Comparing Rehabilitation Interactions Using Social Robot Augmented Telepresence to Classical Telepresence

MICHAEL J SOBREPERA, University of Pennsylvania, USA
AJAY ANAND, University of Pennsylvania, USA
AUTHORS TO ADD: LAURA PROSSER, SALLY EVANS, ANH NGUYEN, JIAYU LI, JESS WHITTINGTON, KYLA MADDEN
MICHELLE J JOHNSON, PHD, University of Pennsylvania, USA

A clear and well-documented LaTeX document is presented as an article formatted for publication by ACM in a conference proceedings or journal publication. Based on the "acmart" document class, this article presents and explains many of the common variations, as well as many of the formatting elements an author may use in the preparation of the documentation of their work.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Michael J Sobrepera, Ajay Anand, Authors to add: Laura Prosser, Sally Evans, Anh Nguyen, Jiayu Li, Jess Whittington, Kyla Madden, and Michelle J Johnson, PhD. 2018. Comparing Rehabilitation Interactions Using Social Robot Augmented Telepresence to Classical Telepresence. 1, 1 (May 2018), 31 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The COVID-19 pandemic has aggravated the existing shortage of neuromotor rehabilitation clinicians and raises questions about how to deliver high quality, standardized, and effective rehabilitation care. Growing trend of shifting from traditional in-person to telepresence rehabilitation, i.e. telerehabilitation, a subcategory of telehealth, the remote delivery of rehabilitation using available telecommunication technology, has provided a great alternative to alleviate the situation that has the potential to both enlarging patients coverage and providing standardized and reliable rehab care to patients that require frequent rehabilitation care. However, telerehabilitation is not capable of providing parallel rehabilitation experience compared to classical in-person rehabilitation due to the limitation of interactions, video and audio only. Augmented rehabilitation with the addition of social rehabilitation robots may have the potential to negate the negative effects in telepresence rehabilitation. The addition of the social rehabilitation robot enables more physical contact with patients and previous work has shown that therapists are optimistic about the addition of social robots

Authors' addresses: Michael J Sobrepera, mjsobrep@seas.upenn.edu, University of Pennsylvania, 1800 Lombard St, Philadelphia, Pennsylvania, USA, 19146; Ajay Anand, University of Pennsylvania, 1800 Lombard St, Philadelphia, Pennsylvania, USA, 19146, anandaj@seas.upenn.edu; Authors to add: Laura Prosser, Sally Evans, Anh Nguyen, Jiayu Li, Jess Whittington, Kyla Madden; Michelle J Johnson, PhD, University of Pennsylvania, 1800 Lombard St, Philadelphia, Pennsylvania, USA, 19146, johnmic@pennmedicine.upenn.edu.

in telepresence scenarios with increase in patient motivation and compliance compared to traditional telepresence rehabilitation.

Patients feedback after interacting with such a system is critical for future development that promotes better care quality and efficacy. In this paper, evaluations and analysis of the effects and outcomes of adding social robot augmented telepresence (SRAT), adding robots as an agent to assist the rehabilitation, has been compared with classical in-person rehabilitation and classical telepresence in experiment.

1.1 Need for New Ways to Provide Care

1.2 Social Robots

1.3 Telepresence

1.4 SRAT - Flo

1.5 Hypotheses

Given evidence from previous work, we believed that social robot augmented telepresence (SRAT) would outperform classical telepresence (CT) as a medium for completing rehab tasks (**H1**). We expected this to manifest across multiple domains: people would find interactions via SRAT more enjoyable (**H1e**), of higher value (**H1v**), and to cause less pressure (**H1p**) than interactions via CT. We expected these features to lead to higher compliance during natural interactions, although not necessarily during a study trial. Further, we expected people would feel more competent interacting via SRAT (**H1c**) and would experience lower task load when completing activities via SRAT (**H1l**) than CT. More generally, we expected subjects to prefer SRAT over CT when doing rehab activities (**H2**). In general, we expected that age, motor function, and cognitive function would affect these outcomes.

1.6 Goals

The primary goals of this study were to (1) demonstrate the feasibility of interactions via SRAT with a wide diversity of subjects, as determined by their ability to complete activities, their perception of safety, and their ability to understand instructions, (2) determine if interaction quality via SRAT is better than via CT, (3) determine if people prefer SRAT over CT, (4) determine if and how age, level of motor function, and level of cognitive function affect perceptions of interaction quality between SRAT or CT, (5) determine if and how age, level of motor function, and level of cognitive function affect preference for SRAT and CT.

2 METHODS

Subjects completed a series of clinical assessments followed by three interactions using different modalities of interaction: face to face, classical telepresence, and social robot augmented telepresence. Their reaction to the interactions was recorded throughout by survey instruments. The University of Pennsylvania (Penn) Institutional Review Board and the Children's Hospital of Philadelphia (CHOP) Office of Research Compliance approved this study.

2.1 Recruitment

Subjects 4 years old and older with or without upper extremity impairment were recruited via relationships with clinicians, public flyers, the Penn and CHOP clinical trial subject databases, subject mailing lists held by the Penn Rehab Robotics Lab, and from a pool of inpatient patients at CHOP in the Division of Rehabilitation Medicine. The participating subjects form a sample of convenience.

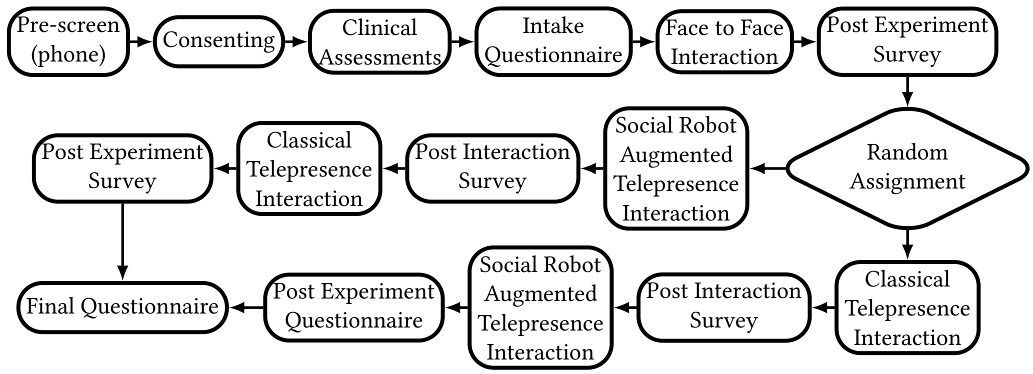


Fig. 1. Experiment flow

2.2 Experiment

Trials took place in one of two environments: (1) a simulated clinical environment (the Penn Rehab Robotics Lab, the Penn Gait Lab, the CHOP Neuromotor Performance Lab, or the Penn Clinical Simulator). (2) a clinical environment (the teen activity room in the CHOP Division of Rehabilitation Medicine). Each trial was conducted by a minimum of two researchers: (1) an operator who delivered the rehab activities and operated the robots and (2) an interviewer who administered the subject surveys. For some trials, additional observers were also present during the study. The flow of the experiment can be seen in fig. 1. For subjects who completed the trial in the clinical setting, the trial occurred over two days, approximately one hour each day. In the first day, the subjects completed the clinical assessments, intake survey, FTF interaction, and first post experiment survey. On the second day, they completed the SRAT and CT interactions, the remaining post experiment survey and the final survey. For subjects who completed the trial in the simulated environment, the entire trial was completed in one day, over approximately two hours.

2.2.1 Pre-Trial. Prior to beginning the trial, a phone pre-screen was performed with all subjects who completed the trial in the simulated clinical environment. Subjects were scheduled for an experiment slot and sent a copy of the consent form. For subjects in the clinical environment, a clinician on the care team completed the pre-screen. The contents of the pre-screen form can be seen in table 2. Consent was gained from all subjects prior to beginning the study.

2.2.2 Clinical Assessments. After consenting, subjects were assessed using the Box and Block Test [7], Color Trails Test 1 and 2 [11, 12], and grip strength test [10] (fig. 2). The Box and Block test measures the subjects' unilateral gross manual dexterity. The Color Trails Test measures executive function and sustained attention. The grip strength test measures hand and forearm strength as a proxy for upper limb strength.

2.2.3 Intake Survey. Prior to beginning the trial interactions, after the clinical assessments, an intake survey was administered to each of the subjects to determine their baseline affect using the Self-Assessment Manikin (SAM) [3], experience with technology, experience with therapy, feelings towards robots, feelings towards telehealth, and demographic information (table 3). For subjects in the real clinical setting, who completed the trial over two days, the SAM was administered again at the beginning of the second day to determine their affect on that day.

2.2.4 Rehab Interactions. Three methods of performing rehab interactions were tested during the trial:

Fig. 2. Subjects performing the clinical assessments at the beginning of the trial. From left to right: the Box and Block test, Color Trails Test, and grip strength test. Subjects shown provided release to publish images of them.

Fig. 3. Subjects participating in the face to face (FTF) interaction, Simon says on the left and target touch on the right. Subjects shown provided release to publish images of them.

Fig. 4. Subjects participating in the classical telepresence (CT) interaction, Simon says on the left and target touch on the right. Subjects shown provided release to publish images of them.

FTF Face to face interaction, where the operator is present in the testing environment with the subject, interacting directly with them (fig. 3).

CT Classical telepresence interaction, where the operator and the subject interact via audio and video (fig. 4).

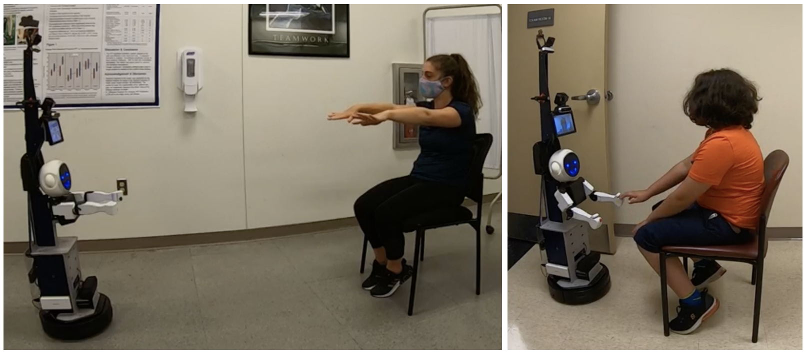


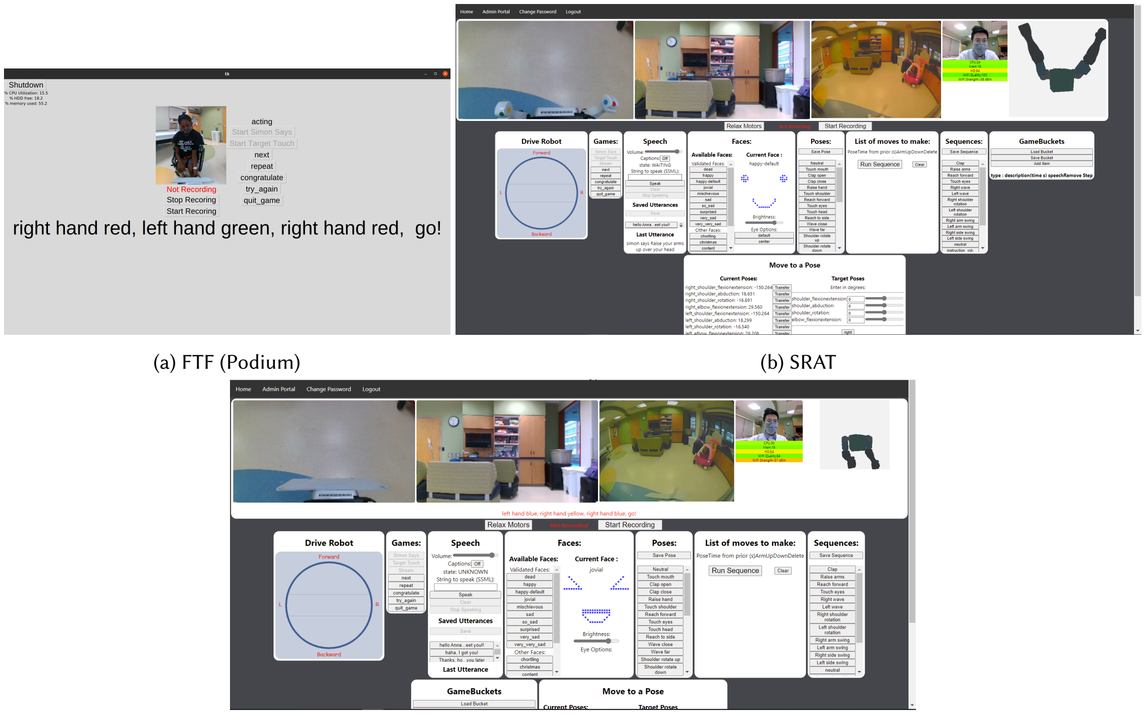
Fig. 5. Subjects participating in the social robot augmented telepresence (SRAT) interaction, Simon says on the left and target touch on the right. Subjects shown provided release to publish images of them.

(a) FTF Configuration (b) SRAT Configuration (c) CT Configuration

Fig. 6. Different configurations for the Flo telepresence platform used throughout the study.

SRAT Social robot augmented telepresence interaction, where the subject is introduced to a humanoid robot mounted on the telepresence system by the operator who is virtually present using audio and video. The interaction is facilitated by the humanoid robot with the operator interjecting, when necessary, as a secondary facilitator (fig. 5).

For all subjects, the FTF interaction was completed first. Prior work has suggested that an initial interaction in-person, prior to interactions via telepresence/with robots creates better engagement and understanding of activities to be completed [5]. It is our expectation that in practice any patient would first be treated in-person, prior to transitioning to a blended remote/face to face therapy regimen. The order of the remaining two interactions (CT, SRAT) was randomly determined to create a balanced study by age and impairment. Randomization was done using stratified permuted block randomization in blocks of four subjects with strata for the cross of age (4-10, 11-17, 18-49, 50+), motor impairment (impaired, not impaired), and cognitive impairment (impaired, not impaired). At least 10 minutes was allowed to elapse after the conclusion of each interaction prior to beginning the next interaction.



(a) FTF (Podium)

(b) SRAT

(c) CT

Fig. 7. Interfaces which the operator used for each condition. In the face-to-face (FTF) condition, the interface was present on a screen on the podium. In the classic telepresence (CT) and social robot augmented telepresence (SRAT) conditions, the interface was viewed on a web browser. In the CT condition, instructions for the operator to say were shown in red under the video feeds.

For the face-to-face interactions, the operator used a podium, consisting of a wooden lectern with 2 Intel RealSense cameras and 1 GoPro camera mounted on a vertical arm, a screen present on the surface, and a target touch board with four colored dots mounted to the front (fig. 6a). Software on the podium instructed the operator on what to say and do to control the flow of the experiment (fig. 7a). The podium was painted using the same color scheme as the Flo robot.

In the social robot augmented telepresence condition, the full Flo system, with the humanoid mounted on the mobile telepresence platform was used (fig. 6b). The operator remained present through the screen on the telepresence system and controlled the robot through a remote web interface (fig. 7b). The humanoid had colored dots on its hands which it could move to place at the same point in space as those on the podium's target touch board, relative to the ground and cameras.

For the classical telepresence condition, the Flo robotic system, using only the telepresence portion of the platform, without the humanoid mounted on it, was used (fig. 6c). Instead of the humanoid, a target touch board with four dots was mounted on the platform. The dots on this target touch board match the position of the dots on the podium relative to the ground and the cameras. The system was operated by the same web interface as used for the full Flo system with the humanoid with the addition of instructions printed on the web interface to tell the operator what to say (fig. 7c).

By using the same color scheme, similar design profiles, and the same positioning of targets for activities, the mechanical aspects of the study are well controlled. By further using software to control how the operator delivered the activities, the core of the interactions was also controlled. The difference between modalities was the physical presence of the operator in the FTF condition, the physical presence of the humanoid robot in the SRAT condition, and no physical presence in the CT condition. In the FTF and CT condition, the operator was able to have natural language interactions with the subject. In the SRAT condition, the robot led the conversation, with the operator still providing commentary and clarification where needed. This represents one point on the spectrum of how SRAT can be used, the operator could provide more or less remote human interaction.

In the CT and SRAT conditions, the operator was in a different room from the subject. The surveyor remained in the room with the subject. Throughout the experiments, the subject was recorded by the Intel RealSense cameras mounted on the podium and robot, by a GoPro camera on the podium/robot, and by a GoPro camera located elsewhere in the room which provided a third person perspective.

Activities: Each interaction was comprised of two different activities, a Simon says game and then a target touch activity. In the FTF and CT conditions, the operator acted as the facilitator. In the SRAT condition, the humanoid robot acted as the facilitator.

In the Simon says game, the facilitator of the trial first provided instructions for the activity:

> In Simon says, I will tell you something to do and show you how to do it, mirrored. If I say Simon says, you should do it with me. If I do not say Simon says, you should not do the action. Watch out, I may try to trick you. After every movement, return to a ready position.

The facilitator then gave a command and demonstrated a task (mirrored) for the subject to do. If the facilitator prefixed the command with "Simon says", the subject had to repeat the demonstrated task, otherwise the subject had to remain in the neutral position. After each action, the subject returned to a neutral position in readiness for the next command. The subject was asked to repeat tasks if the operator observed that the subject did not complete them. The tasks given to the subject were randomized from a bucket of tasks which are designed to test the range of motion of the elbow and shoulder. Some tasks are naturally bimanual, such as clapping. Others are unimanual and were randomly combined to create bi-manual tasks, such as reaching to the side with the left arm and touching the left shoulder with the right hand. The use of composed bi-manual tasks made the activities more interesting and made each repetition of the game unique, preventing memorization. The complete menu of motions used can be seen in table 6. The Simon says activity primarily measures range of motion and reachable workspace, specifically in the context of motions which are relevant to activities of daily living. The bi-manual component also adds some cognitive challenge.

In the target touch activity, the facilitator first provided instructions for the activity:

> In the target touch activity, I will tell you to touch the dots on my hands [board for CT and FTF]. I will tell you which hand to use and which color dot to touch, then tell you to go. No tricks here, just good work!! Let's start in a ready position, return to this position after every touch.

The facilitator then listed a randomized series of colors and corresponding hands to be used to touch the colored dots. In the CT and FTF conditions, the dots were four colored dots placed near the corners of a board. In the SRAT condition, the same colored dots were on the humanoid robot's hands, which the humanoid moved into the same positions as the dots on the target touch boards in the other conditions. The series could range in length from 1–4 touches. Each sequence used no more than one dot from the left side of the touch space and one from the right. The facilitator could repeat the sequence if the subject carried out the sequence incorrectly or incompletely or asked for repetition. The target touch activity is designed to test motor performance, executive function, and short-term memory.

After the conclusion of each modality, the subject was presented with a survey (table 4). To determine the cognitive and physical load placed on the subjects while completing the trial, questions from the NASA Task Load Index (TLX) [6] were used. To determine the subject's level of pressure during the interaction, value attributed to the interaction, competence in completing the activities, and enjoyment during the interaction, scales from the Intrinsic Motivation Inventory (IMI) [14] were used. Questions within the IMI scales were selected based on experience in a pilot trial. Based on results from a pilot trial, the standard IMI scale of seven levels was condensed to five levels to lessen confusion in the target populations. To measure additional relevant constructs, a custom question on each of understanding of instructions, desire to repeat the interaction, safety during the interaction, and enjoyment during the interaction were asked in the style of the TLX and a question on whether the interaction was an effective method of doing rehab was asked in the style of the IMI. Subjects were also asked if they had any additional comments.

2.2.5 Post-Trial. After completing all of the interactions, the subjects were presented with a post-trial survey (table 5). To understand modality preference, the first question asked subjects to rank the three modalities in order of preference. Questions on the subjects' perceptions of CT vs SRAT for the key features of communication, motivation, compliance, and adherence were asked. Subjects were also asked where they think that Flo could be deployed and what other activities could be done with it. To understand whether subjects liked the humanoid robot, the Godspeed Questionnaire number 3, likability [1] was used. Finally, the subject was asked if they had any prior knowledge of Flo and if they had any final comments.

2.3 Data Analysis

All data analysis was completed using R[13]. Data manipulation was done with dplyr and plots were created using the ggplot package from the tidyverse meta-package [15].

2.3.1 Demographics. We first report on the subjects' demographics, experience with various types of technology, and experience with rehabilitation. For reporting, subjects are grouped into four age categories: young children (4 years old to the end of the 11th year), teens and young adults (12 years old to the end of the 20th year), adults (21 years old to the end of the 64th year), and older adults (65 years and older). Sources of impairment were classified into no injury, brain injuries (stroke, TBI, CP), peripheral injuries (SCI, amputation, rotator cuff injury, motor development delay), neurodegenerative diseases (Multiple Sclerosis, Parkinson's), and psychological disorders (autism, conversion disorder).

To determine measured level of impairment across the study sample, the Box and Block Test and Color Trails Test 2 were used. The number of blocks moved in the Box and Block test were age and arm normalized to produce z-scores using healthy population norms from Mathiowetz et al. and Jongbloed-Pereboom et al. [7–9]. The highest z-score for each arm was taken and the arm with the lower score was used for further analysis. Those with weak arm z-scores greater than -1 were determined to have normal gross manual upper extremity dexterity (taken as a proxy for general upper extremity function), those with z-scores of -1 to -2 were taken to have mild impairment, -2 to -3 were taken to have moderate impairment, and less than -3 were taken to have severe impairment.

Similarly, scores from the (Children's) Color Trails Test 2 were normalized by age and, for adults, education per the test manual [11, 12] to generate z-scores. For subjects who fell below the published norms for the Color Trails Tests ($z < -3$), z-scores were interpolated using the lowest three published z-scores to a minimum z-score of -5. For subjects who were within the administrable age range of the test, but could not complete the Color Trails Test, a z-score of -5 was recorded. Levels of impairment cutoffs were simplified from the Color Trails Test manual to: greater than -1 as normal, -2 to -1 as mild impairment, -3 to -2 as moderate impairment, and less than -3 as severely impaired.

For reporting, mildly impaired and unimpaired subjects (z >= -2) are grouped together and moderately and severely impaired subjects (z < -2) are grouped together.

2.3.2 Selecting Relevant Factors to Explore. The central question of this work was whether SRAT would perform better than CT, so the effect of interaction modality was the primary factor which was explored. Additionally, one should expect that the robot operator and the order in which experiments occur could impact results, these are not things we seek to know/quantify, but factors which had to be controlled. We hypothesized that age, motor function, and cognitive function would affect how subjects experience the various interaction modalities.

Of course, other factors, like the affect of subjects, prior telepresence experience, feelings towards robots, etc. could also (probably do) impact results. However, these are not the kinds of things that, as roboticists, we can design around. We cannot make robots and specify that they are only supposed to be used with happy subjects who are alert. We design robots that should interact with people in the populations who need them, as they are. Those populations can, and should, be broken up by age and function. It is very reasonable to design a robot differently for adults vs children or people with high cognitive function vs low cognitive function. By understanding if and how these factors interact with the modalities which we are testing, we can provide design direction for future development and identify the highest value opportunities for social robot augmented telepresence.

2.3.3 Task Load. After the completion of each interaction, questions from the NASA TLX were asked about the just completed interaction. As is typical, we did not perform the optional weighting step as part of the TLX administration. Therefore, we simply averaged the scores for each subject, across the TLX, to generate an aggregate score for task load. To understand how these ratings varied between interaction modalities and how relevant factors (age, motor function, cognitive function) affected ratings, a linear mixed model was used. The model was created using the lme4 package [2] with the equation:

```
TLX~interaction.modality*((Age*BBT*CTT2)+experimental.order+robot.operator)+(1|subject)
```

Which models the task load (average of TLX ratings) as a linear equation on age, Box and Block Test, and Color Trails Test, crossed, along with experimental order and robot operator, all crossed with interaction modality. The subject IDs were treated as a random variable to accommodate for the interactions across the three different modalities. It would make sense for robot operator to be treated as a random effect, since we have clearly not exhaustively sampled the population of possible robot operators and we do not care to understand how different robot operators perform. However, we only have three robot operators, which is too few to use a random effect. The linear model was checked for the assumption of normality of residuals using QQ-plots of the residuals and the random effects and direct visualization of the residual density plot and for Homoscedasticity with a plot of the residuals against the fitted values.

To understand which terms in the linear model were important, an ANOVA with type III Wald chi-square tests was used (using the car package [4]). The significant terms were then visualized to understand how the factors interact to affect task load. Given the sample size, it was not appropriate to run any form of post-hoc analysis. Exploration of the model provides some general intuition as to how different people might interact with social robot augmented telepresence compared to classical telepresence. These should not be taken as final statistical fact, but rather should serve to provoke thought on how different people interact with robots and via telepresence and should be taken as an invitation for further exploration.

2.3.4 Competence, Enjoyment, Pressure, and Value. For the IMI, as suggested by the tool designers [14], we performed a factor analysis to ensure that individual questions were well aligned with their scales in our sample. One question ("The activities did not hold my attention at all") had to be dropped from our usage of the IMI because it did not sufficiently load onto its assigned scale (0.27). All other items showed reasonable (>0.57) loading on their respective scales. The questions assigned to each scale were averaged to generate scale scores for competence, enjoyment, pressure, and value.

To understand how competence, enjoyment, pressure, and value were different among the three interaction modalities and how those differences manifested in relation to the important factors of age, cognitive function, and motor function, the same method as was used for the TLX was used. A single model was constructed independently for each of the four domains: competence, enjoyment, pressure, and value.

2.3.5 Safety, Understanding, And Desire to Repeat. Three questions were used, in the style of the TLX questions, to determine whether subjects felt safe, whether they understood what they were supposed to do, and whether they would want to repeat any of the interactions. Because these questions were asked individually, and not part of larger scales, no analysis was done on them, however the responses are reported using descriptive methods.

2.3.6 Effective Method of Doing Rehab. The IMI asks subjects about how much they value the interactions which they have completed. However, to get to the core of whether the system is valued by people for rehab, a more direct question was asked after the completion of each interaction "This was an effective method of doing rehab" with answer options mirroring those in the IMI. This question is not part of a scale, and so the results are only reported with descriptive methods.

2.3.7 Reported Modality Preference. To understand whether subjects preferred CT or SRAT, they were asked to rank order the interaction modalities by preference. This led to six possible orders, which were simplified to the binary of SRAT better than CT or CT better than SRAT, dropping the FTF condition. Preference counts are reported. To determine for which ages, levels of motor function, and levels of cognitive function SRAT is/is not preferred, a generalized linear model with logit linking function was used. This model looked at the crossed interactions between age, Box and Block Test Scores, and Color Trails Test 2 Scores and included terms to control for experimental order and robot operator:

```
srat.better.than.ct ~ Age * CTT2 * BBT + experimental.order + robot.operator
```

To determine which factors were important, a Type III ANOVA was used. Results were interpreted using probability plots.

2.3.8 Flo Likability. Questions from the Godspeed III: Likability survey were averaged and reported.

3 RESULTS

3.1 Subjects

Forty (40) subjects consented to participate in the trial. One subject dropped out of the study prior to completing either telepresence condition and was therefore excluded. During the telepresence interactions for another subject, a wire broke, preventing the system's audio from working; so that subject's trial was also excluded. Thirty-eight (38) remaining subjects completed the trial, and their data was analyzed.

Two subjects were too young for the Children's Color Trails. One had reported having no-impairment and the other reported having a motor impairment only. They were both recorded as having no cognitive impairment. Due to the lack of a valid Color Trails Test Score, they were excluded from the models which incorporated Color Trails Test 2 scores (sections 2.3.3 and 2.3.4).

	Young Children N = 7	Teens-Young Adults N = 9	Adults N = 15	Older Adults N = 7	Sum N = 38
Reported Impairment					
Cognitive	0 (00%)	1 (11%)	0 (00%)	1 (14%)	2 (05%)
Motor	2 (29%)	4 (44%)	4 (27%)	4 (57%)	14 (37%)
Motor and Cognitive	3 (43%)	3 (33%)	5 (33%)	2 (29%)	13 (34%)
None	2 (29%)	1 (11%)	6 (40%)	0 (00%)	9 (24%)
Measured Impairment					
Motor	5 (71%)	5 (56%)	8 (53%)	4 (57%)	22 (58%)
Motor and Cognitive	1 (14%)	4 (44%)	1 (07%)	1 (14%)	7 (18%)
None	1 (14%)	0 (00%)	6 (40%)	2 (29%)	9 (24%)
Gender					
Male	4 (57%)	5 (56%)	5 (33%)	1 (14%)	15 (39%)
Female	3 (43%)	4 (44%)	10 (67%)	6 (86%)	23 (61%)
Class of Condition					
Brain Injury	2 (33%)	3 (50%)	6 (40%)	4 (57%)	15 (44%)
Neurodegenerative Disorder	0 (00%)	0 (00%)	5 (33%)	2 (29%)	7 (21%)
Peripheral Injury	2 (33%)	1 (17%)	2 (13%)	1 (14%)	6 (18%)
Psychological Disorder	1 (17%)	1 (17%)	0 (00%)	0 (00%)	2 (06%)
No Injury	1 (17%)	0 (00%)	2 (13%)	0 (00%)	3 (09%)
Unknown	0 (00%)	1 (17%)	0 (00%)	0 (00%)	1 (03%)
Race/Ethnicity					
American Indian or Alaska Native	0	0	1	0	1
Asian	0	0	2	0	2
Black or African American	2	5	3	1	11
Hispanic or Latino	1	0	1	0	2
Middle Eastern or North African	0	1	0	0	1
White	5	2	13	6	26
Prefer not to answer	0	2	0	0	2

Table 1. Subject Demographics. Percentages are shown per column for each category. Measured impairment is shown for subjects measured to have moderate or severe impairment.

Another subject with severe cognitive impairment was not able to understand the surveys asked after completing each interaction and so their results on those surveys were excluded. They were able to understand the preference question and likability question from the final survey and so their responses to those questions are preserved.

Aggregate demographics for the subjects can be seen in table 1, showing the subjects' self-reported impairment, measured impairment (using the BBT and CTT2), gender, class of condition, and race and ethnicity. Although some subjects have a condition which could lead to an impairment, they reported not having one, for example in the case of multiple sclerosis, or having one which has recovered, such as a motor impairment from stroke which is no longer present.

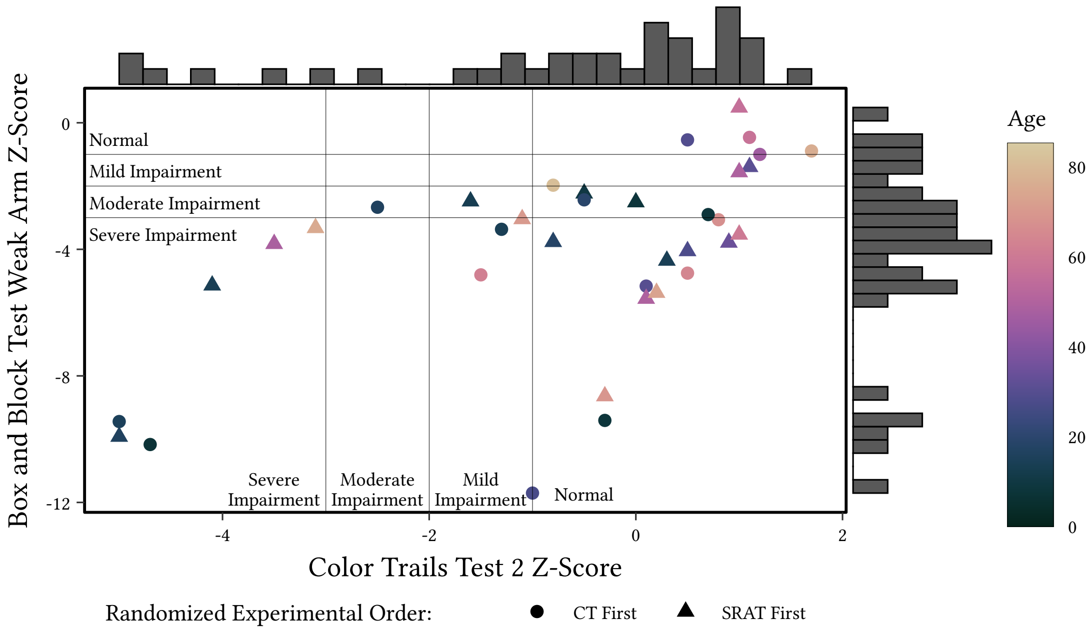


Fig. 8. Box and Block Test z-scores for subjects' weak arms and Color Trails Test 2 z-scores. Age is indicated by color. Levels of impairment for both motor and cognitive impairment are shown. Density for each axis is shown in histograms. Subjects who completed the FTF condition, then CT condition, and finally the SRAT condition, are indicated by a circle. Subjects who did the FTF condition first, followed by the SRAT condition, and finally the CT condition are indicated by a triangle.

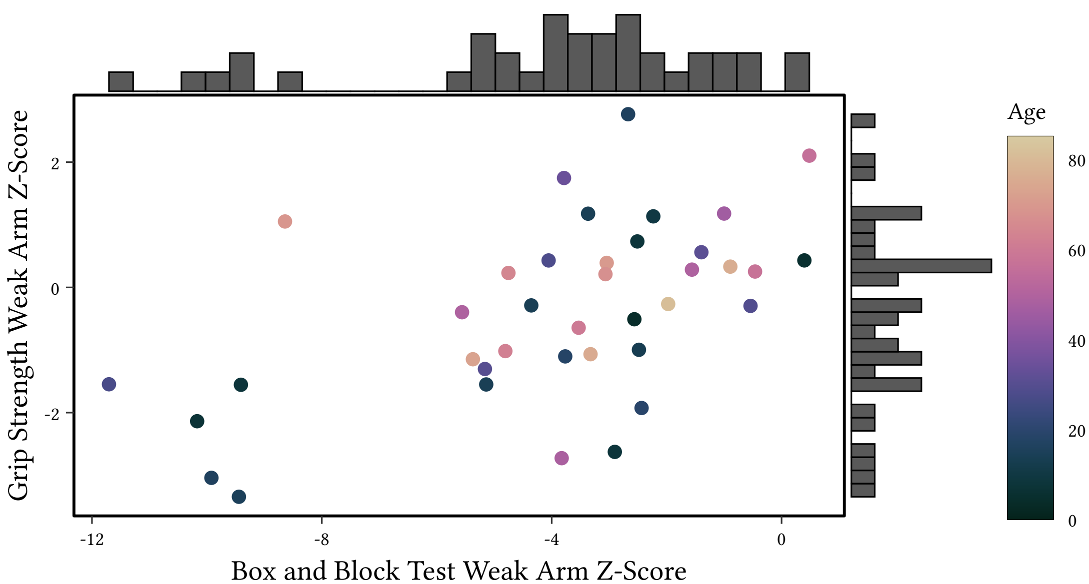


Fig. 9. Grip strength z-scores from the weak arm (as determined by the grip strength test) against Box and Block Test z-scores from the weak arm (as determined by the Box and Block Test). Age is shown in color. Density for each axis is shown in histograms.

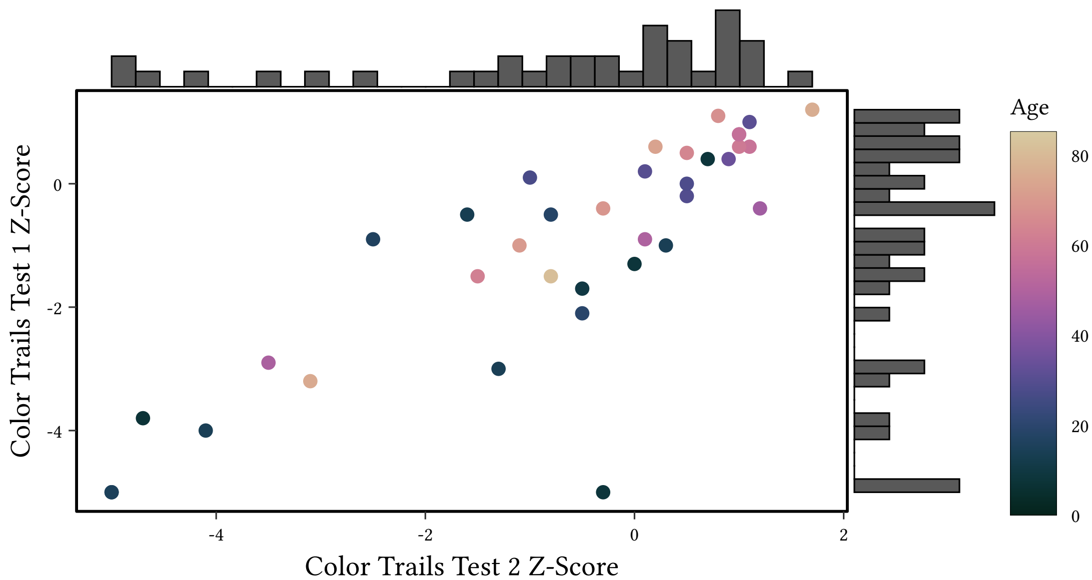


Fig. 10. Color Trails Test (for children, Children's Color Trails Test) part 1 and 2. Age is shown in color. Density for each axis is shown in histograms.

Fig. 11. Ratings on the question "How do you feel about robots?". The mean is shown as a red dashed line.

3.1.1 Cognitive and Motor Performance. Subjects had a variety of levels of motor function and cognitive function, as measured by the Box and Block Test and Color Trails Test 2 (fig. 8). In addition to unilateral gross motor function, strength was also measured using the grip strength test, shown in fig. 9 related to the Box and Block Test. As can be seen in fig. 10, the two parts of the Color Trails Test are well correlated. For the remainder of the analysis, to simplify the number of variables under consideration, Box and Block Test z-scores will be used as the sole measure of motor function and Color Trails Test 2 z-scores will be used as the sole measure of cognitive function.

3.1.2 Experience With and Feelings Towards Technology. Subjects had positive feelings towards robots (fig. 11). They had low levels of experience with robots, mixed prior experience with computers, and high experience with smartphones and tablets (fig. 12).

3.1.3 Experience With Telepresence. Thirty-three (33, 87%) of the subjects reported prior experience making video calls and 18 (47%) reported that they had used video calls for healthcare. Feelings on using video calls for healthcare were mixed, but non-negative (fig. 13).

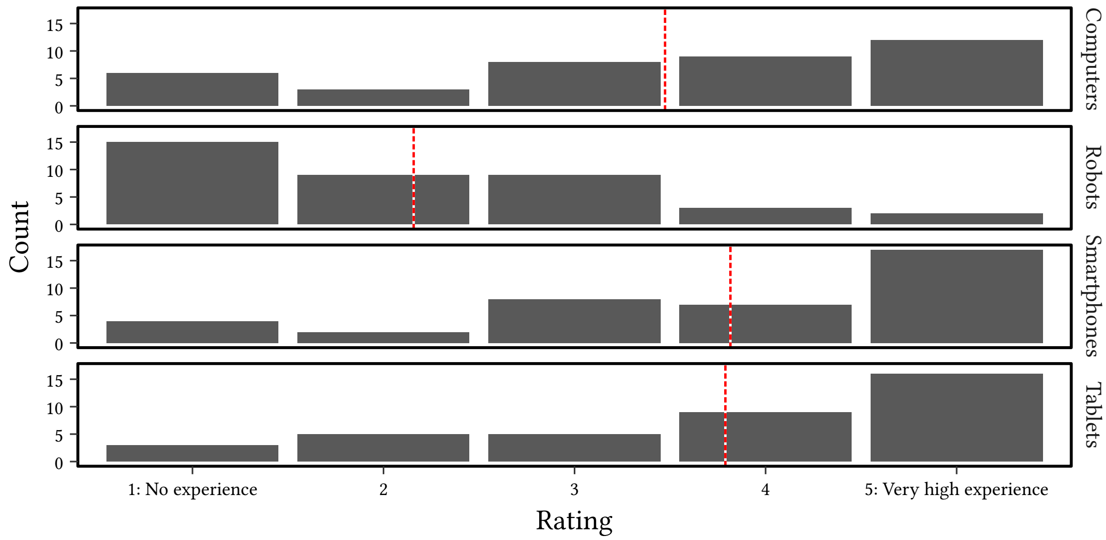


Fig. 12. Ratings on the question “Please rate your level of experience with the following:”. The mean response for each question is shown as a red dashed vertical line.

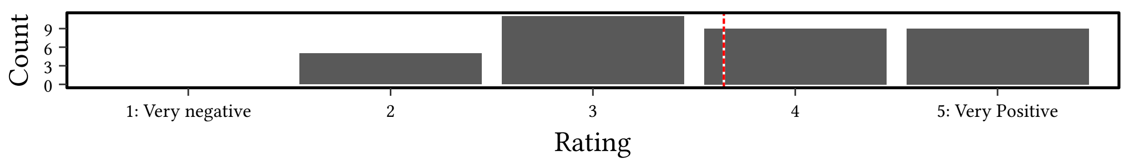


Fig. 13. Responses to the question “How do you feel about using video calls for healthcare?”. The mean response for each question is shown as a red dotted vertical line.

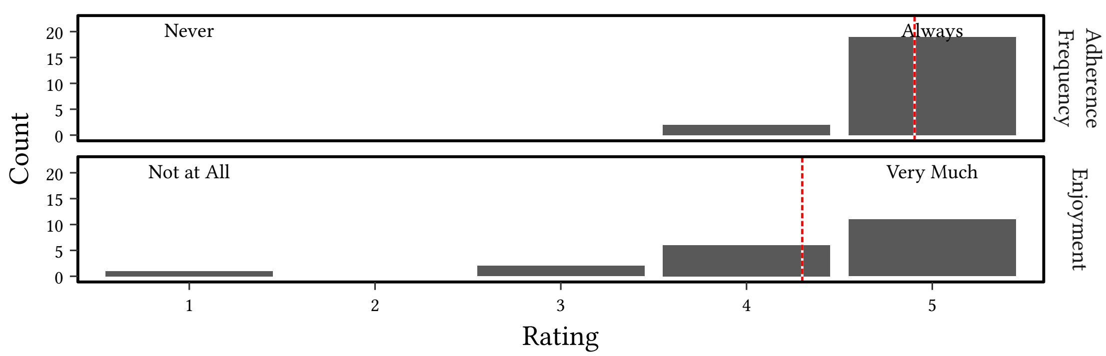


Fig. 14. Responses to the questions “How much do you enjoy your current therapy?” and “How often do you do the therapy you are supposed to do?”. These questions were asked only of subjects who reported that they are currently receiving therapy.

3.1.4 Prior Rehab Experience. Twenty-one (21, 55%) subjects reported that they are currently receiving therapy. Sixteen (16) subjects received physical therapy, 12 occupational therapy, 7 speech and language pathology, and 5 cognitive and behavioral therapy. They were receiving therapy primarily at hospitals for children (10) and rehab centers (9), along

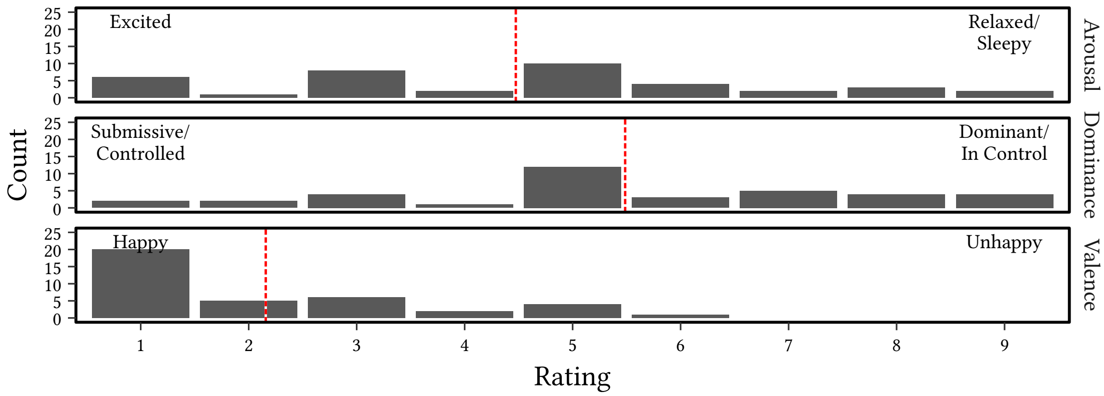


Fig. 15. Results from the Self-Assessment Manikin (SAM) administered prior to the face-to-face condition. Odd number ratings fell on the standard images from the SAM and even number ratings fell between the images. A dashed red line shows the mean for each question.

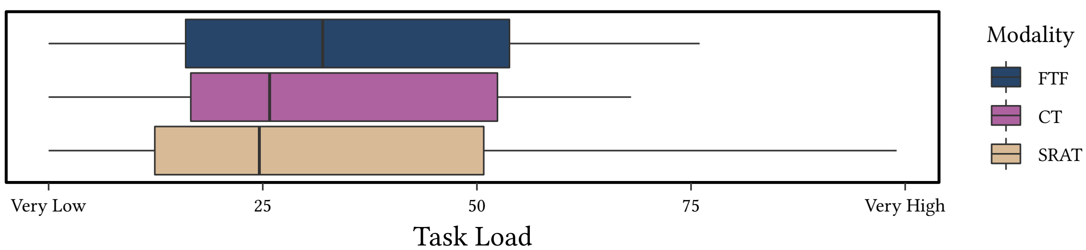


Fig. 16. Task load across the three interaction modalities. Data are shown as box and whiskers plots with the first and third quartiles at the box ends, median as a black line, and whiskers extending to the largest value not further than 1.5 times the interquartile range from the nearest quartile.

with inpatient facilities (4), outpatient facilities (3), at home (2), at school (1), and at a general hospital (1). Subjects enjoyed their therapy and were highly adherent (fig. 14).

3.1.5 Pre-Experiment Affect. On average, subjects started the study with neutral arousal and dominance and high valence (happy) (fig. 15). For subjects who completed the study over two days (n=10), valence and arousal saw small changes from their first day and dominance saw larger changes.

3.2 Task Load

Task load was low across all three modalities (fig. 16). The residuals from fitting the task load linear model are approximately normal with no apparent pattern between residuals and fitted values. The ANOVA on the model shows that several factors were significant: age (p=0.003), BBT (p=0.02), experimental order (p=0.04), Age:BBT (p=0.002), interaction modality:BBT (p=0.007), and interaction modality:age:BBT (p=0.02). The highest order of these, which is of interest, is interaction modality:age:BBT, the effects of which are plotted in fig. 17. As can be seen, there are slight differences between age groups and levels of motor function. Specifically, comparing SRAT to CT, the model predicts that people with normal to mildly impaired motor function, less than around 30 years old, will experience slightly

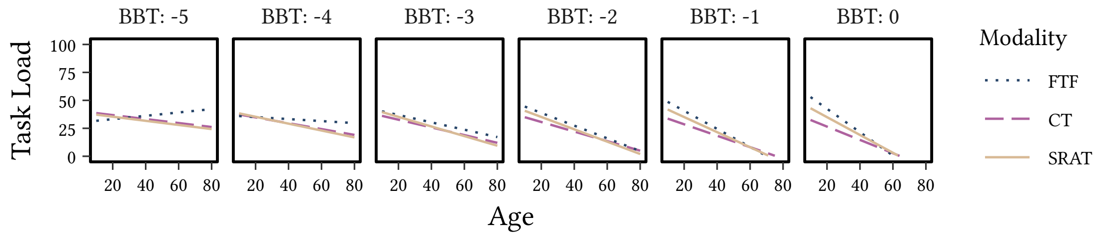


Fig. 17. Plots of the estimated results given the model fit to the task load scale, showing the interaction between motor function (BBT z-score) and age in predicting the task load of the three modalities.

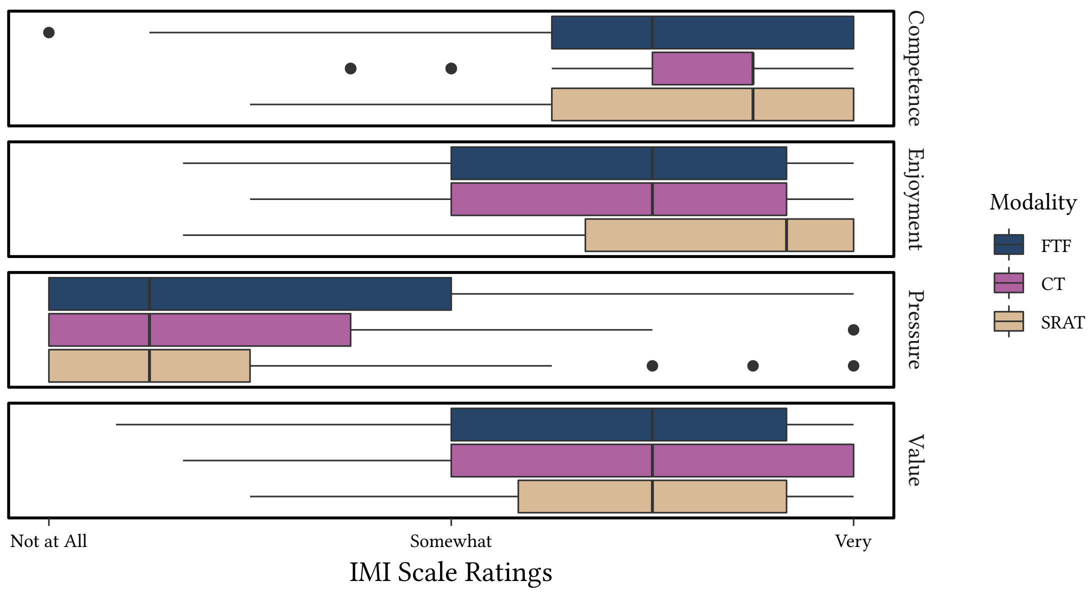


Fig. 18. Ratings on the IMI scales for competence, enjoyment, pressure, and value

higher task load with SRAT than with CT. All other ages/levels of motor function show no visible difference between SRAT and CT.

3.3 Competence

Competence was in general high across all three interaction modalities (fig. 18). The residuals from the linear model fit to the confidence scale showed good normality on a QQ-plot, although the residuals visually show a slight double peak and slightly heavy tails. A plot of residuals to fitted values shows banding due to the scale construction, but no global pattern. The ANOVA on the model shows that age:BBT and BBT:CTT are both significant contributors to the model. There are not however any significant interactions due to interaction modality.

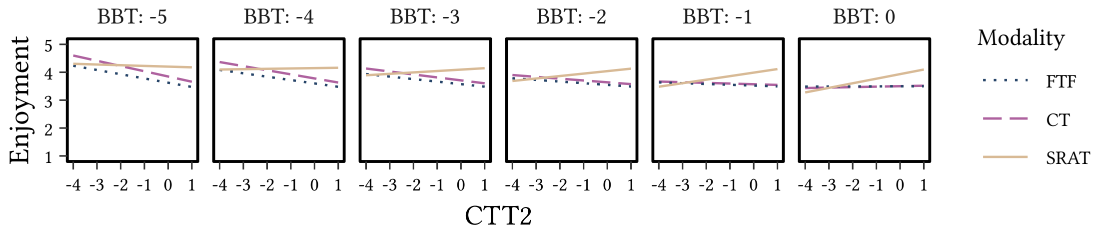


Fig. 19. Plots of the estimated results given the model fit to the IMI enjoyment scale, showing the interaction between motor function (BBT z-score) and cognitive function (CTT2 z-score) in predicting the enjoyment level associated with the three modalities.

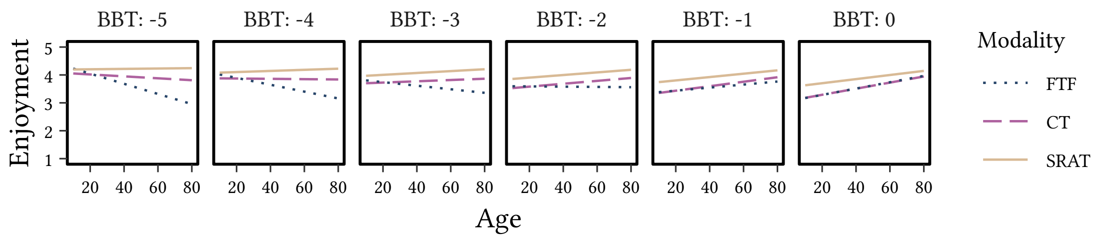


Fig. 20. Plots of the estimated results given the model fit to the IMI enjoyment scale, showing the interaction between motor function (BBT z-score) and age in predicting the enjoyment associated with each of the three modalities.

3.4 Enjoyment

Enjoyment of each interaction was high (fig. 18). Enjoyment of the SRAT interaction was approximately a quartile higher than the other two interactions. The QQ-Plot of the residuals for the enjoyment model shows some deviation from normal, but still within reason, visualizing the residual density, there is slight kurtosis and heavy tails. A plot of residuals to fitted values shows banding due to the scale construction, but no global pattern. The ANOVA shows several significant terms: Age:BBT (p=0.04), interaction modality:robot operator (p=0.003), interaction modality:age:BBT (p=0.03), and interaction modality:BBT:CTT2 (p=0.01).

Examining the interplay first between cognition, motor function, and the type of interaction in determining enjoyment (fig. 19), among all levels of motor impairment, at normal cognitive function, SRAT is more enjoyable than CT. At and beyond severe cognitive impairment (z<-3), but at all levels of motor function, CT is more enjoyable. Between moderate (z=-2) and severe (z=-3) cognitive impairment, at all levels of motor function, enjoyment is equivalent between conditions.

Moving on to the interplay between motor function, age, and interaction modality, in understanding enjoyment (fig. 20), among all ages, SRAT is more enjoyable than CT. For older people, there is a greater difference at lower motor function levels. Among younger people, there is a greater difference at higher motor function levels.

The higher order term of BBT:CTT2:age:interaction modality is not significant, however, evaluating all of these terms together, since they are independently significant, can provide a more complete picture of who enjoys SRAT more than CT and who does not (fig. 21). These interactions are more complex to follow. At normal levels of cognitive function (CTT > -1), and very severe motor impairment (BBT <= -5), adults (age > 40) enjoy SRAT more the CT, younger persons show no difference. At the same cognitive level, but normal motor function (BBT > -2), older adults enjoy SRAT

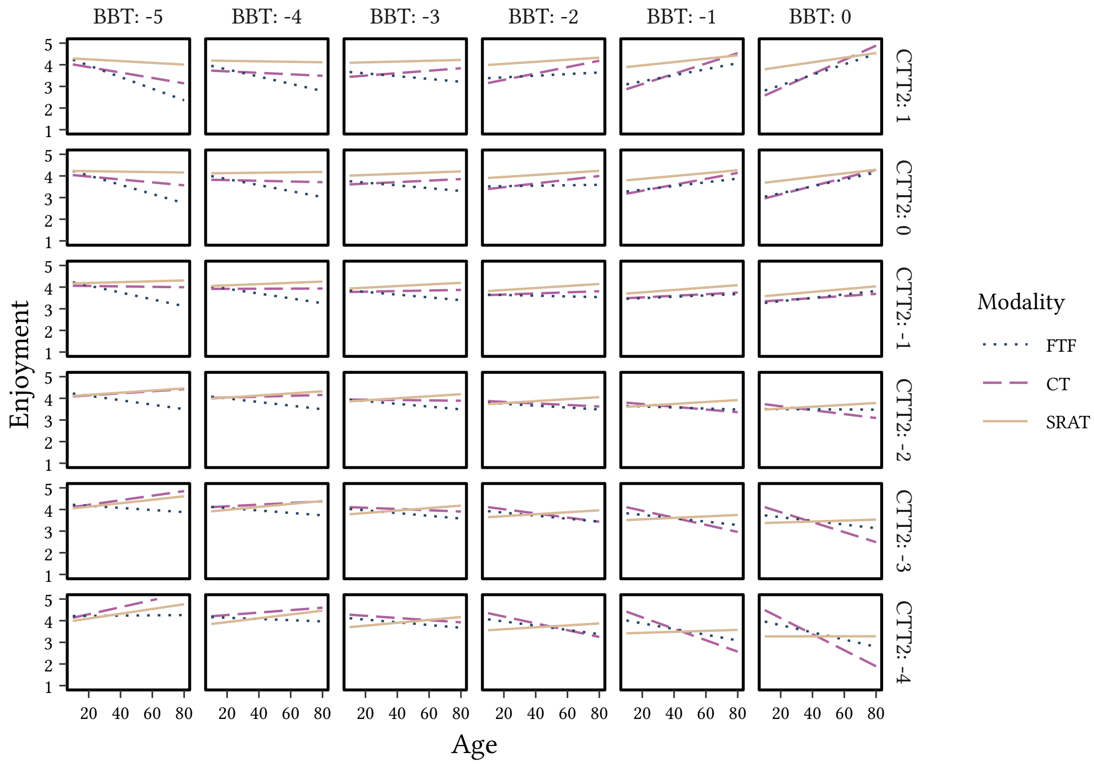


Fig. 21. Plots of the estimated results given the model fit to the IMI enjoyment scale, showing the interaction between motor function (BBT z-score), cognitive function (CTT2) and age in predicting the enjoyment associated with each of the three modalities.

and CT about equally and younger persons (age < 60) enjoy SRAT more than CT. Between high and low motor function, all ages appear to enjoy SRAT slightly more than CT. At mild cognitive impairment (CTT between -1 and -2), across all ages and motor levels, SRAT and CT appear to be enjoyed equally. At severe levels of cognitive impairment (CTT < -3) and high motor function, older persons enjoy SRAT more than CT and younger persons enjoy CT more than SRAT with the split between the two groups at about 40 years old. As motor impairment becomes more severe, the differences in enjoyment level between SRAT and CT for both groups diminish disappearing by around BBT = -3.

3.5 Pressure

The QQ-plot of residuals for the model on the pressure domain of the IMI shows small divergence from normality and the density plot shows mild kurtosis and skew. The plot of residuals vs fitted values shows significant deviation of the residuals from a random pattern, due to the limited number of levels in the pressure domains from having a small number of questions. The ANOVA for this model does not show any factors being significant.

3.6 Value

The QQ-plot of residuals for the model fit to the value domain of the IMI shows only slight deviation from normality and the density plot shows a nearly normal distribution. The plot of residuals against fitted values shows a random

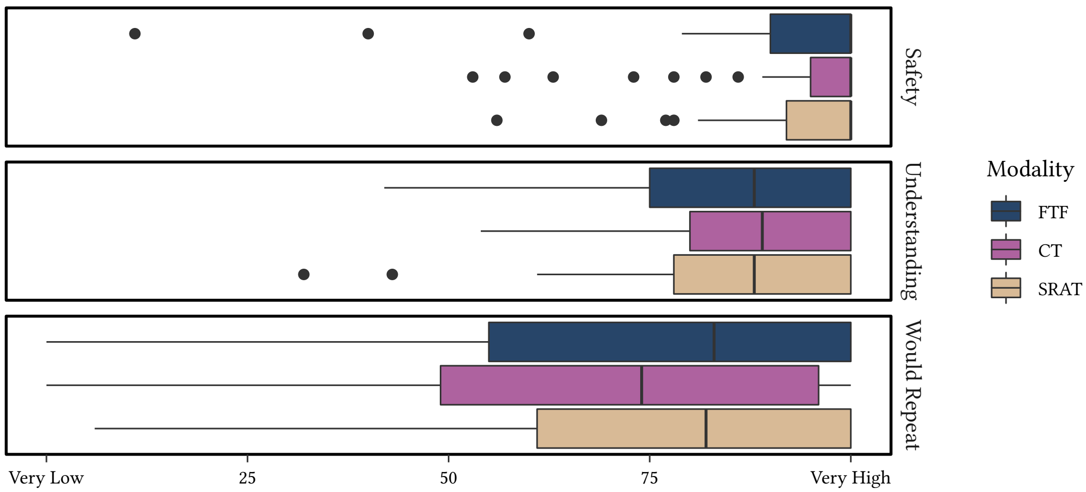


Fig. 22. Boxplots showing ratings to three questions using sliders from 0–100, asking if subjects felt safe, understood what they were supposed to do, and would want to repeat interactions.

distribution of the residuals. The ANOVA shows that several factors are relevant: CTT2 (p=0.05), age:CTT2 (p=0.03), BBT:CTT2 (p=0.03), age:BBT:CTT2 (p=0.05). Given that none of these terms involve the interaction modality, none of them help us shed light on how SRAT compares to CT.

3.7 Safety

Most subjects reported feeling very safe in all interactions (fig. 22).

3.8 Understanding

Most subjects reported that they understood what they were supposed to do during all the interactions (fig. 22).

3.9 Desire To Repeat Interactions

On average subjects reported a desire to repeat all three interactions (fig. 22), although there was considerable spread, with the interquartile ranges extending to a neutral level.

3.10 Effective Method of Doing Rehab

Subjects felt that all three modalities were effective methods of doing rehab with considerable spread in responses (fig. 23).

3.11 Reported Modality Preference

Most subjects (24, 63%) reported that face to face interactions were the best (fig. 24). Social robot augmented telepresence was rated better than classical telepresence by 26 of the subjects (68%) and was rated as the best interaction by 12 (32%) subjects.

The ANOVA on the model for interpreting the preference of subjects, comparing SRAT to CT, had several significant factors: age:CTT2 (p=0.01), CTT2:BBT (p=0.03), age:CTT2:BBT (p=0.01). The interaction of these factors is shown in

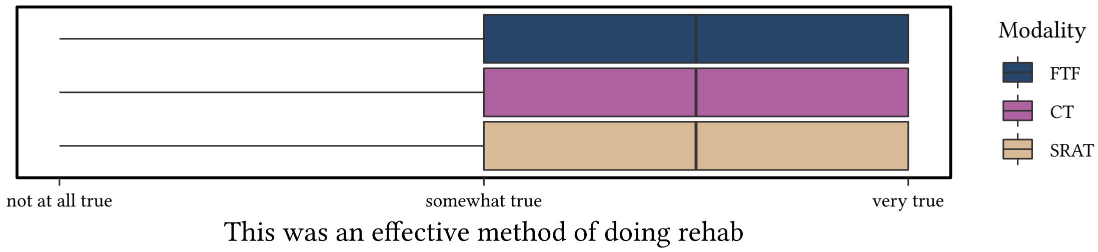


Fig. 23. Boxplots showing ratings to the question of whether the interaction was an effective method of doing rehab.

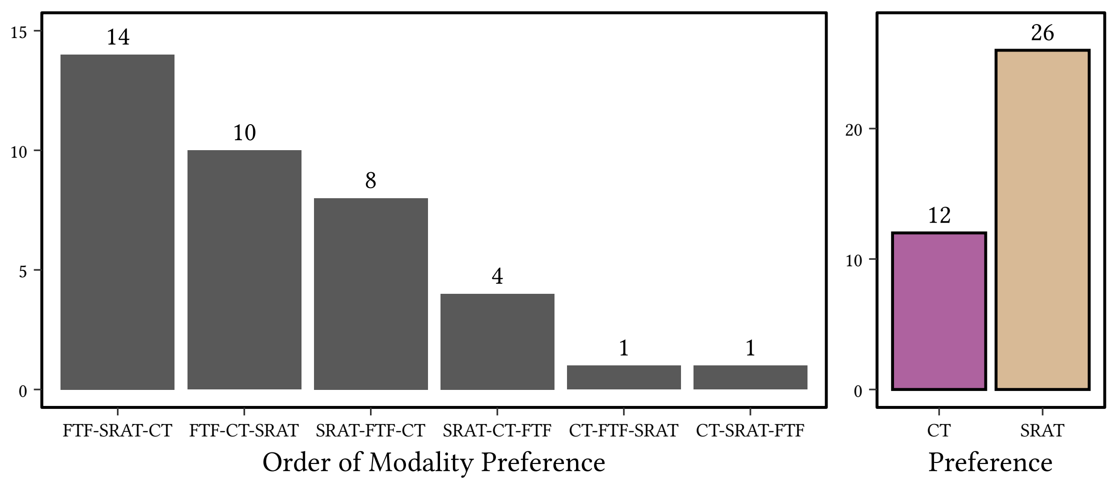


Fig. 24. On the left, the ranking of face to face (FTF), classical telepresence (CT), and social robot augmented telepresence (SRAT) by subjects when asked directly to rank them at the conclusion of the study. On the right, the same data, compressed down to only examine the comparison between classical telepresence and social robot augmented telepresence.

fig. 25. Among persons with high cognitive function (CTT2 >= 1) and no motor impairment, subjects above the age of approximately 60 would be expected to choose classical telepresence, as motor impairment increases, this cutoff age shifts older, exceeding 70 by mild motor impairment (BBT < -2). People with very severe motor impairments (BBT < -4) and moderate cognitive impairment (CTT2 < -2) over the age of about 60 would be expected to choose CT over SRAT. As cognitive impairment worsens, the cutoff age shifts to be younger, reaching less than 40 with very severe cognitive impairment (CTT2 < -4). The same trend is found at more severe motor impairment (BBT < -5), where at mild cognitive impairment (CTT2 < -1), subjects just over 40 years old would be expected to prefer CT over SRAT and the cutoff once again shifts younger as cognitive impairment increases, reaching approximately 20 years with very severe cognitive impairment (CTT < -4). Other groups of subjects are all expected to prefer SRAT, with varying likelihood.

3.12 Likability of Flo

Subjects overwhelmingly liked Flo (fig. 26).

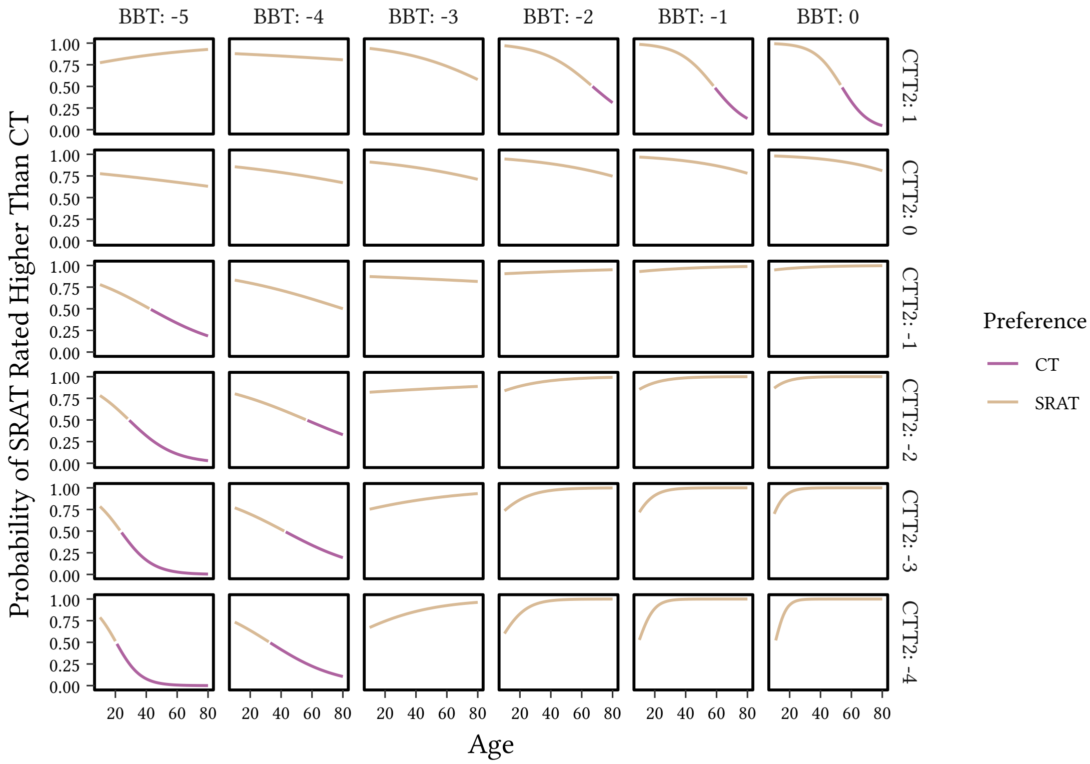


Fig. 25. Estimates from the preference model of which people of different ages, levels of cognitive function (CTT2), and levels of motor function (BBT) will prefer SRAT over CT. The plot is laid out to match fig. 21. Probabilities of choosing SRAT over CT by 0.5 or more are shown as preferring SRAT by color, probabilities less than 0.5 are indicated as preferring CT.

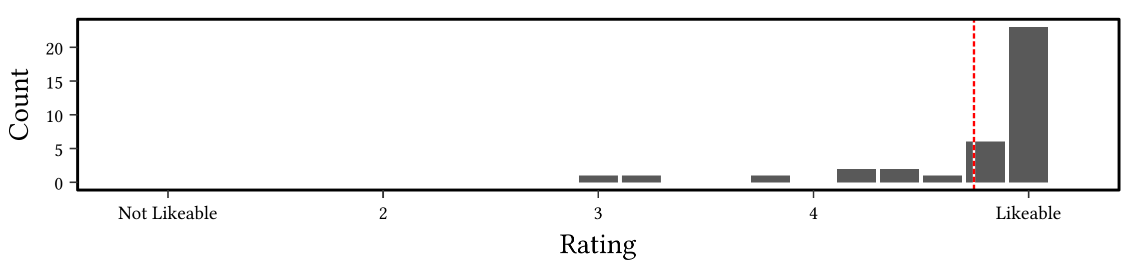


Fig. 26. Ratings on the Godspeed Scale Part III. The mean is shown as a dashed red line.

3.13 Responses to Open Ended Questions

During each post-interaction survey and in the final survey, the subjects were asked for other thoughts on their experience. Some of the responses provide further context on the data and interesting points to ponder:

One of the subjects summed up the utility of telehealth: “If I’m remote this is a good way to go about doing rehab but If I can access a physical person, I would prefer that, but if this were a lower cost alternative, I might still choose tele-rehab.” (49 yrs old, motor impaired, brain injury).

Subjects reported things they liked about Flo. A subject said that Flo was “Very cute and adorable” (16 yrs old, motor and cog impaired, brain injury). Another thought the system was “Gorgeous” (70 yrs old, motor impaired, neurodegenerative disorder). One subject named the robot John. They were very excited to interact with technology in general. After the CT condition, they reported “I enjoyed that but I really love John” (10 yrs old, motor impaired, psychological disorder). Another subject reached out to the study team after the trial to share some further thoughts on the utility of a social robot: “It occurred to me that Flo is perfect for working with folks with disabilities because she is entirely non-judgmental. She is unaware of age, race, intelligence, etc.” (81 yrs old, unimpaired, brain injury).

Some subjects articulated why they preferred CT over SRAT. One said “It’s [(CT)] better than with the robot, easier, more natural.” (28 yrs old, motor impaired, peripheral injury). Some subjects found the humanoid distracting, one said “I was a little distracted by the robot since I’m so used to one on one therapy” (29 yrs old, no impairment) and another reported “In the beginning I was kinda distracted by the robot’s movement and how different it was from a human’s, it was distracting at the beginning” (30 yrs old, motor impaired, brain injury). Yet another said: “For myself I was so distracted by watching the robot move that I was unable to concentrate on the task that I was given. The robot was very fascinating, but it distracted from doing the actual rehab activities”, but later said “this could be very useful for telehealth. [..] Robots like this could really change the way telehealth and telerehab is done in the future.” (49 yrs old, unimpaired, neurodegenerative disorder). One subject was intimidated by the robot “It’s a little bit scary to me because I’m not used to it [(robots)], I’m more used to people but I think I can get used to it” but felt that others would not be intimidated: “All in all a good experience and a whole lot of other people in my group would not be intimidated by the robot” (62 yrs old, motor impaired, brain injury). Another subject felt that the robot made more sense for pediatric use “The robot would probably be more helpful for younger people (or children)” and went on to say that “I’m very extroverted and would much rather have a human being to interact with.” (56 yrs old, unimpaired, neurodegenerative disorder).

A subject complained that the social robot was not as responsive as the operator: “The robot didn’t give feedback as often as [the operator] did in the classical telepresence. Had to wait for her to say go which challenged memory.” (64 yrs old, motor impaired, brain injury). Similarly, another subject felt that the robot did not demonstrate activities long enough: “During Simon says, in classical telepresence or in person, [the operator] kept doing the action such as swinging his arm but with Flo, she stopped so it might cause confusions on whether to continue doing the action.” (29 yrs old, unimpaired).

A few subjects reported challenges with the voice used on Flo. One said “It was a little harder, being able to hear it from a human was a little easier. Hearing and knowing it was a robot made it a little harder.” But then reported that working via CT led to more errors: “I messed up more this time” (14 yrs old, motor impaired, no injury). Similarly, another subject said: “It would be a better if the voice wasn’t choppy, it sounds like a computer. It would be better if it sounded like Siri and Alexa, more like a human” (19 yrs old, motor impaired, psychological disorder). However, other subjects liked the voice for the robot: “I liked the robot, I liked the faces, voices have matured and grown up vs the older (synthesized) voices which were very monotone” (49 yrs old, motor impaired, brain injury). Another subject thought the robot voice was so good that it was higher quality than a human over telepresence: “Could hear it better, very understandable. Maybe Flo’s voice is clearer than [the operator’s]” (16 yrs old, motor and cog impaired, brain injury).

Subjects also observed that Flo’s range of motion is not quite sufficient to complete all of the tasks and sometimes its arms are a bit too large: “while doing some of the physical stuff it should actually be able to get there (like reaching shoulder, instead of that it collided hands)” (19 yrs old, motor impaired, psychological disorder), “Major changes that would help her would be to have the arms not hit/clunk/smaller hands when covering mouth." (34 yrs old, motor impaired, brain injury).

One subject felt the robot should be height adjustable to be able to sit at eye level with everyone who interacts with it. Similarly, they requested that the target touch board be able to move up and down (19 yrs old, motor impaired, psychological disorder).

A number of subjects commented on the small screen size on the Flo system: "The screen was kinda small, would have been nice to have a larger screen where I could see all of his [(the operator's)] body" (30 yrs old, motor impaired, brain injury), "It was hard to tell the difference between shoulder flexion and abduction in the small screen during the telepresence" (29 yrs old, no impairment), "Classical telepresence would be better with at least the full torso in view" (31 yrs old, no impairment), "If the screen on the classical telepresence robot was a little bigger it would increase its efficiency." (64 yrs old, motor impaired, brain injury), "Screen was too small to see the therapist's arms sometimes" (73 yrs old, motor impaired, brain injury).

Subjects also shared ways in in which the activities provided a challenge for them. Some subjects found it challenging to keep straight which hand was their left and right hand for the target touch activity. This was noted for a subject who was four years old and one who was eight years old, both with motor impairments. More generally, multiple subjects reported on the cognitive load presented by the activities (after the FTF condition): "Effort was mostly having to think. Nothing hard, just had to listen before moving. Put your mind to it." (16 yrs old, motor and cog impaired, brain injury), "[the activities] make you think about what you are doing, you have to use your brain a lot" (19 yrs old, motor impaired, psychological disorder).

Two subjects recommended alternative ways in which to use Flo. A subject saw potential to use the social robot in triadic interactions in person: "It would be nice to have the robot with the human as an aid [in-person]" (30 yrs old, motor impairment, brain injury). A different subject was interested in using the humanoid robot by itself: "Would be interesting to not be able to see [operator's] face when working with Flo" (64 yrs old, motor impaired, brain injury).

4 DISCUSSION

This study presents exciting results comparing the classical method of delivering telerehab (classical telepresence, CT) with a new method using social robot augmented telepresence (SRAT). A tightly controlled study allowed CT and SRAT to be compared across several domains. Subjects who participated covered a broad cross-section of ages, levels of motor function, and levels of cognitive function. They had the levels of experience with technology that would be expected from the general population (we did not enroll engineering students in this study). Overall subjects were positive on all interaction modalities in which they participated.

The primary takeaway is that SRAT was preferred by subjects over CT by a ratio of over double (26 preferred SRAT, 12 preferred CT) (**H2**: supported). Across every level of motor and cognitive function, children (<20) are expected to prefer SRAT over CT. Among older adults with severe motor impairment and mild to severe cognitive impairment and adults with high cognitive function and normal motor function, CT is preferred over SRAT. This mirrors the dynamic found for enjoyment. Clearly there is an impact on preference based on impairment level and age and, as expected, older adults are less likely to prefer SRAT, a new unfamiliar technology.

As expected, face-to-face interactions (FTF) were the most preferred modality. This is expected, and is well reflected in the comment by one of the subjects that when FTF interactions are possible, that is preferred, but when they are not, or when they are very costly, that telepresence provides a viable alternative.

Across interaction modalities, subjects reported low task load. Only small differences were observed in task load between SRAT and CT conditions among younger subjects with high motor function (**H1l**: not supported).

Overall, enjoyment was higher in SRAT than CT conditions (**H1e**: supported). The relationship of enjoyment between modalities varied by cognitive impairment, motor impairment, and age. The highlight is that with no more than mild cognitive impairment, SRAT is either more enjoyable or as enjoyable as CT. With normal motor function, older adults find them equivalent. However, as cognitive impairment increases, some subjects find CT to be more enjoyable than SRAT.

One possible reason for the dynamic observed in both modality preference and enjoyment across modalities is that interactions by robot are less resilient to perturbation as a result of cognitive impairment, causing CT, with a flexible human as the center of the interaction, to be more enjoyable. This is not necessarily a reflection of the concept of SRAT, but instead on the specific form of SRAT presented here, in which interactions were very robot centric. It is also possible that with a certain level of cognitive impairment, a robot is simply too foreign for a person to understand. However, people with severe cognitive impairments who are over the age of 50 are expected to enjoy SRAT more than CT. More study is needed to understand this dynamic. Regardless, age, cognitive function, and motor function appear to affect whether a person will enjoy SRAT more than CT or not.

Other measures, competences, pressure, value, safety, understanding, desire to repeat, and effectiveness for rehab were all positive and did not appear to show any difference between SRAT and CT (**H1c**, **H1p**, and **H1v**: not supported).

The subjects, in their open ended responses, highlight some of the positive and negative reactions to each modality. Some subjects adored the humanoid. The idea of humanoid robots being non-judgemental, which a subject reported here, has come up in prior testing as well. It presents an interesting idea for how robots can reduce stress associated with recovery from impairment. Of course, the idea that the robot has infinite time isn't real since the operator's time is limited. But a long term goal for SRAT could be to deploy SRAT, have a clinician remote into the system to begin a session, and then let the robot and patient go through exercises on their own, at the best pace for the patient with the robot acting autonomously. Monitoring would let the clinician know when they were needed back and they could rejoin the session remotely. This sort of autonomy has been demonstrated by projects like the Nao-Therapist and Bandit, although there is much work to be done. This sort of clinician robot teaming could drastically increase the impact a single clinician could have.

A number of subjects said that the humanoid robot is distracting. To some extent the robot is mean to grab the patient's attention. But when it overshadows the activities to be done, that is not a good thing. As robots become more ubiquitous in society, the issue of being distracting will decrease. Even over the course of the trial, some subjects suggest that the distraction decreased, that the robot was most distracting at the beginning of the trial. The same dynamic can be expected for the robot being intimidating. The expectation with robots is generally that older people will reject them. Although the data does show some groups of older people preferring CT over SRAT, it does not show strong rejection of SRAT by elders and among some groups of elders, SRAT was more enjoyable than CT.

Subjects also complained about the fidelity of the arm movements on Flo, the quality of the voice, and the system's responsiveness. All very reasonable observations given the nature of the demonstration system. What is exciting is that even with all of these challenges, inherent in an early prototype, SRAT was still preferred over CT. Clearly SRAT is worth further exploration as a way to improve telerehab to bridge the growing gap in care.

4.1 Limitations

The sample size for this study is good for both the fields of social robotics and rehab robotics. The sample is broad and honest. However, the sample size is still too small to draw definitive conclusions on how different groups react to SRAT. The ratio of samples to parameters in the generalized linear models used to understand how different groups react to SRAT vs CT is non-favorable, which limits the interpretation which can be made from those studies. Linear models were used throughout, this is standard in the field (the ANOVA is a special case of the generalized linear model), however, other more flexible models may be more appropriate. More flexible models would however require a larger sample to fit appropriately. Another challenge is the presence of holes in age/level of motor and cognitive function, there was only one subject with moderate cognitive impairment and no subjects with normal or mild motor impairment and cognitive impairment (fig. 8). It is therefore likely that the models do not capture those sub-populations well. Bigger studies are needed.

The Flo system itself is a prototype, which is imperfect. As technology continues to improve to take load off of the operator through automation and make the system more reactive to the patient, the SRAT experience will improve.

A number of subjects noted that the screen used for the CT condition was too small. This is true, and telepresence systems with larger screens are becoming more common. This screen size is in line with systems like the Vgo robot by Vecna Robotics which are used in healthcare settings and is certainly larger than cellphone screens which are often used for telehealth encounters.

However, even with these limitations, the data are clear: on whole subjects prefer SRAT over CT for rehab activities and enjoyment is significantly impacted by modality preference. Further, 38 subjects were able to complete interactions via SRAT.

5 CONCLUSION

ACKNOWLEDGMENTS

subjects

nih

department

sim center

CHOP

Vera helped build the podium and set up protocol Many helped build/design robot

Some helped do experiments (should we ask to help with paper and make authors?): - Maria - Gary - Jeong Inn

REFERENCES

[1] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. *International Journal of Social Robotics* 1, 1 (Jan. 2009), 71–81. https://doi.org/10/bvsbrx

[2] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[3] Margaret M. Bradley and Peter J. Lang. 1994. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (March 1994), 49–59. https://doi.org/10/bx59sf

[4] John Fox and Sanford Weisberg. 2019. *An R Companion to Applied Regression* (third ed.). Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

[5] Marina Fridin and Mark Belokopytov. 2014. Embodied Robot versus Virtual Agent: Involvement of Preschool Children in Motor Task Performance. *International Journal of Human-Computer Interaction* 30, 6 (June 2014), 459–469. https://doi.org/10/ggdd9p

[6] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology*. Vol. 52. Elsevier, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[7] M. Jongbloed-Pereboom, M.W.G. Nijhuis-Van Der Sanden, and B. Steenbergen. 2013. Norm Scores of the Box and Block Test for Children Ages 3-10 Years. *American Journal of Occupational Therapy* 67, 3 (2013), 312–318. https://doi.org/10.5014/ajot.2013.006643

[8] Virgil Mathiowetz, Susan Federman, and Diana Wiemer. 1985. Box and Block Test of Manual Dexterity: Norms for 6–19 Year Olds. *Canadian Journal of Occupational Therapy* 52, 5 (Dec. 1985), 241–245. https://doi.org/10.1177/000841748505200505

[9] Virgil Mathiowetz, Gloria Volland, Nancy Kashman, and Karen Weber. 1985. Adult Norms for the Box and Block Test of Manual Dexterity. *The American Journal of Occupational Therapy* 39, 6 (June 1985), 386–391. https://doi.org/10/ggdd98

[10] V. Mathiowetz, D.M. Wiemer, and S.M. Federman. 1986. Grip and Pinch Strength: Norms for 6- to 19-Year-Olds. *The American journal of occupational therapy. : official publication of the American Occupational Therapy Association* 40, 10 (1986), 705–711. https://doi.org/10.5014/ajot.40.10.705

[11] PAR 1989. *Children's Color Trails Test*. PAR, Lutz, FL.

[12] PAR 1994. *Color Trails Test*. PAR, Lutz, FL.

[13] R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[14] unknown unknown. 0000. Intrinsic Motivation Inventory (IMI). (0000).

[15] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4, 43 (2019), 1686. https://doi.org/10.21105/joss.01686

A SURVEYS

Question	Input Format
Date of Birth	Date
Gender	Gender
Have you had/do you have any of:	Multiple Choice: None, Other, Stroke, Heart Attack, Cerebral Palsy, Traumatic Brain Injury, Multiple Sclerosis, Parkinsons, Spinal Cord Injury, Peripheral Nerve Injury
Which sides of your body did your stroke (or other brain trauma) affect (often the opposite of the side of the brain the stroke occurred on)	Multiple Choice: Left, Right
Does the subject have a motor impairment?	Yes/No
What is their motor impairment?	Text Entry
Does the subject have a cognitive impairment?	Yes/No
What is their cognitive impairment?	Text Entry
Other notes on diagnostics	Text Entry
Arm Function	No, Somewhat, Yes, No answer
Can touch head?	
Can reach arms out in front?	
Can reach arms out to the side?	
Sitting Function	Yes/No
Does the subject use a wheelchair?	
Can the subject sit without help, with free movement of arms (trunk support ok)?	
Can the subject follow instructions?	No, Somewhat, Yes, No answer

Table 2. Pre-Screen form, excluding contact information questions

Question	Input Format
How are you feeling right now? (Self-Assessment Manikin)	Images from SAM with 9 steps:
Affect	1: Happy – 9: Unhappy
Arousal	1: Excited – 9: Relaxed/Sleepy
Dominance	1: Dominant/In Control – 9: Submissive/Being Controlled
How do you feel about robots?	Likert: 1: Very Negative – 5: Very Positive
Please rate your level of experience with the following:	Likert: 1: No Experience – 5: Very High Experience
Computers	
Tablets	
Smartphones	
Robots	
Do you currently receive therapy?	Yes/No
Where do you currently receive therapy? (If currently receiving therapy)	Checkboxes
School	
Hospital for children	
General hospital	
Elder care hospital	
Rehab center	
Elder care home	
Community center	
At home	
Inpatient facility	
Outpatient facility	
Other	
What other locations? (If Other selected)	Text Entry
What kind of therapy do you receive? (If currently receiving therapy)	True/False Checkbox
Physical Therapy	
Occupational Therapy	
Speech and Language Pathology	
Cognitive Behavioral	
Other	
What other types? (If Other selected)	Text Entry
How much do you enjoy your current therapy? (If currently receiving therapy)	Likert: 1: Not at all – 5: Very much
How often do you do the therapy you are supposed to do? (If currently receiving therapy)	Likert: 1: Never – 5: Always
Do you take any mood or focus-altering medications?	Yes/No
Which mood or focus-altering medications do you take? (If taking)	Text Entry
Have you ever done a video call?	Yes/No
Have you ever done a video call for healthcare?	Yes/No
How do you feel about using video calls for healthcare?	Likert: 1: Very negative – 5: Very positive
How would you describe yourself? (Select all that apply)	Checkboxes
American Indian or Alaska Native	
Asian	
Hispanic or Latino	
Black or African American	
Middle Eastern or North African	
White	
Native Hawaiian or other Pacific Islander	
other	
prefer not to answer	
Please specify other (if other selected):	Text Entry

Table 3. Intake Survey

Question	Input Format
Please answer the following questions based on the interaction you just had using the sliders:	Slider:
How well did you understand what you were supposed to do?	Not at all – Perfectly
Would you want to have this interaction again?	Not at all – Very much
How safe did you feel during the interaction?	Not at all safe – Very safe
Mental Demand: How mentally demanding was the interaction?	Very Low – Very High
Physical Demand: How physically demanding was the interaction?	Very Low – Very High
Performance: How well did you perform the tasks you were asked to do?	Failure – Perfect
Effort: How hard did you have to work to perform the activities asked of you?	Very low – Very high
Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you?	Not at all – Very much
Enjoyment: How much did you enjoy the interaction?	Not at all – Very much
For each of the following statements, please indicate how true it is for you, based on the interaction you just had and activities you just completed, using the following scale:	Likert: 1: not at all true – 3: somewhat true – 5: very true
I was anxious while doing the activities	
The activities were fun to do	
I believe the activities could be of some value to me	
I would describe the activities as very interesting	
I was very relaxed in doing the activities	
I think that doing these activities is useful for rehab	
I thought the activities were quite enjoyable	
I think doing these activities could help me to improve my arm function	
This was an activity that I couldn't do very well	
The activities did not hold my attention at all	
I am satisfied with my performance at these tasks	
This was an effective method of doing rehab	
Do you have any other comments or thoughts about this interaction?	Text entry

Table 4. Post Interaction Survey

Question	Input Format
Please rank which interaction you thought was best, second best, and worst:	Best, Second best, Third best:
Face-to-face	
Telepresence + Social Robot	
Classical Telepresence	
Do you think telehealth would change how you manage your health and medical needs if you and your clinician used it?	Likert: 1 Not at all – 5: Very much
Would you follow your doctor's/therapist's/nurse's advice less or more if they worked with a telehealth system?	Likert: 1: Much Less – 3: No Change – 5: Much More
Would video visits be a convenient form of healthcare delivery for you?	Likert: 1: No – 5: Yes
Please rate how you believe that using the humanoid robot (like Lil'Flo, with arms and a head) with video telepresence will compare to using video telepresence alone:	Likert: 1: Much better with humanoid – 3: No difference – 5: Much better without humanoid
Communication between me and the clinician	
My motivation to do rehab activities	
My compliance with instructions during interactions	
My adherence to treatment plans after interactions	
What locations do you think Lil'Flo could be deployed in?	Checkboxes
Rural outpatient clinics	
Rural inpatient clinics	
Elder care facilities	
Schools	
Patient homes	
Community centers	
Urban inpatient clinics	
Urban outpatient clinics	
None	
Other	
What other locations? (If other selected)	Text input
Are there other activities which you would like to do with Lil'Flo?	Text input
Please rate your impression of Lil'Flo on these scales:	5 Element Likert Scales from 1 – 5
Dislike – Like	
Unfriendly – Friendly	
Unkind – Kind	
Unpleasant – Pleasant	
Awful – Nice	
Do you have any other comments or feedback?	Text entry
Before this study, did you have any prior experience with Lil'Flo?	Checkboxes
No prior knowledge	
I have read a paper on the system	
I have seen the system in person	
I have used the system	
I have some other experience with system	
What other prior experience? (If other selected)	Text entry

Table 5. Final Survey

Movements
Clap your hands
Raise your arms up over your head
Touch your right hand to your left shoulder
Touch your left hand to your right shoulder
Reach forward with your arms
Cover your eyes with your hands
Touch your mouth with your right hand
Touch your mouth with your left hand
Touch your head with your right hand
Touch your head with your left hand
Reach to the side with your right arm
Reach to the side with your left arm
Wave with your right arm
Wave with your left arm
Rotate your right arm like me
Rotate your left arm like me
Swing your right arm up and down like this
Swing your left arm up and down like this
Swing your right arm to the side like this
Swing your left arm to the side like this

Table 6. Movements used in the Simon says game