Personalized Human-Robot Interaction in Companion Social Robots

Bahram Salamat Ravandi[1,2]

[1]Digitalisation, Interaction, Cognition, and Emotion (DICE) Lab, Sweden
[2]Department of Applied IT, University of Gothenburg, Sweden

Abstract

In the near future, it is likely that we will see companion social robots assisting humans in various settings. To be effective assistants, these robots need to be able to adapt their interactions to be more likable and engaging for humans. To explore this area of study, an interaction loop has been proposed that includes a human participant, a social robot, and a gamified task. To maintain human engagement, adjusting interactions through personalized social feedback, and task modulation are defined. So far several evaluative experiments and piloting are conducted. In the first study, the presence of a social robot (Furhat) was evaluated in comparison to two other conditions: the robot's simulator and a control group without any robot involvement. The second study focused on examining how different types of feedback (performance-based feedback versus affective feedback) affected users' perceptions and engagement. Furthermore, in order to enable tailored feedback and task adjustment, an affective-based engagement detection model was developed using Deep Learning methods. Preliminary findings from the first study indicate that participants favored social robots over other conditions, as evidenced by significantly lower arousal levels reported on the SAM scale in the robot condition. The analysis of the second study for evaluating performance-based feedback vs affective feedback is still ongoing. For forthcoming research, we aim to incorporate XAI technology to facilitate the explainability of AI-related modules. This approach is beneficial for ensuring transparency in forthcoming applications, as it helps maintain credibility and trustworthiness.

Keywords

Assistive Social Robots, Companion Social Robots, Engagement, Human-Robot Interaction, Affective Feedback, User Experience

1. Introduction

Over the past decade, there has been an increasing focus on developing autonomous social robots with the ability to interact with humans. Socially Assistive Robots (SARs) have the potential to be valuable tools in both education and healthcare domains, particularly for those suffering from Alzheimer, Dementia, or Autism [1, 2, 3]. Using robots as a factor in human engagement in games or tasks, including gamified tasks, has two main benefits: increasing motivation through social presence effects, where people tend to increase effort or have increased motivation in the

Late-breaking work, Demos and Doctoral Consortium, colocated with The 1st World Conference on eXplainable Artificial Intelligence: July 26–28, 2023, Lisbon, Portugal
*Bahram Salamat Ravandi
bahramsalamat@ait.gu.se (B. S. Ravandi)
https://www.dice-r-lab.com/bahram (B. S. Ravandi)
0009-0008-2712-8684 (B. S. Ravandi)

presence of others under certain conditions, and affective interaction. While socially interactive robots may initially be engaging to humans as a result of the novelty that the human participant experiences, it is possible for individuals to lose engagement during interactions. Nevertheless, there are various strategies that can be employed to enhance user engagement, including the incorporation of gamification elements like challenges, textual and verbal feedback, badges and prizes, avatars, narratives, emotional agents, and interactive agents [4].

In order to explore users' engagement in HRI, an interaction loop has been developed that has three main components of a social robot, a human user, and a fast-paced game or task. Different levels of interaction between these components are planned and require to be implemented gradually. Alongside evaluating the interaction loop, the focus of the current study is on developing an engagement function within the interaction loop for monitoring and evaluating emotional, behavioral, and task-based engagement within the interaction loop and subsequently adapting the interaction based on the engagement model. The approach is required to generalize across related classes of tasks that is maintaining a focused and appropriate level of engagement in the activity, i.e. cognitive therapy tasks. Some primary research questions to be investigated in this study are:

1. What are the suitable modalities for modeling different dimensions of engagement including social engagement, affective engagement, and performance engagement?
2. Which types of feedback, performance-based or affective-based, are preferred from the users' perspective?
3. How should interactions be adapted to deliver personalized feedback and adjust task challenge levels effectively?

2. Related works

This paper adopts a specific definition of engagement in HRI, drawing from the works of O'Brien and Toms [5]. According to them, engagement encompasses various component features and is regarded as the quality of the user experience. Recently, many studies have been carried out to examine the way in which humans and robots interact while working together, as well as when robots provide aid to humans. Andriella and colleagues [1] proposed a platform for an assistive robot to help Alzheimer's patients with memory training exercises through the use of verbal and non-verbal communication. [6] employed the Tega robot as an educational aid for children learning a new language, utilizing reinforcement learning (RL) techniques to provide personalized affective feedback. Rudovic and colleagues [7] developed a multi-modal active learning approach to detect engagement of real-world child-robot interactions. [8] trained CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory) models to detect three classes of engagement/disengagement, mid-engagement, and high engagement on a dataset they collected based on a TEGA robot interacting with Children. [9] established a CultureNet based on CNN to personalize engagement detection for target subjects. Mollahosseini and colleagues [10] used Deep Neural Networks for facial expression recognition. The trained model was used to give empathetic responses by the Ryan companion bot based on the affective state of the user.

While the existing literature has investigated the interaction loop between humans and robots during tasks, there has been limited exploration of adaptive interactions. Moreover, a standardized approach for detecting engagement in rapid immersive tasks is lacking. In addition, there are variations in the feedback mechanisms used by different types of robots, and a significant number of studies demonstrate limited adaptability for robots to give affective feedback.

3. Proposed setup

Figure 1 illustrates the proposed interaction loop, showcasing various planned modules. Users can interact with the game and receive visual or audio feedback from both the robot and the game. There are six components to this interaction loop, which are briefly described as follows:

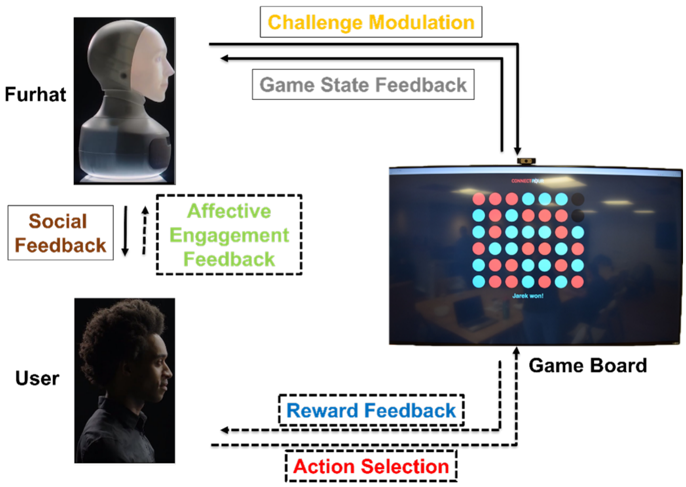

Figure 1: Interaction Loop: User, robot, and the game board are three main components of the proposed setup.

1. **Challenge Modulation:** Adjusting the challenge level of the game or task based on user's expressed state of engagement or disengagement.
2. **Game State Feedback:** This component involves providing information on the human performance considering the current state of the game.
3. **Reward Feedback:** This component provides users with direct feedback on their performance and the outcome of a specific action.
4. **Action Selection:** This component involves the use of touchscreen-based or verbalized inputs for selecting actions.
5. **Affective Engagement Feedback:** Possible inputs that may determine the user's affective engagement.

3.1. Study I

In the initial study, the robot was placed behind the laptop (Figure 2.a) to minimize disturbance while users played the game. The experimental design of the study centered on two independent variables: set up (physical robot, simulated robot, or control group), and challenge level (easy, medium, and hard). The sample for this study consisted of 78 adults, aged 19 to 46 (M = 25.59, SD = 4.86). Prior to initiating the game, the participants received instructions from the experimenter. During the interaction, audio-visual feedback from both robot and the game was provided. Participants were instructed to fill out a SAM scale questionnaire after finishing each level of the game, in addition to recording the eye tracker, audio, and video data. The SAM questionnaire aimed to gather information about users' emotional state, encompassing valence, arousal, and dominance dimensions, concerning both the game itself and the experimental setup.

The first experiment provided valuable information regarding how users perceive and perform on various challenge levels when interacting with both physical and simulator robots. In general, participants had a favorable impression of their interactions with the robot. Task performance was not significantly affected by the presence of either the physical or simulator robot compared to the control group (which had no robot). The results showed that the physical robot group achieved a task performance rate of 71.22%, the simulator group achieved 67.88%, and the control group achieved 66.46%. Statistical analysis ($F(2, 74) = .856$, $p = .429$, $\eta^2 = .023$) indicated that there were no significant differences among the groups. In terms of the SAM scales comparison, participants in the robot group reported lower arousal ratings (2.72) compared to both the control group (3.17, $p = .044$) and the simulated robot group (3.20, $p = .032$), with the statistical analysis, $F(2, 71) = 3.01$, $p = .056$, $\eta^2 = .078$. However, it remains unclear whether these positive outcomes will endure with repeated use or if they were primarily influenced by the novelty of the experience.

3.2. Study II

The focus of the second study was on the examination of the effects of different forms of feedback (affective-based vs performance-based) on users' performance and perception of the robot during a two-way interaction on the game. Affective feedback includes feedback that expresses emotions, acknowledges enjoyment, and asks about the individual's feelings. While performance feedback provides information about the number of correct responses, improvement, and overall performance. As shown in figure 2.b, the robot was placed beside the human participant, allowing for a tilted view angle of both the game and the participant to increase the interaction between the human and the robot. During the game (within the block at the end of each trial) the robot provided feedback similar to the first study (figure 2.a), with the added feature of randomly moving its head toward the participant. At the end of each block, there was a brief interval during which the robot provided personalized feedback based on the participant's performance or affective state, or both. In order to facilitate the two-way communication, questions were asked by the robot and participants were given the opportunity to respond. The study was designed using two independent variables: challenge level (easy and medium) and feedback type. A sample of 58 subjects aged 18 to 24 (M = 20, SD = 1.87) was recruited for the purpose of participating in the second experiment. Participants were instructed

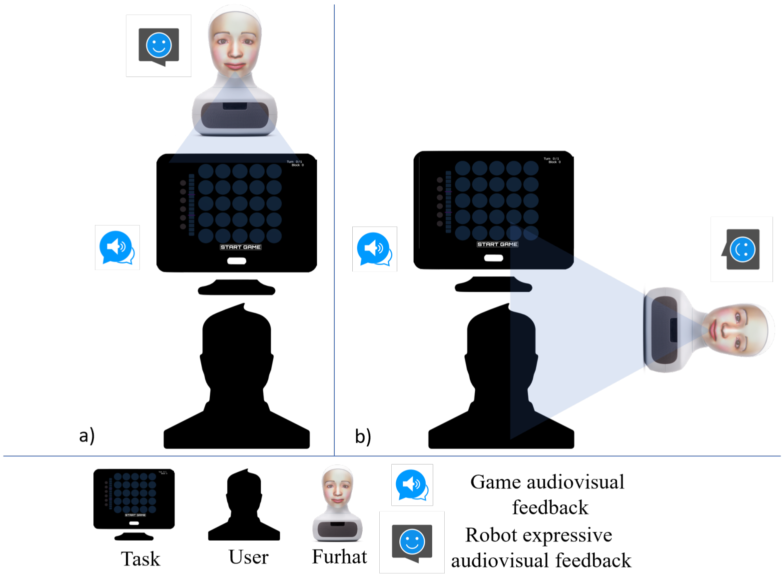

Figure 2: The configuration scheme of two experimental setups, the first study (a) and the second study (b). In both experimental setups, the robot provided audiovisual feedback related to the task at hand, along with expressive audiovisual feedback.

to fill out a SAM scale questionnaire after finishing each block of the game, in addition to recording the eye tracker, audio, and video data. The participants were also interviewed at the end of the game. Furthermore, an affective engagement model was developed to estimate the positive and negative engagement of participants in order to give personalized affective feedback. A CNN architecture was employed to train the affective-based engagement model. The FER2013 dataset [11] was utilized to train the model. The model predicts the probability of a video frame being classified within the seven emotion categories provided in the dataset (Anger, Disgust, Fear, Happy, Sad, Neutral, and Surprise). The probability of the frame being classified as "happy" is employed as a measure of the user's positive engagement. At present, the data that has been gathered is undergoing analysis.

4. Conclusion and Discussion

This paper introduces an established framework aimed at investigating the interaction loop between humans and a social robot on a gamified task. The main objective is to develop a model that can accurately capture user engagement, enabling adaptive and personalized interactions by adjusting tasks and offering feedback based on users' internal state. A sequence of pilot, evaluative, and experimental studies were conducted. The first study indicated that participants had a positive view of the social robot (Furhat) and that the robot (both physical and simulator) did not negatively impact users' performance. The second study's data analysis process is currently ongoing. Moreover, an engagement detection function was developed using a facial expressions-based dataset (FER2013) to provide personalized affective feedback.

However, it's important to acknowledge its limitations because facial expressions may not always accurately depict one's level of engagement, as not everyone displays their engagement through expressions. To overcome this, alternative methods such as physiological measures (e.g. GSR, infrared cameras, and EEG) could be utilized to capture additional factors such as valence and arousal.

In future studies, in order to achieve effective user engagement, it is imperative to modify the difficulty level of the task based on the user's engagement and performance. To minimize any disruptions and maintain the stability of the task or game state, a separate framework like a role-based state machine may be utilized to modify game difficulty at specific intervals. Additionally, the research aims to expand the developed positive affective engagement model by incorporating more modalities. Moreover, considering the fact that repeated phrasal feedback could be unrealistic and unnatural, using a language model to generate feedback would be beneficial. However, ethical considerations regarding the model's outputs need to be taken into account.

Furthermore, the integration of self-explaining social robots holds promise in enhancing the quality of interactions between humans and robots, leading to more effective and satisfying experiences for users [12]. Social robots can achieve this by actively seeking clarifications from users or engaging in dialogues to better comprehend their preferences and concerns. While this approach enhances the trustworthiness of the model, a crucial aspect of establishing trust in autonomous systems, such as robots, is providing users with a clear understanding of the decision-making processes [13]. This can be challenging when trying to detect emotions and engagement, as well as in generating phrasal feedback, particularly if using Large Language Models (LLM) instead of pre-programmed feedback. Techniques such as rule extraction for explaining task modification in case of using a learned policy (for e.g., using Multi-Armed Bandit RL) [14] and differentiating between self-learned features for detecting different facial expressions using methods such as LIME [15] can help to achieve more explainability. Therefore, a potential future research direction is using XAI technology for AI-related modules in the setup such as task modulation, and engagement detection to present the framework in a way that users can understand.

5. Acknowledgments

I would like to express my sincere gratitude to Ana B. Vivas and Sofia Sjöberg for their outstanding dedication in the first study. Additionally, I am profoundly thankful to Imran Khan for his essential contributions to the second study. Moreover, I extend my deepest appreciation to Robert Lowe, Alva Markelius, and Martin Bergström for their invaluable and active involvement in both main studies.

References

[1] A. Andriella, G. Alenyà, J. Hernández-Farigola, C. Torras, Deciding the different robot roles for patient cognitive training, International Journal of Human-Computer Studies 117 (2018) 20–29. URL: https://www.sciencedirect.com/science/article/pii/S1071581918300958. doi:https://doi.org/10.1016/j.ijhcs.2018.03.004, cognitive Assistants.

[2] L. Cuffaro, F. Lorenzo, S. Bonavita, G. Tedeschi, L. Leocani, L. Lavorgna, Dementia care and covid-19 pandemic: a necessary digital revolution, Neurological Sciences 41 (2020) 1977–1979.

[3] J. Goris, M. Brass, C. Cambier, J. Delplanque, J. R. Wiersema, S. Braem, The relation between preference for predictability and autistic traits, Autism Research 13 (2020) 1144–1154. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/aur.2244. doi:https://doi.org/10.1002/aur.2244. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/aur.2244.

[4] T. W. Malone, What makes things fun to learn? heuristics for designing instructional computer games, SIGSMALL '80, Association for Computing Machinery, New York, NY, USA, 1980, p. 162–169. URL: https://doi.org/10.1145/800088.802839. doi:10.1145/800088.802839.

[5] H. L. O'Brien, E. G. Toms, What is user engagement? a conceptual framework for defining user engagement with technology, Journal of the American Society for Information Science and Technology 59 (2008) 938–955. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.20801. doi:https://doi.org/10.1002/asi.20801. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20801.

[6] G. Gordon, C. Breazeal, S. Engel, Can children catch curiosity from a social robot?, HRI '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 91–98. URL: https://doi.org/10.1145/2696454.2696469. doi:10.1145/2696454.2696469.

[7] O. Rudovic, M. Zhang, B. W. Schuller, R. W. Picard, Multi-modal active learning from human data: A deep reinforcement learning approach, volume abs/1906.03098, 2019. URL: http://arxiv.org/abs/1906.03098. arXiv:1906.03098.

[8] O. Rudovic, H. W. Park, J. Busche, B. Schuller, C. Breazeal, R. W. Picard, Personalized estimation of engagement from videos using active learning with deep reinforcement learning, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2019, pp. 217–226. doi:10.1109/CVPRW.2019.00031.

[9] O. Rudovic, Y. Utsumi, J. Lee, J. Hernandez, E. C. Ferrer, B. Schuller, R. W. Picard, Culturenet: A deep learning approach for engagement intensity estimation from face images of children with autism, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 339–346. doi:10.1109/IROS.2018.8594177.

[10] A. Mollahosseini, H. Abdollahi, M. H. Mahoor, Studying effects of incorporating automated affect perception with spoken dialog in social robots, in: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2018, pp. 783–789. doi:10.1109/ROMAN.2018.8525777.

[11] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie,

L. Romaszko, B. Xu, Z. Chuang, Y. Bengio, Challenges in representation learning: A report on three machine learning contests, 2013. arXiv:1307.0414.
[12] A. Di Nuovo, D. Conti, G. Trubia, S. Buono, S. Di Nuovo, Deep learning systems for estimating visual attention in robot-assisted therapy of children with autism and intellectual disability, Robotics 7 (2018). URL: https://www.mdpi.com/2218-6581/7/2/25. doi:10.3390/robotics7020025.
[13] N. Wang, D. V. Pynadath, S. G. Hill, The impact of pomdp-generated explanations on trust and performance in human-robot teams (2016) 997–1005.
[14] R. C. Engelhardt, M. Lange, L. Wiskott, W. Konen, Sample-based rule extraction for explainable reinforcement learning (2023) 330–345.
[15] G. del Castillo Torres, M. F. Roig-Maimó, M. Mascaró-Oliver, E. Amengual-Alcover, R. Mas-Sansó, Understanding how cnns recognize facial expressions: A case study with lime and cem, Sensors 23 (2023). URL: https://www.mdpi.com/1424-8220/23/1/131. doi:10.3390/s23010131.