How transparency impacts the human perception of robotic expressions of emotions?

Mina Marmpena
University of Plymouth, UK
Al Lab, SoftBank Robotics Europe
asimina.marmpena@plymouth.ac.uk

Torbjørn S. Dahl
University of Plymouth, UK
torbjorn.dahl@plymouth.ac.uk

Angelica Lim
Simon Fraser University
Burnaby, British Columbia, Canada
angelica@sfu.ca

ABSTRACT
The integration of affective cues in robot behavior during human-robot interaction (HRI) can enhance engagement and boost collaboration between humans and robots. In our previous work we explored how humans evaluate valence and arousal of robotic non-verbal animations mimicking affective body language in a context-free setup. In our future work we seek to utilize the previous evaluations to examine the robot’s anthropomorphism and empathy impact when it displays the same expressions with respect to emotion inducing context. Furthermore, we will examine how transparency on the mechanism of emotion responses selection influences empathy scores.

KEYWORDS
social human robot interaction, robot affective body language, emotional context, transparency

1 INTRODUCTION
Socially interactive robots can be useful in many different HRI scenarios such as collaboration tasks, care settings, education, or entertainment. Emotional expression is an essential feature for socially interactive robots [1] and appropriate displays of affect can drastically augment engagement on the side of the user [2]. Nonetheless, ethical considerations must also be addressed with respect to embodied agents with highly sophisticated emotional behaviors [3].

In the work presented here, we seek to understand the impact of a humanoid robot that is mimicking emotional body language. Furthermore, we ask how this impact modulated when the emotion expression mechanism is explicitly explained to the human. The emotion representation we use is the dimensional model of non-directional core affect [4], a valid low dimensional structure of the full-blown emotion [5]. Essentially, emotional states are represented on a two-dimensional surface defined by valence (pleasure/displeasure) and arousal.

2 PREVIOUS WORK
In previous work [6], we used as a starting point a set of 36 context-free, non-verbal animations designed by SoftBank Robotics animators for the Pepper robot to convey emotions. The selection was done from the broader animation libraries for Pepper with the objective to reflect emotions in nine different classes of valence/ arousal levels combinations (valence levels: negative, neutral, positive, and arousal levels: tired, calm, excited). Our objective was 1) to increase the usability of the animations by labeling them with dimensional affect ratings, and 2) to explore how humans perceive them.

CONCLUSION
We run an experiment with 20 participants who watched the robot performing the animations, and for each of them they submitted ratings of valence and arousal, as well as the confidence level in their judgment for each rating. Participants responses were facilitated by the question “How does the robot feel”, and it was explicitly stated that the robot was programmed to mimic emotional body language.

We collected the ratings with the Affective Slider interface [7], which displays one bipolar slider for each dimension. The resolution was of 100 data points, and ranged from 0 to 1. For the confidence level, two five-point Likert scales (from 1 = ’Not confident at all’ to 5 = ’Very confident’) were presented after each trial.

For the first objective, we examined the inter-rater reliability which was assessed based on intra-class correlation (ICC) [8]. The result was in a high range (ICC>0.94, where 1 is perfect agreement) for both valence and arousal indicating that raters perceived the emotion displayed by the robot similarly. Based on these results, we derived the final set of the aggregated annotations by taking the mean across raters for each animation. The final annotations are plotted in the valence-arousal space in Fig 1, along with the original tags of the animations and color-coded according to the pre-assigned class.

An interesting trend for valence was observed during data exploration. When confidence in ratings was low, raters tend to evaluate the expression as neutral. This trend might indicate the difficulty to evaluate valence in robots with constrained facial expression, an explanation that can be supported from findings that show facial expression as a more stable visual modality for valence recognition [9]. Another trend that was observed with respect to arousal, was that the lower arousal space was very sparsely populated with ratings. This may be related to the pre-selection of the animations which was conducted by people who are very familiar with the robotic platform as opposed to participants who had minimum or no experience with it. Hence, it might be that the novelty effect contributes to inflated ratings of arousal. Finally, we found that raters’ confidence is significantly higher for arousal than valence ($z=-1.90, p=0.03$), with a medium effect size ($r=0.3$).

3 PROPOSED WORK
The study we propose next follows up from our previous work by utilizing the valence and arousal ratings for selecting emotionally appropriate robotic responses towards emotion inducing pictures that serve as context stimuli. Furthermore, it introduces a transparency factor: the robot’s emotional mechanism is explained to the participants. Our objective is to investigate how humans will evaluate such an agent in terms of empathy and anthropomorphism, and how their evaluations are affected by transparency.
We will use the humanoid Pepper robot, created by SoftBank Robotics (https://www.ald.softbankrobotics.com/en/robots/pepper). The set of the robot’s emotional expressions consists of the 36 animatons: each one lasts for approximately 10 seconds and involves mainly body motion, while some of the animations make use of additional interactive modalities including eye LEDs colorful patterns, sounds or both. Each animation has been labeled with valence and arousal ratings as described in Section 2.

The context representation for this study will be in the form of visual stimuli. We will use a subset of the Geneva Affective Picture E Database (GAPED) [10]. The database consists of 730 emotion inducing pictures labeled with valence and arousal ratings on scales ranging from 0 to 100 points. For our proposed experiment, we will select 36 pictures based on the shortest euclidean distance between their labels and the labels of the animation set.

3.2 Experimental procedure and analysis

For each trial, a picture from the GAPED subset will be presented to the participants, and Pepper will execute an animation as an affective response triggered by the picture. For each trial, the participants will answer a Likert scale asking “How empathic was the robot’s response”. In the end of the experimental session the participants will be asked to fill in a Godspeed questionnaire [7] about the overall degree of anthropomorphism of the robot.

To test how transparency influences human evaluation of empathy and anthropomorphism, we will use two conditions in a between-group design. In the transparency condition, before the task begins, the robot will give a short verbal explanation of the emotion response mechanism it uses, stating the following: 1) the robot is incapable of having emotions, 2) the robot’s emotional body language is designed by animators to mimic human expressions, and 3) each emotional response displayed by the robot is selected with respect to the similarity of the ratings that humans assigned to the pictures and the animations. The second group will only receive instructions on how to do the task, without any explanation about the emotion response behavior.

In the analysis, we will examine the correlation between the empathy scores and the appropriateness of robot’s response. Furthermore, we will test for differences in the empathy and anthropomorphism scores when the transparency factor is applied.

4 CONCLUSIONS

Empathic communication is an essential feature for social HRI. Our objective is to contribute to the integration of such features, taking into consideration the user’s perspective, as well as the potential ethical implications. Initially we aimed to better understand how humans perceive context-free affective expressions performed by a humanoid, and to consolidate this knowledge into usable dimensional affect labels. Following that, we seek to test the impact of this model with respect to emotional context, and to examine how transparency affects the empathic and anthropomorphic weight attributed to the robot by a human. Empathic communication in HRI is an active field of research with benefits and potential risks. Transparency might be the way to mitigate the risks, but could also weaken the positive effects. In any case, further examination of such trade-offs can be proved informative.

ACKNOWLEDGMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No 674868 (APRIL).

REFERENCES


