

Article

Task Engagement as Personalization Feedback for Socially-Assistive Robots and Cognitive Training

Konstantinos Tsiakas* , Maher Abujelala and Fillia Makedon

Heracleia Human-Centered Computing Laboratory, Computer Science and Engineering Department, University of Texas at Arlington, Arlington, TX 76010, USA; maher.abujelala@mavs.uta.edu (M.A.); makedon@uta.edu (F.M.)

* Correspondence: konstantinos.tsiakas@mavs.uta.edu; Tel.: +1-817-805-9043

Received: 22 February 2018; Accepted: 9 May 2018; Published: 14 May 2018



Abstract: Socially-Assistive Robotics (SAR) has been extensively used for a variety of applications, including educational assistants, exercise coaches and training task instructors. The main goal of such systems is to provide a personalized and tailored session that matches user abilities and needs. While objective measures (e.g., task performance) can be used to adjust task parameters (e.g., task difficulty), towards personalization, it is essential that such systems also monitor task engagement to personalize their training strategies and maximize the effects of the training session. We propose an Interactive Reinforcement Learning (IRL) framework that combines explicit feedback (task performance) with implicit human-generated feedback (task engagement) to achieve efficient personalization. We illustrate the framework with a cognitive training task, describing our data-driven methodology (data collection and analysis, user simulation) towards designing our proposed real-time system. Our data analysis and the reinforcement learning experiments on real user data indicate that the integration of task engagement as human-generated feedback in the RL mechanism can facilitate robot personalization, towards a real-time personalized robot-assisted training system.

Keywords: socially-assistive robotics; personalization; interactive reinforcement learning; brain-computer interfaces

1. Introduction

Socially-Assistive Robotics (SAR) is a research area that studies how robots can be deployed to assist users through social interaction, as users perform a cognitive or physical task [1]. The goal of such assistive robots is to build an effective interaction with the user, so as to enhance his/her performance during the training session. Such agents can be deployed for various tasks, such as cognitive and/or physical training [2,3], language tutoring [4], rehabilitation exercises [5] and others. As technology advances and becomes more affordable, SAR systems can be considered as an effective tool for educational and training purposes. A key feature of SAR systems is their ability to provide personalized interaction with the user. Personalization is essential for effective training or tutoring since it can enhance the effectiveness of the session, maximizing user learning potential. Based on the famous Bloom's two sigma result [6], one-to-one tutoring presents better learning effects than group (conventional) tutoring.

An effective robot-based training system should be able to adjust the task parameters in order to provide a training session that fits user's abilities and skills, resulting in an "optimally challenging activity" [7]. One approach is through behavioral and physiological monitoring, i.e., affect detection. Emotion and flow theories have been extensively applied to HRI applications. Considering the flow theory [8], affective states such as boredom, engagement and anxiety can be detected through EEG sensors and used to adjust task difficulty in order to keep users in the flow channel [9].

Socially-assistive robotics has been developed to improve user performance through the use of physiological signals [3], considering the Yerkes–Dodson law, which links human arousal and task performance [10]. From another perspective, recent works define interactive personalization for socially-assistive robotics as “the process by which an intelligent agent adapts to the needs and preferences of an individual user through eliciting information directly from that user about their state” [11,12]. Based on this definition, certain information about the human learner may only be observable through the learner’s direct input, as explicit feedback (e.g., self-report). Recent works focus on combining both implicit and explicit probes from the user considering task engagement, including self-reports, facial expressions and task behavior, towards developing a personalized engagement detection system [13]. Taking all approaches into consideration, personalization is a complex computational problem that requires the training agent to interactively assess, adapt and leverage a model of the user’s abilities, skills, preferences, affect, etc., utilizing different types of feedback [14].

Personalization can be considered as an interaction management problem; the problem of modeling and optimizing the interaction patterns in order to maximize the efficiency of the interaction [15]. Considering the sequential nature of human-robot interactions, interaction management can be seen as a sequential decision making problem, where the system needs to learn the appropriate sequence of actions in order to optimize the interaction, given a utility metric, e.g., user performance and/or satisfaction. Machine Learning (ML) methods can be used to solve this problem and optimize the interaction patterns. Reinforcement Learning (RL) is an appropriate machine learning framework for sequential decision making problems and dynamic environments. Several RL approaches have been successfully applied to model the dialogue manager in adaptive dialogue systems [16].

In this work, we discuss how Interactive Reinforcement Learning (IRL) methods can be used to facilitate personalization for different types of users, in a SAR-based cognitive training scenario. More specifically, we show how task engagement can be used as human-generated feedback through learning from feedback. We present related work on SAR systems and reinforcement learning approaches for robot personalization, as well as methods for measuring and using task engagement through EEG sensors for adaptive and personalized interactions (Section 2). Then, we describe our proposed system for personalized robot-assisted cognitive training (Section 3). We present the data collection procedure, followed by the data analysis (Section 4). Then, we present the experimental procedure, including the user simulation and the interactive reinforcement learning experiments (Section 5), and we conclude with a discussion on possible improvements and future steps towards a real-time personalized SAR system (Section 6).

2. Background and Related Work

2.1. Reinforcement Learning for Socially-Assistive Robotics

Socially-assistive robots can provide personalized assistance through social interaction, by adjusting verbal, non-verbal or mixed behaviors (supportive feedback, attention acquisition, affective behavior, etc.) towards establishing an efficient interaction with the user. In this work, we focus on the personalization procedure of a SAR system, formulating it as a reinforcement learning problem. As mentioned before, reinforcement learning is an appropriate paradigm for learning sequential decision making processes with the potential to develop adaptive robots that adjust their behavior based on human abilities and needs, through either implicit or explicit feedback.

In the context of socially-assistive robotics, RL approaches are used to enable the robot to personalize its behavior (i.e., policy) towards different users. Depending on the application, RL is used to adjust different parameters that can influence the effectiveness of the interaction. For example, in a language learning scenario, a social robot has been deployed to achieve personalization through affective behavior [4]. The presented system uses a camera to capture and analyze facial expressions

and affective features (gaze, smile, engagement, valence, etc.), during a language tutoring application, in order to provide a personalized affective interaction through social verbal behavior (valence and engagement of spoken instructions). The system combines the estimated values (user engagement and affect) into a reward signal. The system learns to adjust its behavior by selecting appropriate motivational strategies (using verbal and non-verbal actions), based on current child's state (affect and performance), in order to keep the child engaged and in a positive affective state. Another example demonstrates how RL-based SAR systems can be deployed as exercise trainers, to enable personalized physical rehabilitation through social behavior adaptation [17]. In this work, the authors investigate three different robot behavior parameters (i.e., interaction distance, speed and vocal content) and their effect on the user, in order to achieve long-term personalization and maximize user performance. The authors proposed a Policy Gradient RL (PGRL) method to learn the combination of the behavior parameters, which maximizes user compliance and performance.

In another work, a social robotic tutor was proposed to assist users in logic puzzle solving [18]. The robotic tutor learns a user model during the interaction, which assesses whether the user is experiencing difficulties in the task. Based on this, the robot decides whether it will perform a supportive behavior or not. An RL-based personalization module learns which specific supportive behavior (tangible, esteem, emotional support) can maximize user performance. RL-based personalization approaches have been also proposed for adaptive storytelling through social signals [19]. More specifically, the authors proposed an RL approach to learn which robot personality parameter (extraversion level) matches user's preferences and keeps them engaged. The proposed system estimates user engagement, through a multisensing framework (SSI), and adjusts the robot's current extraversion level to maximize user engagement during the session. Their simulated experiments show promising results in a small, but noisy state space. Social robots have also been used to engage individuals in cognitively-stimulating activities through task-related assistance, using verbal and non-verbal behaviors [20], where the robot acts as a motivator during a memory card game. A hierarchical RL approach is used to enable the robot to learn when to deploy specific assistive behaviors (assistance, encouragement and celebration) and personalize the interaction based on perceived user states (activity performance and arousal). In a similar application, RL has been proposed to dynamically adapt robot's assistive behavior expressed by different modalities (i.e., speech, gaze, facial expression and head gesture) in a memory card game [21]. The system, under the guidance of a human wizard, decides if the user needs help and selects the appropriate combination of gestures to grab user's attention, guide the user through the task and maximize task progress. In a recent work [22], the author has presented an architecture to learn personalized robot policies, combining human expert demonstrations and reinforcement learning, in order to provide users with personalized assistance during Activities of Daily Living (ADLs).

These works support the effectiveness of reinforcement learning as a personalization framework for SAR-based systems. A main limitation of such RL-based systems is scalability; learning efficiency and convergence speed when the state-action space is large and the environment dynamics (human behavior), even the environment itself (new user), change. Another limitation in designing RL agents for interactive systems, including the definition of a proper state-action space, is defining an appropriate reward function that serves the purpose of the system [23,24]. Our research is motivated by these challenges that arise when different types of users and feedback types are considered for real-time personalization using reinforcement learning [25]. To this end, we illustrate the proposed interactive learning and adaptation framework with a cognitive training task, investigating how interactive RL methods (learning from feedback) can be used to integrate human-generated feedback through EEG data (task engagement) and facilitate personalization.

2.2. Brain-Computer Interfaces

In a learning (tutoring, training) environment, affective and cognitive states are highly correlated with task engagement and learning effects [26]. Positive states, i.e., flow, curiosity and task engagement,

have a positive correlation with learning, in comparison with negative states, such as boredom and/or frustration [27]. Taking into consideration such information is essential in designing an effective learning or training system that estimates and monitors task engagement to adjust the behavior parameters and sustain compliance [28,29]. However, quantifying task engagement and attention is not trivial, since it depends on and overlaps with several user states, such as interest, sustained attention, immersion and (attentional and emotional) involvement [30]. Recently, Brain-Computer Interfaces (BCI) have been used towards this purpose [31].

There is a growing trend towards using passive BCI systems, which implicitly monitor brain activity, to personalize interactive systems through EEG sensors [32]. In our work, we follow the approach of passive BCI to measure and utilize task engagement, using the Muse EEG headset [33]. Muse is a low-cost portable EEG headset, which has been used to detect brain states of concentration and relaxation [34], task enjoyment [35], student cognitive state detection [36], as well as for pain detection through self-calibrating protocols and interactive machine learning [37]. Muse provides four channels of data coming from dry frontal EEG electrodes (TP9, AF7, AF8, TP10). The device provides access to raw EEG signals, as well as to a set of power spectral density measurements extracted from the raw data. The frequency bands provided by the device are δ (1–4 Hz), θ (5–8 Hz), α (9–13 Hz), β (12–30 Hz) and γ (30–50 Hz). Research in EEG analysis for task engagement has offered the following formula for calculating a signal E , based on α , β and θ waves, which is correlated with task engagement: $E = \beta / (\alpha + \theta)$ [38]. This approach has been followed for intelligent interactive systems that monitor task engagement and adjust their behavior to keep users engaged. In an adaptive storytelling application [32], a social robot used behavioral techniques (vocal cues, gestures) to regain user attention during drops in engagement, as estimated by the aforementioned formula. In a similar manner, this engagement index has been used to evaluate task engagement while playing a video game [39]. The engagement index was capable of differentiating high intensity game events (e.g., player death) from general game play.

2.3. Learning from Human Feedback

Learning from feedback is an Interactive Reinforcement Learning (IRL) method that treats human input as a reinforcement signal after the executed action. Several works have considered the use of feedback to facilitate the learning procedure. In [40], they proposed the TAMER framework (Training an Agent Manually via Evaluative Reinforcement), which includes a supervised learner for building a human reward model during the interaction, which enables humans to shape agents during their learning. In [41], they present a learning framework that integrates human feedback, in the form of facial expressions, as an additional reinforcement signal, to guide the agent during learning. In [42], the authors propose a method for personalized information filtering for learning user preferences, by capturing and transforming implicit feedback from the user to a reinforcement signal. These approaches support that IRL methods can facilitate real-time personalization from human-generated feedback.

There are two main approaches based on how feedback is integrated to the RL mechanism: reward shaping and Q -augmentation [40]. Reward shaping uses the feedback as an additional reward component added to the environmental reward ($R'(s, a) = R(s, a) + \beta * H(s, a)$), while in Q -augmentation, feedback is used to directly adjust the policy, by modifying the Q -values ($Q'(s, a) = Q(s, a) + \beta * H(s, a)$). A specific Q -value is an estimate of the long-term expected discounted reward for taking action a in state s . In both techniques, $H(s, a)$ is the shaping function and β is the combination parameter. In this work, we outline the developmental process towards a data-driven SAR system that monitors task engagement and performance during a cognitive training task (sequence learning). We illustrate how interactive reinforcement learning approaches can be used for task engagement, through EEG data, and facilitate personalization. The long-term goal of this research is to develop HRI systems that combine and utilize different types of real-time human-generated feedback (implicit or explicit) to continuously and dynamically adjust their behavior in a lifelong learning setup.

3. Personalized Robot-Assisted Cognitive Training

3.1. The Sequence Learning Task

In this section, we present our experimental testbed, a cognitive task related to working memory and sequencing. Sequencing is the ability to arrange language, thoughts, information and actions in an effective order. It has been shown that many children with learning and attention issues have trouble with sequencing [43]. Influenced by the NIH Toolbox Cognition Battery Working Memory test [44] and SAR-based approaches for cognitive training [2,17], we present the sequence learning task; a cognitive task, related to working memory, that evaluates the ability of an individual to remember and repeat a spoken sequence of letters. For our experimental setup, we deploy the NAO robot [45] as a socially-assistive robot that monitors both behavioral (task performance) and physiological (EEG) data and instructs the user towards a personalized cognitive training session. The sequence learning setup is shown in Figure 1. The user has three buttons in front of them (“A”, “B”, “C”), and the robot asks the user to repeat a given sequence of these letters. We follow the assumption that task difficulty $D = [1, 2, 3, 4]$ is proportional to sequence length $L = [3, 5, 7, 9]$. Based on the outcome ($success = [0, 1]$), the user receives a score, defined as:

$$score = \begin{cases} D, & \text{if } success = 1 \\ -1, & \text{if } success = 0 \end{cases} \quad (1)$$

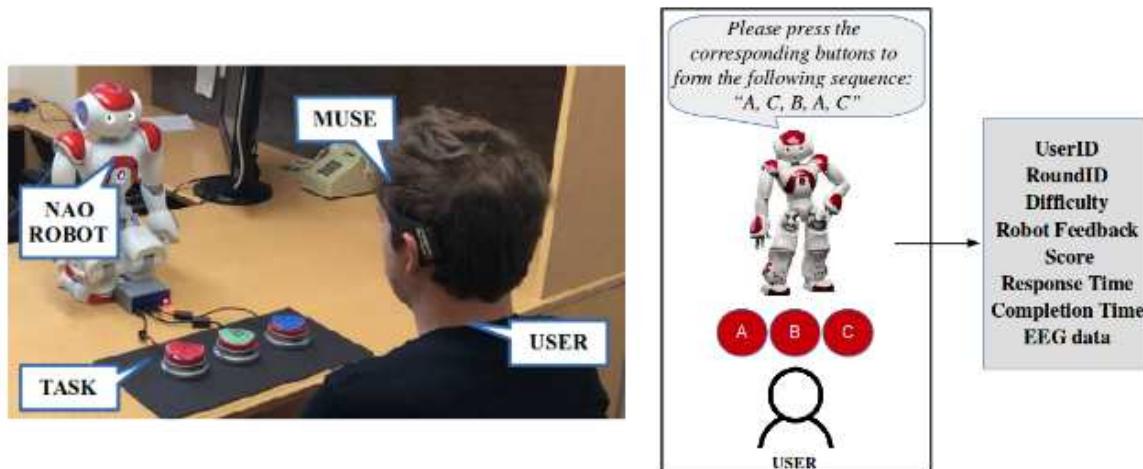


Figure 1. The sequence learning setup.

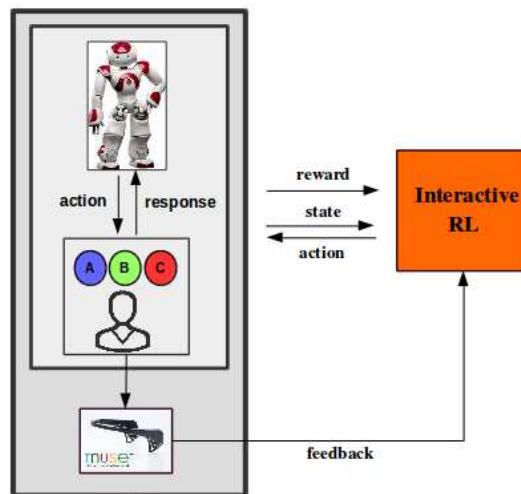
Based on this scoring approach, the user gets more points by succeeding in harder levels, while the negative score is the same for all levels, such that the system does not discourage users from playing harder levels. The robot can also provide feedback after the user completes a sequence. Studies have examined the influence of different feedback styles, including the absence of feedback, on user's engagement and performance [46]. In our system, when the robot provides feedback, it reports the current outcome (success or failure) by providing either encouraging or challenging feedback and continues with a sequence of the same difficulty (length). In the case of no feedback, the robot moves on with the next sequence (same or different difficulty) without reporting on the result. Preliminary results show that absence of feedback can positively affect task performance in certain difficulty levels [47]. Examples of encouraging and challenging feedback are shown in Table 1.

Table 1. Examples of encouraging and challenging feedback.

Encouraging Feedback	
success	“That was great! keep up the good work”
failure	“Oh, that was wrong! But that’s fine! don’t give up!”
Challenging Feedback	
success	“OK, that was easy enough! Let’s see now...”
failure	“Hey! Are you there? Stay focused when I speak!”

3.2. System Architecture

The system consists of two main components: (A) the task (physical component) and (B) the RL agent (computational component), as we show in Figure 2. The robot interacts with the user during the training task; it assigns a sequence to the user, and the user responds by pressing the corresponding buttons. The system keeps track of the current sequence length and robot feedback, task performance, as well as task engagement through EEG (Muse). This information is used by the system to adjust its behavior (state). The robot must learn an efficient training policy, which will dictate how to adjust task difficulty and robot feedback based on the current state (action), in order to maximize task performance and keep the user engaged (reward). A reinforcement learning agent is deployed to learn such personalized strategies.

**Figure 2.** The sequence learning task as an interactive Reinforcement Learning (RL) problem.

Reinforcement learning provides an appropriate framework for interaction modeling and optimization for sequential decision making problems formulated as Markov Decision Processes (MDP). An MDP is described by a tuple $\langle S, A, T, R, \gamma \rangle$ where:

- S is a finite set of states (state space)
- A is the finite set of available actions (action space)
- T is the transition model where $T(s, a, s')$ denotes the probability $P(s'|s, a)$
- $R(s, a, s')$ is a reward function, which evaluates the transition $s, a \rightarrow s'$
- γ is a discount factor

MDP models can capture how human behavior, such as user performance, stochastically changes according to the robot's decisions. The solution of an MDP results in an optimal policy π ; the mapping from states to actions that maximizes the total expected return during the interaction. In our case, the state space includes information about task difficulty and robot feedback, as well as the previous result (previous level and outcome). More specifically, the state features are: sequence Length

$L = [3, 5, 7, 9]$, Robot Feedback $RF = \{0 : \text{None}, 1 : \text{Encouraging}, 2 : \text{Challenging}\}$ and Previous Result $PR = [-4, 4]$.

Based on the current state, the robot selects one of the available system actions (task difficulty, robot feedback), and the system perceives the next state, receiving a reward based on task performance and task engagement, as we describe in our experimental procedure. The transition model captures how user performance varies from state to state. We need to note that the state space is designed to be stochastic; each state might lead to a successful turn (positive score) with some probability, in order to capture different user abilities. Table 2 shows the state-action and reward components.

Table 2. The defined MDP of the problem.

State Features	System Actions	Reward
Sequence Length (SL)	Level 1 (L3)	Current Score
Robot Feedback (RF)	Level 2 (L5)	Task Engagement
Previous Result (PR)	Level 3 (L7)	
	Level 4 (L9)	
	Encouraging Feedback (RF1)	
	Challenging Feedback (RF2)	

4. Data Collection and Analysis

4.1. Data Collection

For the data collection procedure, we recruited 69 Computer Science undergraduate and graduate students, who received extra credits in their class, after agreement with their instructors. Each user completed a predefined session of the sequence learning task, consisted of 25 turns (sequences). Each session was sampled uniformly from a set of predefined sessions, such that the difficulty levels and the robot feedback types were uniformly distributed across all users. Each session lasted for about 20 min, including a post-session user survey.

Before the session, the participant was provided with a verbal and written explanation of the task and the experimental procedure. After the proper placement of the Muse sensor, the NAO robot greeted the user and provided them with a sequence example to get them familiarized with the task and the button setup, as well as to ensure that the user did not require any more clarifications. From a preliminary user study on this task [47], we found out that users prefer to be aware of the upcoming difficulty level, before the robot announces the sequence. Considering this, the robot announces the difficulty level to the user, before each sequence.

At the end of the session, each user completed a user survey, regarding task difficulty and their self-assessment on task engagement and performance for each level. During each session, for each turn, we recorded the task parameters (turn ID, sequence length, robot feedback), user's performance (user response, reaction and completion time, as well as the EEG data as provided by the Muse sensor. The EEG signals were evaluated and filtered based on the Muse headband status indicator, resulting in a dataset of 50 users (41 males and nine females) between the age of 17 and 45 ($M = 23.32$, $SD = 5.88$). The full dataset, including the subset used for the following analysis and experiments, is publicly available online for researchers [48]. The dataset is collected and stored such that it can be explored for several research purposes and approaches, including robot behavior modeling, user modeling, recommendation systems, EEG analysis and others. Considering the EEG data, we separate them based on the task state; while the user listens to the sequence (listening phase) and while the user responds (response phase). For the purpose of this work, we use the EEG data recorded during the listening phase. As an ongoing work, we have demonstrated how EEG data can be used to predict task performance [49].

4.2. User Survey

At the end of the experiment, the participants completed a survey for each level of the task. More specifically, the survey was designed to elicit subjective information from the users about task difficulty, task performance, task engagement and the reason for disengagement, if applicable. Considering task difficulty, more than 95% of the users consider Levels 1 and 2 to be either easy or just right. In contrast, only 44% and 8% of users find Levels 3 and 4 easy or just right, respectively. Taking task performance into consideration, for Level 1, 84% of users reported that they did above average in the task. This value changes to 70% for Level 2 and to 34% for Level 3. For Level 4, only 6% of users reported that their performance was either average or below. Level 2 has the highest percentage of engaged users, as 96% of users reported they were engaged in this level, and this percentage is the lowest for Level 4 with only 60% of users reporting engagement in this level. This percentage goes to 76% for Level 1 and 86% for Level 3. The majority of the disengaged users in Levels 1 and 2 reported that task level easiness was the reason for their disengagement, and disengaged users in Levels 3 and 4 reported task level difficulty as the reason for their disengagement. Figure 3 shows a summary of the user survey results.

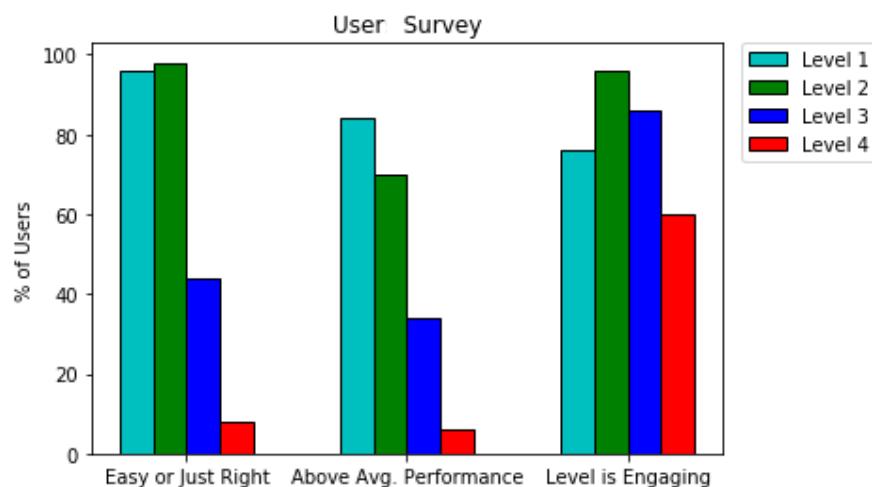


Figure 3. User survey results.

The results of the user survey illustrate how different users perceive and assess task difficulty, engagement and performance across different difficulty levels, denoting the need for learning personalized training policies based on user skills and preferences.

4.3. User Modeling and Clustering

For the purposes of our RL experiments, we analyzed the performance and engagement data, in order to model different user behaviors across different task parameters. Our first step in the data analysis is to perform user clustering and group participants based on their task performance and engagement across the different difficulty levels. For each user, we estimate the engagement signal E using the engagement index formula [32,38]. More specifically, the relative band values for the α , β and θ frequencies were extracted and smoothed, applying an Exponentially-Weighted Moving Average (EWMA) filter, based on which:

$$\tilde{s}(t) = \begin{cases} s(t) = y(t) & t = 0 \\ s(t) = \alpha * \tilde{s}(t-1) + (1 - \alpha) * s(t-1) & t > 0 \end{cases} \quad (2)$$

These smoothed values were then used to estimate the engagement signal $E = \alpha / (\beta + \theta)$, segmented per round and annotated by task difficulty, robot feedback and current result. For each

user, this signal was normalized to $[0, 1]$ and the mean engagement values were estimated for each difficulty level. Each user can be represented with an array $UM = [P_1, P_2, P_3, P_4, E_1, E_2, E_3, E_4]$, where $P_{level=i} = P(success|level = i)$, and $E_{level=i} = \bar{e}_i$, where \bar{e}_i is the mean engagement value for level $i \in [1, 4]$. In Figure 4, we visualize two different users and how they can be described by their performance and engagement values.

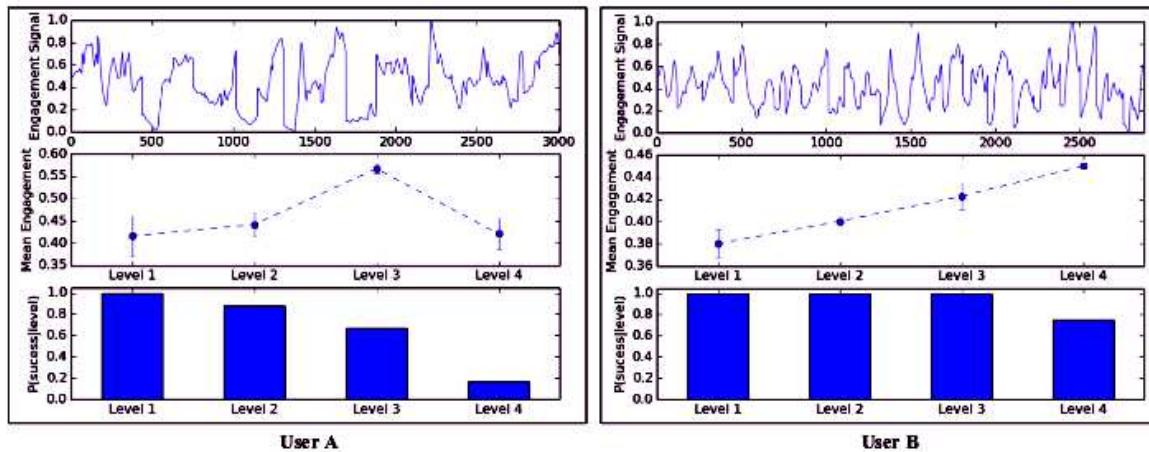


Figure 4. Task performance and engagement for different users in the sequence learning task.

The first row shows the normalized engagement signal during the session (in samples). The second row shows the mean engagement value for each difficulty level, and the third one visualizes task performance as probabilities of success at each level. We observe that User B can perform better in the task, since there is a high probability of success in the hardest level (Level 4), while for User A, this probability is small, with Level 3 being the most difficult level in which this user can probably succeed. We observe that both users show their maximum engagement values during these levels (Level 3 for User A and Level 4 for User B).

In order to gain further insight into the distribution of the participants, considering performance and engagement, we project the User Model arrays UM into a 2D visualization using Multi-Dimensional Scaling (MDS), with each point corresponding to a single user. We then apply K-means clustering to the resulting projection, grouping the users into three clusters. The selection of K ($K = 3$) is such that each cluster has an efficient number of samples (≈ 15). Each cluster can be seen as a group of users that share similar user skills and behaviors. Based on this, we visualize the cluster means; the average probability of success and mean engagement value per level (Figure 5).

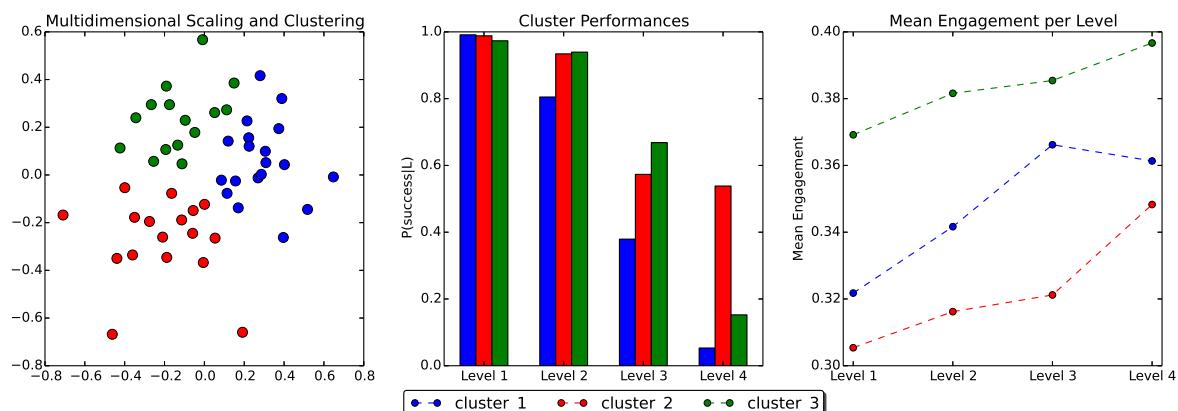


Figure 5. User clustering using Multi-Dimensional Scaling and K-means (left), cluster means as success probabilities at each level (middle) and mean engagement per level (right).

Based on the visualization of user clustering, we could note that users in Cluster 2 show a high probability of success in Level 4, relative to users in Clusters 1 and 3. Moreover, this cluster shows an upward trend in task engagement as difficulty increases. On the other hand, users in Cluster 1 seem to have no probability of succeeding in levels higher than Level 2. However, their maximum task engagement values appear for Level 3 and Level 4. As an example, the resulting clustering labeled User A as a member of Cluster 3 and User B as a member of Cluster 2 (see Figure 4). We follow this baseline clustering approach for our data-driven methodology, assuming that these clusters depict three possible different types of users. In the next section, we present our RL experiments towards learning personalized policies for different user models.

5. Learning Personalized Training Policies for Simulated Users

As mentioned before, we follow a data-driven approach to get insights towards the development of our proposed SAR system. Data-driven methods are being used as methodologies for system development and evaluation of complex interactive systems, e.g., adaptive multimodal dialogue systems [16], including user simulation for offline RL experimentation, training and evaluation. In this section, we present our user simulation modeling and the RL experiments. Using the collected data and their analysis, we build simulation models considering task performance and engagement. Different simulated users allow for offline RL experimentation before the system deployment with real users, as well as establishing a database of user models and their offline personalized policies. In Figure 6, we show our approach to learn personalized policies, using simulation models in a data-driven manner.

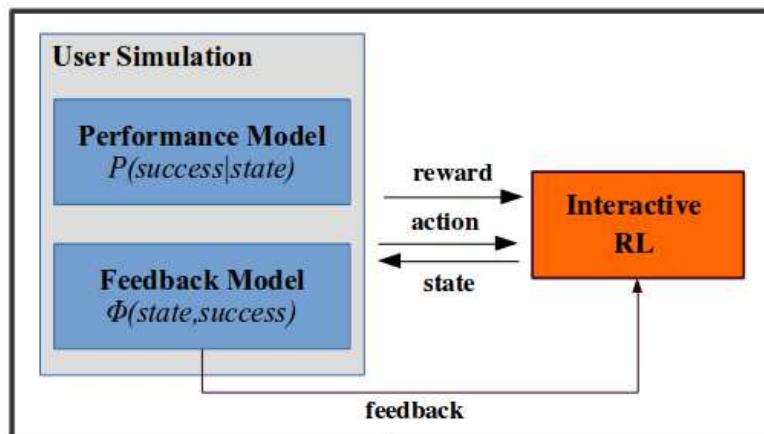


Figure 6. Reinforcement learning setup using simulated users.

5.1. User Simulation

In our data-driven approach, the goal of user simulation is to define a set of different environments (e.g., users) with different dynamics (e.g., skills). For this reason, we follow a supervised approach to learn different user simulation models, based on our clustered data. For each cluster, we define the performance model and the engagement model.

The performance model is defined as $UP_k = P(\text{success}|\text{state})$, which estimates how likely a user is to succeed in a given state. To learn these success probabilities from our collected data, we employ a regression model on the observed data [50]. More specifically, for each cluster k , we apply Maximum Likelihood Estimation (MLE), to learn the probabilities $P(\text{success}|\text{state}) = N(\text{success}, \text{state})/N(\text{state})$ of our observed data. In order to deal with unobserved states, we deployed a neural network with softmax output, as a regressor, which estimates success probability for all possible states. The input for

the performance model is the state features, current level, robot feedback and previous scores, and the output is the probability of success in this state. As an evaluation metric, we used the training error. The RMSE was 0.09, 0.12 and 0.08, for each cluster, respectively.

In a similar way, the feedback model is defined as $\Phi_k(state, success)$, which estimates task engagement for each state and outcome. In order to train the model based on the clustered data, we estimated the mean engagement value per state and outcome. Support Vector Regression (SVR) was used as a regression model to estimate the engagement value $\phi \in [0, 1]$ for each state and outcome. The RMSE was 0.08, 0.08 and 0.11, for each cluster, respectively. When compared to the RMSE scores of a neural network model, the scores were very similar, with the SVR model being slightly better. Since the purpose of the user simulation is not to develop a generalized model for task performance and engagement, but to represent our collected data as accurately as possible, the RMSE training error was selected to evaluate the simulation models.

While further experimentation is needed for the user simulation component, we argue that our simulation approach serves the purpose of learning a set of baseline user simulation models. As ongoing work, we are investigating more advanced methods for EEG analysis [51] and user simulation, such as input-output Hidden Markov Models [15] and Dynamic Bayesian Networks [52], to also encode temporal information, which is out of the scope of this current work. In the next section, we describe the reinforcement learning experiments, using these models.

5.2. Interactive Reinforcement Learning Experiments

The scope of the following reinforcement learning experiments is to investigate if integrating task engagement in the RL mechanism improves learning results, using the defined user models. The outcome of these learning experiments will be a set of User Models UM and their corresponding User-Specific Policies USP . To learn these policies, we apply Q -learning with learning rate $\alpha = 0.08$ and discount factor $\gamma = 0.95$ with the softmax exploration strategy with state-visit-based decreasing temperature parameter τ . The softmax exploration allows for probabilistic policies $\pi(s) = P(a|s)$. The parameters were selected empirically to ensure optimal policy learning. The small learning rate avoids instability in learning under noisy observations. The large discount factor enhances the maximization long-term rewards. The exploration parameter was initially high such that all actions were considered as equal at the start of training and to ensure an efficient state-action space exploration.

The motivation for these experiments was to identify an appropriate method to utilize both task performance and engagement to learn personalized RL-based policies. For each cluster, we use the corresponding simulation models to learn the USP , as we show in Figure 6. Our first approach is to use a performance-based reward, where $r(s, a) = score$ (Equation (1)). In order to integrate task engagement, for reward shaping, we used $r'(s, a) = r(s, a) + \beta_1 * \Phi(s, a)$, and for Q -augmentation, $Q'(s, a) = Q(s, a) + \beta_2 * \Phi(s, a)$. The selection of such a parameter depends on the range of values (rewards and Q -values), as well as on patterns in performance and engagement. In this work, we follow an empirical approach, comparing different β values, concluding that $\beta_1 = 7.5$ and $\beta_2 = 0.8$. Further experimentation is needed in order to learn appropriate shaping functions H and combination parameters β , considering patterns in engagement and performance, as well as scaling and bounding methods to ensure formal lower and upper bounds for feedback values [53].

We present the learning results in Figure 7, visualizing task performance and task engagement for each user model, as the algorithm learns with each method. We observe that integration of task engagement through both feedback techniques increases both task performance and engagement, as the algorithm converges to the optimal policy. Generally, both techniques outperform the performance-based approach for the selected combination parameters.

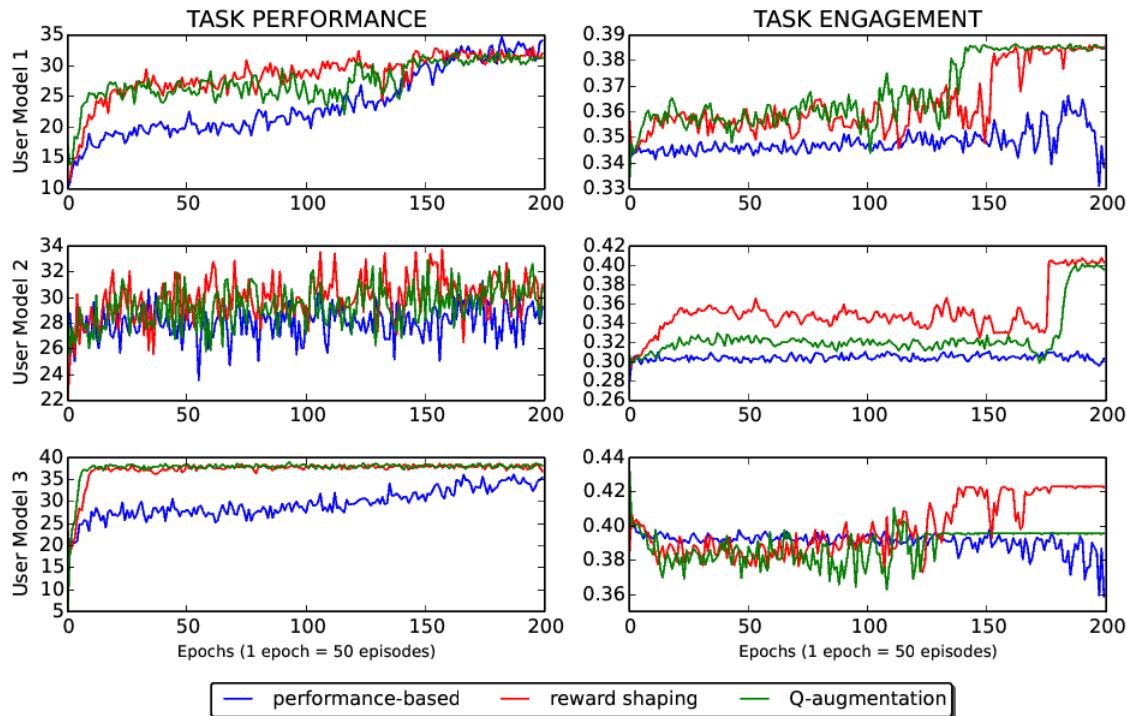


Figure 7. Interactive reinforcement learning for the different user models. We visualize task performance and task engagement for each user model. Task engagement as personalization feedback can facilitate learning.

In order to gain insight into the learned policies, we visualize $p(\text{action}|\text{state})$ for a specific state, in order to compare the decisions across users and methods, as we show in Figure 8. The selected state is $S_0 = (5, 1, -2)$, based on which the current difficulty level is Level 2 ($L = 5$), and the robot has provided encouraging feedback ($RF = 1$) after the user failed in the same level on the previous turn ($PR = -2$). Essentially, the figure visualizes the policies as the probabilities of selecting one of the actions $A = [L1, L2, L3, L4, RF1, RF2]$ in the given state. Each row corresponds to a user model, and each column corresponds to the learning method. Considering a specific method, we can compare the USPs of different users. For example, considering reward-shaping, the policy for User Model 1 chooses action $L2$ with a high probability, while for User Model 2, the policy chooses almost uniformly between $L2$ and $L3$, and for User Model 3, the policy is deterministic, choosing $RF2$. Considering a specific user model, we can observe differences in policies across methods. For example, for User Model 2, we observe that the performance-based approach results in a near-uniform policy, which does not provide enough information. On the contrary, both feedback methods result in a more informative policy, giving high probabilities only on two actions ($L2, L3$ for reward shaping and $L3, RF1$ for Q-augmentation). Both feedback methods propose higher levels to the user, which may result to increase task performance.

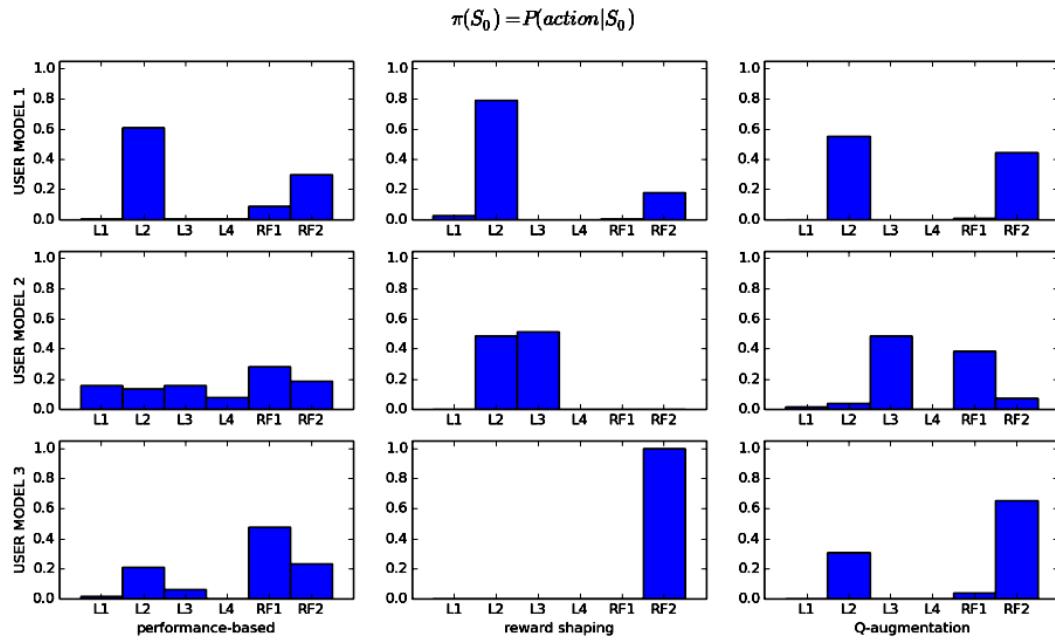


Figure 8. Visualization of the learned policies for a given state: $\pi(state) = P(action|S_0)$. The x-axis shows the possible actions, and the y-axis shows the probability for each action. Each row corresponds to a user model, and each column corresponds to the learning method.

6. Discussion and Future Work

In this work, we presented the developmental process of a data-driven SAR system for personalized robot-assisted training. The process includes data collection and analysis, the user simulation and RL experiments to integrate task engagement in the RL agent. This approach allows for an extensive analysis and experimentation, towards the development of a real-time personalized SAR system. The presented analysis and experiments indicate that users with different skills show different patterns in their task engagement for different difficulty levels. We argue that task engagement is essential information that can be utilized for real-time adaptive SAR systems. Future work includes further experimentation in user performance simulation and EEG modeling, capturing patterns in task engagement under different user skills and difficulty levels, including temporal aspects (input-output HMM) [15]. More sophisticated and accurate user modeling will enable us to learn representative combination parameters and shaping functions.

The long-term goal of this research is to develop an interactive learning and adaptation framework for real-time personalization, as we show in Figure 9. The proposed framework supports the participation of a human supervisor who can observe and control the interaction through an intelligent monitoring and control interface (GUI). These communication channels are integrated with the RL module through interactive reinforcement learning methods. Briefly, the system starts with a user skill assessment policy recording task performance and engagement under different difficulty levels. The challenge is to learn a policy that builds a representative user model of the current user within the initial steps of the interaction. By the end of the assessment mode, the system has an indicative user model *UM* for the current user. The system can use this model to classify the user into one of the existing user models, loading the corresponding *USP*, following the assumption that similar user models result in similar user-specific policies [25]. This policy is loaded as the personalized training policy. At each interaction step, the system performs an action based on this policy, which can be adjusted based on user feedback and human guidance. Prior knowledge (user models and user-specific policies), user feedback (task engagement through EEG) and human expertise (GUI input) are integrated to facilitate the adaptation process, in an interactive reinforcement learning setup. Human guidance can be provided either in the form of demonstrations prior to the interaction [22] or interventions during

the interaction [23]. Ongoing work includes further experimentation towards developing a parametric mapping from user models to policies, allowing the system to handle rare user cases and outliers. User studies will be conducted to evaluate and refine the proposed framework.

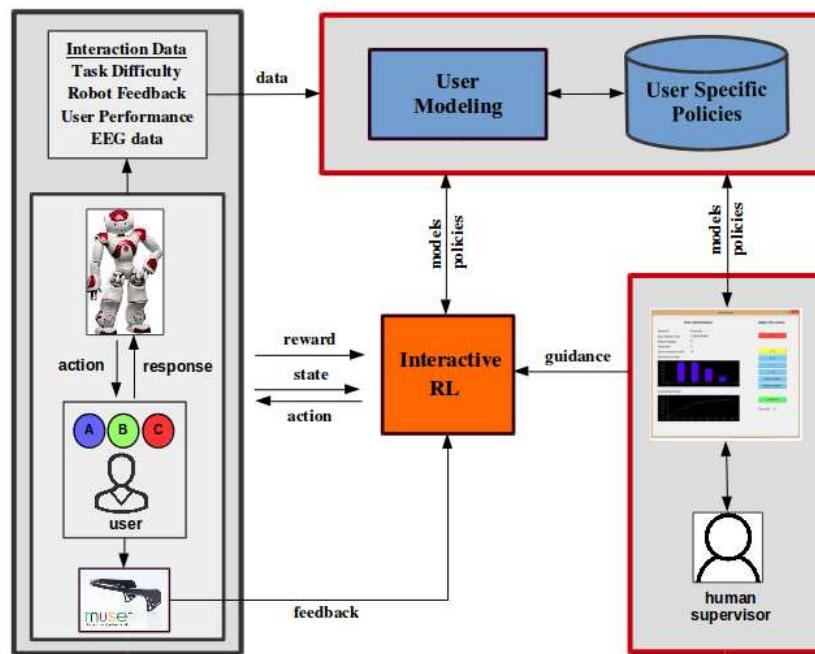


Figure 9. Future work: proposed system for real-time adaptation and personalization through user modeling and interactive RL for implicit feedback (Muse) and human guidance (GUI).

Supplementary Materials: The dataset used in this work is available online [48].

Author Contributions: K.T. conceived of and designed the research study and conducted the research experiments. M.A. administered the data collection procedure and survey data analysis. F.M. supervised the research design and contributed to the overall development of the reported results and the manuscript preparation.

Funding: This work is based on research supported by the National Science Foundation (CHS 1565328, PFI 1719031) and by the educational program of the National Center for Scientific Research "Demokritos" in collaboration with the University of Texas at Arlington.

Acknowledgments: The authors would like to thank Akilesh Rajavenkatanarayanan for his contribution to the data collection procedure and the students of the HERACLEIA lab for their support.

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

References

1. Feil-Seifer, D.; Mataric, M.J. Defining socially assistive robotics. In Proceedings of the 9th International Conference on Rehabilitation Robotics (ICORR 2005), Chicago, IL, USA, 28 June–1 July 2005; pp. 465–468.
2. Fasola, J.; Matarić, M.J. Robot motivator: Increasing user enjoyment and performance on a physical/cognitive task. In Proceedings of the 2010 IEEE 9th International Conference on Development and Learning, Ann Arbor, MI, USA, 18–21 August 2010; pp. 274–279.
3. Matarić, M.J.; Eriksson, J.; Feil-Seifer, D.J.; Weinstein, C.J. Socially assistive robotics for post-stroke rehabilitation. *J. NeuroEng. Rehabil.* **2007**, *4*, 5. [CrossRef] [PubMed]
4. Gordon, G.; Spaulding, S.; Westlund, J.K.; Lee, J.J.; Plummer, L.; Martinez, M.; Das, M.; Breazeal, C. Affective Personalization of a Social Robot Tutor for Children’s Second Language Skills. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016.
5. Mead, R.; Wade, E.; Johnson, P.; Clair, A.S.; Chen, S.; Mataric, M.J. An architecture for rehabilitation task practice in socially assistive human-robot interaction. In Proceedings of the 19th International Symposium in Robot and Human Interactive Communication, Viareggio, Italy, 13–15 September 2010; pp. 404–409.

6. Bloom, B.S. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educ. Res.* **1984**, *13*, 4–16. [[CrossRef](#)]
7. Csikszentmihalyi, M. *Beyond Boredom and Anxiety*; Jossey-Bass: San Francisco, CA, USA, 2000.
8. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*; Harper&Row: New York, NY, USA, 1990.
9. Chanel, G.; Rebetez, C.; Bétrancourt, M.; Pun, T. Boredom, engagement and anxiety as indicators for adaptation to difficulty in games. In Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era, Tampere, Finland, 7–10 October 2008; ACM: New York, NY, USA, 2008; pp. 13–17.
10. Yerkes, R.M.; Dodson, J.D. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **1908**, *18*, 459–482. [[CrossRef](#)]
11. Clabaugh, C.; Matarić, M.J. Exploring elicitation frequency of learning-sensitive information by a robotic tutor for interactive personalization. In Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), New York, NY, USA, 26–31 August 2016; pp. 968–973.
12. Clabaugh, C.E. Interactive Personalization for Socially Assistive Robots. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; ACM: New York, NY, USA, 2017; pp. 339–340.
13. Corrigan, L.J.; Basedow, C.; Küster, D.; Kappas, A.; Peters, C.; Castellano, G. Mixing implicit and explicit probes: Finding a ground truth for engagement in social human-robot interactions. In Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, Bielefeld, Germany, 3–6 March 2014; ACM: New York, NY, USA, 2014; pp. 140–141.
14. Canny, J. *Interactive Machine Learning*; University of California: Berkeley, CA, USA, 2014.
15. Cuayahuitl, H.; Renals, S.; Lemon, O.; Shimodaira, H. Human-computer dialogue simulation using hidden markov models. In Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding, San Juan, Puerto Rico, 27 November–1 December 2005; pp. 290–295.
16. Rieser, V.; Lemon, O. *Reinforcement Learning for Adaptive Dialogue Systems: A Data-Driven Methodology for Dialogue Management and Natural Language Generation*; Springer Science & Business Media: Berlin, Germany, 2011.
17. Tapus, A.; Tăpuş, C.; Matarić, M.J. User—robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy. *Intell. Serv. Robot.* **2008**, *1*, 169–183. [[CrossRef](#)]
18. Gao, A.Y.; Barendregt, W.; Castellano, G. Personalised Human-Robot Co-Adaptation in Instructional Settings using Reinforcement Learning. In Proceedings of the Persuasive Embodied Agents for Behavior Change (PEACH2017) Workshop at the International Conference on Intelligent Virtual Agents (IVA2017), Stockholm, Sweden, 27 August 2017.
19. Ritschel, H.; André, E. Real-Time Robot Personality Adaptation based on Reinforcement Learning and Social Signals. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; ACM: New York, NY, USA, 2017; pp. 265–266.
20. Chan, J.; Nejat, G. Social intelligence for a robot engaging people in cognitive training activities. *Int. J. Adv. Robot. Syst.* **2012**, *9*, 113. [[CrossRef](#)]
21. Hemminghaus, J.; Kopp, S. Towards Adaptive Social Behavior Generation for Assistive Robots Using Reinforcement Learning. In Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; ACM: New York, NY, USA, 2017; pp. 332–340.
22. Moro, C. *Learning Socially Assistive Robot Behaviors for Personalized Human-Robot Interaction*; University of Toronto: Toronto, ON, Canada, 2018.
23. Senft, E.; Baxter, P.; Kennedy, J.; Belpaeme, T. SPARC: Supervised Progressively Autonomous Robot Competencies. In Proceedings of the International Conference on Social Robotics, Paris, France, 26–30 October 2015; Springer: Berlin, Germany, 2015; pp. 603–612.
24. Sugiyama, H.; Meguro, T.; Minami, Y. Preference-learning based inverse reinforcement learning for dialog control. In Proceedings of the Thirteenth Annual Conference of the International Speech Communication Association, Portland, OR, USA, 9–13 September 2012.
25. Tsiakas, K.; Dagioglou, M.; Karkaletsis, V.; Makedon, F. Adaptive robot assisted therapy using interactive reinforcement learning. In Proceedings of the International Conference on Social Robotics, Kansas City, MO, USA, 1–3 November 2016; Springer: Berlin, Germany, 2016; pp. 11–21.
26. Picard, R.W.; Picard, R. *Affective Computing*; MIT Press: Cambridge, UK, 1997; Volume 252.

27. Sabourin, J.L.; Lester, J.C. Affect and engagement in Game-BasedLearning environments. *IEEE Trans. Affect. Comput.* **2014**, *5*, 45–56. [[CrossRef](#)]

28. Bosch, N.; D’Mello, S.; Baker, R.; Ocumpaugh, J.; Shute, V.; Ventura, M.; Wang, L.; Zhao, W. Automatic detection of learning-centered affective states in the wild. In Proceedings of the 20th International Conference on Intelligent User Interfaces, Atlanta, GA, USA, 29 March–1 April 2015; ACM: New York, NY, USA, 2015; pp. 379–388.

29. Fan, J.; Bian, D.; Zheng, Z.; Beuscher, L.; Newhouse, P.A.; Mion, L.C.; Sarkar, N. A Robotic Coach Architecture for Elder Care (ROCare) based on multi-user engagement models. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2017**, *25*, 1153–1163. [[CrossRef](#)] [[PubMed](#)]

30. Peters, C.; Castellano, G.; de Freitas, S. An exploration of user engagement in HCI. In Proceedings of the International Workshop on Affective-Aware Virtual Agents and Social Robots, Boston, MA, USA, 6 November 2009; ACM: New York, NY, USA, 2009; p. 9.

31. George, L.; Léchner, A. An overview of research on “passive” brain-computer interfaces for implicit human-computer interaction. In Proceedings of the International Conference on Applied Bionics and Biomechanics ICABB 2010-Workshop W1 “Brain-Computer Interfacing and Virtual Reality”, Venise, Italy, 14–16 October 2010.

32. Szafir, D.; Mutlu, B. Pay attention!: designing adaptive agents that monitor and improve user engagement. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; ACM: New York, NY, USA, 2012; pp. 11–20.

33. MUSE. Available Online: <http://www.choosemuse.com/> (accessed on 11 May 2018).

34. Li, Z.; Xu, J.; Zhu, T. Prediction of Brain States of Concentration and Relaxation in Real Time with Portable Electroencephalographs. *arXiv* **2015**, arXiv:1509.07642. [[CrossRef](#)]

35. Abujelala, M.; Abellanoza, C.; Sharma, A.; Makedon, F. Brain-ee: Brain enjoyment evaluation using commercial eeg headband. In Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, 29 June–1 July 2016; ACM: New York, NY, USA, 2016; p. 33.

36. Liu, X.; Tan, P.N.; Liu, L.; Simske, S.J. Automated Classification of EEG Signals for Predicting Students’ Cognitive State during Learning. In Proceedings of the International Conference on Web Intelligence, Leipzig, Germany, 23–26 August 2017; ACM: New York, NY, USA, 2017; pp. 442–450.

37. Karydis, T.; Aguiar, F.; Foster, S.L.; Mershin, A. Performance characterization of self-calibrating protocols for wearable EEG applications. In Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, 1–3 July 2015; ACM: New York, NY, USA, 2015; p. 38.

38. Pope, A.T.; Bogart, E.H.; Bartolome, D.S. Biocybernetic system evaluates indices of operator engagement in automated task. *Biol. Psychol.* **1995**, *40*, 187–195. [[CrossRef](#)]

39. McMahan, T.; Parberry, I.; Parsons, T.D. Evaluating player task engagement and arousal using electroencephalography. *Proc. Manuf.* **2015**, *3*, 2303–2310. [[CrossRef](#)]

40. Knox, W.B.; Stone, P. Reinforcement learning from simultaneous human and MDP reward. In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems, Valencia, Spain, 4–8 June 2012; Volume 1, pp. 475–482.

41. Broekens, J. Emotion and reinforcement: Affective facial expressions facilitate robot learning. In *Artificial Intelligence for Human Computing*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 113–132.

42. Seo, Y.W.; Zhang, B.T. A reinforcement learning agent for personalized information filtering. In Proceedings of the 5th International Conference on Intelligent User Interfaces, New Orleans, LA, USA, 9–12 January 2000; ACM: New York, NY, USA, 2000; pp. 248–251.

43. Gathercole, S.E.; Baddeley, A.D. *Working Memory and Language*; Psychology Press: Hoboken, NJ, USA, 2014.

44. Cognition Measures. Available Online: <http://www.healthmeasures.net/explore-measurement-systems/nih-toolbox/intro-to-nih-toolbox/cognition> (accessed on 11 May 2018).

45. Who is NAO? Available Online: <https://www.ald.softbankrobotics.com/en/robots/nao> (accessed on 11 May 2018).

46. Park, E.; Kim, K.J.; Del Pobil, A.P. The effects of a robot instructor's positive vs. negative feedbacks on attraction and acceptance towards the robot in classroom. In Proceedings of the International Conference on Social Robotics, Amsterdam, The Netherlands, 24–25 November 2011; Springer: Berlin, Germany, 2011, pp. 135–141.
47. Tsiakas, K.; Abellanoza, C.; Abujelala, M.; Papakostas, M.; Makada, T.; Makedon, F. Towards Designing a Socially Assistive Robot for Adaptive and Personalized Cognitive Training. In Proceedings of the Robots 4 Learning Workshop R4L@HRI, Vienna, Austria, 6–9 March 2017.
48. Sequence-learning-dataset. Available Online: <https://github.com/TsiakasK/sequence-learning-dataset> (accessed on 11 May 2018).
49. Papakostas, M.; Tsiakas, K.; Giannakopoulos, T.; Makedon, F. Towards predicting task performance from EEG signals. In Proceedings of the International Conference on Big Data, Boston, MA, USA, 11–14 December 2017.
50. Clabaugh, C.; Tsiakas, K.; Mataric, M. Predicting Preschool Mathematics Performance of Children with a Socially Assistive Robot Tutor. In Proceedings of the Synergies between Learning and Interaction Workshop @ IROS, Vancouver, BC, Canada, 24–28 September 2017.
51. So, W.K.; Wong, S.W.; Mak, J.N.; Chan, R.H. An evaluation of mental workload with frontal EEG. *PLoS ONE* **2017**, *12*, e0174949. [CrossRef] [PubMed]
52. Käser, T.; Klingler, S.; Schwing, A.G.; Gross, M. Dynamic Bayesian Networks for Student Modeling. *IEEE Trans. Learn. Technol.* **2017**, *10*, 450–462. [CrossRef]
53. Papudesi, V.N. Integrating Advice with Reinforcement Learning. Ph.D. Thesis, University of Texas at Arlington, Austin, TX, USA, 2002.



© 2018 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 (<http://creativecommons.org/licenses/by/4.0/>).