Assessing Emotion Mitigation through Robot Facial Expressions for Human-Robot Interaction

Luigi D'Arco[1,*], Alessandra Rossi[1] and Silvia Rossi[1]

[1]*Department of Electrical Engineering and Information Technologies, University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy*

Abstract

Affective responses are one of the primary and clearer signals used by agents for communicating their internal state. These internal states can represent a positive or negative acceptance of a robotic agent's behavior during a human-robot interaction (HRI). In these scenarios, it is fundamental for robots to be able to interpret people's emotional responses and to adjust their behaviors accordingly, to appease them, and to provoke an emotional change in them. This research investigates the impact of robot facial expressions on human emotional experiences within HRI, focusing specifically on whether a robot's expressions can amplify or mitigate users' emotional responses when viewing emotion-eliciting videos. To evaluate participants' emotional states, an AI-based multimodal emotion recognition approach was employed, combining analysis of facial expressions and physiological signals, complemented by a self-assessment questionnaire. Findings indicate that participants responded more positively when the robot's facial expressions aligned with the emotional tone of the videos, suggesting that emotion-coherent displays could enhance user experience and strengthen engagement. These results underscore the potential for expressive social robots to influence human emotions effectively, offering promising applications in therapy, education, and entertainment. By incorporating emotional facial expressions, socially assistive robots could foster behavior change and emotional engagement in HRI, broadening their role in supporting human emotional well-being.

Keywords

Emotion elicitation, Socially Assistive Robotics, Human-Robot Interaction, Emotion Recognition

1. Introduction

Socially Assistive Robotics (SAR) is an emerging field of robotics that focuses on developing robots that can assist users with hands-off interaction strategies, providing emotional and cognitive assistance [1, 2]. To improve the Human-Robot Interaction (HRI) experience, SARs must be capable of interpreting, mimicking, and responding to emotional cues, with facial expressions being a primary mode of emotional communication. This ability is essential when robots are used in contexts where emotional engagement can facilitate positive outcomes, such as therapy, learning, and behavior change. In human-human communication, facial expressions are critical for conveying emotions, improving understanding, and guiding social interactions. Several studies showed that facial expressions not only reflect how a person is feeling but also influence how others feel [3]. This phenomenon, known as emotional contagion [4], suggests that emotions can spread from one person to another through non-verbal cues, influencing the emotional state of the observer. If SARs are to be effective in emotionally engaging users, they must be able to use facial expressions in ways that influence the user's emotional experience, particularly in situations where emotional states can influence behavior and decision-making. Staffa et. al [5] investigated whether positive or negative robot personalities can affect the mental state of users during HRI by assessing the Electroencephalogram (EEG) signals of participants. They involved an anthropomorphic robot with two engagement personalities, one more prone to engage the user and the other not, by modeling voice, dialogues, and head and body movements. The results showed that participants felt the robot's personality, affecting their emotional state and engagement. Similarly,

Fiorini et al. [6] explored the impact of a robot's behavior on the emotional state of users during exposure to emotion-eliciting images. The robot displayed emotions that were coherent or incoherent with those experienced by the user to assess the level of influence it could have on the user. The results showed high accuracy, up to 98% in the robot recognizing 3 emotional states, including positive, negative, and neutral, reporting that such states were better identified when the robot was not neutral but performed coherent or incoherent behaviors. Rossi et al. [7] conducted a study to evaluate the impact of non-verbal behaviors of an anthropomorphic robot on the emotional responses in users. By using coherent, incoherent, and neutral behaviors the robot's non-verbal cues were modeled using emotional gestures. Findings revealed that emotional reactions with high arousal could be challenging by using only emotional gestures, and additional interaction strategies are needed.

In light of the different achievements in the literature, the impact of robot facial expressions on human emotions during HRI has yet to be fully investigated. Hence, the present study focuses on assessing whether the facial expressions of a robot can affect the mood of users while watching emotion-eliciting videos. By displaying facial expressions that either match or contrast the user's emotional state, the robot could promote the general effect of mirroring or emotional contagion, whereby an observer tends to covertly and unconsciously mimic the behavior of the person being observed [8]. The study design is based on the approach outlined by Rossi et al. [7], with the modification of including only two conditions: the robot's facial expressions either align with the emotional content of the videos or display opposing emotions. To evaluate the emotion felt by the user, an Artificial Intelligence (AI) approach has been developed that predicts the users' emotional state based on a fusion of facial expressions and physiological signals. Furthermore, participants of the study were provided with a questionnaire at the beginning to ascertain their empathetic capacity and one questionnaire at the end to evaluate their perception of the robot's emotional display. By demonstrating the potential of robots to influence human emotions through facial expressions, the study can contribute to the development of SARs that are more emotionally intelligent and capable of supporting users in emotionally meaningful ways, unleashing their application in scenarios where emotional engagement can facilitate positive outcomes, such as therapy, learning, and behavior change.

2. Materials and Methods

This study evaluates a robot's ability to mitigate emotions in participants watching emotion-eliciting videos. A multimodal emotion recognition system assessed participants' emotional states through facial expression and physiological signal analysis. Pre-experiment questionnaires evaluated participants' empathy levels, while post-experiment questionnaires assessed their perceptions of the robot's emotional displays. The study was conducted in a controlled environment to ensure reliable results.

2.1. Robotic Agent and Sensing Elements

The robotic agent involved in this study is a Furhat robot [9], which is a human-like, rear-projected robotic head that uses computer animations and neck movements to provide facial expressions [10]. The robot is equipped with a camera and a microphone to capture information from the surrounding environment. However, due to the need to show emotion-eliciting videos from the laptop, the laptop camera is used to have a straight view of the user's face, better for identifying the emotion felt. Although facial expression may be the most significant nonverbal form of emotional expression [3], some people can mask their facial emotions by adopting a neutral expression and using non-intuitive human body language that can lead to misinterpretations [11]. Therefore, a multimodal information-based solution for emotion recognition has been pursued to produce a more reliable emotion recognition system. Alongside the facial expressions, physiological signals have been considered, including the Electrocardiogram (ECG) and Galvanic Skin Response (GSR) signals, which can be considered more reliable indicators of emotions, as they are more difficult to cover or alter through human disguise [12]. These signals are acquired by the BITalino biosignal platform provided by PLUX Biosignals [13]. The settings of the equipment involved in the study are shown in Figure 1.
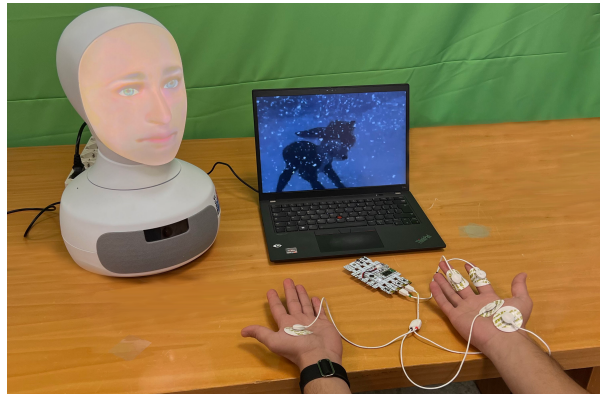
Figure 1: Experimental settings. Example of a user wearing the BITalino biosignal platform while watching emotion-eliciting videos and interacting with the Furhat robot.

2.2. Emotion Recognition Model

The emotion recognition model employed in this study builds upon a baseline architecture previously established in [7]. The model was selected due to its proven effectiveness in recognizing emotions from facial expressions and physiological signals. The model was trained using the AMIGOS dataset [14], which contains multimodal data, including EEG, ECG, GSR, and facial video recordings. The dataset provides affective level annotations for the participants based on the Self-Assessment Manikin (SAM) scale [15] and evaluations made by the dataset's authors. The annotations include valence, arousal, dominance, and basic emotions (Neutral, Disgust, Happiness, Surprise, Anger, Fear, and Sadness) for each participant for every video. The model is based on a multimodal approach that combines facial expressions and physiological signals to predict the user's emotional state. Each modality has been processed individually by an artificial intelligence model, with the facial images processed by a Convolutional Neural Network (CNN)-based architecture, and the physiological signals processed by a Support Vector Machine (SVM) model. The predictions of the extracted emotions from each modality are then fused to provide a final prediction. The model was trained on the AMIGOS dataset using a train-test split approach of 70-30.

2.3. Emotion Elicitation Videos

The videos for emotion elicitation have been selected from the DECAF database [16], which is a multimodal dataset for decoding user physiological responses to affective multimedia content. Videos with a total length not exceeding 120 seconds have been chosen to avoid fatigue and to maintain the participants' attention, but also to ensure that only one emotion is elicited at a time. Three videos were selected for each of the four emotional categories Low Arousal - Negative Valence (LALV), Low Arousal - Positive Valence (LAHV), High Arousal - Negative Valence (HALV), and High Arousal - Positive Valence (HAHV) based on the annotations provided in the DECAF dataset, resulting in a total of 12 videos. For instance, the scene from Bambi where "Bambi's mother gets killed" was categorized under LALV due to its emotionally distressing content, while the scene from Wall-E where "Wall-E and Eve spend a romantic night together" was classified under LAHV to evoke positive but calm emotions. Videos have been presented to the participants in random order to avoid any bias in the results.

2.4. Questionnaires

Two questionnaires have been prepared for the study, one to be completed before the experiment and one after. The pre-experiment questionnaire aimed to collect demographic information about the participants, such as age and gender, as well as information about their previous experience with robots and their empathetic capacity. The empathetic capacity is assessed using the Empathy Quotient test [17], which is a self-report questionnaire designed to measure empathy in adults. The short version of

the test has been chosen, which consists of 40 questions, where each question is scored on a scale from 0 to 2, with higher scores indicating higher levels of empathy. The test provides a total score that ranges from 0 to 80. To distinguish participants' level of empathy, four categories have been identified: low empathy (0-20), medium-low empathy (21-40), medium-high empathy (41-60), and high empathy (61-80). On the other hand, the post-experiment questionnaire aimed to evaluate the participant's perception of the robot's emotional display during the experiment. The post-questionnaire included questions about the robot's facial expressions, the perceived emotions, and the impact of the robot's expressions on the participants' emotional state. The post-questionnaire is created above the System Usability Scale (SUS) score principles [18], and each item is scored on a scale of 1 to 5, with 1 being completely disagree and 5 being completely agree. The post-questionnaire results were scored with a range from 0 to 100, with higher scores indicating a more positive perception of the robot's emotional display.

3. Results

A total of 60 subjects, aged 18 to 34, voluntarily participated in the study. Participants included 34 males, 17 females, and 1 non-binary individual. Of these, 24 participants reported no prior experience with robots. Two participants withdrew from the study before completing the session due to personal commitments. The participants were randomly assigned to two groups: coherent ($n = 29$) and incoherent ($n = 29$). This preliminary analysis aims to assess participants' experiences and perceptions of the robot, comparing post-experiment responses between the two groups. Statistical analyses were conducted using the Student's t-test for independent samples.

Participants in the coherent group rated the robot's behavior as more natural ($\mu = 3.393, \sigma = 1.197$) than those in the incoherent group ($\mu = 2.448, \sigma = 1.213$), with a statistically significant difference ($p < 0.05$). Similarly, participants in the incoherent group reported a higher discomfort level created by the robot ($\mu = 3.000, \sigma = 1.363$) than those in the coherent group ($\mu = 1.786, \sigma = 0.994, p < 0.05$). However, both groups did not perceive a significant influence from the robot while watching the videos ($\mu = 2.579, \sigma = 1.224, p > 0.05$). Furthermore, participants in the coherent group perceived the robot as more aware of the video content ($\mu = 4.143, \sigma = 1.02$) compared to those in the incoherent group ($\mu = 2.414, \sigma = 1.21$). They also rated the robot as less incoherent relative to the video scenes presented ($\mu = 1.964$ for the coherent group, $\mu = 3.793$ for the incoherent group). The robot's expressions were more distracting for participants in the incoherent group ($\mu = 3.000, \sigma = 1.15$) than for those in the coherent group ($\mu = 2.357, \sigma = 1.07$). For other aspects, such as the social acceptability of the robot and its potential utility in communicating emotions, no significant differences were observed between the two groups ($p > 0.05$), with both groups agreeing on the robot's usefulness.

Overall, these findings suggest that coherent emotional expressions in the robot enhance perceptions of it as more natural, aware, and non-intrusive, whereas incoherent expressions increase perceptions of discomfort, distraction, and incoherence. Although participants did not report feeling influenced by the robot, future studies will explore potential unconscious emotional changes in participants using emotion recognition models.

4. Conclusion

This preliminary study explores how robot facial expressions influence human emotional experiences in HRI. In the experiment, participants watched emotion-eliciting videos while interacting with a robot that displays facial expressions either aligned or misaligned with the emotional content of the videos. The findings indicate that participants generally responded positively to interactions where the robot's expressions matched the emotional content of the videos, underscoring the potential of using facial expressions in SARs to enhance user engagement. This study lays a foundation for incorporating emotionally expressive robots in SAR applications across therapeutic, educational, and entertainment settings. Future research will delve further into the collected data to determine whether participants

experienced unconscious emotional changes, advancing the understanding of how emotionally aware robots might foster behavioral change and enhance emotional connection in HRI.

Acknowledgments

References

[1] M. J. Matarić, B. Scassellati, Socially assistive robotics, Springer handbook of robotics (2016) 1973–1994.

[2] L. D'Arco, H. Zheng, H. Wang, Sensebot: A wearable sensor enabled robotic system to support health and well-being, in: 6th Collaborative European Research Conference, 2021, pp. 30–45.

[3] C. Frith, Role of facial expressions in social interactions, Philosophical Transactions of the Royal Society B: Biological Sciences 364 (2009) 3453–3458.

[4] C.-E. Yu, Emotional contagion in human-robot interaction, E-review of Tourism Research 17 (2020).

[5] M. Staffa, S. Rossi, Enhancing affective robotics via human internal state monitoring, in: 31st IEEE Intern. Conf. on Robot and Human Interactive Communication (RO-MAN), 2022, pp. 884–890.

[6] L. Fiorini, F. G. Loizzo, G. D'Onofrio, A. Sorrentino, F. Ciccone, S. Russo, F. Giuliani, D. Sancarlo, F. Cavallo, Can i feel you? recognizing human's emotions during human-robot interaction, in: International Conference on Social Robotics, Springer, 2022, pp. 511–521.

[7] S. Rossi, A. Rossi, S. Sangiovanni, Towards the Evaluation of the Role of Embodiment in Emotions Elicitation, in: 2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), IEEE, 2023, pp. 1–8.

[8] U. Dimberg, M. Thunberg, K. Elmehed, Unconscious facial reactions to emotional facial expressions, Psychological science 11 (2000) 86–89.

[9] F. Robotics, Furhat robot, www.furhatrobotics.com/, 2024. Accessed: 2024-03-01.

[10] S. Al Moubayed, J. Beskow, G. Skantze, B. Granström, Furhat: A back-projected human-like robot head for multiparty human-machine interaction, in: Cognitive Behavioural Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 114–130.

[11] M. Rescigno, M. Spezialetti, S. Rossi, Personalized models for facial emotion recognition through transfer learning, Multimedia Tools and Applications 79 (2020) 35811–35828.

[12] Z. Yu, X. Li, G. Zhao, Facial-video-based physiological signal measurement: Recent advances and affective applications, IEEE Signal Processing Magazine 38 (2021) 50–58.

[13] P. Biosignals, Bitalino, www.pluxbiosignals.com/collections/bitalino, 2024. Accessed: 2024-03-01.

[14] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, I. Patras, Amigos: A dataset for affect, personality and mood research on individuals and groups, IEEE transactions on affective computing 12 (2018) 479–493.

[15] J. D. Morris, Observations: Sam: the self-assessment manikin; an efficient cross-cultural measurement of emotional response, Journal of advertising research 35 (1995) 63–68.

[16] M. K. Abadi, R. Subramanian, S. M. Kia, P. Avesani, I. Patras, N. Sebe, Decaf: Meg-based multimodal database for decoding affective physiological responses, IEEE Transactions on Affective Computing 6 (2015) 209–222.

[17] E. J. Lawrence, P. Shaw, D. Baker, S. Baron-Cohen, A. S. David, Measuring empathy: reliability and validity of the empathy quotient, Psychological medicine 34 (2004) 911–920.

[18] A. Bangor, P. Kortum, J. Miller, Determining what individual sus scores mean: Adding an adjective rating scale, Journal of usability studies 4 (2009) 114–123.