on providing a variety of challenging exercise games, of varying degrees of difficulty, and alternating the games at a regular pace to prevent user boredom and/or frustrations. We also incorporated indirect competition and user autonomy into the system design, which have also been shown to increase intrinsic motivation [18], by having the robot periodically report the user's high score and by giving the user control over the exercise routine in one of the exercise games.

2) *Fluid and Highly Interactive.* For any task to achieve a state of flow, or maximal enjoyment, in the user, it must establish a clear set of goals, combined with immediate and appropriate feedback [17]. A primary goal of our

(a) SAR exercise interaction (b) Physical robot (c) Virtual robot

Fig. 1 Exercise scenario and robot coach embodiments

coaching approach was to provide a fluid interaction, which required the robot to both perceive the user's activity and provide active feedback and guidance in real-time, all with the aim of maintaining user engagement in the task.

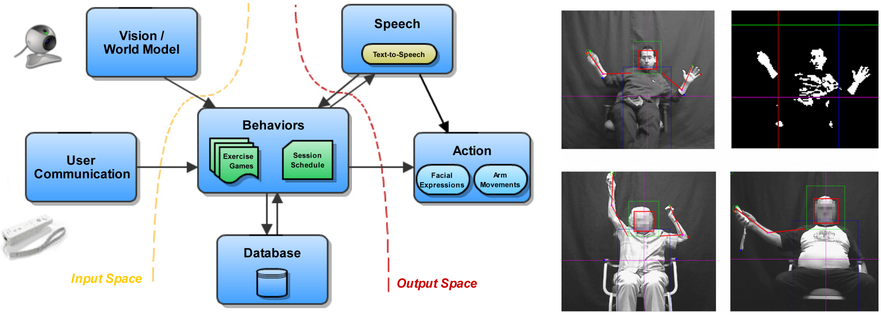
3) *Personable.* The SAR coach employed relationship building characteristics such as praise, empathy, humor, references to mutual knowledge, continuity behaviors, politeness, and trust, among others [11]. Praise, which has been shown to increase intrinsic motivation [19], was given upon successful user completion of exercise gestures. The user's name was also often used to personalize the interaction and to promote the user-robot relationship.

4) *Intelligent.* Trust is a key component to the success of any careprovider-user relationship, and one that is closely linked to the intelligence/helpfulness of the careprovider as perceived by the user [20]. Toward this end, we placed special attention on adding variety to the robot's utterances to minimize any perceived repetitiveness, in addition to accurate monitoring of user activity.

5) *Task-Driven.* In gaining user trust in the system, it is also important that the tasks employed not only be healthcare-driven, but also be successful in achieving the desired therapeutic behavior. In the case of our SAR exercise coach, this means the tasks must elicit consistent physical exercise among the users (measurable through objective quantitative metrics).

3.2 Interaction Scenario

The robot exercise system consists of a socially assistive robot whose purpose is to monitor, instruct, evaluate, and encourage users to perform simple seated physical exercise ("chair aerobics"). The one-on-one interaction scenario consists of the user sitting in a chair across from the robot, with each facing the other. The system's physical robot platform is a 19-DOF humanoid torso with expressive face mounted on a MobileRobots Pioneer base; the virtual robot platform, used in our embodiment comparison user study, is a



(a) System modules (b) Vision output [21]

Fig. 2 SAR system architecure and user activity recognition output

computer simulation of the same robot. The exercise interaction setup and robot platforms are shown in Fig. 1.

Four exercise games are available in our system: the *Workout game* (traditional exercise coach), the *Sequence game* (increased repetitions), the *Imitation game* (user autonomy), and the *Memory game* (cognitive game). In the Workout game, the robot demonstrates the arm exercises with its own arms, and asks the user to imitate. The robot shows only one exercise gesture at a time, and upon successful completion by the user, generates a different gesture, and the process repeats. The robot gives the user feedback in real-time, providing corrections when appropriate (e.g., "Raise your left arm and lower your right arm" or "Bend your left forearm inward a little"), and praise in response to each successful imitation (e.g., "Great job!" or "Now you've got the hang of it."). This game has the fastest pace of all the four exercise games, as the users generally complete the requested gestures quickly. Difficulty increases in the Sequence game, where the robot demonstrates a gesture pair for the user to repeat three times in sequence (i.e., six gestures per sequence). Contrastly in the Imitation game, the robot instead imitates user movements in real-time. Lastly, the goal of the Memory game is for the user to try and memorize an ever-longer sequence of arm gesture poses, and thus compete against his/her own high score.

3.3 System Architecture

The system architecture is comprised of six independent software modules, representing: vision and world model, speech, user communication, behaviors, robot action, and database management. The vision and world model module is responsible for providing information regarding the state of the user to the behavior module for the robot to make task-based decisions during interaction. We developed a novel visual user activity recognition algorithm [21] to recognize user arm poses in real-time, without markers on the user, using only a monocular camera, which can be performed using either color or motion-based segmentation, and is generalizable to other domains. The speech module communicates to the user through a text-to-speech engine, while the user communicates through a Wiimote wireless button interface. The behavior module interfaces with the robot action and database modules, and is responsible for producing all of the coaching behaviors of the system; hence, it is the module that implements most of the motivational techniques outlined in our SAR design principles. A diagram of the system architecture, along with example vision output, is provided in Fig. 2.

4 Study I: Praise and Relational Discourse

We designed and conducted an intrinsic motivation study to investigate the role of praise and relational discourse (politeness, humor, empathy, etc.) in the robot exercise system. Toward that end, the study compared the effectiveness and participant evaluations of two different coaching styles used by our system to motivate elderly users to engage in physical exercise. This section discusses the study methods employed, the dependent measures that were evaluated, and the outcomes of the study with elderly participants.

4.1 Study Design

The study consisted of two conditions, Relational and Non-Relational, to explore the effects of praise and communicative relationship-building techniques on a user's intrinsic motivation to engage in the exercise task with the SAR coach. The study design was within-subject; participants saw both conditions, one after the other, and the order of appearance of the conditions was counter-balanced among the participants. Each condition lasted 10 minutes, totaling 20 minutes of interaction, with surveys being administered after both sessions to capture participant perceptions of each study condition independently. The following describes the two conditions in greater detail:

1) Relational Condition: In this condition the SAR exercise coach employed all of the social interaction and personalization approaches described in Sect. 3. Specifically, the robot always gave the user praise upon correct completion of a given exercise gesture (an example of positive feedback) and provided reassurance in the case of failure (an example of empathy). The robot also displayed continuity behaviors (e.g., by referencing past experiences with the user), humor, and refered to the user by name, all with the purpose of encouraging an increase in the user's intrinsic motivation to engage in the exercise session.

2) Non-relational Condition: In this condition the SAR exercise coach guided the exercise session by providing instructional feedback as needed (e.g., user score, demonstration of gestures, verbal feedback during gesture attempts, etc.), but did not employ explicit relationship building discourse of any kind.

Specifically, the robot did not provide positive feedback (e.g., praise) in the case of successful user completion of an exercise gesture, nor did it demonstrate empathy (e.g., reassurance) in the case of user failure. Furthermore, the SAR coach did not display continuity behaviors, humor, or refer to the user by name.

We recruited elderly individuals to participate in the study through a partnership with be.group, an organization of senior living communities in Southern California, using flyers and word-of-mouth. Thirteen participants (12 female, 1 male) responded and successfully completed both conditions of the study; their ages ranged from 77-92 ($M = 83, S.D. = 5.28$). Half of the participants ($n = 7$) engaged in the Relational condition in the first session, whereas the other half ($n = 6$) engaged first in the Non-Relational condition. The following describes the specific evaluation measures captured in the post-session surveys:

1) Evaluation of Interaction: Two dependent measures were used to evaluate the interaction with the robot exercise system: the *enjoyableness of the interaction* (6 items; Cronbach's $\alpha = .93$), and the perceived *value or usefulness of the interaction* (4 items; Cronbach's $\alpha = .95$). For both measures, participants were asked to rate how well each item described the interaction on a 10-point scale, anchored by "Describes Very Poorly" (1) and "Describes Very Well" (10).

2) Evaluation of Robot: Three dependent measures were used to evaluate the robot coach: as a *companion* (9 items; Cronbach's $\alpha = .86$), as an *exercise coach* (5 items; Cronbach's $\alpha = .88$), and the *social presence* of the robot (7 items; Cronbach's $\alpha = .82$). All items were rated on a 10-point scale.

3) Direct Comparison of Conditions: In addition to the above evaluation measures, at the end of the last exercise session we administered one final survey asking the participants to directly compare the two study conditions (labeled "first" and "second") according to ten evaluation categories.

4.2 Results

Evaluation of Interaction Results. 85% of the participants (11 of 13) rated the Relational condition higher than the Non-Relational condition in terms of enjoyment, and 77% of the participants (10 of 13) rated the Relational condition higher in terms of usefulness than the Non-Relational condition. A Wilcoxon signed-rank test was performed on the data to analyze matched pairs from the sample population's evaluations of both study conditions according to the dependent measures; the results show that the participants evaluated the interaction with the relational robot as significantly more enjoyable/entertaining than the interaction with the non-relational robot ($W[12] = 4, p < .005$), and as somewhat more valuable/useful than the interaction with the non-relational robot, although not to a significant degree ($W[12] = 15.5, p < 0.10$).

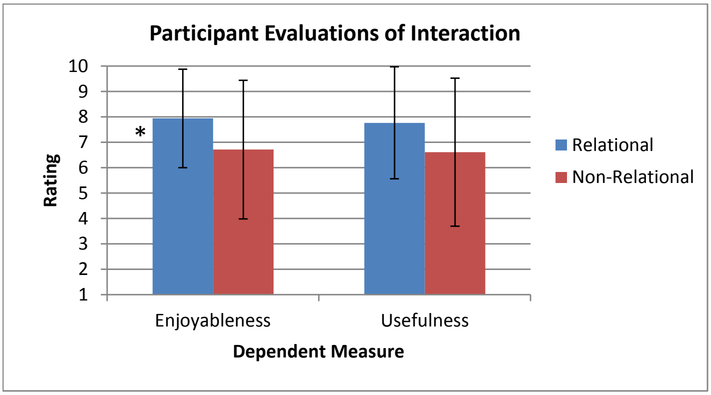
For illustration purposes, Fig. 3(a) shows the average participant ratings of the enjoyableness and usefulness of the interaction for both study conditions.

Evaluation of Robot Results. 77% of the participants (10 of 13) rated the relational robot higher than the non-relational robot in terms of companionship, 77% of the participants (10 of 13) rated the relational robot more positively as an exercise coach, and the comparative ratings of social presence between the robot conditions were approximately equal, as 54% of participants (7 of 13) reported higher social presence for the relational robot. A Wilcoxon signed-rank test was again performed on the data; the results show that the participants rated the relational robot as a significantly better companion than the non-relational robot ($W[13] = 14, p < .05$), and as a significantly better exercise coach than the non-relational robot ($W[11] = 7, p < .02$). There was no significant difference in the participant evaluations of social presence between both robot conditions ($W[12] = 28.5, p > 0.2$), with both robots receiving equally high ratings. The average participant ratings of both robot conditions for all three dependent measures are shown in Fig. 3(b).

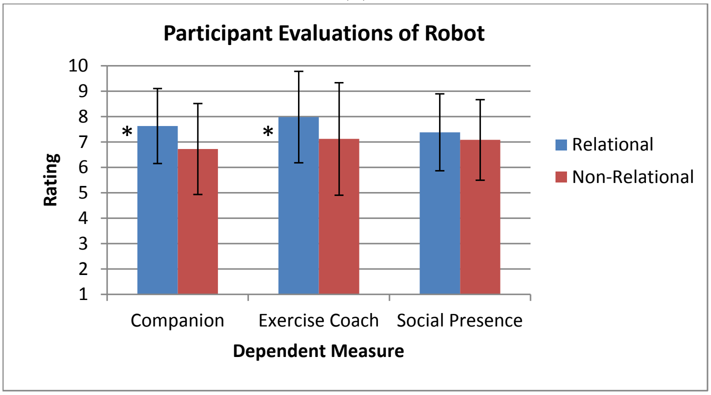
Direct Comparison Results. The direct comparison results demonstrate that, regardless of the order of condition presentation, the participants expressed a strong preference for the relational robot over the non-relational robot. Specifically, the relational robot received 82% of the positive trait votes vs. 16% for the non-relational robot, with the remaining 2% shared equally between them. Notable results include the high number of participants who rated the relational robot as more enjoyable (10 votes, 77%), better at motivating exercise (11 votes, 85%), more useful (11 votes, 85%), and the robot they would choose to exercise with in the future (11 votes, 85%). In contrast, the non-relational robot received a high number of votes for being more frustrating (10 votes, 77%) and more boring (10 votes, 77%) than the relational robot.

4.3 Discussion

The results of the study show a strong user preference for the relational robot over the non-relational robot, demonstrating the positive effects of praise and relational discourse in a healthcare task-oriented human-robot interaction scenario. Comments made by participants after the study further illustrate the positive response to the relational robot, including "It's nice to hear your name, it's personal. I felt more positive reinforcement," and from another participant "The robot encourages you, compliments you; that goes a long way." These results provide significant insight into how people respond to socially assistive robots, and confirm the positive influence that praise and relational discourse have on intrinsic motivation. Furthermore, the results validate our SAR design principles, with particular emphasis on the personable nature of the robot coach, which in turn influences the motivational



(a)



(b)

Fig. 3 Participant evaluations of (a) the interaction and (b) the robot coach, for both study conditions [21]. Note: significant differences are marked by asterisks (*).

capabilities of the system (e.g., by increasing enjoyment) and its perceived usefulness (task-driven), as evidenced by the participant evaluation results. For further analysis and discussion of the study results, the reader is referred to [21].

5 Study II: SAR Evaluation and Embodiment

We designed and conducted a second, larger, user study with older adult participants in order to: 1) evaluate the effectiveness of our SAR approach and system design, and 2) investigate the role of physical embodiment in the robot exercise system. Specifically, the study compared the effectiveness and participant evaluation of our physical humanoid robot to that of a computer simulation of the same robot shown on a flat-panel display.

5.1 Study Design

To analyze the differences between the physical and virtual embodiments in the exercise system, we implemented both a between-subjects and within-subjects study design. Study participants were divided into two groups, physical robot embodiment vs. virtual robot embodiment, and the study consisted of a total of five 20-minute sessions of exercise interaction with the system, conducted over a two-week period. The between-subjects portion of the study constituted the analysis across both conditions of the data pertinent to the first four exercise sessions, where participants in both groups interacted solely with their designated robot embodiment. Distinct from all previous sessions, in the fifth exercise session participants in both groups interacted with the alternative robot embodiment (physical robot group with the virtual robot, virtual robot group with the physical robot). The within-subjects portion of the study consisted of the data analysis between the fourth and fifth exercise sessions within each group.

Surveys were administered at the end of the fourth and fifth sessions for participant evaluation of the SAR system. The same five measures from the first user study were again used to evaluate both the interaction and robot coach, with the addition of three measures: *helpfulness* of the robot (4 items; Cronbach's $\alpha = .96$), *intelligence* of the robot (4 items; Cronbach's $\alpha = .93$), and *social attraction* towards the robot (4 items; 7-point scale; Cronbach's $\alpha = .88$).

To help assess the effectiveness of the SAR exercise system in motivating exercise among the participants, we also collected fifteen different objective measures during the exercise sessions regarding user performance and compliance in the exercise task. Five performance measures were captured during user interaction in the Workout game, including: the *average time to gesture completion* (from the moment the robot demonstrates the gesture, to successful user completion of the gesture), *number of seconds per exercise completed*, *number of failed exercises*, *number of movement prompts* by the robot to the user due to lack of arm movement, and *feedback percentage*. The feedback percentage measure refers to the fraction of gestures, out of the total given, where the robot needed to provide verbal feedback to the user regarding their arm positions in order to help guide them to correct gesture completion. Two measures concerned user activity during the entire exercise session; they were: the *average total number of exercises completed*, and *number of breaks* taken by the user. The eight remaining measures were captured in the other three exercise games, and were similar in nature to those captured in the Workout game.

We recruited thirty-three older adults to participate in the study, again through a partnership with be.group. Half of the participants were placed in the physical robot group ($n = 16$), and the other half in the virtual robot group ($n = 17$). The sample population consisted of 27 female (82%) and 6 male (18%) participants, with ages ranging from 68 to 88 ($M = 76$, $S.D. = 6.32$).

Table 1 Results of between-subjects and within-subjects embodiment comparison

Dependent Measure	Physical Robot	Virtual Robot	P.V.	V.P.
Interaction Evaluation	*Between-Subjects Analysis*		*Within-Subjects Analysis*	
Enjoyable	7.51 (1.77)*	6.00 (2.01)	6.94 (2.21)	7.11 (2.35)†
Valuable/Useful	8.14 (1.66)*	6.19 (2.39)	7.70 (2.13)††	7.51 (2.26)*
Robot Evaluation				
Helpful	8.11 (1.98)*	6.26 (1.98)	8.28 (1.61)	7.44 (2.48)*
Social Attraction	4.70 (1.40)*	3.61 (1.54)	4.31 (1.43)†	4.36 (1.58)††
Social Presence	7.88 (0.94)*	6.47 (2.01)	6.98 (0.97)**	7.22 (1.66)†
Companion	7.48 (2.07)††	6.23 (1.84)	7.12 (1.94)	7.42 (1.87)*
Intelligence	8.17 (2.02)†	6.76 (2.09)	7.61 (1.54)	7.79 (2.66)*
Exercise Partner	7.18 (2.17)††	5.76 (2.18)	6.95 (1.60)	7.01 (2.16)*

P.V. = Physical robot group evaluating the virtual robot (5th session).
V.P. = Virtual robot group evaluating the physical robot (5th session).
$^{\dagger\dagger}p < .10$, $^{\dagger}p < .06$, $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

5.2 Coach Embodiment Comparison Results

A two-tailed independent T-test was performed on the survey data following the fourth exercise session, to compare participant evaluations of the robot embodiments as well as their interactions with them across the two study groups. Survey results from the fourth session were used to perform the comparison analysis as they were less likely to contain scores influenced by the effect of novelty. Table 1 provides the complete set of embodiment comparison results.

The participants evaluated the interaction with the physical robot embodiment as more enjoyable ($t[31] = 2.29, p < .03$) and as more valuable/useful ($t[29] = 2.72, p = .01$) than the interaction with the virtual robot embodiment. In addition, the participants rated the physical robot as significantly more helpful than the virtual robot ($t[31] = 2.66, p = .01$), more socially attractive ($t[30] = 2.09, p < .05$), and as having significantly stronger social presence ($t[23] = 2.59, p < .02$). Evaluations of the remaining measures were also favorable to the physical robot, though not to a significant degree, as the participants rated the physical robot as somewhat more of a companion ($t[30] = 1.81, p < .08$), more intelligent ($t[31] = 1.96, p < .06$), and a moderately better exercise partner ($t[31] = 1.87, p = .07$) than the virtual robot.

To test for significant differences between embodiments within each study group, we used a two-tailed paired T-test to analyze the data gathered from the fourth and fifth session post-session surveys. The within-subjects results largely agree with the results of the between-subjects comparison (see Table 1), with the exception of the physical robot group's ratings of

Table 2 Results of direct comparison survey for all $n = 33$ participants

	Physical	Virtual	Both Equal	Binomial Test
Enjoy More	25 (76%)**	6 (18%)	2 (6%)	$p < .01$
More Intelligent	13 (40%)	6 (18%)	14 (42%)	$p = 0.3$
More Useful	21 (64%)*	7 (21%)	5 (15%)	$p < .05$
Prefer to Exercise with	27 (82%)***	4 (12%)	2 (6%)	$p < .0001$
Better at Motivating	22 (67%)**	4 (12%)	7 (21%)	$p < .01$
More Frustrating	10 (30%)	14 (43%)	9 (27%)	$p = 0.5$
More Boring	4 (12%)	17 (52%)*	12 (36%)	$p < .05$
More Interesting	23 (70%)**	5 (15%)	5 (15%)	$p < .01$
More Entertaining	25 (76%)***	4 (12%)	4 (12%)	$p < .001$
Choice from now on	25 (76%)**	7 (21%)	1 (3%)	$p < .01$

*$p < .05$, **$p < .01$, ***$p < .001$.

the virtual embodiment, which although lower on average than the physical embodiment were not significantly different, indicating possible carryover effects in the evaluation.

As in the first user study, participants were asked to directly compare both embodiments with respect to ten evaluation categories. The results, provided in Table 2, show clear preference for the physical robot, which received 81% of the positive trait votes vs. 19% for the virtual robot among participants who chose one embodiment over the other (85% of the sample); a significant margin as indicated by a two-sided exact binomial test (201 votes vs. 63, $p < .0001$).

5.3 SAR System Evaluation Results

In order to evaluate the effectiveness of the SAR exercise system, we analyzed the data of the physical robot group's fourth session of interaction, together with the data of the virtual robot group's fifth session of interaction. Therefore, the SAR system evaluation results, regarding user perceptions and user exercise performance, were gathered from all 33 older adult participants and consist only of participant interactions with the physically embodied robot, as this condition represented the ideal interaction scenario for users of the system.

User Evaluations of SAR System. To analyze the user evaluations of the SAR exercise system, we performed a two-tailed independent T-test to test for significant differences between the participant ratings of the subjective measures and a neutral evaluation rating. The neutral evaluation rating distribution was obtained from a uniform sampling of the rating scale (integers from 1 to 10) for the approximate number of participants, and has a

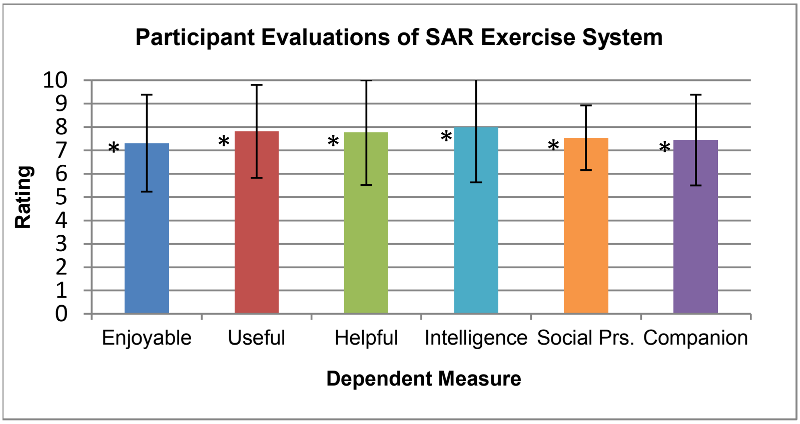


Fig. 4 Participant evaluations of SAR system interaction and robot coach. Note: Significant differences in comparison to neutral rating are marked by asterisks (*).

mean rating of 5.5 ($S.D. = 2.90$). This uniform sampling assumes no prior information regarding user perceptions of the system, and thus is deemed neutral.

The results of the user evaluation of the SAR exercise system were very encouraging, as they showed a notable level of user acceptance of the system, as evidenced by the high ratings across each of the subjective measures, and highlighted the effectiveness of our SAR system design principles. The participants evaluated the interaction with the SAR exercise system as enjoyable ($M = 7.3, S.D. = 2.07$) and valuable/useful ($M = 7.8, S.D. = 1.99$). The ratings for both measures were found to be significantly more positive than a neutral evaluation (enjoyableness: $t[52] = 2.81, p < .01$; usefulness: $t[50] = 3.64, p < .001$). These results illustrate the effectiveness of the system in promoting intrinsic motivation within the users to engage in the healthcare task (intrinsic motivation is characterized by enjoyment [17]), and in guiding the task-driven interaction towards achieving beneficial health outcomes for the user.

Regarding user perceptions of the robot coach of the SAR exercise system, the participants rated the robot highly and significantly more positive than neutral in terms of helpfulness ($M = 7.8, S.D. = 2.25$; $t[54] = 3.42, p < .01$), intelligence ($M = 8.0, S.D. = 2.35$; $t[56] = 3.68, p < .001$), social presence ($M = 7.5, S.D. = 1.38$; $t[41] = 3.48, p < .01$), and as a companion ($M = 7.4, S.D. = 1.94$; $t[50] = 3.08, p < .01$). The participants also rated the robot coach favorably in terms of social attraction ($M = 4.5$ (on a 7-point scale), $S.D. = 1.48$; $t[62] = 1.22, n.s.$), and as an exercise partner ($M = 7.1, S.D. = 2.14$; $t[53] = 2.45, p < .02$). These results illustrate the personable nature and intelligence of the robot coach, as perceived by the participants, both of which aid in the development of trust within the human-robot relationship, and

Table 3 User exercise performance statistics for all $n = 33$ older adult participants

Performance Measure	Mean (std.)
Workout game:	
Time to Gesture Completion (seconds)	2.54 (0.80)
Seconds per Exercise	5.37 (0.88)
Feedback Percentage	7.4% (4.8%)
Number of Failed Gestures	0
Number of Movement Prompts$_W$	0
Sequence game:	
Time to Gesture Completion (seconds)	5.73 (1.37)
Number of Sequences Completed	4.97 (1.16)
Number of Gesture Pairs Completed	14.9 (3.41)
Feedback Percentage	19.6% (11.2%)
Memory game:	
Maximum Score	6
Average Maximum Score	3.52 (1.25)
Time per Gesture Attempt (seconds)	7.62 (3.98)
Imitation game:	
Number of Movement Prompts$_I$	0.37 (0.63)
Entire Session:	
Total Number of Exercises Completed	107.75 (18.1)
Number of Breaks Taken	1 (1.26)

were design goals of our SAR system approach towards providing successful therapeutic interventions. A plot showing participant evaluations of the SAR system interaction and robot coach is shown in Fig. 4.

User Exercise Performance Statistics. The collected statistics regarding participant performance in the exercise task were also very encouraging, as they demonstrated a consistently high level of user exercise performance and compliance within the exercise task.

User compliance and performance in the Workout game were high. The average gesture completion time was 2.54 seconds ($S.D. = 0.80$), and the overall exercise performance averaged 5.37 seconds per exercise ($S.D. = 0.88$), which also includes time taken for verbal praise, feedback, and score reporting from the robot. The low percentage of necessary corrective feedback, averaging 7.4%, combined with zero failures, and zero movement prompts during the interaction sessions, are all very encouraging results, as they suggest that the participants were motivated to do well on the exercises consistently throughout the interaction.

A summary of all statistics regarding user performance, including those from the Sequence, Memory, and Imitation games, can be found in Table 3.

The consistently high level of exercise performance achieved by the study participants, as evidenced by the results, validates the effectiveness of the SAR exercise system approach and design methodology in motivating elderly users to engage in physical exercise, and demonstrates the potential of the technology to provide guided care and to help elderly users achieve beneficial health outcomes.

5.4 Study Expansion with Young Adults

To analyze and compare user evaluations and embodiment effects across age groups, we expanded the study to include 33 young adult participants (6 female, 27 male), yielding a combined sample of 66 participants, all of whom engaged in five sessions of interaction with our SAR exercise system (330 sessions total). The results of the study with young adults were largely consistent with those observed with the older adult participants. Among the combined results, a two-sided exact binomial test showed the physical robot coach received significantly more positive votes than the virtual robot coach upon direct comparison (425 votes vs. 103, $p < .0001$). For further discussion of the results of both the young adult and combined populations, we refer the reader to [22].

6 Conclusions

We have presented a set of design principles for socially assistive robots in therapeutic contexts, and a novel robot system that embodies and validates those principles, designed to motivate and engage elderly users in physical exercise. Our SAR system approach, design methodology, and implementation details were discussed, including five SAR design principles which can be applied to a variety of human-robot interaction-based healthcare interventions.

Results of the user evaluation of the SAR exercise system in two user studies with older adults were presented, which together showed strong participant preferences for relational and physically embodied coaches. The successful acceptance of the SAR exercise system from elderly users, as evidenced by the high participant evaluations of the system and consistent exercise performance in both user studies, validates our SAR system approach, design, algorithms, and effectiveness, and illustrates the potential of SAR technology to help older adults achieve beneficial health outcomes and improve quality of life.

Acknowledgements. We are grateful to be.group, their staff and participating residents, and to Aras Akbari and Farrokh Langroodi for the study data collection. This research was supported by a grant from the Robert Wood Johnson Foundation's Pioneer Portfolio through its national program, Health Games Research, and National Science Foundation grants IIS-0713697, CNS-0709296, and IIS-1117279.

References

1. United States Census Bureau: U.S. Census Bureau Report (Issue Brief No. CB09-97) (2009), http://www.census.gov/population/international
2. Buerhaus, P.: Current and future state of the US nursing workforce. J. Amer. Med. Assoc. 300(20), 2422–2424 (2008)
3. Dawe, D., Moore-Orr, R.: Low-intensity, range-of-motion exercise: invaluable nursing care for elderly patients. J. Adv. Nursing 21, 675–681 (1995)
4. Moak, Z.B., Agrawal, A.: The association between perceived interpersonal social support and physical and mental health: results from the national epidemiological survey on alcohol and related conditions. J. Public Health 32, 191–201 (2010)
5. George, L., Blazer, D., Hughes, D., Fowler, N.: Social support and the outcome of major depression. British J. Psychiatry 154, 478–485 (1989)
6. Dubowsky, S., Genot, F., Godding, S., Kozono, H., Skwersky, A., Yu, H., Shen Yu, L.: PAMM - a robotic aid to the elderly for mobility assistance and monitoring: a helping-hand for the elderly. In: Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), vol. 1, pp. 570–576 (2000)
7. Montemerlo, M., Pineau, J., Roy, N., Thrun, S., Verma, V.: Experiences with a mobile robotic guide for the elderly. In: Proc. AAAI Nat. Conf. on Artificial Intelligence, pp. 587–592 (2002)
8. Wada, K., Shibata, T., Saito, T., Tanie, K.: Analysis of factors that bring mental effects to elderly people in robot assisted activity. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), vol. 2, pp. 1152–1157 (2002)
9. Kidd, C., Taggart, W., Turkle, S.: A sociable robot to encourage social interaction among the elderly. In: Proc. IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 3972–3976 (2006)
10. Matsusaka, Y., Fujii, H., Okano, T., Hara, I.: Health exercise demonstration robot TAIZO and effects of using voice command in robot-human collaborative demonstration. In: Proc. IEEE Int. Symp. on Robot and Human Interactive Communication (RO-MAN), pp. 472–477 (2009)
11. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. Trans. Comput.-Hum. Interact. 12(2), 293–327 (2005)
12. Kidd, C.D., Breazeal, C.: Robots at home: understanding long-term human-robot interaction. In: Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS), pp. 3230–3235 (2008)
13. Powers, A., Kiesler, S., Fussell, S., Torrey, C.: Comparing a computer agent with a humanoid robot. In: Proc. ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI), pp. 145–152 (2007)
14. Jung, Y., Lee, K.M.: Effects of physical embodiment on social presence of social robots. In: Proceedings of Presence, pp. 80–87 (2004)
15. Heerink, M., Kröse, B., Evers, V., Wielinga, B.: Assessing acceptance of assistive social agent technology by older adults: the almere model. Int. J. Social Robotics 2(4), 361–375 (2010)
16. Dienstbier, R.A., Leak, G.K.: Effects of monetary reward on maintenance of weight loss: an extension of the overjustification effect. Am. Psychol. Assoc. Conv. (1976)
17. Csikszentmihalyi, M.: Beyond boredom and anxiety. Jossey-Bass, San Francisco (1975)

18. Fisher, C.D.: The effects of personal control, competence, and extrinsic reward systems on intrinsic motivation. Organizational Behavior and Human Performance 21, 273–288 (1978)
19. Vallerand, R.J., Reid, G.: On the causal effects of perceived competence on intrinsic motivation: a test of cognitive evaluation theory. J. Sport Psychol. 6, 94–102 (1984)
20. Tway, D.C.: A construct of trust. Dissertation, University of Texas at Austin (1994)
21. Fasola, J., Matarić, M.J.: Using socially assistive human-robot interaction to motivate physical exercise for older adults. In: Kanade, T. (ed.) Proc. of the IEEE, Special Issue on Quality of Life Technology (2012)
22. Fasola, J., Matarić, M.J.: Comparing physical and virtual embodiment in a socially assistive robot exercise coach for the elderly. University of Southern California CRES Technical Report, CRES-11-003 (2011)