This is the author's version of the article that has been published at the 2019 IEEE International Conference on Systems, Man, and Cybernetics, Bari, Italy.

You Are Doing Great! Only One Rep Left: An Affect-Aware Social Robot for Exercising

Mingyang Shao, *Student Member, IEEE*, Silas Franco Dos Reis Alves, *Member, IEEE*, Omar Ismail, Xinyi Zhang, Goldie Nejat, *Member, IEEE* and Beno Benhabib

***Abstract*— Regular exercise has immediate and long-term benefits for people of all ages. Maintaining an adequate amount of daily exercise is important to overall health and wellbeing. Our research focuses on the development of a socially assistive robot, Salt, to facilitate different upper body exercises. During the exercises, the robot is uniquely able to autonomously detect a user's affect and engagement as well as measure their heart rate to prevent overexertion. A robot emotion model using an *n*th order Markov Chain is used to determine the robot's appropriate emotions during interactions based on user affect and engagement, and its own emotion history. Human-robot interaction experiments were conducted to investigate perceived usefulness and acceptance. The results showed that most users were engaged and had positive valence towards the robot during the interactions. Post-experiment questionnaire results also showed they were able to detect the robot's emotions and enjoyed interacting with it.**

I. INTRODUCTION

Regular exercise is important to improve overall health. Namely, for children and adolescents, exercising can improve strength, body composition, as well as the cardiovascular and cardio-respiratory system [1]. For adults, exercising can lower the risk of early death, coronary heart disease, stroke, high blood pressure, and certain cancers [2]. Seniors engage in less exercise due to a decrease in mobility and the fear of accidents, increasing their risk of developing chronic degenerative diseases, such as cardiovascular disease, type 2 diabetes, obesity, and osteoporosis [2]. Regular exercise can also provide mental and psychosocial benefits, such as a better sense of wellbeing and an overall reduction in the symptoms associated with anxiety and/or depression [2].

In recent years, social robots have been designed to be incorporated into our daily lives. They are expected to have social norms and follow accepted behaviors and rules [3]. It is important for social robots to communicate naturally with humans in order to build long-term relationships [3],[4]. To achieve more natural and engaging human-robot interactions (HRI), it is essential to recognize, understand, and respond to human emotions [3],[5]. A handful of social robots have been designed for facilitating exercising [6]–[11], but they do not consider bidirectional emotional interactions to promote activity engagement.

Our research focuses on developing socially assistive robots as autonomous assistants for activities of daily living including meal preparation and eating [12],[13], dressing [14],[15], and facilitating cognitively stimulating leisure activities such as Bingo, Trivia and card games [16],[17]. In this paper, we present the development of a socially assistive robot as an autonomous exercise facilitator to encourage physical activity. The robot Salt is used to interact with users and guide them through different exercises. The novelty is in the robot's ability to autonomously detect a user's affect and engagement during exercise sessions using multi-modal inputs such as Electroencephalogram (EEG) signals, and facial features, and adapt its own behaviors using an *n*th order Markov Chain robot emotion model. The emotion model uses affect and engagement as inputs as well as the robot's emotional history to determine the robot's appropriate emotions during interactions. Heart rate is also monitored to ensure a user's heart rate is within the target range for the exercises.

II. RELATED WORK

A. Robots for Exercising

1) Social Robots for Rehabilitation

In [6], the QT robot assisted and motivated patients with limited movements to perform upper limb exercises and maintain their adherence to a rehabilitation program. The robot displayed gestures, and emoji-like pictorial facial expressions. An RGB camera and color segmentation were used to track a ball the user was holding to detect the exercises. At a pre-defined repetition frequency set by a therapist, the robot provided feedback including positive reinforcement, happy gestures, or positive facial expressions. Experiments with healthy adults with and without the robot suggested that most users favored the robot as it provided feedback.

In [7], the NAO robot was used to perform upper body rehabilitation exercises with people with impaired upper limb functioning. NAO provided instructions and demonstrated different exercises to the users, however no user study was reported. In [8], a NAO robot was also used to autonomously instruct upper-limb rehabilitation exercises for patients with cerebral palsy and obstetric brachial plexus palsy. A Kinect sensor tracked user joints and the robot showed correct motions if needed. Experiments were conducted with healthy schoolchildren and pediatric patients. Results showed that most of the children were engaged and enjoyed the exercises, and would participate in future interactions.

This research is supported by AGE-WELL Inc., the Natural Sciences and Engineering Council of Canada (NSERC), and the Canada Research Chairs Program.

All authors are with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada (e-mail: mingyang.shao@mail.utoronto.ca).

2) Social Robots for Coaching Exercise

In [9], a NAO robot was used to autonomously coach older adults in stretching exercises. The robot learned the joint angles for the exercises from a human demonstrator using an RGB-D camera. Simple verbal instructions were also recorded for motions it could not physically implement. The robot tracked user motion and provided corrective or positive feedback. Experiments with older adults showed that the majority of them enjoyed the exercise sessions but did not consider the robot as a friend. The limitations on the social abilities of the robot, including displaying the same robot behavior for all users, negatively affected user experience.

In [10], the robot "Double" served as an exercise trainer for older adults. A tablet mounted on the robot displayed videos of exercises with audio explanations. User joints were tracked by a Kinect sensor and compared with the desired outputs to determine if they were performing the exercises correctly. At the end of each exercise scenario, emoji-like pictorial facial expressions were displayed as visual feedback based on the correctness of the user's poses. Experiments with older adults showed that the robot was able to engage and motivate participants to take part in the exercises.

In [11], an upper-limb exercise coach named HOBBIT was developed. The robot's tablet displayed videos of a human trainer performing the exercises. User joints were tracked by a Kinect sensor to monitor movements and count repetitions. The detected joints were also displayed on the screen side-by-side to the movements of the trainer. In addition, textual feedback was used for user performance and potential corrections to the movements. An experiment with older adults revealed that most users enjoyed exercising with the robot and some mentioned that their level of exercise increased after using it.

The aforementioned robots show the potential for using them to motivate and assist users in performing exercises for improved health. Even though two of them expressed pre-set emotions as visual feedback based on user performance [6],[10], they did not consider incorporating robot emotion models to explicitly respond to varying human affect via emotional behaviors to enhance HRI. Incorporating human affect to influence robot behaviors can better engage a user in a person-centered activity and improve user performance [3], which is essential for our exercise facilitation application.

B. *Affect-Aware Robots*

A handful of social robots have been developed to detect human affect to promote natural HRI [3],[18]–[22]. For example, in [18], the stuffed-animal-like robot CuDDler used a support vector machine (SVM) classifier to detect affective states from facial information. In addition, emotional sounds (e.g., laughing) and non-voiced acts (e.g., punching) were detected using a sound recognizer. The robot determined its own discrete emotions with a look-up table, using the affective states, verbal acts, and non-voiced events as inputs. It expressed its emotions through body movements, sounds and eye blinking. Questionnaires from experiments showed that participants thought the robot understood their emotional acts and could display appropriate emotional behaviors.

In [19], the companion robot CONBE was developed for natural HRI. Human affect states were detected using a hidden Markov model (HMM) classifier from facial features. Using the user's affect, the robot regulated its own emotions via the online machine learning method TopoART-R. The robot's emotions were displayed using pictorial eye expressions. A preliminary experiment conducted with one user showed that the robot was able to display expected eye expressions based on that user's affect.

In [3] and [20], the human-like socially assistive robot Brian was developed to assist people in different activities. In [3], Brian helped users to create an activity schedule. A robot emotion model determined the robot's discrete emotions using an online updating HMM based on human affect and the drives the robot needed to satisfy. Human affect was classified based on static body language. The robot emotions were displayed using a combination of facial expressions, head gestures, body language, and vocal intonation. Experiments showed that Brian was able to adapt its own emotional behaviors to human affect with positive responses from the participants. In [20], Brian was used to assist older adults in meal eating. Dynamic body language features were identified using a combination of a Mixture of Gaussians (MOG) approach and a 3D human upper body model, which were then successfully classified into valence and arousal. Experiments with older adults showed that natural body language can be used for affect classification during HRI.

A robot that is capable of human affect-awareness and in turn responding to varying human affect using robot emotional behaviors can promote natural HRI [3] and engagement with users for extended periods of time [5]. The majority of existing affect-aware robots detect user affect using facial expressions [18], [19]. However, facial expressions are difficult to detect during exercising, as they are perturbed when the user performs physical activity due to the increase of effort and muscle fatigue [23]. Furthermore, body language cannot be detected as the users are engaged in upper body exercises [3],[20]. EEG signals, on the other hand, can be used to detect affect while participating in physical activities [24]. In this paper, we propose for the first time the integration of EEG signals for an affect-aware robot. Namely, we present the development of a social robot exercise facilitator that can autonomously adapt its emotional behaviors based on user affect during exercises.

III. An Affect-Aware Social Robot Exercise Coach

The robot used in our work is Salt, a humanoid Pepper robot from Softbank. The proposed autonomous robot exercise facilitator architecture for Salt, Fig. 1, is comprised of four modules: 1) User State Detection, 2) Robot Emotion, 3) Exercise Monitoring, and 4) Robot Interaction. The EEG signals and heart rate data from the sensor headband, which includes both the EEG sensor and heart rate sensor, are inputs to the User State Detection Module for classifying the user's affect in terms of valence and monitoring heart rate activity during the interaction. The user's head poses are detected from the RGB camera in the Kinect sensor to classify user engagement also via the User State Detection Module. User affect and engagement are then used by the Robot Emotion Module to determine the robot emotion using an *n*th order Markov Model (MM). During the exercises, 3D data from the Kinect sensor is used by the Exercise Monitoring Module to determine the user's arm poses during exercising. The robot's emotion, detected arm poses, and heart rate activity are used as inputs into the Robot Interaction Module to determine the robot exercise-specific behaviors. The details of each module

are discussed below.

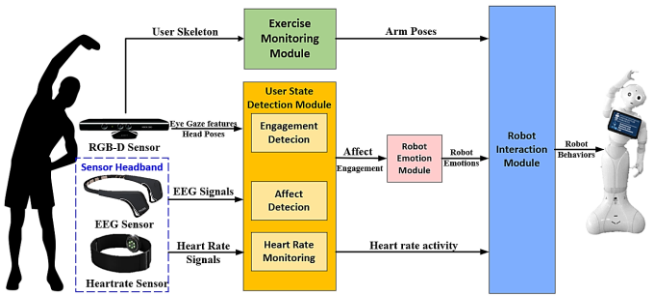


Fig. 1. Robot Exercise Facilitator Architecture

A. User State Detection Module

The User State Detection Module determines 1) user affect defined by valence, 2) engagement, and 3) heart rate.

1) Affect Detection using EEG

A low-cost four-channel dry electrode EEG sensor, InteraXon Muse 2016, is utilized for measuring EEG signals. EEG signals are measured at four electrode locations located at TP9 (above the left ear), AF7 (left side of the forehead), AF8 (right side of the forehead), and TP10 (above the right ear) with a sampling rate of 256 Hz using the International 10-20 system to describe the location of EEG electrodes [25].

EEG Feature Extraction

In our work, each EEG signal is decomposed into three distinct frequency bands: θ (4-8 Hz), α (8-13 Hz), and β (13-30 Hz) [25]. The θ, α, and β bands acquired from each electrode location (TP9, AF7, AF8, TP10) are used as Power Spectral Density (PSD) features [26]. The frontal EEG asymmetry, defined as the power difference between the frontal hemispheres of the brain within each frequency band, has been previously correlated to the emotional state of a person [27]. Namely, higher activation in the left hemisphere represents positive valence while higher activation in the right hemisphere represents negative valence, which can be measured by the ratio of the alpha and beta bands [27]. Valence represents a user's level of pleasure with respect to a stimulus (robot), and is an important metric to determine how much a user enjoys an interaction [20]. The two frontal EEG asymmetry features for measuring the hemispherical activation are adapted from [27] and are calculated as:

$$f1 = \frac{\alpha_{AF8}}{\beta_{AF8}} - \frac{\alpha_{AF7}}{\beta_{AF7}} \quad (1)$$

$$f2 = \frac{\alpha_{AF7}+\alpha_{AF8}}{\beta_{AF7}+\beta_{AF8}} \quad (2)$$

where α_{AF7} , α_{AF8} , β_{AF7}, and β_{AF8} are α and β bands measured at AF7 and AF8 locations.

In total, 14 features which include 12 PSD features measured at the locations TP9, AF7, AF8, TP10 and two frontal EEG asymmetry features are utilized to classify the valence of the users during the activity.

Classification Techniques

To effectively identify the valence of the users during HRI, several learning-based classifiers were investigated. We compared the following classes of learning techniques using the Scikit-Learn toolbox [28]: 1) biologically inspired models, 2) non-linear models, 3) decision trees, 4) nearest neighbors, and 5) probabilistic techniques. The grid search strategy was used to optimize the parameters of each classification technique. The objective of the classification task is to predict the user's positive or negative valence from EEG signals, which is later used as an input into the Robot Emotion Module to determine the robot emotions.

EEG training data was collected from ten individuals (4 female and 6 male) between 22 and 40 years old. The participants were students or staff at the university. Stimuli consisting of pictures and videos designed for emotion elicitation from publicly validated datasets [29],[30] were presented to each participant while they wore the EEG headband. The pictures and video were a priori categorized into inducing positive valence or negative valence. Each set of emotion eliciting videos/images was 5 minutes and was followed by a 2-minute break. During the break, participants were asked to report their valence level using the Self-Assessment Manikin (SAM) scale [31].

A ten-fold cross validation was performed on the training data to evaluate the prediction results. The classification rates for the techniques are presented in Table I. The three hidden layer Multilayer Perceptron Neural Network model was selected based on having the highest classification rate.

TABLE I. CLASSIFICATION RATES FOR DIFFERENT CLASSIFIERS

Class	Technique	Classification Rate
Biologically inspired model	Neural Network	71.50%
Non-linear	SVM	67.96%
Decision trees	Random Forest	69.72%
Nearest neighbor	k-NN	68.83%
Probabilistic technique	Naïve Bayes	52.89%
	Gaussian Process	69.59%

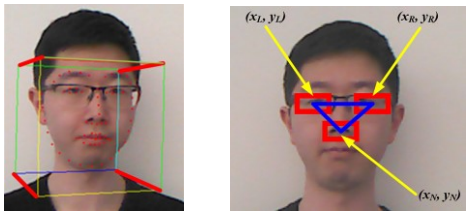
2) Engagement Detection

By detecting and responding to the level of engagement of the user, the robot aims to maintain quality of experience [32]. In our work, a Microsoft Kinect sensor is used for the Engagement Detection Module. This module detects the user's engagement as either engaged or not engaged based on their visual focus of attention (VFOA) via their head poses. Namely, engaged when the user has a VFOA towards the robot for more than 50% of time; not engaged when the user does not focus on the robot.

Head poses are determined using the facial landmarks detected by the Dlib library detector. The detector first finds faces in an image using HOG + Linear SVM, which creates a bounding box with (x, y) coordinates of the centroid of the face [33]. The spatial positions (x, y) of the mouth, eyebrows, eyes, nose, and jaw are localized using an ensemble of regression trees that are trained using images from the HELEN face database [34]. The head poses are estimated from facial landmarks by projecting them onto an image plane and mapping these points to 3D locations, which is done by using the known camera intrinsic matrix to compute the extrinsic matrix via the Levenberg-Marquardt algorithm [35]. The distances between the front of the face and back of the face (labeled in red in Fig. 2(a)) are computed as head pose features. For improved robustness, a second method computes the head pose features from the distances and angles between the left eye (x_L, y_L), right eye (x_R, y_R), and the nose (x_N, y_N) as shown in Fig. 2(b), which has been used in our previous work [36]. The head pose features determined from both methods are used for training a *k-NN* classifier for engagement detection.

Five volunteers recorded two robot exercise sessions, engaged and not engaged with the robot, to build the training data. In total, 1,945 data samples were collected. By

performing a ten-fold cross validation on the training data, the *k-NN* classifier was able to achieve an accuracy of 95%.



(a) (b)

Fig. 2. Using the Dlib library detector to identify facial landmarks and head pose features

3) Heart Rate Monitoring

According to the American Heart Association (AHA) [37], the maximum heart rate (MHR) of an individual can be estimated by:

$$MHR = 220 - age \tag{3}$$

In addition, AHA suggests that the target heart rate during anaerobic exercises (e.g., strength building and flexibility) is between 50% to 85% of the maximum heart rate [37]. In our work, the user's heart rate in bpm is measured at 1 Hz by an optical heart rate sensor, Polar OH1, during the HRI. Heart rate is measured to ensure that during exercising it is below the user's upper threshold (85%) to prevent overexertion.

B. Robot Emotion Module

We adapt herein the *n*th order MM developed in our previous work [38] to determine robot emotions. The model uses both the robot emotional history and the user state as inputs. The robot's emotional state at time *t*, E'_t, for *m* robot emotions and *l* user states is represented as:

$$\boldsymbol{E_t} = w_1 \boldsymbol{H_t} + W_2 \boldsymbol{A_t} \tag{4}$$

$$E'_t = f(\boldsymbol{E_t}) \tag{5}$$

where $\boldsymbol{E_t}$ is the robot emotion output vector, and $\boldsymbol{H_t}$ is the robot emotional state vector based on the emotional history at time t. $\boldsymbol{A_t}$ represents the user state input vector based on both the user's valence and engagement. w_1 is a scalar which represents the weight of the influence of the robot emotional history on the current robot emotion. W_2 represents the robot emotion state-human affect probability distribution. $f(\boldsymbol{E_t})$ represents a winner takes all function to determine the emotion for the robot to display.

Robot Emotion History Model

The following property is satisfied by the robot's emotional state for an nth order MM [38]:

$$P(H_t = e_0 | H_{t-1} = e_1, \dots, H_1 = e_{t-1}) = P(H_t = e_0 | H_{t-1} = e_1, \dots, H_{t-n} = e_n) \tag{6}$$

which represents the probability of the current emotion e_0 being dependent on the previous emotion history. $e_0, \dots, e_t \in \{1, \dots, m\}$ represent the displayed robot emotions.

As time passes, the influence of a past emotion should decrease [38]. Therefore, the weight of each past emotion in discrete time is reduced by a decay function:

$$\lambda_i = e^{-ai}, 0 < a < -\ln(\varepsilon) \tag{7}$$

where λ_i is the weight for the emotion at discrete time $i \in T^+$, a is the rate of decay, $n = \lceil T \rceil - 1$, ε is the lower threshold of the decay function. The value of w_1 of the emotion model is determined as λ_0.

The robot emotion transition probability is modeled as:

$$P(H_t = e_0 | H_{t-1} = e_1, \dots, H_{t-n} = e_n) = \sum_{i=1}^{n} \lambda_i q_{e_i e_0} \tag{8}$$

where $q_{e_i e_0}$ is an element of Q_i, the $m \times m$ robot emotion transition probability. The robot emotion history model can be represented as:

$$\boldsymbol{H_t} = \sum_{i=1}^{n} \lambda_i Q_i \boldsymbol{H_{t-i}} \tag{9}$$

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

$$F^{(i)} = \begin{pmatrix} f_{11}^{(i)} & \cdots & f_{1m}^{(i)} \\ \vdots & \ddots & \vdots \\ f_{m1}^{(i)} & \cdots & f_{mm}^{(i)} \end{pmatrix} \tag{10}$$

Therefore, the estimated Q_i is represented as:

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

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

$$q_{kj}^{(i)} = \begin{cases} \frac{f_{kj}^{(i)}}{\sum_{j=1}^{m} f_{kj}^{(i)}} & \text{if } \sum_{j=1}^{m} f_{kj}^{(i)} \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

C. Exercise Monitoring Module

To determine if the user is performing the requested exercise, the 3D spatial position of the shoulder, elbow, and wrist joints of each arm are detected using the OpenNI and NITE frameworks [39], which use random forests to classify the body parts using depth-invariant features obtained by the Kinect sensor. Five arm poses are trained and classified using a *k-NN* classifier: the resting pose; and the complete pose for each of the four exercises. To collect the training data for the *k-NN* classifier, volunteers (two males and one female) recorded one session of each exercise providing 150 samples of the five arm poses. The classifier achieved an accuracy of 93% when compared to a human expert coder.

D. Robot Interaction Module

The Robot Interaction Module uses a Finite State Machine, Fig. 3, to determine the corresponding robot behaviors, based on the exercise goals and user state input. Prior to the exercise activities, the user's resting heart rate is measured, and their maximum heart rate and heart rate target range are estimated. If the user's heart rate is above the upper limit at any time during the exercise session, the session will be stopped, and the robot will request the user to rest.

At the beginning of each exercise session, Salt greets the user and explains the exercise interaction. Then, Salt shows each exercise set and asks the user to imitate it for *r* number of repetitions. Salt can facilitate a number of different arm exercises including (a) biceps curls, (b) arm raises, (c) standing fly, and (d) lateral trunk stretches. A video of the robot facilitating these exercises with participants can be found here on our YouTube Channel. These exercises were selected as they can strengthen arm and shoulder muscles, and can be performed both while standing and sitting, allowing users with mobility restrictions to partake in them.

During the exercise session, the user's valence, engagement, and arm poses are determined, and the heart rate is monitored. Based on the user's affect (positive or negative valence) and engagement (engaged or not engaged), Salt will determine its own emotion using the robot emotion model to provide both verbal and nonverbal feedback. For the robot, we selected a set of two positive (high valence) and two negative (low valence) emotions with different intensities [40]. The positive emotions are happy and interested while the negative emotions are worried and sad. Each emotion is displayed by using a combination of eye color, body language, and vocal intonation. Our emotional expression designs were adapted based on our previous work [41]. Fig. 4 and Table II detail the emotional expressions designed for Salt.

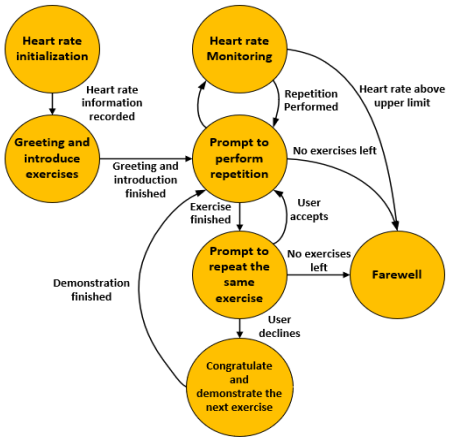


Fig.3. Finite State Machine of the Interaction Module.

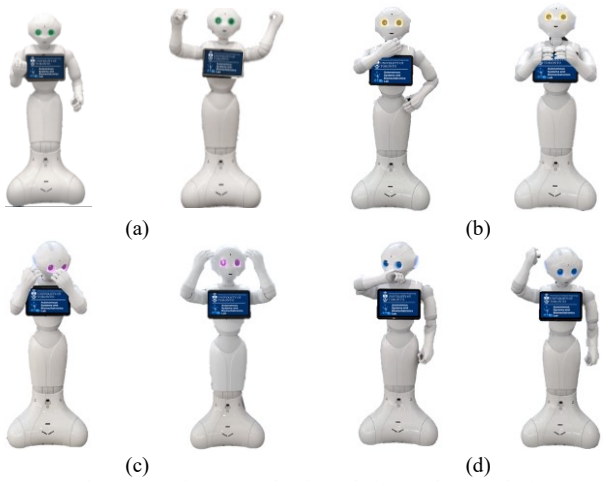


Fig. 4. Emotion expression examples for Salt for: (a) happy; (b) interested; (c) worried; (d) sad.

TABLE II. ROBOT EMOTION DISPLAY

Emotion	Eye Color	Vocal Intonation	Body Language
Happy	Green	High speed Medium pitch High volume	Arms open, fist pump, dancing, swaying
Interested	Yellow	Medium speed Medium pitch Medium volume	Nodding, holding chin, tapping fingers, leaning forward and opening arms
Worried	Purple	Medium speed High pitch Low volume	Covering face, pretending to bite nails, hands on ears, akimbo
Sad	Blue	Low speed High pitch Low volume	Sobbing on arms, fidgeting, scratching head, lowering body

After the user finishes the repetitions for each set, Salt congratulates the user and asks them if they want to do the same exercise again. If the user accepts, Salt repeats the same exercise for r repetitions. Otherwise, Salt performs the next exercise. After performing all the exercises, Salt congratulates the user and says farewell. The details of the robot activity-specific behaviors are shown in Table III.

TABLE III. ROBOT BEHAVIORS FOR EXERCISE SESSIONS

Stage	Non-Verbal	Verbal
Introduction	Waves to the user	Hello, my name is Salt, your personal exercise coach. We are going to do four 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?
Introduce Exercise	Performs the poses for the exercise	First, we will do an exercise called biceps curl. Let me show you how to do it. Start by putting your arms down beside you. Then, curl your arms up, like this. And then bring your arms down again.
Prompt to perform repetitions	Performs the poses for the exercise	We are going to do 1 cycle. Let's get started. *r,r-1*…last one!
Prompt to repeat the exercise	Motions to user	Do you want to continue doing this exercise?
Congratulate (Happy)	Happy robot emotion display	Well done! Doing exercise with you is really enjoyable.
Congratulate (Interested)	Interested robot emotion display	You finished this exercise. You are doing great.
Congratulate (Worried)	Worried robot emotion display	I hope you feel alright after finishing this exercise.
Congratulate (Sad)	Sad robot emotion display	It's too bad you did not like this exercise. Hopefully you will like the next one more.
Farewell	Waves goodbye to the user	I hope the rest of your day goes well. Let's do this again sometime. Bye for now!

IV. TRAINING OF THE EMOTION MODEL

Prior to the experiments, a training stage was used to determine the initial values for the parameters of the robot's state-human affect probability W_2 using 5 different participants in the same age group as those in the experiments, where W_2 is a $m \times l$ matrix that represents the likelihood of user compliance for each robot emotion. Initially, the probabilities of robot emotional states were uniformly distributed to allow each emotion to have the same probability to be chosen. The robot interacted with the user by displaying an emotion i while asking the user to follow a number of exercise repetitions. If the user followed a repetition successfully, the corresponding element of the frequency matrix W_T was updated:

$$W_T^{i,j} = W_T^{i,j} + 1 \tag{13}$$

where $i = \{1, \dots, m\}$ and $j = \{1, \dots, l\}$.

The matrix W_2 was obtained by normalizing the matrix W_T after all repetitions were completed for all users:

$$W_2^{i,j} = \frac{W_T^{i,j}}{\sum_{i=1}^{m} W_T^{i,j}} \tag{14}$$

V. EXPERIMENTS

In order to investigate user affect and engagement, an HRI experiment was conducted with 15 students and staff at our university between the ages of 22 to 36 (μ=25, σ = 3.87). The one-on-one interaction, Fig. 5, consisted of the robot autonomously motivating a user through the exercises. The

number of repetitions for each exercise was chosen to be 8, based on recommendations by the AHA [42]. During the interaction, user valence and engagement were determined, and heart rate was monitored. After the experiment was completed, each participant provided their perceived affect for the entire interaction on a scale of -1 (negative valence) to +1 (positive valence) using the SAM scale [36]. A 5-point Likert questionnaire (1-strongly disagree, 3-neutral, 5-strongly agree) was also administered to the users to investigate their overall experience with the robot.

Fig. 5. HRI Scenario

A. User State Detection Results

The robot was able to appropriately determine the correct valence for 14 of the 15 participants. The average valence and engagement results for all users across the stages of the interaction are presented in Fig. 6(a). On average, the users had positive valence and were engaged towards the robot. Both users 8 and 12 self-rated negative valence. User 8 mentioned that he/she had a muscle strain in his/her neck, and user 12 mentioned that he/she was already tired prior to starting the exercises. Our affect detection system was able to detect the negative valence of user 12, but not user 8. This discrepancy is due to the fact that our system was trained using visual stimuli to evoke positive and negative feelings, and not from internal pain. This would need to be considered for future work. An overall success rate of 93.33% was determined for valence detection. In addition, the users were engaged towards the robot on average for 93.21% (standard deviation of 7.36%) of the time during the interactions while spending the remaining time observing their environment. Finally, none of the user's heart rate exceeded the upper limit of the target range (85% of their MHR) as they did not find the exercises that intense.

The detailed results for two of the users are presented in Fig. 6. In Fig. 6(b), the user was detected to have positive valence and be engaged towards the robot for 92.3% of time, and the robot displayed happy and interested emotions based on this positive valence and high engagement, as well as its own emotion history. In Fig. 6(c), the user was detected to have negative valence, and the robot transitioned to worried and sad emotions due to this negative valence.

B. Questionnaire Results

The 5-point Likert questionnaire results are presented in Table IV. Based on the results, most users liked using Salt as an exercise facilitator (μ=4.07, σ=0.8). Four users found the sensor headband uncomfortable to wear (rating of 1 or 2 on the Likert scale) while the remaining majority were relatively neutral (rating of 3). The majority of users thought that the robot displayed appropriate emotions during the interactions (μ=4.33, σ=0.9). Moreover, Salt had high ratings on questions related to perceived usefulness with respect to the activity (Questions 5-10). All users were able to follow the robot's instructions (negatively coded μ=1.07, σ=0.26) and found it had a clear voice (μ=4.73, σ=0.46). The users were able to clearly understand the robot's instructions (negatively coded, μ=1.40, σ=0.83). However, one user did mention that they paid more attention towards the robot's movements rather than verbal instructions.

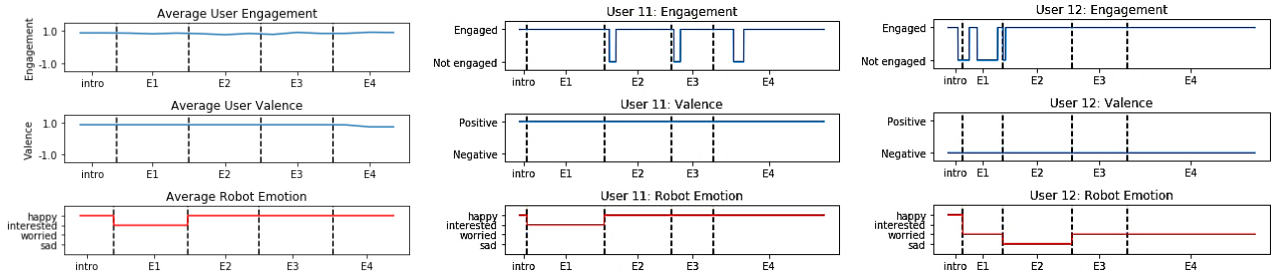
TABLE IV. USER ACCEPTANCE RESULTS

	Question	Min†	Max†	Mean	SD*
Interaction: Acceptance	1. I like using the robot to do exercise	3	5	4.07	0.80
	2. I would use the robot again	2	5	3.73	0.96
	3. The sensor headband is uncomfortable to wear	1	4	2.87	0.92
	4. The robot displays appropriate emotions	2	5	4.33	0.90
Interaction: Perceived Usefulness	5. The exercises the robot got me to do are good for my overall health	3	5	4.47	0.74
	6. The robot is not helpful for doing exercise	1	4	2.07	1.03
	7. The robot clearly displays each exercise	3	5	4.60	0.63
	8. The robot is difficult to use	1	2	1.33	0.49
	9. The robot motivates me to exercise	2	5	3.67	0.72
	10. I don't trust the robot's advice	1	3	1.67	0.72
Robot Appearance and Movements	11. The robot moves too fast for me to follow	1	2	1.07	0.26
	12. I think the robot has a clear voice	4	5	4.73	0.46
	13. I don't understand the robot's instructions	1	4	1.40	0.83
	14. I think the robot's size is appropriate for exercising	2	5	3.53	1.13

* SD: Standard Deviation
† Min and max: the minimum and maximum value of all users' responses

VI. CONCLUSIONS

In this paper, we present the development of the socially assistive Salt robot as an autonomous exercise facilitator. The robot is able to autonomously detect the user's affect and engagement during exercising in order to determine its own emotions using an *n*th order Markov model. In addition, the



(a) (b) (c)

Fig. 6. Results of engagement, valence and robot emotions: (a) average for all users; (b) a positive valence user (User 11); and (c) a negative valence user (User 12) across the stages of the interaction, including introduction ("intro") and exercise sets from one (E1) to four (E4).

robot measures the user's heart rate to prevent overexertion. Experiments were conducted to investigate the performance, acceptance and perceived usefulness of the robot. Results showed that the users, in general, had positive valence during the interaction and were engaged towards the robot. Furthermore, the robot was able to display appropriate emotions based on the detected user states. Most users also enjoyed using the robot. In the future, we will design HRI experiments to have Salt facilitate exercises with older adults living in one of our partner long-term care facilities to investigate efficacy and usability with this population. The robot will adapt an exercise program based on the preference of individual user (i.e., type of exercise and exercise speed).

REFERENCES

[1] B. Landry and S. Driscoll, "Physical activity in children and adolescents," *PM R*, vol. 4, no. 11, pp. 826–832, 2012.

[2] U.S. Department of Health and Human Services, "2008 Physical Activity Guidelines for Americans," 2008.

[3] M. Ficocelli, J. Terao, and G. Nejat, "Promoting Interactions Between Humans and Robots Using Robotic Emotional Behavior," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2911–2923, 2016.

[4] F. Cavallo, F. Semeraro, L. Fiorini, G. Magyar, P. Sinčák, and P. Dario, "Emotion Modelling for Social Robotics Applications: A Review," *J. Bionic Eng.*, vol. 15, no. 2, pp. 185–203, 2018.

[5] I. Leite, C. Martinho, and A. Paiva, "Social Robots for Long-Term Interaction: A Survey," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 291–308, 2013.

[6] F. J. Rodriguez-lera, L. Gomes, P. Ziafati, A. Nazarikhorram, A. Stefanetti, and A. Schuller, "Emotional Robots for Coaching : Motivating Physical Rehabilitation using Emotional," *HRI Conference*, 2018.

[7] M. Assad-Uz-Zaman, M. R. Islam, and M. H. Rahman, "Upper - Extremity Rehabilitation with NAO Robot," *5th International Conference of Control, Dynamic Systems, and Robotics*, 2018.

[8] J. C. Pulido, J. C. Gonz´alez, and F. Fern´andez, "NAOTherapist : Autonomous Assistance of Physical Rehabilitation Therapies with a Social Humanoid Robot," *International Workshop on Assistive & Rehabilitation Technology (IWART)*, 2016, pp. 15–16.

[9] B. Görer, A. A. Salah, and H. L. Akın, "An autonomous robotic exercise tutor for elderly people," *Auton. Robots*, vol. 41, pp. 657–678, 2017.

[10] A. Lotfi, C. Langensiepen, and S. W. Yahaya, "Socially Assistive Robotics : Robot Exercise Trainer for Older Adults," *Technologies*, vol. 6, no. 1, 2018.

[11] M. Foukarakis *et al.*, "A Robot-based Application for Physical Exercise Training," *Int. Conf. Info. Commun. Tech. for Ageing Well and e-Health*, 2016, pp. 45–52.

[12] D. McColl and G. Nejat, "Meal-time with a socially assistive robot and older adults at a long-term care facility," *Spec. Issue HRI Syst. Stud. J. Human-Robot Interact.*, vol. 2, no. 1, pp. 152–173, 2013.

[13] D. McColl and G. Nejat, "A Socially Assistive Robot that can Monitor Affect of the Elderly during Meal-Time Assistance," *J. Med. Devices-Transactions ASME*, vol. 8, no. 3, 2014.

[14] L. Woiceshyn, Y. Wang, G. Nejat, and B. Benhabib, "Personalized Clothing Recommendation by a Social Robot," *IEEE 5th International Symposium on Robotics and Intelligent Sensors (IRIS 2017)*, 2017, pp. 179–185.

[15] L. Woiceshyn, Y. Wang, G. Nejat, and B. Benhabib, "A Socially Assistive Robot to Help with Getting Dressed," *Design of Medical Devices Conference*, 2017.

[16] W. G. Louie, J. Li, C. Mohamed, F. Despond, V. Lee, and G. Nejat, "Tangy the Robot Bingo Facilitator: A Performance Review," *J. Med. Devices-Transactions ASME*, vol. 9, no. 2, 2015.

[17] W. G. Louie, J. Li, T. Vaquero, and G. Nejat, "A Focus Group Study on the Design Considerations and Impressions of a Socially Assistive Robot for Long - Term Care," *IEEE International Symposium on Robot and Human Interactive Communication*, 2014, pp. 237–242.

[18] K. L. Dilip *et al.*, "Affective Social Interaction with CuDDler Robot," *6th IEEE Conf. Robot. Autom. Mechatronics*, pp. 179–184, 2013.

[19] S. Chumkamon, E. Hayashi, and K. Masato, "Intelligent emotion and behavior based on topological consciousness and adaptive resonance theory in a companion robot," *Biol. Inspired Cogn. Archit.*, vol. 18, pp. 51–67, 2016.

[20] D. McColl and G. Nejat, "Determining the Affective Body Language of Older Adults during Socially Assistive HRI," *IEEE/RSJ Int. Conf. Intell. Robot. Syst. (IROS 2014)*, pp. 2633–2638, 2014.

[21] A. Hong, Y. Tsuboi, G. Nejat, and B. Benhabib, "Multimodal Affect Recognition for Assistive Human-Robot Interactions," *Design of Medical Devices Conference*, 2017.

[22] D. McColl, A. Hong, N. Hatakeyama, G. Nejat, and B. Benhabib, "A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI," *J. Intell. Robot. Syst.*, vol. 82, pp. 101–133, 2016.

[23] M. C. Uchida *et al.*, "Identification of muscle fatigue by tracking facial expressions," *PLoS One*, vol. 13, no. 12, 2018.

[24] E. K. Broelz, P. Enck, A. M. Niess, P. Schneeweiss, S. Wolf, and K. Weimer, "The neurobiology of placebo effects in sports: EEG frontal alpha asymmetry increases in response to a placebo ergogenic aid," *Sci. Rep.*, vol. 9, no. 1, 2019.

[25] InteraXon, "Technical specifications, validation, and research use."

[26] A. Al-Nafjan, M. Hosny, A. Al-Wabil, and Y. Al-Ohali, "Classification of Human Emotions from Electroencephalogram (EEG) Signal using Deep Neural Network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, pp. 419–425, 2017.

[27] R. Ramirez, Z. Vamvakousis, "Detecting Emotion from EEG Signals Using the Emotive Epoc Device," *Brain Inf.* 2012, pp. 175–184.

[28] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825--2830, 2011.

[29] E. S. Dan-Glauser and K. R. Scherer, "The Geneva affective picture database (GAPED): A new 730-picture database focusing on valence and normative significance," *Behav. Res. Methods*, vol. 43, no. 2, pp. 468–477, 2011.

[30] M. K. Uhrig *et al.*, "Emotion Elicitation: A Comparison of Pictures and Films," *Front. Psychol.*, vol. 7, 2016.

[31] M. Bradley and P. J. Lang, "Measuring emotion : the Self-Assessment Manikin and the semantic differential," *J Behav Ther Exp Psychiatry*, vol. 25, no. I, pp. 49–59, 1994.

[32] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the Engagement with Social Robots," *Int. J. Soc. Robot. Springer*, pp. 1–14, 2015.

[33] A. Rosebrock, "Histogram of Oriented Gradients and Object Detection," 2014. [Online]. Available: https://www.pyimagesearch.com/2014/11/10/histogram-oriented-gradients-object-detection/. [Accessed: 31-Mar-2019].

[34] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1867–1874, 2014.

[35] Z. Zhang, "A Flexible New Technique for Camera Calibration æ," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[36] D. McColl, W. G. Louie, and G. Nejat, "Brian 2.1: A Socially Assistive Robot for the Elderly and Cognitively Impaired," *Spec. Issue Assist. Robot. IEEE Robot. Autom. Mag.*, vol. 20, no. 1, 2013.

[37] American Heart Association, "Know Your Target Heart Rates for Exercise, Losing Weight and Health." [Online]. Available: https://www.heart.org/en/healthy-living/fitness/fitness-basics/target-heart-rates. [Accessed: 31-Mar-2019].

[38] X. Zhang, S. Alves, G. Nejat, and B. Benhabib, "A Robot Emotion Model with History," *IEEE International Symposium on Robotics and Intelligent Sensors*, 2017, pp. 5–7.

[39] PrimeSense, "OpenNI/OpenNI: OpenNI." [Online]. Available: https://github.com/OpenNI/OpenNI. [Accessed: 15-Mar-2019].

[40] J. Poser, J.A. Russell, and B.S. Peterson, "The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology," *Dev Psychopathol*, vol. 17, no. 3, pp. 715–734, 2005.

[41] D. McColl and G. Nejat, "Recognizing Emotional Body Language Displayed by a Human-like Social Robot," *Int. J. Soc. Robot.*, vol. 6, no. 2, pp. 261–280, 2014.

[42] American Heart Association, "Strength and Resistance Training Exercise." [Online]. Available: https://www.heart.org/en/healthy-living/fitness/fitness-basics/strength-and-resistance-training-exercise. [Accessed: 31-Mar-2019].

This is the author's version of the article that has been published at the 2019 IEEE International Conference on Systems, Man, and Cybernetics, Bari, Italy.