

Article

Long-Term Exercise Assistance: Group and One-on-One Interactions between a Social Robot and Seniors

Mingyang Shao ¹, Michael Pham-Hung ¹, Silas Franco Dos Reis Alves ¹, Matt Snyder ², Kasra Eshaghi ¹, Beno Benhabib ¹ and Goldie Nejat ^{1,3,4,*}

¹ Autonomous Systems and Biomechanics Laboratory, Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Rd., Toronto, ON M5S 3G8, Canada

² Yee Hong Centre for Geriatric Care, 5510 Mavis Rd., Mississauga, ON L5V 2X5, Canada

³ Toronto Rehabilitation Institute, 550 University Ave., Toronto, ON M5G 2A2, Canada

⁴ Rotman Research Institute, Baycrest Health Sciences, 3560 Bathurst St., North York, ON M6A 2E1, Canada

* Correspondence: nejat@mie.utoronto.ca

Abstract: For older adults, regular exercises can provide both physical and mental benefits, increase their independence, and reduce the risks of diseases associated with aging. However, only a small portion of older adults regularly engage in physical activity. Therefore, it is important to promote exercise among older adults to help maintain overall health. In this paper, we present the first exploratory long-term human–robot interaction (HRI) study conducted at a local long-term care facility to investigate the benefits of one-on-one and group exercise interactions with an autonomous socially assistive robot and older adults. To provide targeted facilitation, our robot utilizes a unique emotion model that can adapt its assistive behaviors to users' affect and track their progress towards exercise goals through repeated sessions using the Goal Attainment Scale (GAS), while also monitoring heart rate to prevent overexertion. Results of the study show that users had positive valence and high engagement towards the robot and were able to maintain their exercise performance throughout the study. Questionnaire results showed high robot acceptance for both types of interactions. However, users in the one-on-one sessions perceived the robot as more sociable and intelligent, and had more positive perception of the robot's appearance and movements.

Keywords: autonomous exercise facilitation; human–robot interaction; intelligent robot; social robotics; older adults; long-term study



Citation: Shao, M.; Pham-Hung, M.; Alves, S.F.D.R.; Snyder, M.; Eshaghi, K.; Benhabib, B.; Nejat, G. Long-Term Exercise Assistance: Group and One-on-One Interactions between a Social Robot and Seniors. *Robotics* **2023**, *12*, 9. <https://doi.org/10.3390/robotics12010009>

Academic Editor: Marco Ceccarelli

Received: 27 October 2022

Revised: 16 December 2022

Accepted: 31 December 2022

Published: 6 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

For older adults, regular exercise can reduce the risk of depression, cardiovascular disease, type 2 diabetes, obesity, and osteoporosis [1]. Older adults who regularly exercise are also more likely to be able to engage in instrumental activities of daily living such as meal preparation and shopping with increased independence [2]. Furthermore, they are less likely to fall and be injured [3]. Even when taken up later in life, older adults can still gain from the benefits of exercise with a decrease in risk of cardiovascular disease mortality [4]. Despite the overwhelming evidence, however, for example, only 37% of older adults aged over 65 in Canada perform the recommended 150 min of weekly physical activities including aerobic (e.g., walking), flexibility exercises (e.g., stretching), and muscle strengthening (e.g., lifting weights) [5]. For the goal of exercise promotion, a handful of social robots have shown potential for use with older adults [6–9]. These robots can extend the capabilities of caregivers by providing exercise assistance when needed, autonomously tracking exercise progress for multiple individuals over time, and facilitating multiple parallel exercise sessions.

In general, human–robot interaction (HRI) can be often conducted in: (1) group interactions, or (2) one-on-one interaction settings. The ability to meet new people and other social aspects of group sessions have shown to be a major motivator in older adults to

participate in exercise [10]. The increase in social interaction among participants, which can be further increased by the facilitation by a robot [11], can stimulate the prefrontal cortex of the participants, regions that are traditionally associated with executive functions such as working memory and attention [12]. This results not only in physical health benefits, but also may add to their cognitive functions by participating in group exercise sessions. One-on-one exercise sessions, on the other hand, can also increase participation as they enable user experience to be directly tailored to the participant by providing individualized feedback [13]. Individualized experiences also increase participants' attentiveness and retention of the information as they perceive it to be more relevant to them, resulting in increased physical performance results [14].

In human–human interactions, human affect plays a significant role as it guides people's thoughts and behaviors, and in turn influences how they make decisions and communicate [15]. To effectively interact with people and provide assistance, robots need to recognize and interpret human affect as well as respond appropriately with their own emotional behaviors. This promotes natural interactions in human-centered environments by following accepted human behaviors and rules during HRI, leading to acceptance of the robot in order to build long-term relationships with its users [16].

During exercising, as people are required to perform physical movements, which can also lead to perturbed facial expressions due to the increase in effort and muscle fatigue [17], common physical affective modes such as body movements and facial expressions are not always available for the robot to detect. Furthermore, these physical modes are difficult to use with older adults as they have age-related functional decline in facial expression generation [18], and body movements and postures [19]. On the other hand, Electroencephalography (EEG) signals, which are largely involuntarily, and activated by the central nervous system (CNS) and the autonomic nervous system (ANS) of the human body can be used to detect both affective cognitive states of older adults [20]. EEGs have also been successfully utilized to detect user affect during physical activities (e.g., cycling) [21].

In this paper, we present the first long-term robot exercise facilitation HRI study with older adults investigating the benefits of one-on-one and group sessions with an autonomous socially assistive robot. Our autonomous robot uses a unique emotional model that adapts its assistive behavior to the user affect during exercising. The robot can also track users' progression towards exercise goals using the Goal Attainment Scale (GAS) while monitoring their heart rate to prevent overexertion. Therefore, herein, we investigate and compare one-on-one and group intervention types to determine their impact on older adults' experiences with the robot and their overall exercise progress. A long-term study (i.e., 2 months) was conducted to determine the challenges for achieving a successful exercise HRI while directly observing the robot as it adapts to user behaviors over time. The aim was to investigate any improved motivation and engagement in the activity through repeated multiple exercise sessions over time and in different interaction settings with an adaptive socially assistive robot.

2. Related Works

Herein, we present existing socially assistive robot exercise facilitation studies that have been conducted in (1) one-on-one sessions [7,9,22,23], (2) group sessions [6,24], and/or (3) a combination of the two scenarios [8]. Furthermore, we discuss HRI studies which have compared group vs. one-on-one interactions in various settings and for various activities.

2.1. Socially Assistive Robots for Exercise Facilitation

2.1.1. Socially Assistive Robots for Exercise Facilitation in One-on-One Settings

In [7], the robot Bandit was used to facilitate upper body exercise with older adults in a one-on-one setting with both a physical and virtual robot during four sessions over two weeks. The virtual robot was a computer simulation of the Bandit robot that was shown on a 27-inch flat-panel display. The study evaluated the users' acceptance and perception of the robot as well as investigated the role of embodiment. In comparison, participants in the

physical robot sessions evaluated the robot as more valuable/useful, intelligent, helpful, and having more social presence and being a companion than those in the virtual robot sessions. No significant differences were observed in participant performance between the two study groups.

In [9], a NAO robot learned exercises performed by human demonstrators in order to perform them in exercise sessions with older adults in one-on-one settings. The study, completed over a 5-week period, found that participants had improved exercise performance after three sessions for most of the exercises. They had high ratings for the enjoyment of the exercise sessions and accepted the robot as a fitness coach. However, the acceptance of the robot as a friend slightly decreased over the sessions as they reported that they recognize the robot as a machine and they only wanted to consider humans as friends. They also showed confused facial expressions during more complicated exercises (e.g., exercises with a sequence of gestures); however, the occurrence of the confused facial expressions decreased by the last session.

In [22], the NAO robot was used to facilitate the outpatient phase of a cardiac rehabilitation program for 36 sessions over an 18-week period. Participants ranged in age from 43 to 80. Each participant was instructed to exercise using a treadmill, while the robot monitored their heart rate using an electrocardiogram and alerted medical staff if it exceeded an upper threshold, and monitored their cervical posture using a camera and provided the participant with verbal feedback if a straight posture was not maintained. The robot also provided the participants with periodic pre-programmed motivational support through speech, gestures, and gaze tracking. The robot condition was compared to a baseline condition without the robot. The results showed participants that used the robot facilitator had a lower program dropout rate and achieved significantly better recovery of their cardiovascular functioning than those that did not use the robot.

In [23], a NAO robot was also used for motor training of children by playing personalized upper-limb exercise games. Healthcare professionals assessed participants' motor skills throughout the study at regular intervals using both the Manual Ability Classification Systems (MACS) scale to assess how the children handle objects in daily activities and the Mallet scale to assess the overall mobility of the upper limb. Questionnaire results showed positive ratings over time, indicating that participants considered the robot to be very useful, easy to use, and operating correctly. There was no change in the MACS scale results for all participants over the three sessions. However, participants slightly improved over time on the Mallet score on their motor skills.

2.1.2. Socially Assistive Robots for Exercise Facilitation in Group Settings

In [6], a Pepper robot was deployed in an elder-care facility to facilitate strength building, flexibility exercises, and to play cognitive games. Six participants were invited to a group session with Pepper twice a week for a 10 week-long study. Interviews with the participants revealed that many older adults were originally fearful of the robot, but became comfortable around the robot by the end of the study. In [24], a NAO robot was used to facilitate seated arm and leg exercises with older adults in a group session with 34 participants. Feedback from both the staff and older adults showed that the use of the NAO robot as an exercise trainer was positively received.

2.1.3. Socially Assistive Robots for Exercise Facilitation in Both One-on-One and Group Settings

In [8], a remotely controlled robot Vizzy was deployed as an exercise coach with older adults in both one-on-one sessions and group sessions. The robot has an anthropomorphic upper torso and head with eye movements to mimic gaze. Vizzy would lead participants from a waiting room to an exercise location, give them instructions to follow a separate interface that showed the exercises, and then provided corrective instructions if necessary—the robot did not demonstrate the exercises itself. A camera in each of Vizzy's eyes was controlled by an operator for gaze direction when the robot spoke. The results demonstrated

that the participants perceived the robot as competent, enjoyable, and had high trust in the robot. They found Vizzy looked artificial and had a machine-like appearance, but thought its gaze was responsive and liked the robot.

The aforementioned robots showed the potential in using a robot facilitator for upper-limb strength and flexibility exercises [6–9,23,24] and for cardiac rehabilitation [22]. The studies focused on perceptions and experiences of the user through questionnaires and interviews [6,8,9,23,24], with the exception of [7,9], which tracked body movements to determine if users correctly performed specific exercises. In general, the results showed acceptance for robots as a fitness coach, as well as a preference for physical robots over virtual ones [7]. Only in [8], were both group and one-on-one sessions considered, however, the two interaction types were not directly compared to investigate any health and HRI benefits between the two. Additionally, the aforementioned studies have only investigated the perception of the robots over short-term durations (one to three interactions) [7–9,24], lack quantitative results when performed over a long-term duration (10 weeks) [6], or have not directly focused on the older adult population in their long-term studies [22]. Furthermore, no other user feedback was used by the robots to engage older users in the long-term. In particular, human affect has been shown to promote engagement and encouragement during HRI by adapting robot emotional behaviors to users. Herein, we investigate for the first time a long-term robot exercise facilitation study with older adults to compare and determine the benefits of one-on-one and group sessions with an intelligent and autonomous socially assistive robot which adapts its behaviors to the users.

2.2. General HRI Studies Comparing Group vs. One-on-One Interactions

To-date, only a handful of studies have investigated and compared group and one-on-one interactions for social robots. For example, in [25], the mobile robot Robovie was deployed in a crowded shopping mall to provide directions to visitors in group or one-on-one settings. Results showed that groups in general, especially entitative (family, friends, female) groups of people, interacted longer with the Robovie, and were more social and positive towards it than individuals. In addition, they found that participants who would not typically interact with the robot based on their individual characteristics were more likely to interact with Robovie if other members of their group did.

In [26], intergroup competition was investigated during HRI. A study was performed where participants played dilemma games, where they had the chance to exhibit competitive and cooperative behaviors in four group settings with varying numbers of humans and robots. Results indicated that groups of people were more competitive towards the robots as opposed to individuals. Furthermore, participants were more competitive when they were interacting with the same number of robots (e.g., three humans with three robots or one human with one robot).

In [27], two MyKeepon baby chick-like robots were used in an interactive storytelling scenario with children in both one-on-one and groups settings. The results showed that individual participants had a better understanding of the plot and semantic details of the story. This may be due to the children being more attentive as there were no distractions from peers in the one-on-one setting. However, when recalling the emotional content of the story, there was no difference between individuals and the groups.

These limited studies show differences in user behaviors and overall experience between group and one-on-one interactions. In some scenarios, interactions were more positive in group settings than individual settings as peers were able to motivate each other during certain tasks [25,26], a phenomenon also reported in non-robot-based exercise settings [10]. Individual interactions, however, were less distracting and allowed individuals to focus on the task at hand [27]. As these robot studies did not focus directly on older adults nor the exercise activity, it is important to explore the specific needs and experiences of this particular user group in assistive HRI. This further motivates our HRI study.

3. Social Robot Exercise Facilitator

In our long-term HRI study, we utilized the Pepper robot to autonomously facilitate upper-body exercises in both group sessions and one-on-one sessions. The robot is capable of facilitating nine different upper-limb exercises, Figure 1: (1) open arm stretches, (2) neck exercises, (3) arm raises, (4) downward punches, (5) breaststrokes, (6) open/close hands, (7) forward punches, (8) lateral trunk stretches (LTS), and (9) two-arm LTS. These exercises were designed by a physiotherapist at our partner long-term care (LTC) home, the Yee Hong Centre for Geriatric Care. The exercises are composed of strength building and flexibility exercises [28]. Benefits include improving stamina and muscle strength, functional capacity, helping maintain healthy bones, muscle, and joints, and reducing the risk of falling and fracturing bones [29].

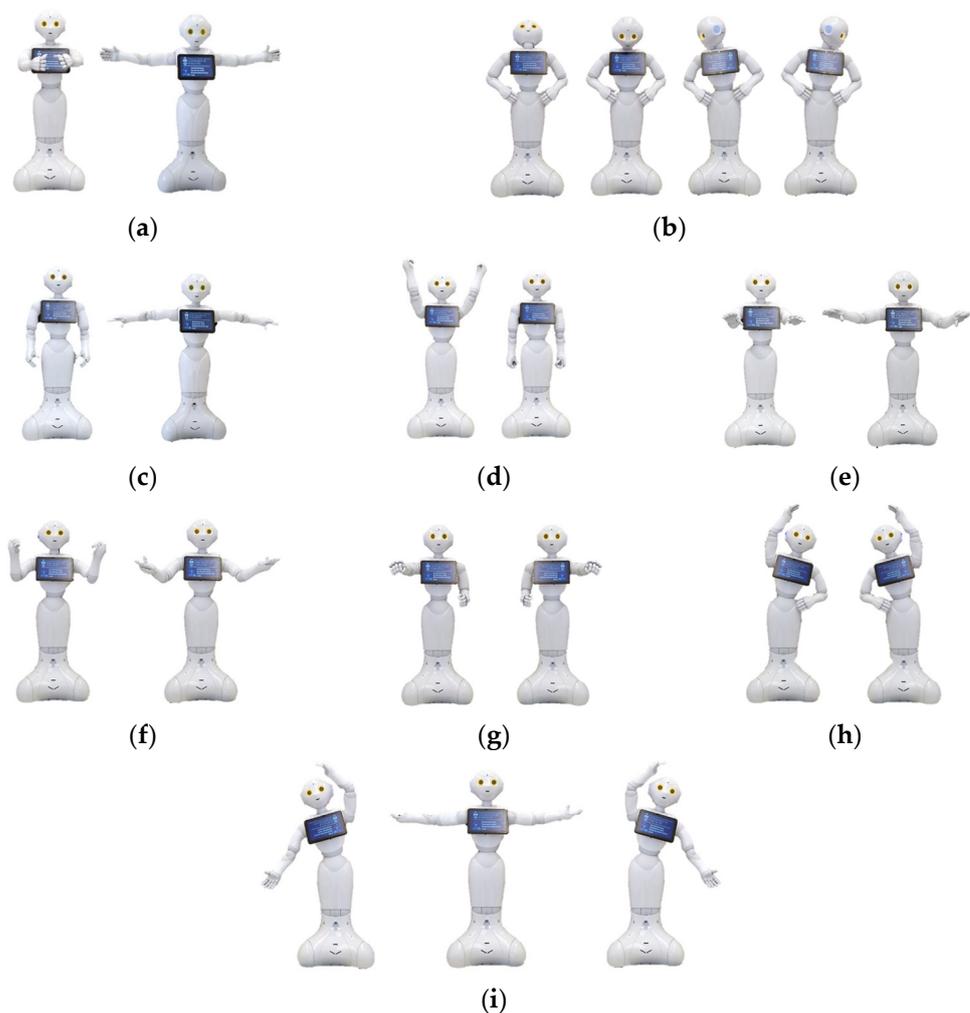


Figure 1. The nine upper-body exercises designed for Pepper: (a) open arm stretches; (b) neck exercises; (c) arm raises; (d) downward punches; (e) breaststrokes; (f) open/close hands; (g) forward punches; (h) lateral trunk stretches; and (i) two-arm lateral trunk stretches.

Our proposed robot architecture for exercise facilitation, Figure 2, is comprised of five modules: (1) Exercise Monitoring, (2) Exercise Evaluation, (3) User State Detection, (4) Robot Emotion, and (5) Robot Interaction. The Exercise Monitoring module tracks a user's skeleton using a Logitech BRIO webcam with 4K Ultra HD video and HDR to estimate the user's body poses during exercise. The detected poses, in turn, are used as input into the Exercise Evaluation Module, which uses the Goal Attainment Scale (GAS) [30] to determine and monitor the user's exercise goal achievements to determine performance

over time. User states are determined via User State Detection Module in order for the robot to provide feedback to the user while they exercise. User state comprises three submodules: (1) Engagement Detection, which detects the user engagement through visual focus of attention (VFOA) from the 4K camera; (2) Valence Detection to determine the user valence from an EEG headband sensor; and (3) Heart Rate Detection, which monitors the user’s heart rate from a photoplethysmography (PPG) sensor embedded into a wristband. The detected user valence and engagement are then used as inputs to the Robot Emotion Module to determine the robot emotions using an n -th order Markov Model. Lastly, the Robot Interaction Module determines the robot exercise-specific behaviors based on the robot’s emotion, the user’s detected exercise poses, and heart rate activity, and displays them using a combination of nonverbal communication consisting of eye color and body gestures, vocal intonation as well as speech. The details of each module are discussed below and in Appendix A.

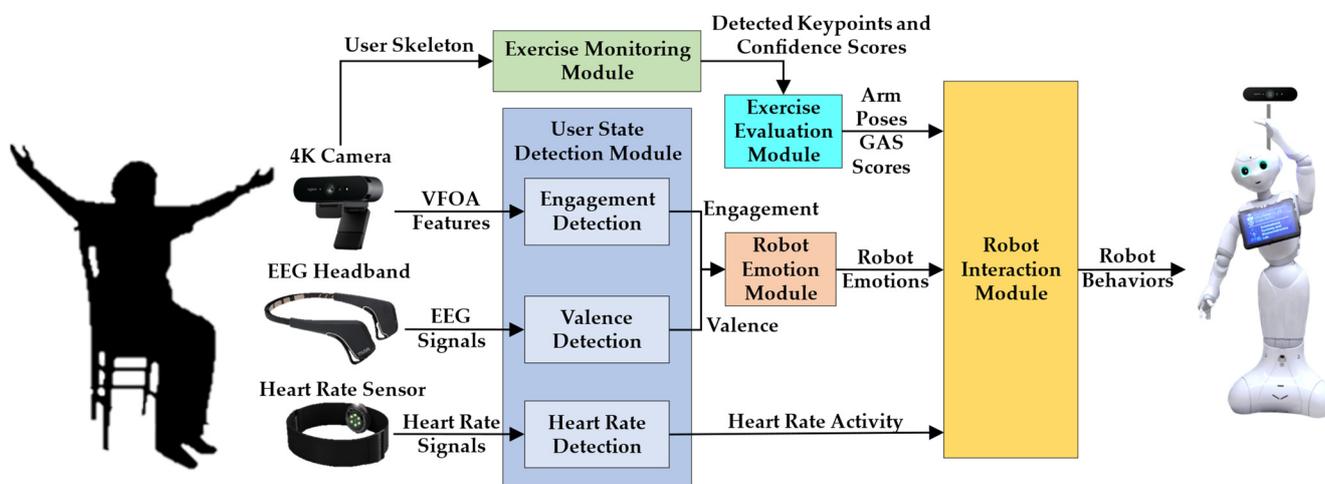


Figure 2. Robot exercise facilitation architecture.

3.1. Exercise Monitoring Module

The Exercise Monitoring Module detects if the user performs the requested exercise poses using the 4K camera placed behind the robot. It tracks the spatial positions (p_x, p_y) of keypoints on the user body detected by the OpenPose model [31] including: (1) five facial keypoints, (2) 20 skeleton keypoints, and (3) 42 hand keypoints. The hand keypoints detection is only enabled when the user is implementing open/close hand exercises. In addition, the OpenPose model also provides the confidence score, S , for each detected keypoint, which can be used to represent the visibility of each keypoint [31]. The OpenPose model is a convolutional neural network model trained on the MPII Human Pose dataset, COCO dataset, and images from the New Zealand Sign Language Exercises [31,32]. Once keypoints are acquired, the pose of the user is classified using Random Forest classifier, which achieves an average pose classification accuracy of 97.1% over all exercises considered. Details of the keypoints used, the corresponding features extracted from the keypoints for each exercise, and pose classification are presented in Appendix A.1.

OpenPose can detect up to 19 people at once [31]. However, based on the field of view of the Logitech Brio 4K camera and the participants of the group exercise sessions sitting in a semi-circle around the robot such that they would not occlude each other, the maximum number of users that can be monitored in a single exercise session is 10 participants.

3.2. Exercise Evaluation Module

In order to evaluate exercise progress over time, the Goal Attainment Scale (GAS) is used. GAS is a measurement used in occupational therapy to quantify and assess a person’s progress on goal achievement [33]. GAS is a well-known assessment which uses measures

that are highly sensitive to evaluating change over time [34,35]. Concurrent validity of GAS was assessed in [36] by correlation with the Barthel Index ($r = 0.86$) and the global clinical outcome rating ($r = 0.82$) for older adults. The repeatable measures can not only provide insight on individual exercise progress but can also be scaled to allow for comparison of change within and between groups of older adults with possibly unique goals [37]. GAS has been used in therapy sessions with robots to evaluate user performance during social skills improvement [38], and motor skills development [39]. The advantages of using GAS are that the goals can be customized based on the specific needs of an individual or a group, GAS then converts these goals into quantitative results to easily evaluate a person's progress towards them [33].

In this work, GAS is used to evaluate a person's progress on exercise performance. Each goal is quantified by five GAS scores ranging from -2 to $+2$, based on a user completing the number of repetitions for each exercise [33]; with -2 not performing all of the repetitions; -1 and 0 all repetitions were implemented with partially pose completion or completed pose repetitions were only performed for less than half of repetitions; $+1$ and $+2$ completed poses were performed for more than half of total repetitions or all repetition were performed correctly. The number of repetitions for each exercise is defined as 8 for the first week and 12 for subsequent weeks and is based on the recommendation by the U.S. National Institute of Aging (NIA) [40]. The details of the GAS score criteria for each exercise are summarized in Table 1.

Table 1. GAS score for performing exercises.

Score	Predicted Attainment
-2	Perform less than 8 (12) repetitions
-1	Perform at least 8 (12) repetitions with partially complete poses only
0	Perform at least 8 (12) repetitions and achieve complete poses for less than 4 (6) repetitions
$+1$	Perform at least 8 (12) repetitions and achieve complete poses for at least 4 (6) repetitions
$+2$	Perform at least 8 (12) repetitions and achieve complete poses for at least 8 (12) repetitions of the total repetitions

The robot computes one GAS score for each exercise, g_i , where i is the index of each exercise. This computation is completed by monitoring the exercise of the user using the Exercise Evaluation Module and estimating its GAS score. The performance of the Exercise Evaluation Module in estimating the GAS score for each exercise is detailed in Appendix A.2.

After the scores are determined for each exercise, a GAS T-Score, T , which is a singular value that quantifies the overall performance of a user during an exercise session based on all GAS scores combined, can be computed for each user at the end of each exercise session [33]:

$$T = 50 + \frac{10 \sum w_i g_i}{\sqrt{0.7 \sum w_i^2 + 0.3 \sum w_i^2}} \quad (1)$$

where w_i is the corresponding weight for each score. In our work, an equal weight (i.e., $w_i = 1$) is selected for each exercise score as we considered each exercise as equally important. T-scores range from 30 to 70, with 30 indicating that the user did not perform any exercises at all and 70 indicating complete poses for all repetitions.

3.3. User State Detection Module

The User State Detection Module determines: (1) user valence, (2) engagement, and (3) heart rate. The quality of user experience with social robots during HRIs can be determined based on their positive or negative affect (i.e., valence) and the ability of the robot to engage users in the activity (i.e., engagement). Furthermore, to ensure users do not

overexert themselves, we use heart rate to determine that it is not above the upper limit that their cardiovascular system can handle during physical activities [41].

3.3.1. Valence

The user valence refers to the user's level of pleasantness during the interaction, which can represent whether the interaction with the robot is helpful or rewarding [42]. The valence detection model is adapted from our previous work [43]. The model uses a binary classification method to determine positive or negative valence from the users, which is consistent with the literature [44–47]. The EEG signals are measured using a four-channel dry electrode EEG sensor, InteraXon Muse 2016. Appendix A.3.1 describes the extracted features obtained from the EEG signals, and the three hidden layer Multilayer Perceptron Neural Network classifier used to classify valence. A classification accuracy of 77% was achieved for valence.

3.3.2. Engagement

For engagement, we use visual focus of attention (VFOA) to determine whether a user is attentive to the robot and exercise activity. VFOA is a common measure of engagement used in numerous HRI studies [48–50], including studies with older adults [51,52]. The robot is able to detect the user engagement as either engaged or not engaged towards the robot based on their VFOA. Two different VFOA features are used for classifying engagement: (1) the orientation of the face, θ_f ; and (2) the visibility of the ears, measured through the confidence scores of their respective keypoints (S_3 and S_4) as detected by the OpenPose model [31]. The orientation of the face is estimated using the spatial positions of the facial keypoints (i.e., eyes and nose) detected by the OpenPose model. Appendix A.3.2 explains the features and k-NN classifier for engagement detection. A classification accuracy of 93% was achieved for engagement.

3.3.3. Heart Rate

The maximum heart rate (MHR) of a person can be estimated by [53]:

$$MHR = 220 - age \quad (2)$$

The maximum target heart rate during anaerobic exercises (e.g., strength building and flexibility) is determined to be 85% of the MHR [53]. We use an optical heart rate sensor, Polar OH1, to measure the user heart rate in bpm at 1 Hz throughout each exercise session. During exercise facilitation, the measured heart rate signals are sent directly via Bluetooth from the heart rate sensor to the Heart Rate Detection sub-module in the User State Detection Module of the Robot Exercise Facilitation Architecture. In this submodule, heart rate measurements are monitored to ensure they remain below the upper threshold (85%) of the MHR to prevent overexertion. If such a condition occurs, the experiment is stopped, and the user is requested to rest by the robot. The heart rate measurements are also saved to a file for post-exercise analysis.

3.4. Robot Emotion Module

This module utilizes a robot emotion model that we have previously developed [54,55], which considers the history of the robot's emotions and the user states (i.e., user valence and engagement) to determine the robot's emotional behavior. This model has been adapted herein for our HRI study.

We use an n th order Markov Model with decay to represent the current robot emotion based on the previous emotional history of the robot [55]. An exponential decay function is used to incorporate the decreasing influence of past emotions as time passes.

The robot emotion state–human affect probability distribution matrix was trained with 30 adult participants (five older adults) prior to the HRI study to determine the transition probability values of the robot's probability distribution matrix. The model is trained in such a way that given the robot's emotional history and the user state, it chooses the emotion

that has the highest likelihood of engaging the user in the current exercise step. Initially, the probabilities of the robot's emotional states are uniformly distributed to allow each emotion to have the same probability to be chosen and are then updated during training for the exercise activity. From our training, when the user has negative valence, the robot would display a sad emotion, which had the highest probability of user engagement for this particular state. Other patterns we note from training are if the robot's previous emotion history was sad and the user had negative valence or low engagement, the robot displays a worried emotion which had the highest probability of user engagement. Furthermore, when the user has positive valence and/or high engagement, the robot displays positive emotions such as happy and interested based on the likelihood of these emotions being the most engaging for these scenarios while considering the robot's emotion history. Additional details of the Robot Emotion Module are summarized in Appendix A.4.

3.5. Robot Interaction Module

The Robot Interaction Module utilizes Finite State Machines (FSMs) to determine the robot behaviors for both types of interactions with the users: (1) one-on-one, and (2) group interaction, Figure 3.

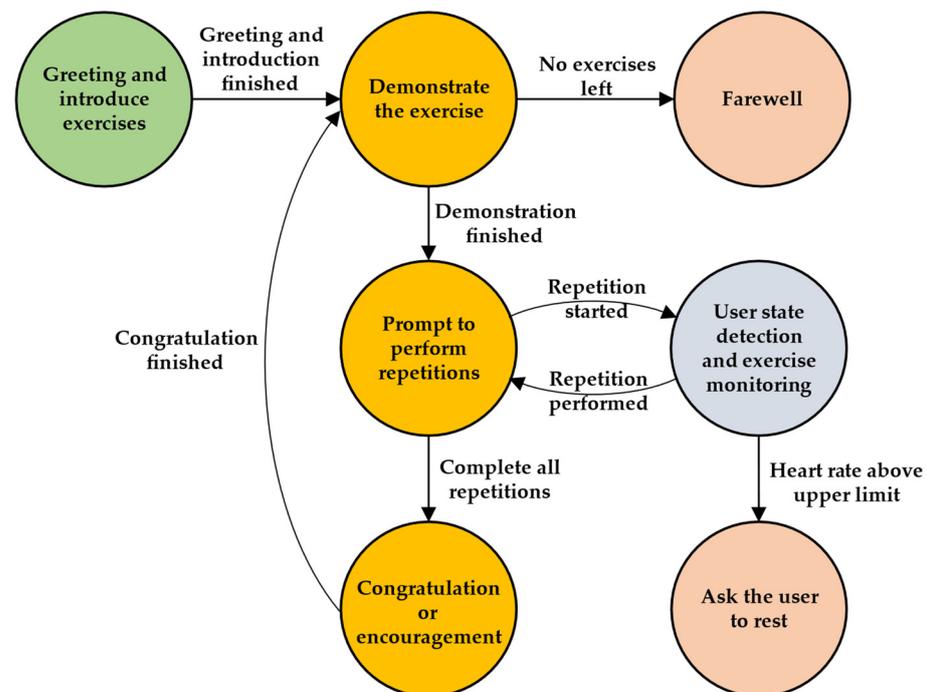


Figure 3. FSM of the robot interaction module.

During the exercise session, the user states (i.e., valence, engagement, and heart rate) are estimated to determine the robot emotions (represented by happy, interested, worried, and sad) via the Robot Emotion Module, and body poses are tracked using both the Exercise Evaluation Module and User State Detection Module to determine performance via GAS and engagement, respectively. The robot displays its behaviors using a combination of speech, vocal intonation, body gestures, and eye colors. If no movement is detected, the robot prompts the user to try again to perform the exercises. In addition, if the user's heart rate is above the upper threshold (85% of their MHR), the robot would terminate the exercise sessions and ask the user to rest. At the end of each exercise, the robot congratulates or encourages the user based on performance and affect. After finishing all the exercises, the robot says farewell to the users. For the group interaction, the interaction scenario is similar. The overall response time of the robot during exercise facilitation is approximately

33 s. This is the time it takes for the robot to respond with its corresponding emotion-based behavior based on user states once the user or user group performs a specific exercise.

In a group exercise session, if one individual user’s heart rate exceeds the upper threshold of the MHR, the exercise session is paused, and all participants are asked to rest until that user’s heart rate is reduced below this threshold. This approach is chosen to promote group dynamics among the participants. Namely, encouraging those in a group setting to exercise (and to rest) together can further motivate participation in and adherence to exercise [56].

In HRI, positive and negative user valence has been correlated with liking and disliking certain HRI scenarios [57,58]. We have designed the robot’s verbal feedback to, therefore, validate its awareness of this user state, and in turn encourage the user to continue exercising to promote engagement [54,59]. Table 2 presents examples of the robot activity-specific behaviors during the exercise sessions.

Table 2. Robot behavior for exercise sessions.

Stage	Non-Verbal	Verbal
Greeting	<p>Waves arms to the user</p> 	<p>“Hello, my name is Pepper, your personal exercise coach. We are going to do nine different exercises together. Each cycle of an exercise has n repetitions. If you are tired, please stop doing the exercise, don’t force yourself!” “Are you ready?”</p>
Introduce exercise (every week, visual shows exercise)	<p>Performs the poses for the exercise</p> 	<p>“First, we will do an exercise called open arm stretches. Let me show you how to do it.”</p>
Introduce exercise (only week one shows with detailed instructions)	<p>Performs the poses for the exercise</p> 	<p>“Start by bringing your arms up to the middle of your chest. And open them sideways, like this. And then close your arms. Finally put your arms back down.”</p>
Prompt to perform repetitions	<p>Performs the poses for the exercise</p> 	<p>“We are going to do eight repetitions.” “Let’s get started.” “Eight, seven . . . , last one!”</p>
Congratulate (happy)	<p>Happy robot emotion display Example: Arms open</p> 	<p>“Excellent job! I really enjoy doing this exercise with you.”</p>
Congratulate (interested)	<p>Interested robot emotion display Example: Nodding</p> 	<p>“You did the exercise really well!”</p>
Encouragement (worried)	<p>Worried robot emotion display Example: Covering face</p> 	<p>“Let me know if you are feeling tired. Otherwise, we are going to move to the next exercise.”</p>

Table 2. Cont.

Stage	Non-Verbal	Verbal
Encouragement (sad)	<p>Sad robot emotion display Example: Scratching head</p> 	<p>“It’s too bad you did not like this exercise. Hopefully you will like the next one more.”</p>
Farewell	<p>Waves goodbye to the user</p> 	<p>“I hope the rest of your day goes well. Let’s do this again sometime. Bye for now!”</p>

4. Exercise Experiments

A long-term HRI study was conducted at the Yee Hong Centre for Geriatric Care in Mississauga to investigate the autonomous facilitation of exercises with Pepper and older adult residents for both one-on-one and group settings. The Yee Hong Centre has 200 seniors living in the facility with an average age of 87, who require 24-h nursing supervision and care to manage frailty and a range of complex chronic and mental illnesses [60]. The majority of residents speak Cantonese or Mandarin and very few can speak English as a second language. In addition, 60% have a clinical diagnosis of dementia [60].

4.1. Participants

The following inclusion criteria for the participants was used: residents who (1) were at least 60 years or older; (2) were capable of understanding English and/or Mandarin with normal or corrected hearing levels; (3) were able to perform light upper body exercise based on the Minimum Data Set (MDS) Functional Limitation in Range of Motion section with a score of 0 (no limitation) or 1 (limitation on one side) [61]; (4) had no other health problems that would otherwise affect their ability to perform the task; (5) were capable of providing consent for their participation; and (6) had no or mild cognitive impairment (e.g., Alzheimer’s or other types of dementia) as defined by the Cognitive Performance Scale (CPS) with a score lower than 3 (i.e., intact or mild impairment) [62] or the Mini-Mental State Exam (MMSE) with a score greater than 19 (i.e., normal or mild impairment) [63]. Before the commencement of the study, a survey was conducted to determine participants’ prior experience with a robot. All of the participants had either never seen a robot or had seen one through a robot demonstration. None had interacted with a robot previously.

A minimum sample size of 25 participants was determined using a two-tailed Wilcoxon signed-rank test power analysis with an α of 0.05, power of 0.8, and effect size index of 0.61. Our effect size is similar to other long-term HRI studies with older adults [64,65], which had effect sizes of 0.61 and 0.6, respectively.

In general, women outnumber men in long-term care homes, especially in Canada and several Western Countries [66], with more than 70% of residents being women [67]. This was also evident in our participant pool. Using our inclusion criteria, 31 participants were recruited of which 27 (3 male and 24 female) completed the study. The participants ranged in age from 79 to 97 ($\bar{x} = 88.93$, $s = 5.62$), and participated on average in 15 sessions. Written consent was obtained from each participant prior to the start of the experiments. Ethics approval was obtained from both the University of Toronto and the Yee Hong Centre for Geriatric Care.

4.2. Experimental Design

At long-term care facilities, exercise programs with older adults are often delivered in group-based or individual-based settings [68]. The participants in the study were organized

randomly into one-on-one or group-based exercise sessions with the robot. The exercise sessions took place twice a week for approximately two months (16 exercise sessions in total for both one-on-one and group settings). Each exercise session was approximately one hour in duration. Eight participants participated in the one-on-one sessions, whereas 19 participants were in the group sessions, which were further split into a group of 9 and a group of 10 participants, respectively. The size of the group sessions was consistent with the group size of the exercise sessions established in the long-term care home.

For the group sessions, participants were seated in a half circle with an approximate radius of 2 m in front of the robot in each session, Figure 4. Two random participants from each group session, who sat in the center of the half circle, wore the EEG headband and heart rate sensor for the robot to detect their valence and heart rate. This allowed for direct line-of-sight to the camera for estimating body poses during exercising; however, any two participants could wear the sensors. For the one-on-one sessions, each participant was seated 1.5 m directly in front of the robot while wearing the EEG headband and heart rate sensor, Figure 5. We used adjustable EEG headbands. The headband was adjusted in size for each participant and secured to their forehead by the research team prior to each exercise session. As the exercises did not include any fast or high-impact neck or upper body motions, movements to the band itself were minimized with no anomalies in the sensory data obtained during interactions.



Figure 4. Group exercise facilitation.

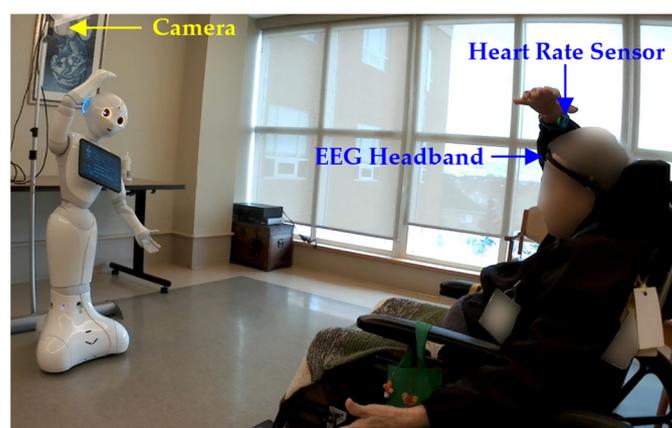


Figure 5. One-on-one exercise facilitation.

4.3. Experimental Procedure

During each single one-on-one and group exercise session, the robot autonomously facilitated the aforementioned upper body exercises consecutively for the full duration, with no breaks in between exercises. The robot facilitated the exercise sessions mostly in Mandarin for both group and one-on-one sessions. There were two participants in the one-on-one sessions who interacted in English based on their preference.

As the same exercise sessions were repeated every week, new exercises were added to them to increase complexity and exercise variation as time went by. These were in the form of exercises that combined 2 to 3 of the individual exercises together—for example, lateral trunk stretches, forward punches, and downward punches. The combination of two exercises were added during the third week while the combination of three exercises were added during the sixth week. Increasing the complexity of the exercises over time has been commonly implemented in exercise programs with older adults in order to gradually improve their functional capacity [69]. In addition, adding variation to the exercises is beneficial in engaging users over time and improving their adherence to the program [70].

For the one-on-one interaction, the exercise sessions begin with the robot greeting the user and providing exercise information (e.g., number of exercises and repetitions). Next, the robot demonstrates each exercise and asks the user to follow it for n number of repetitions. In the first week, the robot would demonstrate each exercise with step-by-step instructions. Subsequent weeks as the users become familiar with the exercises, the robot only visually demonstrates one repetition of the exercises without the step-by-step instructions. In addition, the number of the repetitions of each exercise increases from 8 to 12 in order to increase the level of difficulty after the first week. Video recordings were taken of every one-on-one and group session, with sensor recordings taken from the same two participants in the group sessions and all participants in the one-on-one sessions.

4.4. Measures

The user study is evaluated based on the: (1) user performance over time via GAS; (2) measured valence and engagement during the activity; (3) the robot's adapted emotions based on user valence and engagement; (4) user self-reported valence during the interaction using the Self-Assessment Manikin (SAM) scale [71]; and (5) a five-point Likert scale robot perception questionnaire adapted from the Almere model [72] with the following constructs: acceptance (C1), perceived usefulness and ease of use (C2), perceived sociability and intelligence (C3), robot appearance and movements (C4), and the overall experience with the robot, Table A6 in Appendix B.

Self-reported valence is obtained during the first week, after one month, and at the end of the study (i.e., two months) and the robot perception questionnaire was administered after one month and two months to investigate any changes in user valence towards the robot and perceptions of the robot as the HRI study progressed.

5. Results

5.1. Exercise Evaluation Results

The average GAS T-scores for the one-on-one sessions, group sessions, and all users combined for the first week, one month, two months, and the entire duration are detailed in Table 3. Participants, in general, from both one-on-one sessions and group sessions were able to achieve an average GAS T-score of 64.11, which indicates that they were able to follow all repetitions and perform complete exercise poses for more than half of the total repetitions by the end of the study.

Table 3. Average GAS T-score for different exercise session types during the first week, week of one-month questionnaire, week of two-months questionnaire, and entire duration of study.

Session Type	First Week ($\bar{x} \pm s$)	One Month ($\bar{x} \pm s$)	Two Month ($\bar{x} \pm s$)	Entire Duration ($\bar{x} \pm s$)
One-on-One	62.92 ± 6.03	64.29 ± 6.20	63.19 ± 5.13	64.03 ± 4.92
Group	64.72 ± 2.46	67.50 ± 1.40	67.78 ± 0.91	66.67 ± 2.11
All users	63.28 ± 5.50	64.95 ± 5.69	64.11 ± 4.94	64.27 ± 4.80

5.1.1. One-on-One Sessions

All eight participants in the one-on-one session complied with the robot exercises and had an average GAS T-score of 64.03 ± 4.92 over the duration of two months. User 7,

however, only participated in 12 of the 16 exercise sessions and watched the robot in the remaining four. The user indicated that the robot moved too fast for them to always follow it. This resulted in an average GAS T-score of 51.24 based on the days they complied with the robot.

5.1.2. Group Sessions

The users in the group sessions complied with the robot exercises throughout the sessions and had an average GAS T-score of 66.67 at the end of two months.

We investigated if there was any improvement in GAS T-score between the first week and after one month, and the first week and after two months. In general, an increase in the average GAS T-score was observed between the first week for one-on-one sessions ($\bar{x} = 62.92$, $s = 6.03$) and group sessions ($\bar{x} = 64.72$, $s = 2.46$) and after one month for both one-on-one ($\bar{x} = 64.29$, $s = 6.20$) and group sessions ($\bar{x} = 67.50$, $s = 1.40$), as well as at the end of two months for one-on-one sessions ($\bar{x} = 63.19$, $s = 5.13$) and group sessions ($\bar{x} = 67.78$, $s = 0.91$). This increase in GAS T-score indicates that users achieved more repetitions of the exercises, and that more of the completed repetitions were achieved with a complete pose as opposed to a partially complete pose. These improvements may suggest improved muscle strength and range of motion of the users. Statistical significance was found in the group sessions using a non-parametric Friedman test; $X^2 = 6.5$, $p = 0.039$. However, post-hoc non-parametric Wilcoxon Signed rank tests with Bonferroni correction of $\alpha = 0.016$ showed no statistical significance between the first week and one month ($W = 0.00$, $Z = -1.633$, $p = 0.102$, $r = -0.24$), between the first week and after two months, ($W = 0.00$, $Z = -1.841$, $p = 0.66$, $r = -0.13$), or between one month and after two months ($W = 0.00$, $Z = -1.00$, $p = 0.317$, $r = -0.26$). We postulate that the lack of significance could be due to participants becoming familiar with the exercises. As time went by, they were required to perform the exercises for longer and engage in more challenging sequences of repeated exercises, as new combinations of exercises were added to the exercise sessions for variation. This was especially true at 6 weeks where the complexity of the exercises was the highest. Nonetheless, there was an overall increase in the GAS T-scores from week one to the end of the two-months, which demonstrates exercise goal achievement. A non-parametric Friedman test showed no statistical significance was found in the overall GAS T-scores between the group and one-on-one sessions after one week, after one month, and after two months; $X^2 = 5.4$, $p = 0.067$.

We also investigated if there were differences in GAS T-scores between the one-on-one and the group sessions during the entire study. In general, the group sessions had a higher GAS T-score ($\bar{x} = 66.67 \pm 2.11$) compared to the one-on-one session ($\bar{x} = 64.03 \pm 4.92$) throughout the study. A statistical significant difference was found in GAS T-scores using a non-parametric Mann–Whitney U-test: $U = 174.5$, $Z = 2.10$, $p = 0.018$, $r = 0.10$. We postulate that the higher GAS T scores in the group sessions were due to high task cohesiveness within the group and people's general preference to exercise in groups.

5.2. User State Detection and Robot Emotion Results

The average detected valence measured using the EEG sensor, engagement based on the users' VFOA, and heart rate for users in the one-on-one, group, and for all participants are presented for each time period in Table 4. Participants, on average, from both one-on-one sessions and group sessions had positive valence towards the robot for 87.73% of the interaction of the time. For the group sessions, due to the large number of participants we were unable to accurately measure engagement for every participant, and therefore focus our discussions on the one-on-one sessions. All participants remained engaged towards the robot, regardless of the level of complexity of the exercises, for 98.41% of the interaction time. In addition, participants had an average heart rate of 82.37 bpm during the interactions, and none of their heart rate exceeded the upper limit of the target range (i.e., 120 bpm for 79 years old and 105 bpm for 97 years old).

Table 4. The average and standard deviation of the percentage of time of the detected positive valence, engagement towards the robot, and heart rate for one-on-one sessions, group sessions, and all users combined.

User States	Time Period	One-on-One ($\bar{x} \pm s$) Percent of Interaction Time	Group ($\bar{x} \pm s$) Percent of Interaction Time	All Users ($\bar{x} \pm s$) Percent of Interaction Time
Positive Valence (%)	First Week	93.04 ± 3.30	92.26 ± 2.94	92.89 ± 3.09
	One Month	90.34 ± 5.65	90.26 ± 11.04	90.32 ± 6.26
	Two Month	89.90 ± 13.69	97.49 ± 1.94	91.42 ± 12.51
	Entire Duration	86.88 ± 17.76	91.13 ± 10.39	87.73 ± 16.68
Engagement (%)	First Week	98.49 ± 0.85	N/A	98.49 ± 0.85
	One Month	98.93 ± 0.91	N/A	98.93 ± 0.91
	Two Month	98.41 ± 1.85	N/A	98.41 ± 1.85
	Entire Duration	97.29 ± 6.82	N/A	97.29 ± 6.82
Heart rate (bpm)	Entire Duration	81.56 ± 3.91	85.61 ± 2.96	82.37 ± 3.97

As most users had positive valence and were engaged towards the robot, the robot displayed happy emotions for the exercise sessions for the majority of these interactions. Figure 6 presents the overall percentage of time the robot displayed each of the four emotions. As can be seen from the figure, the happy emotion was displayed for 83.3%, 85%, and 62% of time during the first week, after one month, and after two months, respectively. The change in the emotions displayed by the robot facilitator can be attributed to the increase in the difficulty of the exercises over time, as detailed in Section 4.2. This increase in exercise difficulty resulted in more users displaying varying valence for which the robot adopted other emotions to encourage the users.

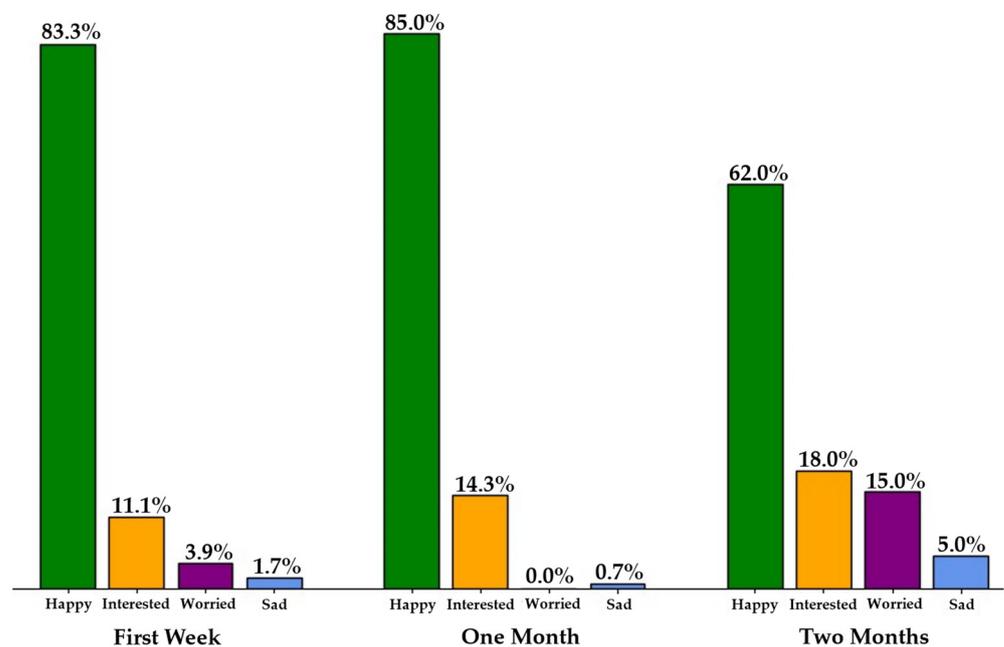


Figure 6. Percentage of time each emotion was displayed by the robot while giving feedback to the users during all exercise sessions for each time period.

Detailed examples of user valence, engagement, and robot emotion for eight users in one-on-one sessions, valence and robot emotions for the two users in the group sessions, and results for all users during the first week, after one month, and after two months are discussed below. We found no statistically significant difference in detected user valence between the three time periods when conducting a non-parametric Friedman test; $X^2 = 0.40, p = 0.819$.

5.2.1. One-on-One Sessions

In all three time periods, the users were engaged above 98% of the time throughout the one-on-one interactions. In the one-on-one sessions after the first week, Figure 7, Users 2–6 had positive valence throughout the session. The robot in turn displayed happy and interested emotions. User 1 had positive valence for the majority of the session, however, they had negative valence only during the open/close hands exercise (denoted as E7). This negative valence was observed to be harder for this user due to the exercise being faster than other exercises (an average of 1.16 s/repetition for E7 versus 3.6 s/repetition for the others). The robot detected this negative valence and in turn displayed a sad emotion to encourage the user to keep trying the exercise.

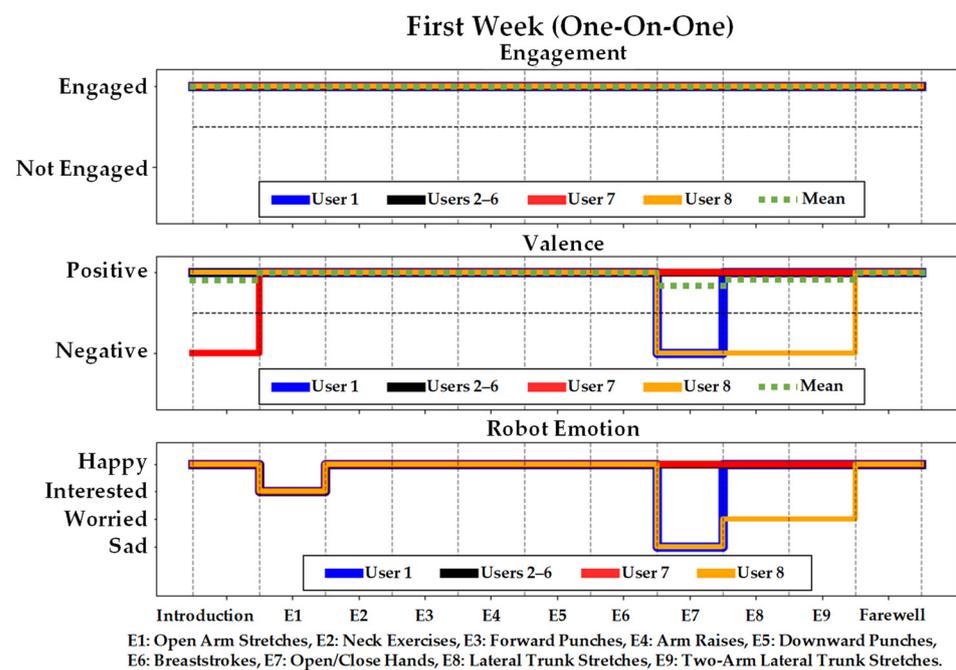


Figure 7. User engagement, valence, and the corresponding robot emotion for one-on-one sessions during the first week.

Alternatively, User 7 first had negative valence during the introduction stage and then had positive valence for the rest of the session. In general, Pepper first displayed the happy emotion during the introduction and transitioned to interested. User 8 also had negative valence during the open/close hands exercise (E7), in addition to the lateral trunk stretches (E8), and two-arm lateral trunk stretches (E9) exercises. The robot displayed a combination of sad and worried emotions during these exercises to encourage the user to continue. In general, User 8 had difficulties performing arm exercises that were faster and had larger arm movements due to their observed upper-limb tremors.

After one month, Figure 8, all users had positive valence, with the exception of User 8 during the introduction and E1 stages. User 8 started with negative valence but then increased to positive valence for the rest of the session. The robot responded by displaying the sad emotion in E1 and then transitioned to interested and happy emotions after detecting positive valence similar to the other users.

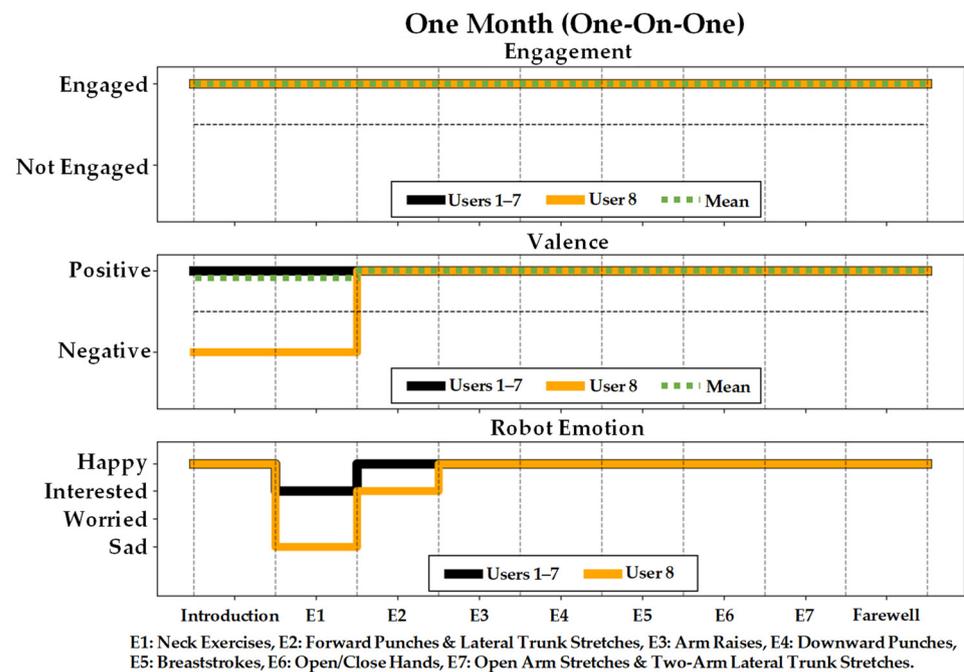


Figure 8. User engagement, valence, and the corresponding robot emotion for one-on-one sessions after one month.

After two months, the users, on average, displayed positive valence throughout the exercise sessions, Figure 9. Four of the eight users always displayed positive valence (Users 2, 3, 5, and 7), for which the robot displayed both interested and happy emotions, Figure 9. Furthermore, two of the four users who experienced negative valence only did so for some exercises and not the entire two-month duration (Users 6 and 8). The overall positive valence of the users was also reported in Table 4, where it was shown that users experience positive valence for approximately 90% of the time after two months of interactions. This result is also consistent with the self-reported positive valence of the users (Section 5.3).

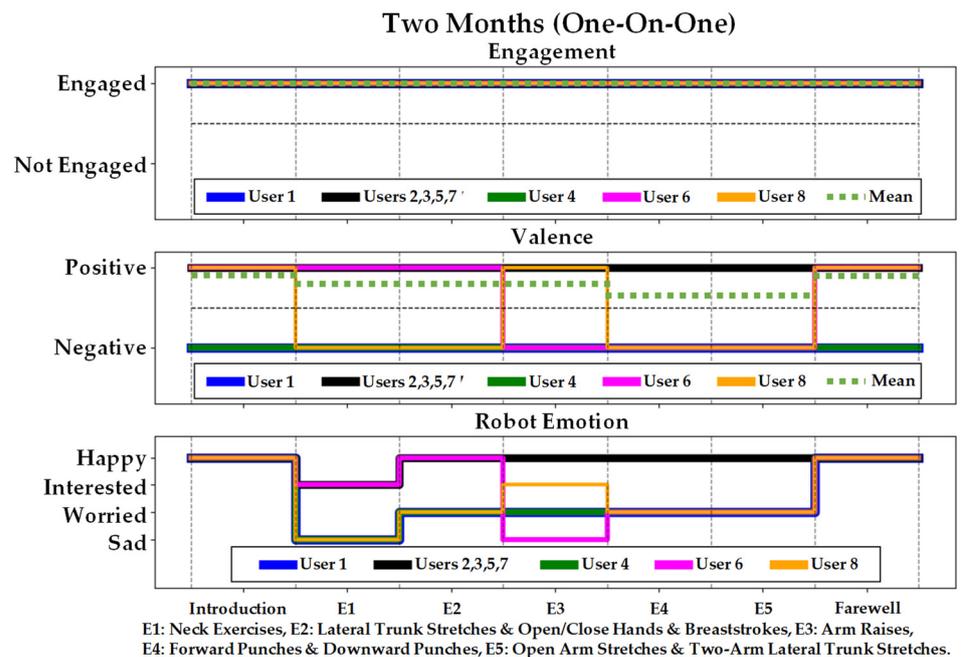


Figure 9. User engagement, valence, and the corresponding robot emotion for one-on-one sessions after two months.

Both Users 1 and 4 had negative detected valence during this entire two-month duration for which the robot displayed sad and worried emotions for encouragement. This increased the robot’s expressions of the worried and sad emotions. User 6 had positive valence except when performing combined arm raises, combined forward and downward punches, and combined open arm stretches with two-arms LTS (E3-E5), for which the robot emotion transitioned to a worried emotion. User 8, again, had negative valence during the neck exercises (E1) and the combination of LTS and open/close hands and breaststrokes (E2), and thus the robot displayed sad and worried emotions. This user also showed negative valence during the neck exercises after one month but not during the first week, which could be due to the fact that it was more strenuous to perform this exercise with the 12 repetitions than with only 8 repetitions. When User 8 transitioned from negative valence to positive valence in arm raises (E3), the robot transitioned to interested. However, User 8 transitioned back to negative valence while performing the combination of forward and downward punches (E4), and also the combination of open arm stretches and two-arm LTS (E5), and the robot displayed worried emotions. User 8 was the only participant to consistently have negative valence during all three sessions, which we believe could be attributed to their physical impairment in completing the exercises.

5.2.2. Group Sessions

For the group sessions, as previously mentioned due to the large number of participants we were unable to accurately measure engagement for every participant, we therefore focus our discussion on user valence. However, it can be noted from our video analysis that the majority of the group focused their attention on the robot throughout each session, as they were consistently performing each exercise. Valence is discussed here for two participants as a representative of the group, Figures 10–12. The participants had an average detected positive valence of 92.26% of the interaction time after the first week, 90.26% after one month, and 97.49% after two months, respectively.

User 9, like most participants in the group session, had positive valence during all time periods. During the first week, User 10 started with positive valence, and transitioned to negative valence during several exercises including arm raises, downward punches, breaststrokes, open/close hands, LTS, and two-arm LTS (E4–E9), which resulted in the robot displaying sad and worried emotions.

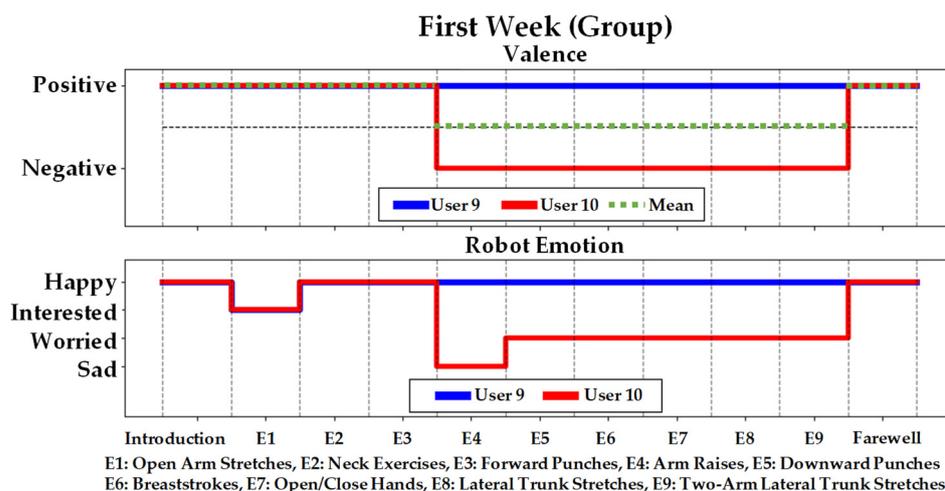


Figure 10. User valence and the corresponding robot emotion for group sessions during the first week.

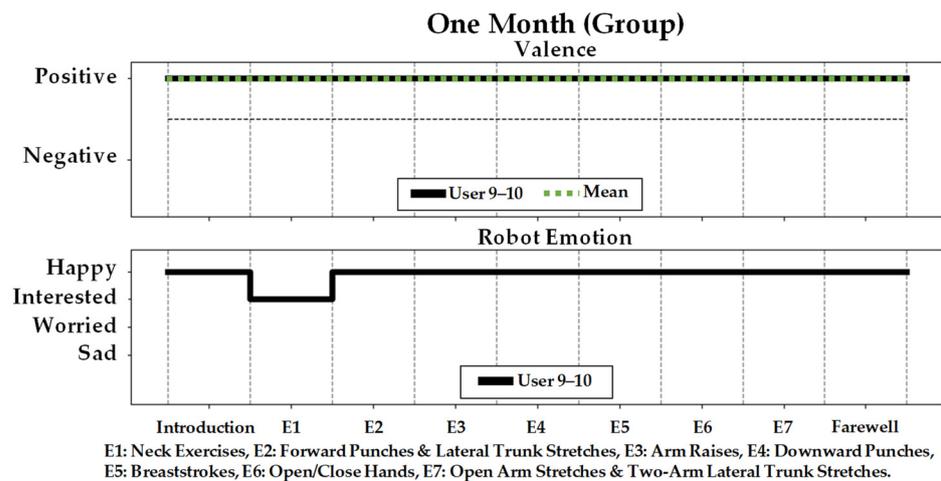


Figure 11. User valence and the corresponding robot emotion for group sessions after one month.

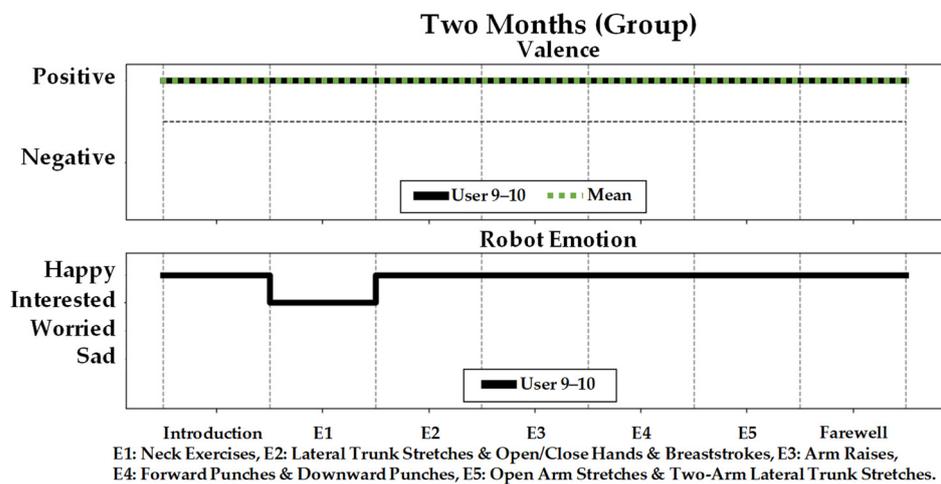


Figure 12. User valence and the corresponding robot emotion for group sessions after two months.

We investigated if there was any difference in detected valence between the one-on-one sessions and the group sessions. In general, detected valence for the group sessions ($\bar{x} = 91.13, s = 10.39$) was higher than the one-on-one sessions ($\bar{x} = 90.31, s = 6.26$). However, no statistical significance was found using a non-parametric Mann–Whitney U-test: $U = 1449.0, Z = -1.070, p = 0.284, r = 1.42$.

5.3. Self-Reported Valence (SAM Scale)

The reported valence from the five-point SAM scale questionnaire measured during the first week, one month, and two months for one-on-one sessions, group sessions, and all users combined are presented in Table 5. The valence is on a scale of -2 (very negative valence), -1 (negative valence), 0 (neutral), $+1$ (positive valence), and $+2$ (very positive valence). All users, in general, reported positive valence throughout the study with the average valence of 1.33, 1.30, and 1.19 after the first week, one month, and two months, respectively.

Table 5. SAM scale results.

Session Type	First Week			One Month			Two Months		
	\bar{x}	s	\tilde{x}	\bar{x}	s	\tilde{x}	\bar{x}	s	\tilde{x}
One-on-One	1.25	0.83	1.50	1.25	0.83	1.50	1.38	0.70	1.50
Group	1.37	0.81	2.00	1.32	1.08	2.00	1.11	0.91	1.00
All	1.33	0.82	2.00	1.30	1.01	2.00	1.19	0.86	1.00

This lower reported valence in the group sessions was due to participants stating they would like the robot to be taller and bigger for them to see it better, as in the group sessions the robot was placed further away from them to accommodate more participants. However, none of the users in the one-on-one sessions had this concern since they were interacting with the robot at a closer distance, Figure 5.

We also investigated if there were any differences between the one-on-one and group sessions during different time periods. One-on-one sessions, in general, had lower reported valence than group sessions during the first week (one-on-one: $\bar{x} = 1.25, s = 0.83, \tilde{x} = 1.50$; group: $\bar{x} = 1.37, s = 0.81, \tilde{x} = 2.00$) and one month (one-on-one: $\bar{x} = 1.25, s = 0.83, \tilde{x} = 1.50$; group: $\bar{x} = 1.32, s = 1.08, \tilde{x} = 2.00$). However, one-on-one sessions had higher valence ($\bar{x} = 1.38, s = 0.70, \tilde{x} = 1.50$) than the group sessions ($\bar{x} = 1.11, s = 0.91, \tilde{x} = 1.00$) after two months. No statistically significant difference was observed for these differences using a non-parametric Mann–Whitney U-test for the first week: $U = 70.0, Z = 0.33, p = 0.37, r = 0.08$, one month: $U = 68.0, Z = 0.45, p = 0.33, r = 0.105$, and two months: $U = 65.0, Z = 0.60, p = 0.27, r = -0.15$.

5.4. Robot Perception Questionnaire

The results from the five-point Likert scale robot perception questionnaire measured after one month and two months for one-on-one, group sessions, and all users combined are summarized in Table A7 (in Appendix B) and Figure 13. For each construct, questions that were negatively worded were reverse-scored for analysis. The internal consistency of each construct measured at each month was determined using Cronbach’s α [73]. The α coefficient for constructs C1–C4, ranged from 0.75 to 0.81 for one-month, and 0.68–0.88 for two-month, respectively. A value of 0.5 and above can be considered as acceptable for short tests [74,75].

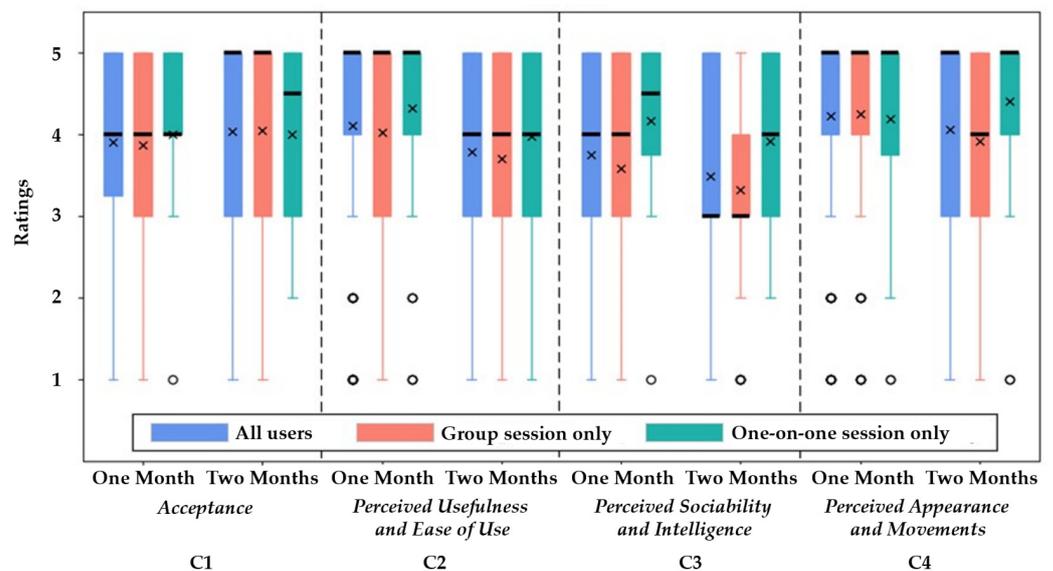


Figure 13. Questionnaire results in box plots for each construct with different user groups with quartiles (box), min-max (whisker), median (black line), and mean (x), and outliers (circles).

5.4.1. Acceptance

Robot Acceptance (C1) results showed that participants from both user groups (one-on-one and group-based) had high acceptance of the robot at the end of two months ($\tilde{x} = 4.0$, $IQR = 1.75$). The participants enjoyed using the robot for exercising ($\tilde{x} = 5.0$, $IQR = 2.00$) and more than half of them (67%) reported they would use the robot again ($\tilde{x} = 5.0$, $IQR = 2.00$) after two months of interaction. They found the sensors comfortable to wear (negatively worded, $\tilde{x} = 4.0$, $IQR = 2.00$).

In general, an increase in this construct was observed from one month ($\tilde{x} = 4.0$, $IQR = 1.75$) to two months ($\tilde{x} = 5.0$, $IQR = 2.00$) for all participants. However, there was no statistically significant difference found using a non-parametric Wilcoxon Signed-rank test: $W = 238.5$, $Z = -0.49$, $p = 0.63$, $r = -0.10$.

On average, participants in the one-on-one sessions had similar Likert ratings as the group sessions for one month (one-on-one: $\tilde{x} = 4.0$, $IQR = 1.00$; group: $\tilde{x} = 4.0$, $IQR = 2.00$) but slightly lower ratings than the group sessions for two months (one-on-one: $\tilde{x} = 4.5$, $IQR = 2.00$; group: $\tilde{x} = 5.0$, $IQR = 2.00$). No statistically significant difference was observed between these two interaction types using a non-parametric Mann–Whitney U-test for the one month: $U = 295.5$, $Z = 0.16$, $p = 0.44$, $r = 0.03$, and two months: $U = 284.0$, $Z = 0.41$, $p = 0.34$, $r = 0.07$.

5.4.2. Perceived Usefulness and Ease of Use

Participants found the perceived usefulness and ease of use (C2) of the robot to be positive after one month ($\tilde{x} = 5.0$, $IQR = 1.00$) with a slight decrease after two months ($\tilde{x} = 4.0$, $IQR = 2.00$). A statistically significant difference between these months was observed using a non-parametric Wilcoxon Signed-rank test: $W = 1950.5$, $Z = -2.71$, $p = 0.01$, $r = 0.18$. Users agreed that the exercises with the robot were good for their overall health ($\tilde{x} = 5.0$, $IQR = 1.00$), more than half of them (67%) found the robot was helpful for doing exercises (negatively worded, $\tilde{x} = 1.0$, $IQR = 2.00$) and 63% also believed the robot motivated them to exercise ($\tilde{x} = 4.0$, $IQR = 2.00$). The majority also (96%) found the robot clearly displayed each exercise ($\tilde{x} = 4.0$, $IQR = 1.00$) and trusted (74%) its advice (negatively worded, $\tilde{x} = 2.0$, $IQR = 1.50$). However, as users reported that they could not set up the equipment (e.g., robot and computer) by themselves, they were neutral about the ease of use of the robot (negatively worded, $\tilde{x} = 3.0$, $IQR = 2.00$).

In general, participants in the one-on-one sessions had similar ratings for this construct as the group sessions both after one month (one-on-one: $\tilde{x} = 5.0$, $IQR = 1.00$; group: $\tilde{x} = 5.0$, $IQR = 2.00$) and two months (one-on-one $\tilde{x} = 4.0$, $IQR = 2.00$; group: $\tilde{x} = 4.0$, $IQR = 2.00$). However, there was no statistically significant difference for perceived usefulness and ease of use using a non-parametric Mann–Whitney U-test for both one month: $U = 3355.5$, $Z = 1.21$, $p = 0.11$, $r = -0.10$, and two months: $U = 3351.0$, $Z = 1.14$, $p = 0.13$, $r = -0.10$.

5.4.3. Perceived Sociability and Intelligence

In general, participants in one-on-one sessions (one month: $\tilde{x} = 4.5$, $IQR = 1.25$; two months: $\tilde{x} = 4.0$, $IQR = 2.00$) found the robot more sociable and intelligent than those in the group sessions (one month $\tilde{x} = 4.0$, $IQR = 2.00$; two months: $\tilde{x} = 3.0$, $IQR = 1.00$). Statistically significant differences were found for both time periods using a non-parametric Mann–Whitney U-test after one month: $U = 514.0$, $Z = 1.84$, $p = 0.03$, $r = -0.25$, and two months: $U = 490.0$, $Z = 2.10$, $p = 0.01$, $r = -0.28$. As there are more participants in the group sessions and the robot did not respond to the user states of all of them, in general, they were neutral about whether the robot understood what they were doing during the exercises ($\tilde{x} = 3.0$, $IQR = 1.00$) and if the robot displayed appropriate emotions ($\tilde{x} = 3.0$, $IQR = 1.00$). However, the two participants that wore the sensors were able to identify the robot's emotions through eye color (worded negatively, $\tilde{x} = 1.0$, $IQR = 0.00$) and vocal intonation ($\tilde{x} = 4.5$, $IQR = 0.50$). In the one-on-one sessions, they thought the robot understood what they were doing ($\tilde{x} = 4.5$, $IQR = 1.25$) as a result of direct feedback

from the robot, and were able to identify the robot's emotions (70% of participants) mainly through its vocal intonation ($\tilde{x} = 4.0$, $IQR = 1.25$), and 50% through the display of eye colors (negatively worded, $\tilde{x} = 2.0$, $IQR = 3.00$), respectively.

Overall, the perceived sociability and intelligence from both interaction types were positive for one month ($\tilde{x} = 4.0$, $IQR = 2.00$) and then became neutral after two months ($\tilde{x} = 3.0$, $IQR = 2.00$). Though, no statistically significant difference was found between one to two months using a non-parametric Wilcoxon Signed-rank test: $W = 475.5$, $Z = -1.82$, $p = 0.069$, $r = 0.14$.

5.4.4. Robot Appearance and Movements

In general, participants reported positive perception of the robot's appearance and movements ($\tilde{x} = 5.0$, $IQR = 2.00$). Seventy percent of them were able to follow the robot movements (negatively worded, $\tilde{x} = 1.0$, $IQR = 2.00$), found it had a clear voice ($\tilde{x} = 5.0$, $IQR = 1.50$) and were able to understand the robot's instructions (negatively worded, $\tilde{x} = 1.0$, $IQR = 1.00$). In addition, more than half (59%) found the robot's size appropriate for exercising ($\tilde{x} = 4.0$, $IQR = 2.00$) with only 19% from the group sessions preferring the robot to be larger and taller like an adult human.

Participants in the group sessions initially had a high positive rating for this construct ($\tilde{x} = 5.0$, $IQR = 1.00$) similar to those in the one-on-one sessions ($\tilde{x} = 5.0$, $IQR = 1.25$) after one month, but as they engaged more with the robot their ratings became less positive in the group sessions ($\tilde{x} = 4.0$, $IQR = 2.00$), in contrast to the one-on-one sessions, which remained highly positive ($\tilde{x} = 5.0$, $IQR = 1.00$) after two months. A statistically significant difference was found after two months between these sessions: $U = 948.5$, $Z = 1.98$, $p = 0.02$, $r = -0.22$. Some participants (19%) in the group sessions noted that the robot's size could be larger, as they reported slightly less positive rating when asked if the robot's size was appropriate for exercising ($\tilde{x} = 4.0$, $IQR = 1.25$) than the participants in the one-on-one sessions ($\tilde{x} = 4.5$, $IQR = 2.50$).

5.4.5. Overall Experience

At the end of the study, the participants' overall experience (in both one-on-one and group sessions) showed that more than half of them (i.e., 56%) thought their physical health was improved ($\tilde{x} = 4.0$, $IQR = 2.00$) and were motivated to perform daily exercise ($\tilde{x} = 4.0$, $IQR = 2.00$). Overall, for the three participants who had a negative rating (i.e., 15%), two of them reported that they hoped the exercise sessions could be longer with more repetitions while the other one stated that they did not think they would get healthier due to their age-related decline in physical functions. Participants, in general, did not find the weekly sessions confusing (negatively worded, $\tilde{x} = 1.0$, $IQR = 1.00$). In addition, more than half of them (i.e., 56%) reported that they were more motivated about performing daily exercises with the robot ($\tilde{x} = 4.0$, $IQR = 2.00$). In our study, participants were motivated to exercise and to continue exercising on a daily basis. This is consistent with other multi-session robot exercise facilitation studies [7,9] which have shown that motivation leads to improved user performance, such as decrease in exercise completion time [7] and executing exercise movements correctly [9].

5.4.6. Robot Features and Alternative Activities

Participants ranked from 1 to 6 (1 being most preferred and 6 being least preferred) which features of the robot they preferred the most, Table 6, as well as what other activities they would like to do with the robot, Table 7. For both one-on-one and group sessions, the most common preferred feature of the robot was its human-like arms and movements. The group sessions also preferred the robot eyes (tied with number one preference). The least preferred feature was the lower body as participants preferred to see the robot have legs to engage in leg exercises as well. These findings are consistent with previous studies comparing older age groups' preferences in robots against other groups [76–78]. It was

consistently found that older adults preferred features that were more human-like as they were more familiar with human appearances.

Table 6. Ranking of preferred robot features.

Robot Features	Group Sessions	One-on-One Sessions	All Users Combined
Eyes	1 (tied)	3	2
Arms and movement	1 (tied)	1	1
Voice	3 (tied)	2	3
Assistance	5	4 (tied)	5
Size	3 (tied)	4 (tied)	4
Lower body	6	6	6

Table 7. Ranking of preferred activities for the robot to assist with.

Activities	Rank
Dressing	4
Meal eating	5
Meal preparation	6
Play games	1
Reminder	3
Escorting	2

Users ranked playing physical or cognitive stimulating games such as Pass the Ball and Bingo games the highest as they wanted the robot to provide helpful interventions to keep them active in daily life. In addition, as some users had difficulty going to their rooms or to the washroom by themselves, Escorting was ranked next to ensure their safety while going to places.

6. Discussions

6.1. User State Detection and Robot Emotion Results

For the one-on-one sessions, in general, users showed a slight decrease in valence for the entire duration. We postulate that this decrease in valence was related to the change of difficulty/complexity of the exercises in later sessions. The participants found the robot increasingly more difficult to use from one month ($\bar{x} = 1.0$, $IQR = 2.00$) to two months ($\bar{x} = 3.0$, $IQR = 2.00$); however, the robot's behavior functionality remained the same and acceptance of the robot was also found to be high throughout the duration of the study. Several exercise studies have reported that an increase in exercise intensity is correlated with a decrease in affective valence [79,80]. For example, a study on aerobic exercise, such as running, found that people's affective valence began to decline as intensity of the exercise increased [79]. A similar study with older adults using the treadmill found that affect declined across the duration of the one-on-one session and became increasingly more negative as participants became more tired [80]. As mentioned, the intensity of the exercises demonstrated in our study increased from 8 repetitions to 12 repetitions in the second week, a combination of two exercises introduced in the third week and a combination of three exercise introduced in the sixth week. The increase in complexity was noted by the participants as additional comments in the two-month questionnaire, with statements such as "I noticed the exercises are getting more difficult" from User 5, one of the healthier participants.

In the group sessions, the negative valence during the first week for User 10 could be due to the participant not being familiar or comfortable exercising with a robot at first, as this user initially reported before the study that they did not have any prior experience with a robot. Over time and through repetitions, however, the exercises became easier for this user to follow (as observed in the videos) and resulted in them having positive valence during the entire exercise session, Figure 12. Similar robot studies have found that

robot exercise systems become easier to understand over repeated sessions (i.e., from 1 to 3 sessions) based on the number of user help requests and an increase in average exercise completion time [7]. We also note that the mean GAS T-scores of User 10 increased from 61.1 (87.3% exercise completion) in the user's first session to 65.5 (93.5% completion) after the user's second session, showing a 6% overall improvement when exercising with the robot. This is consistent with the literature where user performance measures and exercise completion times have improved through some repetition (i.e., 1 to 4 sessions) [9].

The robustness of the robot's emotion-based behaviors can be defined, herein, as its ability to perform exercise facilitation while adapting to the user states in both the one-on-one and group sessions. The changes in the emotions displayed by the robot, Figures 7–12, illustrated its ability to adapt to changes in user state—allowing the robot to maintain, on average, positive valence and high engagement from the users.

6.2. Self-Reported Valence (SAM Scale)

Both the detected (Figures 7–12) and self-reported valence (Table 5) results show that the majority of users maintained a positive valence throughout the study. Namely, the users generally felt positive valence during the (one-on-one and group) exercise sessions and when having the opportunity to reflect on their experience after the sessions. Our results are consistent to other studies that have also used EEG and self-reported to measure affective responses during and after exercise sessions, however, without a robot [81,82].

Overall, we had a high participation rate for our study with 27 of 31 (87%) participants completing the entire two-month sessions. We postulated that the positive valence with the high participation rate, motivates the use of Pepper as an exercise companion. In [83], it was found that positive affect responses during exercise was consistently linked to a high likelihood in performing exercises in the future. Furthermore, in a non-robot study reporting the role of affective states in the long-term participation of exercise specifically in older adults, similar evidence was found [84]. The study cited positive affective responses during exercise as an influential factor on participant belief on gaining benefit from and participating in regular exercise.

6.3. Acceptance

The overall positive acceptance ratings can be attributed to the clear health benefits of using the robot to motivate and engage users in exercise. In general, the level of acceptance of a technology is high when older adults clearly understand the benefits of using the technology [85,86]. For example, in [87] a robot's ability to motivate users to dance and improve health outcomes was identified as a facilitator for robot acceptance. In our study, User 1 noted that “[they] hope[d] [they] will be healthier” by participating in the robot study, and User 2 noted that “[they] think exercising is important” and reported that they believed a robot that motivated them to exercise was good for them. The aforementioned users had a clear understanding that Pepper was developed to autonomously instruct and motivate them to engage in exercising, resulting in positive acceptance ratings.

6.4. Perceived Usefulness and Ease of Use

The perceived usefulness of the robot was expected to be similar between successive exercise sessions (both one-on-one and group) as Pepper performed the same exercises. Over time, however, there was a decrease. One of the main reasons for such a decrease in this construct was that after getting familiar with the robot, four of the participants reported they wished the robot could have had legs to facilitate leg exercises as well. As the participants in our study became familiar with the robot, their expectations evolved into wanting the robot to perform additional exercises.

6.5. Perceived Sociability and Intelligence

When validating the robot's emotion module, after one month, users strongly agreed that the robot displayed appropriate emotions ($\bar{x} = 5.0$, $IQR = 1.25$), while after two

months, their responses were more neutral ($\tilde{x} = 3.0$, $IQR = 1.00$), however, they agreed that the robot's overall feedback was appropriate ($\tilde{x} = 4.0$, $IQR = 2.00$). We believe this could be due to the robot using the same variation of social dialogue during the two-month duration. In other studies that have observed a similar decrease in perceived sociability over time have noted that this was due to the robot not engaging in on-the-fly and spontaneous dialogue [88,89].

6.6. Robot Appearance and Movements

The position of the participants with respect to Pepper may have influenced their opinion of the robot's appearance and movements. Participants from the group sessions that interacted with the robot were sitting in a semi-circle around the robot. As a result, some participants were further back from the robot than others in the group session and may have also not faced the robot directly, Figure 4, compared to those in the one-on-one sessions. In the literature, the preferred robot size by older adults is strongly dependent on the robot's functionality [90–92]. In our study, as the robot's main functionality is to facilitate exercise, we postulate that some participants from the group sessions preferred the robot to be larger so that they could more clearly see the robot and its movements.

6.7. Considerations and Limitations

We consider our HRI study to be long-term as it is repeated for multiple weeks or months, as compared to short-term studies which consider a single session. This is consistent with the literature, where three weeks was explicitly declared as long-term HRI in [93], or two months in [94,95]. The latter is consistent with our study length. Furthermore, length of time is not the only consideration when determining a study is long-term, as the user group and interaction length are also considered. Our study is considered to be long-term based on the older adult user group and the one-hour duration of each interaction session. Other existing studies that have used a robot to facilitate exercise with older adults have either only done so in a single user study session [7,8,23,24], or had users participate in on average of in a total of three sessions [9] lasting on average of 15 min. The only other long-term exercise study we are aware of with older adults is in [6], which was for a 10-week period.

We investigated if group exercise sessions would foster task cohesion. From analysis of the videos, the participants were fully engaged with the robot and showed high compliance with the exercises (there were no drop-outs observed). Participants were focused on following the exercises the robot displayed as closely as possible and continued to do so as the exercise became more complex later in the study. As future work, it would be interesting to identify existing group dynamics prior to the study and after the study to explore if such dynamics change or directly affect group cohesion and participant affective responses. Research has shown that people prefer to exercise in groups and that group cohesion is directly related to exercise adherence and compliance [96,97].

It is possible that participants were extrinsically motivated to participate in the study from factors that we have not examined. For example, social pressure from peers or care staff could have influenced participation in our study; however, participation was completely voluntary, and no rewards were given; participants were allowed to withdraw at any time. We postulate that the participants that completed this two-month HRI study were more intrinsically driven based on their high ratings in their overall experience and perceived usefulness of the robot, as older adults in have been found to be intrinsically motivated by personal health and benefit, in addition to altruistic reasons [98,99].

In our study, participants in both types of interactions had comparable and high acceptance of the robot and high perceived usefulness and ease of use of the robot, which validates the use and efficacy of an exercise robot for both scenario types. The main difference was that users in the one-on-one sessions perceived the robot as more sociable and intelligent, and had provided higher ratings for the robot's appearance and its movements. We believe that this difference is mainly due to: (1) the group setting design, which di-

rectly resembles an exercise session with the human facilitator, participants were seated in a half circle with an approximate radius of 2 m in front of the Pepper robot. The size and distance to the robot may have affected their opinions on these constructs as some participants mentioned they preferred the robot to be larger so that they could more clearly; and, (2) the ability for the robot to provide personalized feedback to each member of the group in addition to the overall group. We do note, however, that the questionnaire results were relatively consistent in the group interactions with an average IQR of 1.48 for all the constructs. This demonstrated that the robot's emotional behavior being responsive to only a small number of participants in the group did not have a significant impact on their overall experience.

Our recommendation is to consider these aforementioned factors when considering group-based interactions to improve perceptions of the robot as a facilitator. In general, in non-robotic studies, research has shown that people prefer to exercise in groups, including older adults [100,101], and this is worth investigating further in HRI studies with groups while considering the above recommendations. For a robot to be perceived as sociable or intelligent in a group session compared to one-on-one interactions, the robot should provide general group feedback and personalized individual feedback for those in the group that need it. Effective multi-user feedback in HRI remains a challenge due to the majority of research focusing on feedback from one-on-one interactions [102]. Furthermore, group-based emotion models require additional understanding of inter-group interactions, as individuals identifying with the group and having in-group cohesion can also influence user affect [103,104].

The user states of the participants not wearing an EEG headband and heart rate sensor in the group session were not monitored. This was due to the limitations with these Bluetooth devices during deployment, where there would be interference between multiple concurrent connections to the host computer when too many devices were on the same frequency. As we obtained self-reported valence, regardless of this challenge, we found that the self-reported valence was consistent within the overall group, and furthermore the self-reported valence was consistent with the measured valence from the two participants wearing the sensors. In the future, this limitation can be addressed by using a network of intermediary devices that connect to the sensors via Bluetooth and connect to a central host computer through a local network. Other solutions include the use of sensors with communication technology that reduces the restrictions on the number of connected devices (e.g., Wi-Fi-enabled EEG-sensors [105]). Other forms of user state estimation that do not require wireless communication can also be considered. For example, thermal cameras have been used to classify discrete affective states by correlating directional thermal changes of areas of interest of the skin (i.e., portions of the face) with affective states [106,107]. However, this relationship has been mainly with discrete states such as joy, disgust, anger, and stress, and additional investigations would be necessary to determine such a relationship with respect to the continuous scale of valence as considered in this work. Thermal cameras have also been used for heart rate estimation by tracking specific regions of interest of the body [108].

Lastly, the experimental design of our study was developed to accommodate the age-related mobility limitations of older adults. For example, the exercises were conducted with the users in a seated position and their range of motion and speed were designed by a physiotherapist at our partner LTC home. We did note, however, that two participants in the one-on-one sessions experienced some difficulties. As discussed in Section 5.1, User 7 participated in 12 of the 16 exercise sessions, and watched the robot in the remaining four sessions. The robot moved too fast for them to always follow it. Similarly, as discussed in Section 5.2, User 8 had difficulties in performing faster and larger motion arm exercises due to their observed upper-limb tremors.

6.8. Future Research Directions

Future research consists of longer-term studies to investigate health outcomes in long-term care with an autonomous exercise robot and the impact of different robot platforms with varying appearances and functionality on exercise compliance and engagement. Literature has shown that upper-limb exercises can provide benefits in functional capacity, inspiratory muscle strength, motor performance, range of motion, and cardiovascular performance [109–111]. It will be worth exploring if robot-facilitated exercising can have the same outcomes.

Robot attributes such as size and type, adaptive (also reported in [112]) and emotional (also reported [113,114]) behaviors, and physical embodiment [7], can influence robot performance and acceptance when interacting with users. Thus, these factors should be further investigated for robot exercise facilitation with older adults.

Furthermore, we will also investigate incorporating other needed activities of daily living with robots to study their benefits with the aim of improving quality of life of older adults in the long-term care home settings.

7. Conclusions

In this paper, we present the first long-term exploratory HRI study conducted with an autonomous socially assistive robot and older adults at a local long-term care facility to investigate the benefits of one-on-one and group exercise interactions. Results showed that participants, in general, had both measured and self-reported positive valence and had an overall positive experience. Participants in both types of interactions had high acceptance of the robot and high perceived usefulness and ease of use of the robot. However, the participants in the one-on-one sessions perceived the robot as more sociable and intelligent and had higher ratings for the robot's appearance and its movements than those in the group sessions.

Author Contributions: Conceptualization, M.S. (Mingyang Shao), S.F.D.R.A. and G.N.; Methodology, M.S. (Mingyang Shao), S.F.D.R.A. and G.N.; Software, M.S. (Mingyang Shao), S.F.D.R.A. and G.N.; Validation, M.S. (Mingyang Shao), M.P.-H., K.E. and G.N.; Formal analysis, M.S. (Mingyang Shao), M.P.-H. and G.N.; Investigation, M.S. (Mingyang Shao), M.P.-H. and G.N.; Resources, M.S. (Matt Snyder) and G.N.; Data curation, M.S. (Mingyang Shao) and S.F.D.R.A.; Writing—original draft preparation, M.S. (Mingyang Shao) and G.N.; Writing—review and editing, M.S. (Mingyang Shao), M.P.-H. M.S. (Matt Snyder), K.E., B.B. and G.N.; Visualization, M.S. (Mingyang Shao), S.F.D.R.A. and G.N.; Supervision, G.N. and B.B.; Project administration, G.N.; Funding acquisition, G.N. and B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by AGE-WELL Inc., the Natural Sciences and Engineering Council of Canada (NSERC), the Canadian Institute for Advanced Research (CI FAR), the Canada Research Chairs Program, and the NSERC CREATE HeRo program.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank our partner long-term care facility, the Yee Hong Centre for Geriatric Care in Mississauga, and our experiment participants.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Details of the Social Robot Exercise Facilitator

Appendix A.1. Exercise Monitoring Module

Appendix A.1.1. Keypoints and Features

The keypoints acquired from the OpenPose model and their corresponding indices are labeled in Figure A1a.

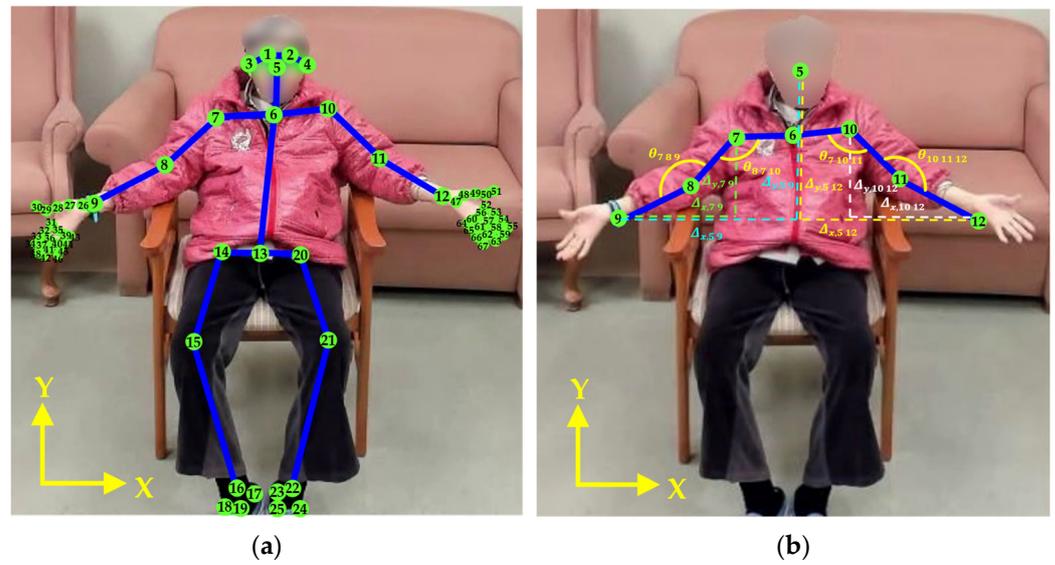


Figure A1. (a) Eye keypoints (1–2), ear keypoints (3,4), nose keypoint (5), body keypoints (6–25), and hand keypoints (26–67) detected by the OpenPose model; and (b) examples of the computed relative distance features between two keypoints, Δ_{ij} , and angle features between three keypoints, θ_{ijk} .

The spatial position of each detected keypoint ($p_{x,i}, p_{y,i}$) is normalized using the L2 normalization:

$$\begin{bmatrix} p_{x,i} \\ p_{y,i} \end{bmatrix} = \frac{[p_{x,i} \ p_{y,i}]^T}{\sqrt{\sum(p_{x,i}^2 + p_{y,i}^2)}} \quad (A1)$$

where $i \in \{1, 2, \dots, 67\}$ are the index of each keypoint in Figure A1a.

In order to classify different poses during exercise, three input features are utilized by the classification model: (1) relative distance features, Δ_{ij} ; (2) angle features, θ_{ijk} ; and (3) confidence score features, S_i , where $i, j, k \in \{1, 2, \dots, 67\}$ are the index of each keypoint in Figure A1b. The relative distance and angle features are computed based on the spatial positions of the keypoints while the confidence score features are acquired directly from the OpenPose model. Examples of the relative distance features and angle features are labeled in Figure A1b.

The relative distance features, Δ_{ij} , are determined in the X and Y axes between two keypoints ($p_{x,i}, p_{y,i}$) and ($p_{x,j}, p_{y,j}$):

$$\Delta_{ij} = \begin{bmatrix} \Delta_{x,ij} \\ \Delta_{y,ij} \end{bmatrix} = \begin{bmatrix} p_{x,i} \\ p_{y,i} \end{bmatrix} - \begin{bmatrix} p_{x,j} \\ p_{y,j} \end{bmatrix} \quad (A2)$$

The angle features, θ_{ijk} , are computed between three keypoints, ($p_{x,i}, p_{y,i}$), ($p_{x,j}, p_{y,j}$) and ($p_{x,k}, p_{y,k}$) [115]:

$$\theta_{ijk} = \arccos \left(\frac{\left(\begin{bmatrix} p_{x,j} \\ p_{y,j} \end{bmatrix} - \begin{bmatrix} p_{x,i} \\ p_{y,i} \end{bmatrix} \right) \cdot \left(\begin{bmatrix} p_{x,j} \\ p_{y,j} \end{bmatrix} - \begin{bmatrix} p_{x,k} \\ p_{y,k} \end{bmatrix} \right)}{\sqrt{(p_{x,j} - p_{x,i})^2 + (p_{y,j} - p_{y,i})^2} \sqrt{(p_{x,j} - p_{x,k})^2 + (p_{y,j} - p_{y,k})^2}} \right) \quad (A3)$$

where ($p_{x,j}, p_{y,j}$) is the vertex of the angle.

The confidence score features, S_i , are used to estimate the head rotation from the front facing view during the neck exercises based on the confidence score of the ears such that a cervical rotation of greater than 52° can be estimated by a confidence score of 0 in one of the ears (i.e., $S_3 = 0$ or $S_4 = 0$) [116].

To determine which arm is performing the exercises (e.g., forward punch and LTS), the total displacement of each arm joints (i.e., wrist, elbow, and shoulder) are compared in sequential frames during each repetition:

$$\Delta_d = \sum_f \sum_j \|P_{l,f}^i - P_{l,f-1}^i\| - \sum_f \sum_j \|P_{r,f}^j - P_{r,f-1}^j\| \tag{A4}$$

where $P_{l,f-1}^i$ and $P_{r,f-1}^j$ denote the 2D vectors with the (p_x, p_y) coordinates for the left and right arm joint j in the previous frame $f - 1$, Figure A2a; $P_{l,f}^i$ and $P_{r,f}^j$ denote the 2D vectors in the current frame f , Figure A2b; and $i \in \{7, 8, 9\}$ and $j \in \{10, 11, 12\}$ which are the index of the wrist, elbow, and shoulder joints for left and right arms, respectively. If Δ_d is greater than 0, then the left arm is performing the exercises. Otherwise, the right arm is performing the exercises.

During each exercise repetition, the body poses are classified using the aforementioned three input features (i.e., relative distance, angle, and confidence score features) based on the completion of the required movements and range of motion performed as: (1) complete pose, (2) partially complete pose, and (3) resting pose. The desired ranges of motion are determined from motion ranges from healthy older adults [117,118]. The body pose completion criteria for each exercise is presented in Table A1, and the specific features used for the classification of each exercise are detailed in Table A2.

Table A1. Complete, partially complete, and resting poses for each exercise.

Exercises	Complete Pose	Partially Complete Pose	Resting Pose
Open arm stretches	Open both arms from the center of the chest and raise them at least 65° from the sides of the body (i.e., $\theta_{8,7,10}$ and $\theta_{7,10,11} \geq 155^\circ$) while opening both arms [117].	Open both arms from the center of the chest but not raise at least one arm to or above 65° from the sides of the body (i.e., $\theta_{8,7,10}$ and/or $\theta_{7,10,11} < 155^\circ$) while opening both arms [117].	
Neck exercises (up and down)	Move the neck in both up and down directions.	Move the neck in only up or down direction.	
Neck exercises (left and right)	Rotate the neck to the left or right to achieve a cervical rotation for at least 52° [118], which can be estimated when one of the ears is invisible from the front view of the person (i.e., confidence score S_3 or $S_4 = 0$) [116].	Move the neck left or right but not to achieve a cervical rotation for at least 52° (i.e., confidence score S_3 and $S_4 > 0$) [118].	
Arm raises	Raise both arms to at least 65° from the sides of the body (i.e., $\theta_{8,7,10}$ and $\theta_{7,10,11} \geq 155^\circ$) [117].	Raise at least one arm but not to at least 65° from the sides of the body (i.e., $\theta_{8,7,10}$ and/or $\theta_{7,10,11} < 155^\circ$) [117].	Arms resting beside the waistline, on the armrest, or on the lap.
Downward punches	Raise both hands above the head (i.e., normalized $\Delta_{y,5,9}$ and $\Delta_{y,5,12} \geq 0.025$) before each downward punch [118].	Raise at least one wrist but not above the head (i.e., normalized $\Delta_{y,5,9}$ or $\Delta_{y,5,12} < 0.025$) [118].	
Breaststrokes	Sweep both arms to the sides of the body (i.e., $\Delta_{x,7,9}$ and $\Delta_{x,10,12} > 0.05$).	Sweep only one arm to the side of the body (i.e., $\Delta_{x,7,9}$ or $\Delta_{x,10,12} < 0.05$).	
Open/close hands	Open and close both hands.	Open and close at least one hand.	
Forward punches	Extend each arm straight (i.e., $\theta_{7,8,9}$ or $\theta_{10,11,12} \geq 155^\circ$) while punching [119].	Punch forward but do not fully extend the arm straight (i.e., $\theta_{7,8,9}$ or $\theta_{10,11,12} < 155^\circ$) [119].	
LTS	Raise each hand above the head (i.e., normalized $\Delta_{y,5,9}$ and $\Delta_{y,5,12} \geq 0.025$) while stretching [118].	Raise each hand but not above the head (i.e., normalized $\Delta_{y,5,9}$ or $\Delta_{y,5,12} < 0.025$) [118].	
Two-arm LTS	Raise each hand above the head (i.e., normalized $\Delta_{y,5,9}$ and $\Delta_{y,5,12} \geq 0.025$) while the other arm is extending to the side of the body [118].	Raise each hand but not above the head (i.e., normalized $\Delta_{y,5,9}$ or $\Delta_{y,5,12} < 0.025$) [118].	

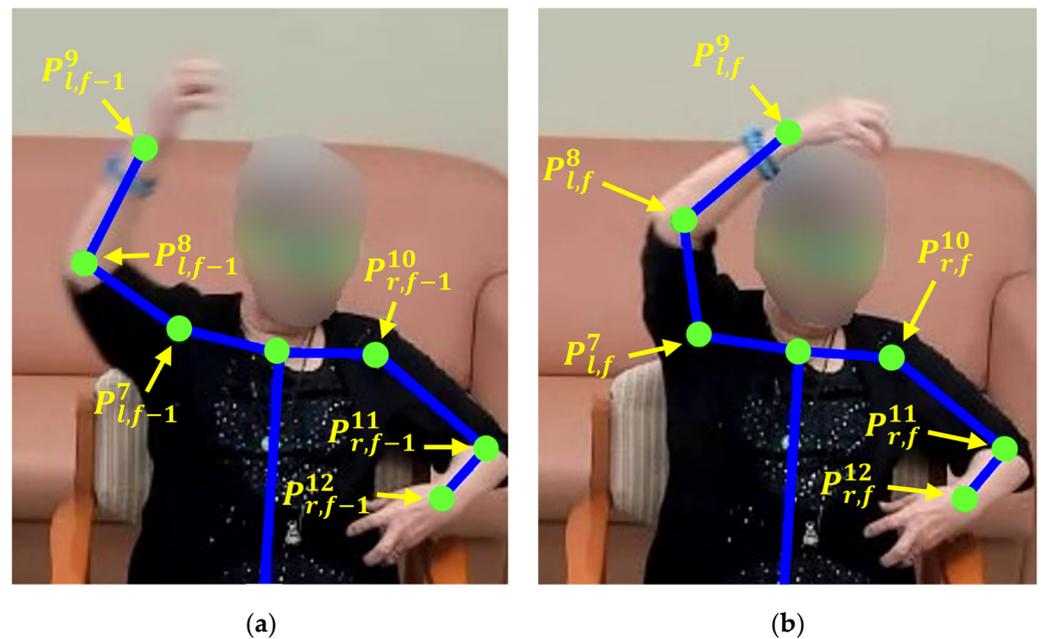


Figure A2. Arm joint position in: (a) the previous frame $f - 1$, and (b) the current frame f .

Appendix A.1.2. Pose Classification

To create the training dataset for the exercise pose classification, two volunteers recorded two exercise sessions with the robot. The volunteers were instructed to perform each exercise with all three outcomes. Then two expert coders selected 30 samples for each pose recorded from both volunteers, similar to [54] and [120]. To effectively identify pose completion during exercise, multiple learning-based classifiers—including k-Nearest Neighbor (k-NN), Multilayer Perceptron Neural Network (NN), Random Forest, and Support Vector Machine (SVM)—were investigated using the scikit-learn library [121]. The grid search strategy was used to optimize the parameters of each classification technique. A standard 10-fold cross validation was used to evaluate each classifier for each exercise. The classification rates are presented in Table A3. The Random Forest classifier was selected as it achieved the highest average classification rate of classifying the exercise poses for each exercise.

Appendix A.2. Exercise Evaluation Module

The performance of the Exercise Evaluation Module was evaluated by measuring the GAS T-score of each user as measured by expert coders and comparing the results to those estimated through the developed model. Namely, the GAS T-scores of each participant were obtained from two expert coders independently coding the user body poses for each exercise repetition during the sessions. For independent coding, both coders were presented with videos of the interactions. Then they met to discuss their coded results to obtain inter-coding consensus for reducing coder bias [122]. A Cohen's kappa, $\kappa = 0.75$, was determined between the two coders, indicating substantial agreement between these coders. In Table A4, an overview of the (dis)agreements between the coders is presented for each exercise type. In general, the two coders had an agreement rate of 90%, according to the body pose completion criteria (e.g., complete, partially complete) outlined in Table A1. The classification rates (defined by the number of classified repetition poses in agreement with coded poses divided by the total number of repetitions) are presented in Table A5 for all repetitions of each exercise. In general, the robot was able to track and estimate exercise poses correctly with an average classification rate of 89.44% using the Random Forest model developed for the Exercise Monitoring Module trained for each exercise.

Table A2. Classification features for each exercise.

Exercise	Features
Open arm stretches	<ol style="list-style-type: none"> Distances of the wrists-shoulders ($\Delta_{9\ 7}$, $\Delta_{12\ 10}$) and elbows-shoulders ($\Delta_{8\ 7}$, $\Delta_{11\ 10}$) for each arm in both x and y direction using Equation (A2). Angles of elbows ($\theta_{7\ 8\ 9}$, $\theta_{10\ 11\ 12}$) and shoulders ($\theta_{8\ 7\ 10}$, $\theta_{7\ 10\ 11}$) for each arm using Equation (A3).
Neck exercises (up and down)	<ol style="list-style-type: none"> Distances of the eyes-shoulders ($\Delta_{y\ 2\ 10}$, $\Delta_{y\ 1\ 7}$), nose-shoulders ($\Delta_{y\ 5\ 10}$, $\Delta_{y\ 5\ 7}$), and nose-neck ($\Delta_{y\ 5\ 6}$) in the y direction using Equation (A2). Angles of the nose ($\theta_{1\ 5\ 2}$) using Equation (A3).
Neck exercises (left and right)	<ol style="list-style-type: none"> Distances of the eyes-shoulders ($\Delta_{x\ 2\ 10}$, $\Delta_{x\ 1\ 7}$), nose-shoulders ($\Delta_{x\ 5\ 10}$, $\Delta_{x\ 5\ 7}$), and nose-neck ($\Delta_{x\ 5\ 6}$) in the x direction using Equation (A2). Confidence scores of the ears (S_3, S_4).
Arm raises	<ol style="list-style-type: none"> Distances of the wrists-shoulders ($\Delta_{9\ 7}$, $\Delta_{12\ 10}$) and elbows-shoulders ($\Delta_{8\ 7}$, $\Delta_{11\ 10}$) for each arm in both x and y direction using Equation (A2). Angles of elbows ($\theta_{7\ 8\ 9}$, $\theta_{10\ 11\ 12}$) and shoulders ($\theta_{8\ 7\ 10}$, $\theta_{7\ 10\ 11}$) for each arm using Equation (A3).
Downward punches	<ol style="list-style-type: none"> Distances of the wrists-shoulders ($\Delta_{y\ 9\ 7}$, $\Delta_{y\ 12\ 10}$), and wrists-nose ($\Delta_{y\ 9\ 5}$, $\Delta_{y\ 12\ 5}$) for each arm in the y direction using Equation (A2).
Breaststroke	<ol style="list-style-type: none"> Distances of the wrists-shoulders ($\Delta_{x\ 9\ 7}$, $\Delta_{x\ 12\ 10}$), elbows-shoulders ($\Delta_{x\ 8\ 7}$, $\Delta_{x\ 11\ 10}$), elbows-nose ($\Delta_{x\ 8\ 5}$, $\Delta_{x\ 11\ 5}$), and wrists-nose ($\Delta_{x\ 9\ 5}$, $\Delta_{x\ 12\ 5}$) for each arm in the x direction using Equation (A2) Angles of the elbows ($\theta_{7\ 8\ 9}$ or $\theta_{10\ 11\ 12}$) for each arm using Equation (A3)
Open/close hands	<ol style="list-style-type: none"> Average distances between fingertips and the wrists ($\Delta_{x\ 26\ 30}$, $\Delta_{x\ 26\ 34}$, $\Delta_{x\ 26\ 38}$, $\Delta_{x\ 26\ 42}$, $\Delta_{x\ 26\ 46}$ for the left hand and $\Delta_{x\ 47\ 51}$, $\Delta_{x\ 47\ 55}$, $\Delta_{x\ 47\ 59}$, $\Delta_{x\ 47\ 63}$, $\Delta_{x\ 47\ 67}$ for the right hand) using Equation (A2)
Forward punches	<ol style="list-style-type: none"> Distances of the elbows-shoulders ($\Delta_{y\ 8\ 7}$, $\Delta_{y\ 11\ 10}$), wrists-shoulders ($\Delta_{y\ 9\ 7}$, $\Delta_{y\ 12\ 10}$), elbows-nose ($\Delta_{y\ 8\ 5}$, $\Delta_{y\ 11\ 5}$), and wrists-nose ($\Delta_{y\ 9\ 5}$, $\Delta_{y\ 12\ 5}$) for each arm in the y direction using Equation (A2) Angles of the elbows ($\theta_{7\ 8\ 9}$ or $\theta_{10\ 11\ 12}$) for each arm using Equation (A3)
Lateral trunk stretch	<ol style="list-style-type: none"> Distances of the wrists-shoulders ($\Delta_{9\ 7}$, $\Delta_{12\ 10}$), elbows-shoulders ($\Delta_{8\ 7}$, $\Delta_{11\ 10}$), elbows-nose ($\Delta_{8\ 5}$, $\Delta_{11\ 5}$), and wrists-nose ($\Delta_{9\ 5}$, $\Delta_{12\ 5}$) for each arm in both x and y direction using Equation (A2) Angles of the elbows ($\theta_{7\ 8\ 9}$ or $\theta_{10\ 11\ 12}$) and shoulders ($\theta_{8\ 7\ 10}$, $\theta_{7\ 10\ 11}$) for each arm using Equation (A3)
Two-arm lateral trunk stretch	<ol style="list-style-type: none"> Distances of the wrists-shoulders ($\Delta_{9\ 7}$, $\Delta_{12\ 10}$), elbows-shoulders ($\Delta_{8\ 7}$, $\Delta_{11\ 10}$), elbows-nose ($\Delta_{8\ 5}$, $\Delta_{11\ 5}$), and wrists-nose ($\Delta_{9\ 5}$, $\Delta_{12\ 5}$) for each arm in both x and y direction using Equation (A2) Angles of the elbows ($\theta_{7\ 8\ 9}$ or $\theta_{10\ 11\ 12}$) and shoulders ($\theta_{8\ 7\ 10}$, $\theta_{7\ 10\ 11}$) for each arm using Equation (A3)

Appendix A.3. User State Detection Module

Appendix A.3.1. Valence

The user's valence is detected using data obtained through the EEG headband (Interaxon Muse 2016); the four electrode locations are TP9, AF7, AF8, and TP10 described using the International 10–20 system, Figure A3 [123].

Table A3. Classification rates for each classifier and exercise.

Exercise	Classifier			
	K-Nearest Neighbor (k-NN)	Multilayer Perceptron Neural Network (NN)	Random Forest (RF)	Support Vector Machine (SVM)
Open arm stretches	97.8%	96.7%	97.8%	91.1%
Neck exercises (up and down)	65.6%	74.4%	92.2%	54.4%
Neck exercises (left and right)	98%	98.0%	98.0%	78%
Arm raises	96.7%	98.9%	98.9%	87.8%
Downward punches	96.7%	94.4%	97.8%	94.4%
Breast-stroke	62.2%	88.9%	98.9%	64.4%
Open/close hands	88.3%	50.0%	97.5%	94.9%
Forward punches	71.4%	80.0%	92.7%	54.3%
Lateral trunk stretch	98.9%	98.3%	98.9%	97.8%
Two-arm lateral trunk stretch	97.8%	96.7%	98.4%	95.6%
Average	87.3%	87.6%	97.1%	81.3%

Table A4. Agreement and disagreement rates between coders.

Exercise	Agreed Completion (%)	Agreed Incompletion (%)	Disagreement (%)
Open arm stretches	21.5	68.76	9.74
Neck exercises (up and down)	71.72	16.01	12.27
Neck exercises (left and right)	68.05	18.86	13.09
Forward punches	51.52	35.48	13
Arm raises	69.24	19.56	11.2
Downward punches	69.35	25.04	5.61
Open/close hands	77.33	12.35	10.32
Breaststrokes	95.89	1.03	3.08
LTS	77.85	13.41	8.74
Two-arm LTS	79.85	11.80	8.35

Table A5. Exercise classification rate.

Exercise	Classification Rate (%)
Open arm stretches	86.86
Neck exercises (up and down)	85.12
Neck exercises (left and right)	87.17
Forward punches	88.34
Arm raises	89.07
Downward punches	92.98
Breaststrokes	96.87
Open/close hands	90.14
LTS	89.79
Two-arm LTS	88.01
Average	89.44

The EEG signals are processed using the Muse LSL package [124], and two types of EEG frequency domain features. With respect to the latter, the power spectral density (PSD) feature and the frontal asymmetry features, are extracted from the EEG data for valence classification. These two types of features have been used in real-time valence detection [125]. The Fast Fourier Transform (FFT) is utilized to decompose the EEG signal to extract the PSD [124]. This process is implemented through a 1 s sliding window with

an overlap of 80% to reduce spectral leakage and minimize data loss [126]. Then, from each electrode location, the PSD features are acquired in four distinct frequency bands: θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–40 Hz) [125]. Frontal EEG asymmetry refers to the difference in power between the left and right frontal hemispheres of the brain within the α and β frequency bands [46]. The frontal EEG asymmetry features can be computed as v_1 to v_4 , through Equations (A5)–(A8), to determine valence [125].

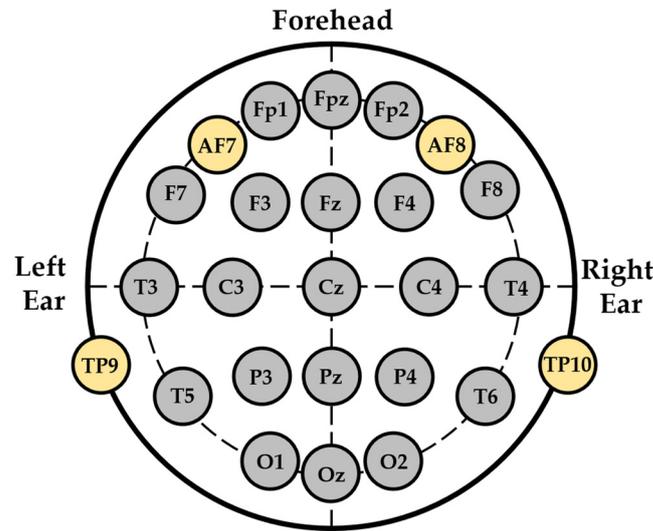


Figure A3. Muse sensor four electrode locations on the International 10–20 system [123].

$$v_1 = \frac{\alpha_{AF8}}{\beta_{AF8}} - \frac{\alpha_{AF7}}{\beta_{AF7}} \tag{A5}$$

$$v_2 = \ln(\alpha_{AF7}) - \ln(\alpha_{AF8}) \tag{A6}$$

$$v_3 = \frac{\beta_{AF7}}{\alpha_{AF7}} - \frac{\beta_{AF8}}{\alpha_{AF8}} \tag{A7}$$

$$v_4 = \alpha_{AF8} - \alpha_{AF7}, \tag{A8}$$

Above, α_{AF7} , α_{AF8} , β_{AF7} , and β_{AF8} are the α and β band powers measured at the AF7 and AF8 locations shown in Figure A3.

In total, 20 features are utilized, which include 16 PSD features from the four frequency bands; θ , α , β , and γ measured at the locations TP9, AF7, AF8, and TP10; and four frontal EEG asymmetry features obtained from Equations (A5)–(A8).

During HRI, it was found that user valence is directly related to the interaction with the robot itself [127–129]. This is based on the assumption that the user is engaged with the robot, as disengaging and averting VFOA facilitates remembrance of other memories that would influence user valence [130]. To accurately detect valence, it is important for the valence detection model to be trained with valence that occurs with a robot [129]. For valence elicitation, stimuli consisting of robot body movements to music were used to induce positive and negative valence, which were designed in our previous work [43].

To obtain training data for valence elicitation, we recruited six older adults between 81 and 96 years old from the Yee Hong Centre for Geriatric Care. All were healthy older adults with no or mild cognitive impairment with a Cognitive Performance Scale (CPS) score of lower than three (i.e., intact or mild impairment) [62] and the Mini-Mental State Exam (MMSE) score of greater than 19 (i.e., normal or mild impairment). The robot displayed positive and negative valence stimuli to each participant while they wore the EEG headband to better interpret and respond to users during HRI [129]. Each participant

had two sessions; one was for inducing positive valence while the other was for inducing negative valence. The stimuli were presented to each participant in a random order. Each session was approximately 4 min in duration with a 5-min break after each session. During the break, participants were asked to report their valence level using the Self-Assessment Manikin (SAM) scale [71] to label the EEG training data. Among these older adults, three of them were able to perceive both the positive valence and negative valence stimuli correctly as intended, and their data were used as training data for valence classification. These EEG signals share a similar pattern in PSD features of θ , α , β , and γ bands, which are used for valence classification, among those in the same age range (e.g., 66 years and older) with similar cognitive impairment levels [131,132]. A three hidden layer Multilayer Perceptron Neural Network using the scikit-learn library [121] was used to classify user valence. A 10-fold cross validation was performed on the training data to evaluate the prediction results, achieving a 77% classification rate. Our classification rate is comparable to or higher than several other learning-based classification methods that also use EEG signals for valence detection, which have reported rates between 57% and 76%, e.g., [133–137].

Appendix A.3.2. Engagement

The user’s engagement is estimated based on the orientation of their face and the visibility of their ears. The user is considered engaged when their face is oriented within 45° , Figure A4a, and not engaged when their face is oriented greater than 45° in either left or right direction away from the robot, Figure A4b [138]:

$$\theta_f = \sin^{-1}\left(\frac{l_{en} - r_{en}}{l_{en} + r_{en}}\right) \tag{A9}$$

where θ_f is the orientation of the face, r_{en} is the distance from the center of the right eye to the nose, and l_{en} is the distance from the center of the left eye to the nose.

The visibility of the ears can be represented by the confidence scores of the ear keypoints obtained from the OpenPose model [31]. Therefore, the confidence scores of the ears (S_3 and S_4) are also used as VFOA features for engagement detection such that the VFOA features, f_{VFOA} , can be expressed as:

$$f_{VFOA} = [\theta_f \ S_3 \ S_4]. \tag{A10}$$

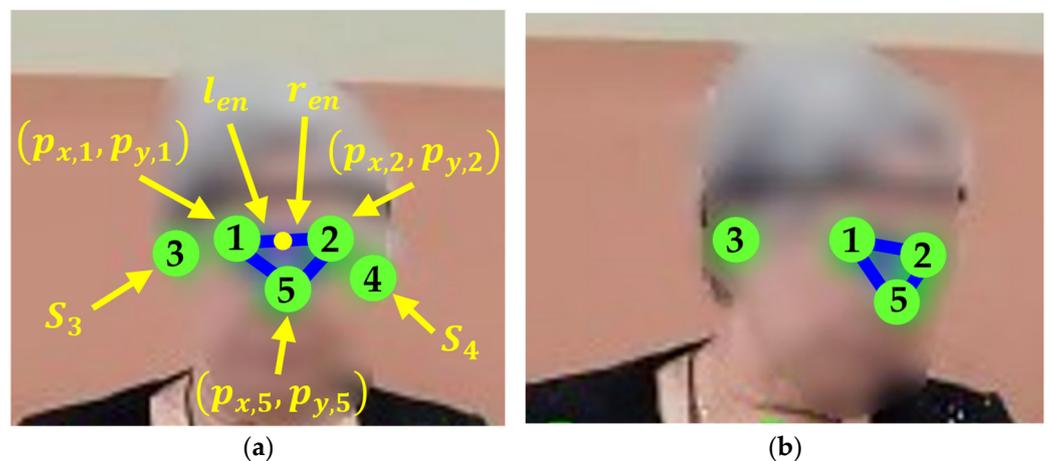


Figure A4. Engagement detection based on the user VFOA: (a) engaged; and (b) not engaged towards the robot.

Two volunteers recorded two robot exercise sessions, engaged (i.e., looking at the robot with $\theta_f \leq 45^\circ$) and not engaged (i.e., not looking at the robot with $\theta_f > 45^\circ$) with the robot, for creating a training dataset using the aforementioned VFOA features, f_{VFOA} . The dataset was used for training a k-NN classifier for engagement detection. Forty head pose

samples for both engaged and not engaged were coded and selected by two experts. The data from both volunteers were used to train the classifiers. By performing a 10-fold cross validation on the training data, the k-NN classifier was able to achieve a classification rate of 93% [31].

Appendix A.4. Robot Emotion Module

This module utilizes a robot emotion model that we have developed in our previous work, which takes in the user states (i.e., user valence and engagement) to determine the robot’s emotional behavior. This model has been adapted herein for our HRI study [55].

The robot emotional state for m robot emotions and l user states at time t , E'_t can be represented as:

$$E_t = w_1 H_t + W_2 A_t, \tag{A11}$$

$$E'_t = f(E_t) \tag{A12}$$

where E_t represents the robot emotion output vector, H_t is the robot emotional state vector based on the emotional history at time t , and A_t is the user state input vector based on both the user valence and engagement. In addition, w_1 is a scalar that is the weight of the influence of the robot emotional history on the current robot emotion. W_2 is the robot emotion state-human affect probability distribution. Finally, $f(E_t)$ is a winner-takes-all function to determine the robot emotion to display.

Robot Emotion History Model

Human emotions are time-related processes such that the current emotion is often influenced by past emotions [139]. Therefore, we integrate the n th order MM property we have used in our previous work [54] to model the robot emotional state, which represents the probability of the current emotion e_0 is dependent on the previous emotional history, e_1, \dots, e_n [55]:

$$P(H_t = e_0 | H_{t-1} = e_1, \dots, H_1 = e_{t-1}) = P(H_t = e_0 | H_{t-1} = e_1, \dots, H_{t-n} = e_n) \tag{A13}$$

where $e_0, \dots, e_t \in \{1, \dots, m\}$ represent the displayed robot emotions.

In addition, the influence of a past emotion on the current emotion should decrease as time passes [55]. A decay function is utilized to reduce the weight of each past emotion in discrete time:

$$\lambda_i = e^{-ai}, 0 < a < -\ln(\varepsilon) \tag{A14}$$

where λ_i is the weight of the robot emotion at discrete time $i \in T^+$, a is the rate of the decay, and ε is the lower threshold of the decay function.

The robot emotion transition probability, which represents the probability of the current robot emotion based on the past n robot emotional steps, is modeled as:

$$P(H_t = e_0 | H_{t-1} = e_1, \dots, H_{t-n} = e_n) = \sum_{i=1}^n \lambda_i q_{e_i e_0} \tag{A15}$$

where $q_{e_i e_0}$ is an element of the $m \times m$ robot emotion transition probability matrix, Q_i , and $n = T - 1$.

Then, the robot emotion history model can be modeled as:

$$H_t = \sum_{i=1}^n \lambda_i Q_i H_{t-i} \tag{A16}$$

To estimate the robot emotion transition probability matrix, Q_i , the transition frequency $f_{kj}^{(i)}$ from emotional state j to emotional state k is considered with history i :

$$F^{(i)} = \begin{pmatrix} f_{11}^{(i)} & \dots & f_{1m}^{(i)} \\ \vdots & \ddots & \vdots \\ f_{m1}^{(i)} & \dots & f_{mm}^{(i)} \end{pmatrix} \tag{A17}$$

Therefore, the estimated Q_i can be represented as:

$$\hat{Q}_i = \begin{pmatrix} \hat{q}_{11}^{(i)} & \cdots & \hat{q}_{1m}^{(i)} \\ \vdots & \ddots & \vdots \\ \hat{q}_{m1}^{(i)} & \cdots & \hat{q}_{mm}^{(i)} \end{pmatrix} \tag{A18}$$

where $q_{kj}^{(i)}$ is:

$$q_{kj}^{(i)} = \begin{cases} \frac{f_{kj}^{(i)}}{\sum_{j=1}^m f_{kj}^{(i)}} & \text{if } \sum_{j=1}^m f_{kj}^{(i)} \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{A19}$$

Appendix B. Robot Perception Questionnaire

Table A6. Robot Perception Questionnaire and results.

Construct	Question	One Month				Two Month			
		\bar{x}	s	\tilde{x}	IQR	\bar{x}	s	\tilde{x}	IQR
C1: Acceptance	Q1. I like using the robot to do exercise	4.04	1.07	4.00	2.00	4.11	1.13	5.00	2.00
	Q2. I would use the robot again	3.78	1.40	4.00	2.00	3.96	1.40	5.00	1.00
	Q3. The sensor headband is uncomfortable to wear * †	1.40	0.92	1.00	1.00	1.60	0.80	1.00	0.00
C2: Perceived Usefulness and Ease of Use	Q4. The exercises the robot got me to do are good for my overall health	4.30	0.97	5.00	1.00	4.26	1.11	5.00	1.00
	Q5. The robot is not helpful for doing exercise †	1.70	1.15	1.00	2.00	2.04	1.35	1.00	1.25
	Q6. The robot clearly displays each exercise	4.44	1.10	5.00	1.00	4.37	0.82	4.00	0.25
	Q7. The robot is difficult to use †	2.04	1.37	1.00	2.00	2.52	1.29	3.00	2.00
	Q8. I can use the robot without any help	3.11	1.81	4.00	2.50	2.63	1.28	3.00	4.00
	Q9. I don't trust the robot's advice †	1.67	1.09	1.00	1.50	2.04	1.07	2.00	1.25
C3: Perceived Sociability and Intelligence	Q10. The robot motivates me to exercise	4.33	1.09	5.00	2.00	3.85	1.24	4.00	1.00
	Q11. After each exercise, the feedback the robot provided is appropriate	3.82	1.09	4.00	2.00	3.70	1.24	4.00	2.00
	Q12. The robot understands what I am doing during exercising	3.44	1.32	3.00	1.00	3.44	1.07	3.00	2.00
	Q13. The robot displays appropriate emotions	4.00	1.33	5.00	1.00	3.33	1.09	3.00	1.25
	Q14. I am not able to identify the robot's emotions through eye colors * †	2.30	1.19	2.00	2.75	2.40	1.50	2.00	1.75
C4: Robot Appearance and Movements	Q15. I am able to identify the robot's emotions from vocal intonation *	4.40	0.92	5.00	0.75	3.70	1.01	4.00	1.00
	Q16. The robot moves too fast for me to follow †	1.82	1.36	1.00	2.00	2.00	1.31	1.00	2.00
	Q17. I think the robot has a clear voice	4.33	1.16	5.00	1.50	4.19	1.25	5.00	1.00
	Q18. I don't understand the robot's instructions †	1.82	1.31	1.00	1.00	1.67	1.16	1.00	1.25
Overall Experience (2 months)	Q19. I think the robot's size is appropriate for exercising	4.22	1.23	5.00	2.00	3.74	1.29	4.00	1.25
	Q20. I feel my physical health is improved from the exercise sessions with the robot	N/A	N/A	N/A	2.00	3.70	1.24	4.00	2.00
	Q21. I find what I am doing in the weekly sessions confusing †	N/A	N/A	N/A	1.00	1.67	1.12	1.00	1.00
	Q22. As a result of these sessions, I am more motivated to perform daily physical exercises	N/A	N/A	N/A	2.00	3.70	1.08	4.00	0.00
	Q23. The robot always seemed interested in interacting with me	N/A	N/A	N/A	1.50	3.56	1.13	3.00	1.00

* Questions that were only administrated to participants wearing the sensors. † Questions that were negatively worded (i.e., 1 or 2 represent positive responses).

Table A7. Robot perception questionnaire results for each construct for one-on-one sessions, group sessions, and all users combined measured after one month and two months. The minimum (x_{min}), median (\tilde{x}), mode (x_{MO}), and interquartile range (IQR) are presented.

Construct	Session Type	One Month					Two Months				
		x_{min}	x_{max}	\tilde{x}	x_{MO}	IQR	x_{min}	x_{max}	\tilde{x}	x_{MO}	IQR
C1: Acceptance (One-Month: $\alpha = 0.75$, Two-Months: $\alpha = 0.88$)	One-on-One	1	5	4.0	4	1.00	2	5	4.5	5	2.00
	Group	1	5	4.0	5	2.00	1	5	5.0	5	2.00
	All	1	5	4.0	5	1.75	1	5	5.0	5	2.00
C2: Perceived Usefulness and Ease of Use (One-Month: $\alpha = 0.81$, Two-Months: $\alpha = 0.83$)	One-on-One	1	5	5.0	5	1.00	1	5	4.0	5	2.00
	Group	1	5	5.0	5	2.00	1	5	4.0	5	2.00
	All	1	5	5.0	5	1.00	1	5	4.0	5	2.00
C3: Perceived Sociability and Intelligence (One-Month: $\alpha = 0.79$, Two-Months: $\alpha = 0.68$)	One-on-One	1	5	4.5	5	1.25	2	5	4.0	5	2.00
	Group	1	5	4.0	5	2.00	1	5	3.0	3	1.00
	All	1	5	4.0	5	2.00	1	5	3.0	3	2.00
C4: Robot Appearance and Movements (One-Month: $\alpha = 0.80$, Two-Months: $\alpha = 0.72$)	One-on-One	1	5	5.0	5	1.25	1	5	5.0	5	1.00
	Group	1	5	5.0	5	1.00	1	5	4.0	5	2.00
	All	1	5	5.0	5	1.00	1	5	5.0	5	2.00

References

- Panton, L.; Loney, B. Exercise for Older Adults. Available online: https://file.lacounty.gov/SDSInter/dmh/216745_ExerciseforOlderAdultsHealthCareProviderManual.pdf (accessed on 27 September 2022).
- Nelson, M.E.; Rejeski, W.J.; Blair, S.N.; Duncan, P.W.; Judge, J.O.; King, A.C.; Macera, C.A.; Castaneda-Sceppa, C. Physical Activity and Public Health in Older Adults: Recommendation from the American College of Sports Medicine and the American Heart Association. *Med. Sci. Sport. Exerc.* **2007**, *39*, 1435–1445. [\[CrossRef\]](#)
- Piercy, K.L.; Troiano, R.P.; Ballard, R.M.; Carlson, S.A.; Fulton, J.E.; Galuska, D.A.; George, S.M.; Olson, R.D. The Physical Activity Guidelines for Americans. *JAMA* **2018**, *320*, 2020–2028. [\[CrossRef\]](#) [\[PubMed\]](#)
- Blair, S.N.; Kohl, H.W.; Barlow, C.E.; Paffenbarger, R.S.; Gibbons, L.W.; Macera, C.A. Changes in Physical Fitness and All-Cause Mortality. A Prospective Study of Healthy and Unhealthy Men. *JAMA* **1995**, *273*, 1093–1098. [\[CrossRef\]](#) [\[PubMed\]](#)
- Statistics Canada. Physical Activity, Self Reported, Adult, by Age Group (Table 13-10-0096-13). Available online: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310009613> (accessed on 27 September 2022).
- Carros, F.; Meurer, J.; Löffler, D.; Unbehau, D.; Matthies, S.; Koch, I.; Wieching, R.; Randall, D.; Hassenzahl, M.; Wulf, V. Exploring Human-Robot Interaction with the Elderly: Results from a Ten-Week Case Study in a Care Home. In Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–12.
- Fasola, J.; Matarić, M.J. A Socially Assistive Robot Exercise Coach for the Elderly. *J. Hum.-Robot Interact.* **2013**, *2*, 3–32. [\[CrossRef\]](#)
- Avelino, J.; Simão, H.; Ribeiro, R.; Moreno, P.; Figueiredo, R.; Duarte, N.; Nunes, R.; Bernardino, A.; Čaić, M.; Mahr, D. Experiments with Vizzy as a Coach for Elderly Exercise. In *Workshop on Personal Robots for Exercising and Coaching-HRI Conference*; PREC: Chicago, IL, USA, 2018; pp. 1–6.
- Görer, B.; Salah, A.A.; Akin, H.L. An Autonomous Robotic Exercise Tutor for Elderly People. *Auton. Robot.* **2017**, *41*, 657–678. [\[CrossRef\]](#)
- Costello, E.; Kafchinski, M.; Vrazel, J.; Sullivan, P. Motivators, Barriers, and Beliefs Regarding Physical Activity in an Older Adult Population. *J. Geriatr. Phys. Ther.* **2011**, *34*, 138–147. [\[CrossRef\]](#)
- Wada, K.; Shibata, T. Living with Seal Robots—Its Sociopsychological and Physiological Influences on the Elderly at a Care House. *IEEE Trans. Robot.* **2007**, *23*, 972–980. [\[CrossRef\]](#)
- Ybarra, O.; Burnstein, E.; Winkelman, P.; Keller, M.C.; Manis, M.; Chan, E.; Rodriguez, J. Mental Exercising through Simple Socializing: Social Interaction Promotes General Cognitive Functioning. *Pers. Soc. Psychol. Bull.* **2008**, *34*, 248–259. [\[CrossRef\]](#)
- van Stralen, M.M.; de Vries, H.; Mudde, A.N.; Bolman, C.; Lechner, L. The Long-Term Efficacy of Two Computer-Tailored Physical Activity Interventions for Older Adults: Main Effects and Mediators. *Health Psychol. Off. J. Div. Health Psychol. Am. Psychol. Assoc.* **2011**, *30*, 442–452. [\[CrossRef\]](#)
- Marcus, B.H.; Bock, B.C.; Pinto, B.M.; Forsyth, L.H.; Roberts, M.B.; Traficante, R.M. Efficacy of an Individualized, Motivationally-Tailored Physical Activity Intervention. *Ann. Behav. Med. Publ. Soc. Behav. Med.* **1998**, *20*, 174–180. [\[CrossRef\]](#)
- Trampe, D.; Quoidbach, J.; Taquet, M. Emotions in Everyday Life. *PLoS ONE* **2015**, *10*, e0145450. [\[CrossRef\]](#)
- Cavallo, F.; Semeraro, F.; Fiorini, L.; Magyar, G.; Sinčák, P.; Dario, P. Emotion Modelling for Social Robotics Applications: A Review. *J. Bionic Eng.* **2018**, *15*, 185–203. [\[CrossRef\]](#)
- Uchida, M.C.; Carvalho, R.; Tessutti, V.D.; Bacurau, R.F.P.; Coelho-Júnior, H.J.; Capelo, L.P.; Ramos, H.P.; Dos Santos, M.C.; Teixeira, L.F.M.; Marchetti, P.H. Identification of Muscle Fatigue by Tracking Facial Expressions. *PLoS ONE* **2018**, *13*, e0208834. [\[CrossRef\]](#) [\[PubMed\]](#)

18. Tanikawa, C.; Takata, S.; Takano, R.; Yamanami, H.; Edlira, Z.; Takada, K. Functional Decline in Facial Expression Generation in Older Women: A Cross-Sectional Study Using Three-Dimensional Morphometry. *PLoS ONE* **2019**, *14*, e0219451. [[CrossRef](#)] [[PubMed](#)]
19. Ketcham, C.J.; Stelmach, G.E. *Movement Control in the Older Adult*; Pew, R.W., Van Hemel, S.B., Eds.; National Academies Press (US) Steering Committee for the Workshop on Technology for Adaptive Aging: Washington, DC, USA, 2004.
20. Al-Nafjan, A.; Hosny, M.; Al-Ohali, Y.; Al-Wabil, A. Review and Classification of Emotion Recognition Based on EEG Brain-Computer Interface System Research: A Systematic Review. *Appl. Sci.* **2017**, *7*, 1239. [[CrossRef](#)]
21. Broelz, E.K.; Enck, P.; Niess, A.M.; Schneeweiss, P.; Wolf, S.; Weimer, K. The Neurobiology of Placebo Effects in Sports: EEG Frontal Alpha Asymmetry Increases in Response to a Placebo Ergogenic Aid. *Sci. Rep.* **2019**, *9*, 2381. [[CrossRef](#)]
22. Céspedes, N.; Irfan, B.; Senft, E.; Cifuentes, C.A.; Gutierrez, L.F.; Rincon-Roncancio, M.; Belpaeme, T.; Múnera, M. A Socially Assistive Robot for Long-Term Cardiac Rehabilitation in the Real World. *Front. Neurobotics* **2021**, *15*, 633248. [[CrossRef](#)]
23. Pulido, J.C.; Suarez-Mejias, C.; Gonzalez, J.C.; Duenas Ruiz, A.; Ferrand Ferri, P.; Martinez Sahuquillo, M.E.; Ruiz De Vargas, C.E.; Infante-Cossio, P.; Parra Calderon, C.L.; Fernandez, F. A Socially Assistive Robotic Platform for Upper-Limb Rehabilitation: A Longitudinal Study with Pediatric Patients. *IEEE Robot. Autom. Mag.* **2019**, *26*, 24–39. [[CrossRef](#)]
24. Back, I.; Makela, K.; Kallio, J. Robot-Guided Exercise Program for the Rehabilitation of Older Nursing Home Residents. *Ann. Long-Term Care* **2013**, *21*, 38–41.
25. Fraune, M.R.; Šabanović, S.; Kanda, T. Human Group Presence, Group Characteristics, and Group Norms Affect Human-Robot Interaction in Naturalistic Settings. *Front. Robot. AI* **2019**, *6*, 48. [[CrossRef](#)]
26. Fraune, M.R.; Sherrin, S.; Šabanović, S.; Smith, E.R. Is Human-Robot Interaction More Competitive between Groups than between Individuals? In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Daegu, Republic of Korea, 11–14 March 2019; pp. 104–113.
27. Leite, I.; McCoy, M.; Lohani, M.; Ullman, D.; Salomons, N.; Stokes, C.; Rivers, S.; Scassellati, B. Emotional Storytelling in the Classroom: Individual versus Group Interaction between Children and Robots. In Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction, Portland, OR, USA, 2–5 March 2015.
28. Harvard Health Publishing. The 4 Most Important Types of Exercise. Available online: <https://www.health.harvard.edu/exercise-and-fitness/the-4-most-important-types-of-exercise> (accessed on 27 September 2022).
29. United States Department of Health and Human Services. *Physical Activity and Health: A Report of the Surgeon General*; Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion: Atlanta, GA, USA, 1996.
30. Kiresuk, T.J.; Sherman, R.E. Goal Attainment Scaling: A General Method for Evaluating Comprehensive Community Mental Health Programs. *Community Ment. Health J.* **1968**, *4*, 443–453. [[CrossRef](#)] [[PubMed](#)]
31. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.-E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [[CrossRef](#)] [[PubMed](#)]
32. Simon, T.; Joo, H.; Matthews, I.; Sheikh, Y. Hand Keypoint Detection in Single Images Using Multiview Bootstrapping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4645–4653.
33. Krasny-Pacini, A.; Hiebel, J.; Pauly, F.; Godon, S.; Chevignard, M. Goal Attainment Scaling in Rehabilitation: A Literature-Based Update. *Ann. Phys. Rehabil. Med.* **2013**, *56*, 212–230. [[CrossRef](#)] [[PubMed](#)]
34. Jones, M.C.; Walley, R.M.; Leech, A.; Paterson, M.; Common, S.; Metcalf, C. Using Goal Attainment Scaling to Evaluate a Needs-Led Exercise Programme for People with Severe and Profound Intellectual Disabilities. *J. Intellect. Disabil.* **2006**, *10*, 317–335. [[CrossRef](#)] [[PubMed](#)]
35. Rockwood, K.; Stolee, P.; Fox, R.A. Use of Goal Attainment Scaling in Measuring Clinically Important Change in the Frail Elderly. *J. Clin. Epidemiol.* **1993**, *46*, 1113–1118. [[CrossRef](#)]
36. Stolee, P.; Rockwood, K.; Fox, R.A.; Streiner, D.L. The Use of Goal Attainment Scaling in a Geriatric Care Setting. *J. Am. Geriatr. Soc.* **1992**, *40*, 574–578. [[CrossRef](#)]
37. Toto, P.E.; Skidmore, E.R.; Terhorst, L.; Rosen, J.; Weiner, D.K. Goal Attainment Scaling (GAS) in Geriatric Primary Care: A Feasibility Study. *Arch. Gerontol. Geriatr.* **2015**, *60*, 16–21. [[CrossRef](#)]
38. Valadao, C.T.; Goulart, C.; Rivera, H.; Caldeira, E.; Bastos Filho, T.F.; Frizzera-Neto, A.; Carelli, R. Analysis of the Use of a Robot to Improve Social Skills in Children with Autism Spectrum Disorder. *Res. Biomed. Eng.* **2016**, *32*, 161–175. [[CrossRef](#)]
39. Cook, A.M.; Bentz, B.; Harbottle, N.; Lynch, C.; Miller, B. School-Based Use of a Robotic Arm System by Children with Disabilities. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2005**, *13*, 452–460. [[CrossRef](#)]
40. National Institute on Aging. *Exercise: A Guide from the National Institute on Aging*; National Institute on Aging: Bethesda, MD, USA, 2001.
41. Su, S.W.; Huang, S.; Wang, L.; Celler, B.G.; Savkin, A.V.; Guo, Y.; Cheng, T.M. Optimizing Heart Rate Regulation for Safe Exercise. *Ann. Biomed. Eng.* **2010**, *38*, 758–768. [[CrossRef](#)]
42. Barrett, L.F. Valence Is a Basic Building Block of Emotional Life. *J. Res. Pers.* **2006**, *40*, 35–55. [[CrossRef](#)]
43. Shao, M.; Snyder, M.; Nejat, G.; Benhabib, B. User Affect Elicitation with a Socially Emotional Robot. *Robotics* **2020**, *9*, 44. [[CrossRef](#)]
44. Spezialetti, M.; Placidi, G.; Rossi, S. Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives. *Front. Robot. AI* **2020**, *7*, 532279. [[CrossRef](#)]

45. Apicella, A.; Arpaia, P.; Mastrati, G.; Moccaldi, N. EEG-Based Detection of Emotional Valence towards a Reproducible Measurement of Emotions. *Sci. Rep.* **2021**, *11*, 21615. [[CrossRef](#)] [[PubMed](#)]
46. Ramirez, R.; Vamvakousis, Z. Detecting Emotion from EEG Signals Using the Emotive Epoc Device. In *International Conference on Brain Informatics*; Zanzotto, F.M., Tsumoto, S., Taatgen, N., Yao, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 175–184.
47. Chen, M.; Han, J.; Guo, L.; Wang, J.; Patras, I. Identifying Valence and Arousal Levels via Connectivity between EEG Channels. In *Proceedings of the International Conference on Affective Computing and Intelligent Interaction*, Xi'an, China, 21–24 September 2015; pp. 63–69.
48. Sidner, C.L.; Lee, C.; Kidd, C.D.; Lesh, N.; Rich, C. Explorations in Engagement for Humans and Robots. *Artif. Intell.* **2005**, *166*, 140–164. [[CrossRef](#)]
49. Sidner, C.L.; Kidd, C.D.; Lee, C.; Lesh, N. Where to Look: A Study of Human-Robot Engagement. In *Proceedings of the 9th International Conference on Intelligent User Interfaces*; Association for Computing Machinery: New York, NY, USA, 2004; pp. 78–84.
50. Michalowski, M.P.; Sabanovic, S.; Simmons, R. A Spatial Model of Engagement for a Social Robot. In *Proceedings of the IEEE International Workshop on Advanced Motion Control*, Istanbul, Turkey, 27–29 March 2006; pp. 762–767.
51. Moro, C.; Lin, S.; Nejat, G.; Mihailidis, A. Social Robots and Seniors: A Comparative Study on the Influence of Dynamic Social Features on Human–Robot Interaction. *Int. J. Soc. Robot.* **2019**, *11*, 5–24. [[CrossRef](#)]
52. Li, J.; Louie, W.-Y.G.; Mohamed, S.; Despond, F.; Nejat, G. A User-Study with Tangy the Bingo Facilitating Robot and Long-Term Care Residents. In *Proceedings of the IEEE International Symposium on Robotics and Intelligent Sensors*, Tokyo, Japan, 17–20 December 2016; pp. 109–115.
53. American Heart Association. Target Heart Rates Chart. Available online: <https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates> (accessed on 27 September 2022).
54. Shao, M.; Alves, S.F.D.R.; Ismail, O.; Zhang, X.; Nejat, G.; Benhabib, B. You Are Doing Great! Only One Rep Left: An Affect-Aware Social Robot for Exercising. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Bari, Italy, 6–9 October 2019; pp. 3811–3817.
55. Zhang, X.; Alves, S.; Nejat, G.; Benhabib, B. A Robot Emotion Model with History. In *Proceedings of the IEEE International Symposium on Robotics and Intelligent Sensors*, Ottawa, ON, Canada, 5–7 October 2017; pp. 230–235.
56. Stødle, I.V.; Debesay, J.; Pajalic, Z.; Lid, I.M.; Bergland, A. The Experience of Motivation and Adherence to Group-Based Exercise of Norwegians Aged 80 and More: A Qualitative Study. *Arch. Public Health* **2019**, *77*, 26. [[CrossRef](#)]
57. Yang, E.; Dorneich, M.C. The Effect of Time Delay on Emotion, Arousal, and Satisfaction in Human-Robot Interaction. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2015**, *59*, 443–447. [[CrossRef](#)]
58. Rudovic, O.; Lee, J.; Mascarell-Maricic, L.; Schuller, B.W.; Picard, R.W. Measuring Engagement in Robot-Assisted Autism Therapy: A Cross-Cultural Study. *Front. Robot. AI* **2017**, *4*, 36. [[CrossRef](#)]
59. Fasola, J.; Mataric, M.J. Using Socially Assistive Human–Robot Interaction to Motivate Physical Exercise for Older Adults. *Proc. IEEE* **2012**, *100*, 2512–2526. [[CrossRef](#)]
60. Yee Hong Centre for Geriatric Care. *MDS CIHI Data Jan 2020 for Yee Hong Centre Mississauga*; Yee Hong Centre for Geriatric Care: Mississauga, ON, Canada, 2020.
61. National Cancer Institute Division of Cancer Control and Population Sciences. *SEER-Medicare: Minimum Data Set (MDS)—Nursing Home Assessment*; National Cancer Institute: Bethesda, MD, USA, 2010.
62. Morris, J.N.; Fries, B.E.; Mehr, D.R.; Hawes, C.; Phillips, C.; Mor, V.; Lipsitz, L.A. MDS Cognitive Performance Scale. *J. Gerontol.* **1994**, *49*, M174–M182. [[CrossRef](#)]
63. Kurlowicz, L.; Wallace, M. The Mini-Mental State Examination (MMSE). *J. Gerontol. Nurs.* **1999**, *25*, 8–9. [[CrossRef](#)]
64. Müller, B.C.N.; Chen, S.; Nijssen, S.R.R.; Kühn, S. How (Not) to Increase Older Adults' Tendency to Anthropomorphise in Serious Games. *PLoS ONE* **2018**, *13*, e0199948. [[CrossRef](#)]
65. Werner, C.; Kardaris, N.; Koutras, P.; Zlatintsi, A.; Maragos, P.; Bauer, J.M.; Hauer, K. Improving Gesture-Based Interaction between an Assistive Bathing Robot and Older Adults via User Training on the Gestural Commands. *Arch. Gerontol. Geriatr.* **2020**, *87*, 103996. [[CrossRef](#)]
66. Gruneir, A.; Forrester, J.; Camacho, X.; Gill, S.S.; Bronskill, S.E. Gender Differences in Home Care Clients and Admission to Long-Term Care in Ontario, Canada: A Population-Based Retrospective Cohort Study. *BMC Geriatr.* **2013**, *13*, 48. [[CrossRef](#)] [[PubMed](#)]
67. American Association for Long-Term Care Insurance Long Term Care—Important Information for Women. Available online: <https://www.aaltci.org/long-term-care-insurance/learning-center/for-women.php> (accessed on 4 December 2022).
68. Burke, S.; Carron, A.; Eys, M.; Ntoumanis, N.; Estabrooks, P. Group versus Individual Approach? A Meta-Analysis of the Effectiveness of Interventions to Promote Physical Activity. *Sport Exerc. Psychol. Rev.* **2006**, *2*, 19–35.
69. Cadore, E.L.; Rodríguez-Mañas, L.; Sinclair, A.; Izquierdo, M. Effects of Different Exercise Interventions on Risk of Falls, Gait Ability, and Balance in Physically Frail Older Adults: A Systematic Review. *Rejuvenation Res.* **2013**, *16*, 105–114. [[CrossRef](#)] [[PubMed](#)]
70. Glaros, N.M.; Janelle, C.M. Varying the Mode of Cardiovascular Exercise to Increase Adherence. *J. Sport Behav.* **2001**, *24*, 42–62.
71. Bradley, M.M.; Lang, P.J. Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential. *J. Behav. Ther. Exp. Psychiatry* **1994**, *25*, 49–59. [[CrossRef](#)]

72. Heerink, M.; Kröse, B.; Evers, V.; Wielinga, B. Assessing Acceptance of Assistive Social Agent Technology by Older Adults: The Almere Model. *Int. J. Soc. Robot.* **2010**, *2*, 361–375. [[CrossRef](#)]
73. Cronbach, L.J. Coefficient Alpha and the Internal Structure of Tests. *Psychometrika* **1951**, *16*, 297–334. [[CrossRef](#)]
74. Kehoe, J. Basic Item Analysis for Multiple-Choice Tests. *Pract. Assess. Res. Eval.* **2019**, *4*, 10. [[CrossRef](#)]
75. Pallant, J. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*, 4th ed.; Open University Press/McGraw-Hill: Maidenhead, Australia, 2011; ISBN 978-1-952533-63-1.
76. Prakash, A.; Rogers, W.A. Why Some Humanoid Faces Are Perceived More Positively than Others: Effects of Human-Likeness and Task. *Int. J. Soc. Robot.* **2015**, *7*, 309–331. [[CrossRef](#)] [[PubMed](#)]
77. Bedaf, S.; Marti, P.; De Witte, L. What Are the Preferred Characteristics of a Service Robot for the Elderly? A Multi-Country Focus Group Study with Older Adults and Caregivers. *Assist. Technol.* **2019**, *31*, 147–157. [[CrossRef](#)]
78. Tu, Y.-C.; Chien, S.-E.; Yeh, S.-L. Age-Related Differences in the Uncanny Valley Effect. *Gerontology* **2020**, *66*, 382–392. [[CrossRef](#)]
79. Ekkekakis, P.; Hall, E.E.; Petruzzello, S.J. The Relationship between Exercise Intensity and Affective Responses Demystified: To Crack the 40-Year-Old Nut, Replace the 40-Year-Old Nutcracker! *Ann. Behav. Med. Publ. Soc. Behav. Med.* **2008**, *35*, 136–149. [[CrossRef](#)] [[PubMed](#)]
80. Smith, A.E.; Eston, R.; Tempest, G.D.; Norton, B.; Parfitt, G. Patterning of Physiological and Affective Responses in Older Active Adults during a Maximal Graded Exercise Test and Self-Selected Exercise. *Eur. J. Appl. Physiol.* **2015**, *115*, 1855–1866. [[CrossRef](#)]
81. Bixby, W.R.; Spalding, T.W.; Hatfield, B.D. Temporal Dynamics and Dimensional Specificity of the Affective Response to Exercise of Varying Intensity: Differing Pathways to a Common Outcome. *J. Sport Exerc. Psychol.* **2001**, *23*, 171–190. [[CrossRef](#)]
82. Woo, M.; Kim, S.; Kim, J.; Petruzzello, S.J.; Hatfield, B.D. The Influence of Exercise Intensity on Frontal Electroencephalographic Asymmetry and Self-Reported Affect. *Res. Q. Exerc. Sport* **2010**, *81*, 349–359. [[CrossRef](#)]
83. Rhodes, R.E.; Kates, A. Can the Affective Response to Exercise Predict Future Motives and Physical Activity Behavior? A Systematic Review of Published Evidence. *Ann. Behav. Med.* **2015**, *49*, 715–731. [[CrossRef](#)]
84. McAuley, E.; Jerome, G.J.; Elavsky, S.; Marquez, D.X.; Ramsey, S.N. Predicting Long-Term Maintenance of Physical Activity in Older Adults. *Prev. Med.* **2003**, *37*, 110–118. [[CrossRef](#)]
85. Oppenauer, C.; Preschl, B.; Kalteis, K.; Kryspin-Exner, I. Technology in Old Age from a Psychological Point of View. In *HCI and Usability for Medicine and Health Care*; Holzinger, A., Ed.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4799, pp. 133–142, ISBN 978-3-540-76805-0.
86. Tacken, M.; Marcellini, F.; Mollenkopf, H.; Ruoppila, I.; Széman, Z. Use and Acceptance of New Technology by Older People. Findings of the International MOBILATE Survey: “Enhancing Mobility in Later Life”. *Gerontechnology* **2005**, *3*, 126–137. [[CrossRef](#)]
87. Chen, T.L.; Bhattacharjee, T.; Beer, J.M.; Ting, L.H.; Hackney, M.E.; Rogers, W.A.; Kemp, C.C. Older Adults’ Acceptance of a Robot for Partner Dance-Based Exercise. *PLoS ONE* **2017**, *12*, e0182736. [[CrossRef](#)] [[PubMed](#)]
88. Wu, Y.-H.; Wrobel, J.; Cornuet, M.; Kerhervé, H.; Damnée, S.; Rigaud, A.-S. Acceptance of an Assistive Robot in Older Adults: A Mixed-Method Study of Human-Robot Interaction over a 1-Month Period in the Living Lab Setting. *Clin. Interv. Aging* **2014**, *9*, 801–811. [[CrossRef](#)] [[PubMed](#)]
89. Hebesberger, D.; Koertner, T.; Gisinger, C.; Pripfl, J. A Long-Term Autonomous Robot at a Care Hospital: A Mixed Methods Study on Social Acceptance and Experiences of Staff and Older Adults. *Int. J. Soc. Robot.* **2017**, *9*, 417–429. [[CrossRef](#)]
90. Cavallo, F.; Esposito, R.; Limosani, R.; Manzi, A.; Bevilacqua, R.; Felici, E.; Di Nuovo, A.; Cangelosi, A.; Lattanzio, F.; Dario, P. Robotic Services Acceptance in Smart Environments with Older Adults: User Satisfaction and Acceptability Study. *J. Med. Internet Res.* **2018**, *20*, e264. [[CrossRef](#)] [[PubMed](#)]
91. Broadbent, E.; Tamagawa, R.; Kerse, N.; Knock, B.; Patience, A.; MacDonald, B. Retirement Home Staff and Residents’ Preferences for Healthcare Robots. In Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009; pp. 645–650.
92. Chu, L.; Chen, H.-W.; Cheng, P.-Y.; Ho, P.; Weng, I.-T.; Yang, P.-L.; Chien, S.-E.; Tu, Y.-C.; Yang, C.-C.; Wang, T.-M.; et al. Identifying Features That Enhance Older Adults’ Acceptance of Robots: A Mixed Methods Study. *Gerontology* **2019**, *65*, 441–450. [[CrossRef](#)]
93. Deshmukh, A.; Lohan, K.S.; Rajendran, G.; Aylett, R. Social Impact of Recharging Activity in Long-Term HRI and Verbal Strategies to Manage User Expectations during Recharge. *Front. Robot. AI* **2018**, *5*, 23. [[CrossRef](#)] [[PubMed](#)]
94. Kanda, T.; Sato, R.; Saiwaki, N.; Ishiguro, H. A Two-Month Field Trial in an Elementary School for Long-Term Human–Robot Interaction. *IEEE Trans. Robot.* **2007**, *23*, 962–971. [[CrossRef](#)]
95. Sung, J.; Christensen, H.I.; Grinter, R.E. Robots in the Wild: Understanding Long-Term Use. In Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction, La Jolla, CA, USA, 11–13 March 2009; pp. 45–52.
96. Spink, K.S.; Carron, A.V. Group Cohesion Effects in Exercise Classes. *Small Group Res.* **1994**, *25*, 26–42. [[CrossRef](#)]
97. Burke, S.M.; Carron, A.V.; Shapcott, K.M. Cohesion in Exercise Groups: An Overview. *Int. Rev. Sport Exerc. Psychol.* **2008**, *1*, 107–123. [[CrossRef](#)]
98. Marcantonio, E.R.; Aneja, J.; Jones, R.N.; Alsop, D.C.; Fong, T.G.; Crosby, G.J.; Culley, D.J.; Cupples, L.A.; Inouye, S.K. Maximizing Clinical Research Participation in Vulnerable Older Persons: Identification of Barriers and Motivators. *J. Am. Geriatr. Soc.* **2008**, *56*, 1522–1527. [[CrossRef](#)]
99. Soule, M.C.; Beale, E.E.; Suarez, L.; Beach, S.R.; Mastromauro, C.A.; Celano, C.M.; Moore, S.V.; Huffman, J.C. Understanding Motivations to Participate in an Observational Research Study: Why Do Patients Enroll? *Soc. Work Health Care* **2016**, *55*, 231–246. [[CrossRef](#)] [[PubMed](#)]

100. Beauchamp, M.R.; Carron, A.V.; McCutcheon, S.; Harper, O. Older Adults' Preferences for Exercising Alone versus in Groups: Considering Contextual Congruence. *Ann. Behav. Med.* **2007**, *33*, 200–206. [[CrossRef](#)] [[PubMed](#)]
101. Cohen-Mansfield, J.; Marx, M.S.; Biddison, J.R.; Guralnik, J.M. Socio-Environmental Exercise Preferences among Older Adults. *Prev. Med.* **2004**, *38*, 804–811. [[CrossRef](#)]
102. Fan, J.; Beuscher, L.; Newhouse, P.A.; Mion, L.C.; Sarkar, N. A Robotic Coach Architecture for Multi-User Human-Robot Interaction (RAMU) with the Elderly and Cognitively Impaired. In Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication, New York, NY, USA, 26–31 August 2016; pp. 445–450.
103. Correia, F.; Mascarenhas, S.; Prada, R.; Melo, F.S.; Paiva, A. Group-Based Emotions in Teams of Humans and Robots. In Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction, Chicago, IL, USA, 5–8 March 2018; pp. 261–269.
104. Kessler, T.; Hollbach, S. Group-Based Emotions as Determinants of Ingroup Identification. *J. Exp. Soc. Psychol.* **2005**, *41*, 677–685. [[CrossRef](#)]
105. Affanni, A.; Aminosharieh Najafi, T.; Guerci, S. Development of an EEG Headband for Stress Measurement on Driving Simulators. *Sensors* **2022**, *22*, 1785. [[CrossRef](#)]
106. Filippini, C.; Perpetuini, D.; Cardone, D.; Chiarelli, A.M.; Merla, A. Thermal Infrared Imaging-Based Affective Computing and Its Application to Facilitate Human Robot Interaction: A Review. *Appl. Sci.* **2020**, *10*, 2924. [[CrossRef](#)]
107. Abd Latif, M.H.; Yusof, H.; Sidek, S.N.; Rusli, N. Thermal Imaging Based Affective State Recognition. In Proceedings of the IEEE International Symposium on Robotics and Intelligent Sensors, Langkawi, Malaysia, 18–20 October 2015; pp. 214–219.
108. Manullang, M.C.T.; Lin, Y.-H.; Lai, S.-J.; Chou, N.-K. Implementation of Thermal Camera for Non-Contact Physiological Measurement: A Systematic Review. *Sensors* **2021**, *21*, 7777. [[CrossRef](#)]
109. Silva, C.M.d.S.E.; Gomes Neto, M.; Saquetto, M.B.; da Conceição, C.S.; Souza-Machado, A. Effects of Upper Limb Resistance Exercise on Aerobic Capacity, Muscle Strength, and Quality of Life in COPD Patients: A Randomized Controlled Trial. *Clin. Rehabil.* **2018**, *32*, 1636–1644. [[CrossRef](#)]
110. Whittall, J.; Waller, S.M.; Silver, K.H.C.; Macko, R.F. Repetitive Bilateral Arm Training with Rhythmic Auditory Cueing Improves Motor Function in Chronic Hemiparetic Stroke. *Stroke* **2000**, *31*, 2390–2395. [[CrossRef](#)]
111. Vigorito, C.; Giallauria, F. Effects of Exercise on Cardiovascular Performance in the Elderly. *Front. Physiol.* **2014**, *5*, 51. [[CrossRef](#)] [[PubMed](#)]
112. Schneider, S.; Kummert, F. Comparing Robot and Human Guided Personalization: Adaptive Exercise Robots Are Perceived as More Competent and Trustworthy. *Int. J. Soc. Robot.* **2021**, *13*, 169–185. [[CrossRef](#)]
113. Hong, A.; Lunscher, N.; Hu, T.; Tsuboi, Y.; Zhang, X.; Franco dos Reis Alves, S.; Nejat, G.; Benhabib, B. A Multimodal Emotional Human–Robot Interaction Architecture for Social Robots Engaged in Bidirectional Communication. *IEEE Trans. Cybern.* **2021**, *51*, 5954–5968. [[CrossRef](#)] [[PubMed](#)]
114. Ficocelli, M.; Terao, J.; Nejat, G. Promoting Interactions between Humans and Robots Using Robotic Emotional Behavior. *IEEE Trans. Cybern.* **2016**, *46*, 2911–2923. [[CrossRef](#)] [[PubMed](#)]
115. Agrawal, Y.; Shah, Y.; Sharma, A. Implementation of Machine Learning Technique for Identification of Yoga Poses. In Proceedings of the IEEE 9th International Conference on Communication Systems and Network Technologies, Gwalior, India, 10–12 April 2020; pp. 40–43.
116. Dias, P.A.; Malafronte, D.; Medeiros, H.; Odone, F. Gaze Estimation for Assisted Living Environments. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, 1–5 March 2020; pp. 279–288.
117. Lazowski, D.A.; Ecclestone, N.A.; Myers, A.M.; Paterson, D.H.; Tudor-Locke, C.; Fitzgerald, C.; Jones, G.; Shima, N.; Cunningham, D.A. A Randomized Outcome Evaluation of Group Exercise Programs in Long-Term Care Institutions. *J. Gerontol. A Biol. Sci. Med. Sci.* **1999**, *54*, M621–M628. [[CrossRef](#)]
118. Swank, A.M.; Funk, D.C.; Durham, M.P.; Roberts, S. Adding Weights to Stretching Exercise Increases Passive Range of Motion for Healthy Elderly. *J. Strength Cond. Res.* **2003**, *17*, 374–378. [[CrossRef](#)]
119. Fiebert, I.; Fuhri, J.R.; New, M.D. Elbow, Forearm, and Wrist Passive Range of Motion in Persons Aged Sixty and Older. *Phys. Occup. Ther. Geriatr.* **1993**, *10*, 17–32. [[CrossRef](#)]
120. Alves, S.F.; Shao, M.; Nejat, G. A Socially Assistive Robot to Facilitate and Assess Exercise Goals. In Proceedings of the IEEE International Conference of Robotics and Automation Workshop on Mobile Robot Assistants for the Elderly, Montreal, QC, Canada, 20–24 May 2019; pp. 1–5.
121. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
122. Trace, J.; Janssen, G.; Meier, V. Measuring the Impact of Rater Negotiation in Writing Performance Assessment. *Lang. Test.* **2017**, *34*, 3–22. [[CrossRef](#)]
123. InteraXon Inc. *Technical Specifications, Validation, and Research Use*; InteraXon Inc.: Toronto, ON, Canada, 2016.
124. Barachant, A.; Morrison, D.; Banville, H.; Kowaleski, J.; Shaked, U.; Chevallier, S.; Tresols, J.J.T. Muse-lsl. Available online: <https://github.com/alexandrebarachant/muse-lsl> (accessed on 31 May 2020).
125. Al-Nafjan, A.; Hosny, M.; Al-Wabil, A.; Al-Ohali, Y. Classification of Human Emotions from Electroencephalogram (EEG) Signal Using Deep Neural Network. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 419–425. [[CrossRef](#)]

126. Zhao, G.; Zhang, Y.; Ge, Y. Frontal EEG Asymmetry and Middle Line Power Difference in Discrete Emotions. *Front. Behav. Neurosci.* **2018**, *12*, 225. [[CrossRef](#)] [[PubMed](#)]
127. Castellano, G.; Leite, I.; Pereira, A.; Martinho, C.; Paiva, A.; McOwan, P.W. Affect Recognition for Interactive Companions: Challenges and Design in Real World Scenarios. *J. Multimodal User Interfaces* **2010**, *3*, 89–98. [[CrossRef](#)]
128. Riether, N.; Hegel, F.; Wrede, B.; Horstmann, G. Social Facilitation with Social Robots? In Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction, Boston, MA, USA, 5–8 March 2012; pp. 41–47.
129. Sanghvi, J.; Castellano, G.; Leite, I.; Pereira, A.; McOwan, P.W.; Paiva, A. Automatic Analysis of Affective Postures and Body Motion to Detect Engagement with a Game Companion. In Proceedings of the 6th International Conference on Human-Robot Interaction, Lausanne, Switzerland, 6–9 March 2011; pp. 305–312.
130. Glenberg, A.M.; Schroeder, J.L.; Robertson, D.A. Averting the Gaze Disengages the Environment and Facilitates Remembering. *Mem. Cogn.* **1998**, *26*, 651–658. [[CrossRef](#)]
131. Al Zoubi, O.; Ki Wong, C.; Kuplicki, R.T.; Yeh, H.-W.; Mayeli, A.; Refai, H.; Paulus, M.; Bodurka, J. Predicting Age from Brain EEG Signals-A Machine Learning Approach. *Front. Aging Neurosci.* **2018**, *10*, 184. [[CrossRef](#)]
132. Zappasodi, F.; Marzetti, L.; Olejarczyk, E.; Tecchio, F.; Pizzella, V. Age-Related Changes in Electroencephalographic Signal Complexity. *PLoS ONE* **2015**, *10*, e0141995. [[CrossRef](#)]
133. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.-S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijholt, A.; Patras, I. DEAP: A Database for Emotion Analysis Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [[CrossRef](#)]
134. Abadi, M.K.; Subramanian, R.; Kia, S.M.; Avesani, P.; Patras, I.; Sebe, N. DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. *IEEE Trans. Affect. Comput.* **2015**, *6*, 209–222. [[CrossRef](#)]
135. Katsigiannis, S.; Ramzan, N. DREAMER: A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-Cost off-the-Shelf Devices. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 98–107. [[CrossRef](#)]
136. Lin, Y.-P.; Yang, Y.-H.; Jung, T.-P. Fusion of Electroencephalographic Dynamics and Musical Contents for Estimating Emotional Responses in Music Listening. *Front. Neurosci.* **2014**, *8*, 94. [[CrossRef](#)]
137. Pandey, P.; Seeja, K.R. Subject Independent Emotion Recognition from EEG Using VMD and Deep Learning. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 1730–1738. [[CrossRef](#)]
138. McColl, D.; Louie, W.-Y.G.; Nejat, G. Brian 2.1: A Socially Assistive Robot for the Elderly and Cognitively Impaired. *IEEE Robot. Autom. Mag.* **2013**, *20*, 74–83. [[CrossRef](#)]
139. Verduyn, P.; Delaveau, P.; Rotgé, J.-Y.; Fossati, P.; Van Mechelen, I. Determinants of Emotion Duration and Underlying Psychological and Neural Mechanisms. *Emot. Rev.* **2015**, *7*, 330–335. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.