Through NAO's Emotions: How a Robot Can Express Them Without Words

Ilaria Consoli, Claudio Mattutino and Cristina Gena

[1] Dept. of Computer Science, University of Turin, Corso Svizzera 185, 10149, Turin, Italy

In the context of our research activities on affective computing and human-robot interaction, we are working on both the recognition of human emotions and the expression of emotions by robots. In our vision, robots will be increasingly present in schools, factories, and homes, and their empathetic behaviour may foster their acceptance. In particular, in one of our research, we sought to replicate gestures associated with specific emotions on a social robot, NAO. Our focus was on Ekman's six primary emotions, along with five emotions selected from Plutchik's wheel of emotions. In our opinion, the cultural component linked to the expression of emotions through gestures certainly influenced both us and the participants. The aim of the experiment was to find out whether emotions expressed by robots could be recognised and classified correctly by users.

Keywords

Emotion recognition, Human robot Interaction, Affective computing

1. Introduction

Human communication can be divided into verbal, which involves speech, and non-verbal, often referred to as 'body language.' Non-verbal communication is considered the most vital aspect of human interaction, as 38% of a person's messages are conveyed through para-verbal communication (tone, volume, rhythm...), 55% through body language, and only 7% through speech [13].

The human face is a remarkably expressive part of our body that can convey a wide range of emotions, including sadness, anger, fear, happiness, surprise and disgust. These emotions are known as Ekman's six primary emotions [2]. For example, a big smile is a clear sign of happiness, while wide-open eyes and raised eyebrows convey surprise. Moreover, the face can express a combination of emotions, such as fear and surprise, as seen in expressions like fright.

It is indeed true that body posture and gestures can tell us a lot about a person's emotional state. For instance, crossed arms can indicate defensiveness, while a slouched posture may suggest sadness. Basic gestures like smiling and frowning are universally understood, and even more specific ones like nodding in agreement have evolved through natural selection. On the other hand, extrinsic gestures such as turning away from someone may be learned during childhood [9].

Certain emotions are more effectively expressed through facial expressions, while others find better communication through body movements, or a combination of these two. Gestures can be a useful way to detect the emotional state of a user, especially when combined with voice and facial recognition. Gestures can be simple, reflexive responses, like shrugging shoulders when you the answer to a question is unknown, or they can be complex and meaningful, like using sign language to communicate. Humans can use gestures without any object or environment, like waving our hands, clapping, or practicing a sign. In the field of affective computing, a machine should be able to recognize these gestures, analyze the context, and respond meaningfully to

[1]Joint Proceedings of the ACM IUI Workshops 2024, March 18-21, 2024, Greenville, South Carolina, USA
email1@mail.com (I. Consoli); email2@mail.com (C. Mattutino); email3@mail.com (C. Gena)
0009-0008-6176-0683 (I. Consoli); 0000-0002-0413-2436 m(C. Mattutino); 0000-0003-0049-6213 (C. Gena)

CEUR Workshop Proceedings (CEUR-WS.org)

effectively interact with humans. However, in the context of Human Robot Interaction (HRI), a robot, equipped with arms and hands, should not only be able to recognize and classify user’s gestures, but also to use gestures to express its emotions and its communicative acts.

In the context of our research activities on human-centered AI and human-robot interaction we are working on both the recognition of human’s emotions [7] and the expression of emotions by robots [6]. In our vision, robots will be increasingly present in schools [5], factories [1], and homes [12], and their empathetic behavior may foster their acceptance [8].

In particular, in one of our research projects, we sought to replicate gestures associated with specific emotions on a social robot, NAO. Our focus was on Ekman's six primary emotions [2], thus the six universally recognizable emotions, along with five common emotions selected from Plutchik's wheel of emotions [14]. The aim of the experiment was to find out whether emotions expressed by robots could be recognised and classified correctly by users.

2. The experiment

In one of our research projects, we wanted to replicate gestures associated with specific emotions on a social robot, NAO. Our goal was to create easily recognizable gestures for users to associate with certain emotions. To achieve this, we conducted a guessability study where we directly asked users to associate emotions with gestures. Our references were Ekman's six primary emotions [2], along with five emotions selected from Plutchik's wheel of emotions [14] and Figure 1.

2.1. Background

According to Ekman [2] specific facial expressions are culturally universal and closely related to what he calls 'basic emotions'. These basic emotions, namely joy, sadness, anger, surprise, fear and disgust evolved as physiological reactions and specific expressive signals because of their utility for individual and group survival. Thus, a person pervaded by an emotion such as anger will tend to display an emblematic facial expression, i.e. a sulky face. Ekman and his colleagues are credited with developing a Facial *Action Coding System (FACS*), which provides mappings between muscles and an emotional space. Currently, most attempts to automate facial expression recognition are based on Ekman's system.

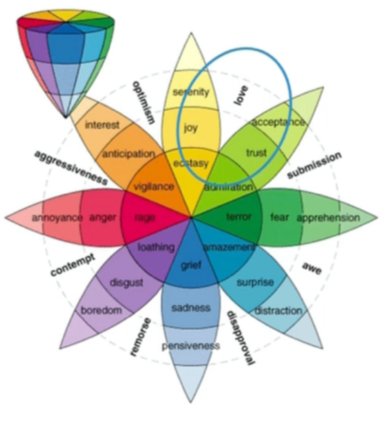

Figure 1. The Plutchik’s wheel of emotions

According to Plutchik's work [14], the category of 'Emotion' is the foundation for all other emotional constructs. The Plutchik’s model of emotions combines a categorial approach to emotions, with distinct emotion types such as joy, awe or fear, with a dimensional approach that sets emotions into similarity and opposition relations, useful to explore diversity. According to this theory, emotions, and their interconnections, can be represented on a spatial structure, a wheel (as reported in the left of the Figure 1), in which the affective distance between different emotional states is a function of their radial distance. The Plutchik's ontology, formalizing such a theory, encodes emotional categories in a taxonomy, representing: basic or primary emotions; complex (or compound) emotions; opposition between emotions; and similarity between emotions. In particular, by following Plutchik's account, complex emotion are considered as resulting from the composition of two basic emotions (where the pair of basic emotions involved in the composition is called a dyad). The compositions occurring between similar emotions (adjacent on the wheel) are called primary dyads. Pairs of less similar emotions are called secondary dyads (if the radial distance between them is 2) or tertiary dyads (if the distance is 3), while opposites cannot be combined [10].

1.1. Design

Initially, eleven emotions were considered, with six being Ekman's main emotions:

- Disgust;
- Happiness;
- Fear;
- Anger;
- Surprise;
- Sadness.

Additionally, five emotions were chosen from Plutchik's wheel of emotions:

- Love;
- Interest;
- Disapproval;
- Boredom;
- Thoughtfulness (Pensiveness) .

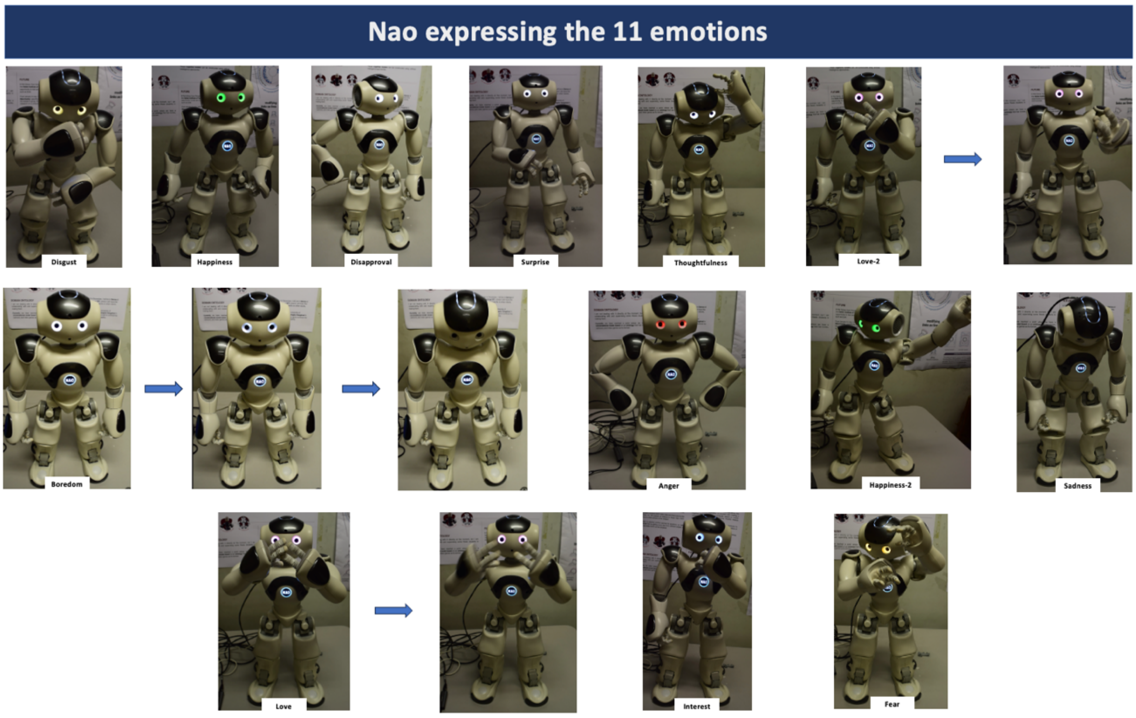

Figure 2. Nao expresses the 11 emotions, some of them are performed twice.

In deciding on gestures representative of emotions, we relied on these sources reported by Github [18] and then readjusted the gestures to those typical of Italian culture [15]. The total number of

emotions represented was consistently more than eleven due to some emotions being depicted multiple times, see Figure 2. The sensory aspect most engaged in the experiment was sight, with the only restriction being the absence of sound. The exclusion of sound aimed to emphasize the role of visual perception in the participants' interpretation of emotions through robot animations. However, it's acknowledged that involving the auditory sense could have expedited participants' recognition of emotions.
For each emotion, a specific color was associated with eye movements to enhance identification. The color selection process incorporated both a rational approach, considering the meaning of colors in art, and playful sources like cartoons, exemplified by 'Inside Out.'
The gestures and behaviors to reproduce emotions in NAO were programmed using the robot's development environment (the NAOqi Framework and the Choregraphe multi-platform desktop application within) and were directly triggered by the experimenter through the same software.

1.2. Subjects

Participants included a total of 20 students, comprising 11 males and 9 females. Among males, 5 were in the 18-24 age group, and 6 were in the 25-34 age group. For females, 8 were in the 18-24 age group, and 1 was in the 25-34 age group. The 20 people were divided into two groups of 10; the initial 10 participants were presented with preliminary prototype animations, while the subsequent 10 were exposed to revised animations, incorporating feedback provided by the first group.

1.3. Procedure

The experiment was conducted as a *guessability* study [17], in which participants were asked to guess the emotion expressed by the NAO robot, choosing from a list of given emotions.
The experiment was conducted in two stages, creating two groups of ten participants each. The first group was shown the initial animations created with NAO, while the second group was presented with revised versions of the same animations.
Participants were given a list of emotions to guide them. Once the animations were introduced, the participants began to watch them. The order of presentation was different from the order on the sheets, and duplicate animations were not communicated to the participants to avoid any exclusion or ambiguity in the results. In both experiments, duplicates were represented with different characteristics.
Participants were asked to observe the emotions mimicked by NAO and then try to identify them. If the response did not coincide with the intended emotion, the participant was then asked to suggest changes.
The feedback from the first ten participants was used to clarify which animations corresponded to which emotions. Based on these conclusions, improvements were made to the second part of the experiment, which aimed to be more intuitive. The final gestures used in the second experiment are shown in Figure 2.

1.4. Results

As shown in Figure 3, the results of the first experiment did not fully meet the initial expectations. There was a total of 15 first animations, including duplicates. However, most of the emotions conveyed by the animations were not particularly intuitive, and some of them were ambiguous, possibly due to the contrasting colour of the eyes. The approach of using colors to express emotions did not prove to be very effective. For example, the colour green was used to represent disgust, but many users associated it with something positive, making it more appropriate for expressing happiness.

Among all the animations, Thoughtfulness, Happiness, and Disapproval were the most successful ones. These animations were also proposed to the second group without making any changes, while the others were modified based on the users' suggestions.

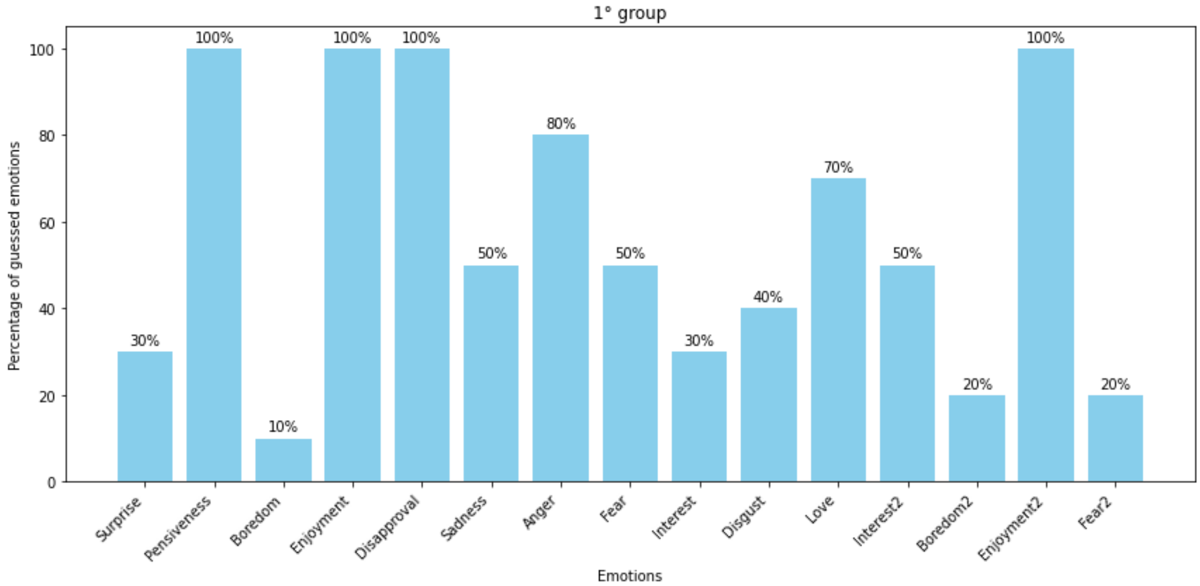

Figure 3. Experimental results (first group) of the guessability study. The percentage represents the number of participants who guessed the emotion.

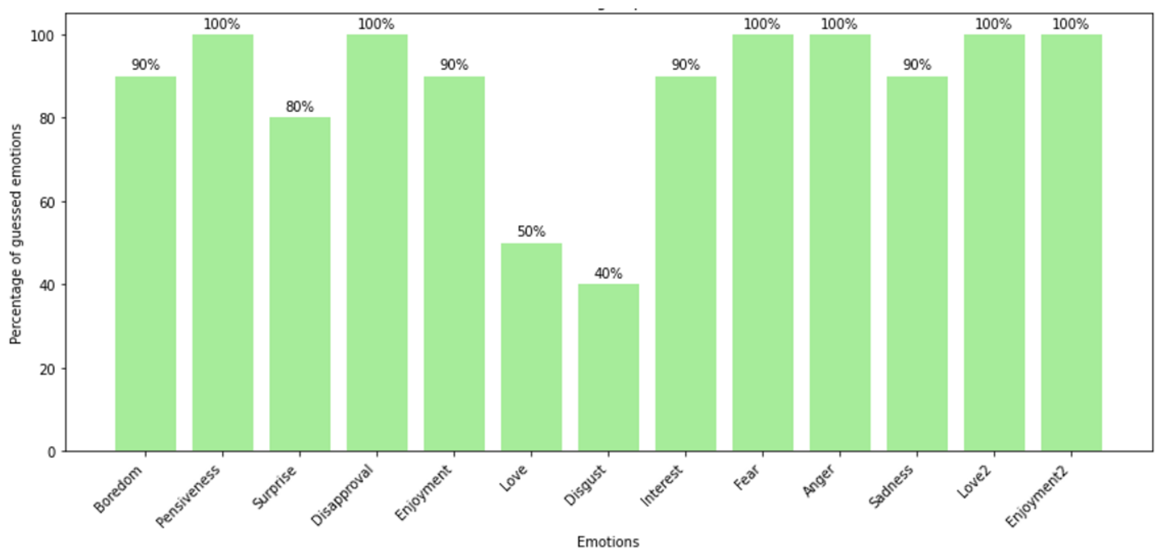

Figure 4. Experimental results (second group) of the guessability study. The percentage represents the number of participants who guessed the emotion.

Results of the second experiment were quite encouraging: apart from Disgust (40%) and one version of Love (50%), participants quite easily guessed the emotion mimicked with gestures, as it can be observed in Fig. 4.

3. Discussion

The obtained results are interesting, but we need to replicate the experiment with a different sample of users. We believe that the Italian cultural component, which is linked to expressing

emotions through gestures, had an impact on both us and the participants. Indeed Italian gestures often act as a secondary language that goes beyond any linguistic barriers. Over time, they have developed into a unique form of communication (as mentioned in [15]), and are sometimes used to depict stereotypes of Italian culture [4].
However, further investigation is needed to determine their universal applicability and relevance in different cultures, as suggested in [3]. Therefore, it would be interesting to replicate our study in other countries to determine whether the emotions conveyed by our 'Italian' robot are similarly recognized in settings with cultural similarities.
This proposed cross-cultural and cross-domain approach [16] not only enhances the generalizability of our findings but also provides an opportunity to determine cultural influences in interpreting robotic gestures.

References

[1] Brunetti, D., Gena, C., & Vernero, F. (2022). Smart Interactive Technologies in the Human-Centric Factory 5.0: A Survey. *Applied Sciences*, *12*(16), 7965.

[2] Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

[3] Esposito, A., Riviello, M. T., & Bourbakis, N. (2009). Cultural specific effects on the recognition of basic emotions: A study on Italian subjects. In HCI and Usability for e-Inclusion: 5th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2009, Linz, Austria, November 9-10, 2009 Proceedings 5 (pp. 135-148). Springer Berlin Heidelberg.

[4] Gena, C., & Ardissono, L. (2001, July). On the construction of TV viewer stereotypes starting from lifestyles surveys. In *Workshop on Personalization in Future TV.*

[5] Gena, C., Mattutino, C., Perosino, G., Trainito, M., Vaudano, C., & Cellie, D. (2020, May). Design and development of a social, educational and affective robot. In *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)* (pp. 1-8). IEEE.

[6] Gena, C., Mattutino, C., Maieli, A., Miraglio, E., Ricciardiello, G., Damiano, R., & Mazzei, A. (2021, June). Autistic Children's Mental Model of an Humanoid Robot. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 128-129).

[7] Cristina Gena, Alberto Lillo, Claudio Mattutino, and Enrico Mosca. 2022. Wolly: an affective and adaptive educational robot. In Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '22 Adjunct). Association for Computing Machinery, New York, NY, USA, 146–150. https://doi.org/10.1145/3511047.3537684

[8] Gena, C., Manini, F., Lieto, A., Lillo, A., & Vernero, F. (2023). Can empathy affect the attribution of mental states to robots?. In Proceedings of the 25th International Conference on Multimodal Interaction, ICMI 2023, Paris, France, October 9-13, 2023 (pp. 94-103). Association for Computing Machinery.

[9] Liao, J., & Wang, H. C. (2019). Gestures as intrinsic creativity support: Understanding the usage and function of hand gestures in computer-mediated group brainstorming. *Proceedings of the ACM on Human-Computer Interaction*, *3*(GROUP), 1-16.

[10] Lieto Antonio, Manuel Striani, Cristina Gena, Enrico Dolza, Anna Maria Marras, Gian Luca Pozzato & Rossana Damiano (2024) A sensemaking system for grouping and suggesting stories from multiple affective viewpoints in museums, Human–Computer Interaction, 39:1-2, 109-143, DOI: 10.1080/07370024.2023.2242355

[11] Lutz, C., & White, G. M. (1986). The Anthropology of Emotions. *Annual Review of Anthropology*, *15*, 405–436. http://www.jstor.org/stable/2155767

[12] Macis, D., Perilli, S., & Gena, C. (2022, July). Employing Socially Assistive Robots in Elderly Care. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 130-138).

[13] Albert Mehrabian. 2017. Nonverbal communication. Routledge.

[14] Plutchik, R. (2001). The nature of emotions. *American Scientist*, *89*(4), 344–350.
[15] Poggi, I. (2002). Symbolic gestures: The case of the Italian gestionary. Gesture, 2(1), 71-98.
[16] Giuseppe Sansonetti, Fabio Gasparetti, and Alessandro Micarelli. 2019. Cross-Domain Recommendation for Enhancing Cultural Heritage Experience. In Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP'19 Adjunct). Association for Computing Machinery, New York, NY, USA, 413–415. https://doi.org/10.1145/3314183.3323869
[17] Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005, April). Maximizing the guessability of symbolic input. In *CHI'05 extended abstracts on Human Factors in Computing Systems* (pp. 1869-1872).
[18] https://github.com/mikecheninoulu/Emotional-gesture-papers