Multimodal Affective Behaviour Expression: Can It Transfer Intentions?

Christiana Tsiourti
University of Geneva
Geneva, Switzerland

Astrid Weiss
Vienna University of Technology
Vienna, Austria

Abstract— **Autonomous interactive robots developed for social human–robot interaction scenarios, should have cognitive architectures supporting social capabilities that enable naturalistic interactions with humans. The cognitive sciences and HRI literature support the use of facial expressions, bodily expressions and vocal expressions as a way to convey how external events influence a robot's internal affective state so that humans can interpret and predict its intentions and behaviors. Nevertheless, HRI researchers have yet to establish to what degree, and how precisely, each of these modalities is involved in the perception of a robot's affective state and the attribution of robot intentions. In this position paper, we propose a series of empirical studies to prove the influence of facial, bodily and verbal expressions, on the accurate recognition of a robot's affective expression. Once the influence of each modality is established, we can use and combine facial, bodily and verbal cues, to design expressions for robots that interact naturally and effectively with people. Our work is grounded in psychological research on human expression and perception of affective stimuli, and our empirically validated findings will contribute towards the establishment of common ground and standard guidelines for affect expression in social HRI scenarios.**

Keywords—Social HRI, robot affective expression, verbal/non-verbal Human-Robot-Interaction, facial expression body motion, multimodality.

I. Introduction

Humans have evolved social-cognitive mechanisms that promote fluid and effective social interactions. Affect is a key mechanism of human social-cognition that is significant for everyday functioning and interaction. At the heart of any emotionally charged event are the so-called affective states, which influence our reflexes, perception, cognition, and behaviour and are influenced by many internal and external causes [1]. Autonomous interactive robots, developed for social human–robot interaction (HRI) scenarios, should have cognitive architectures supporting social capabilities that enable naturalistic bi-directional interactions with humans and other robots [2]. Such robots need to have the ability to perceive and identify complex human social behaviours and, in turn, be able to express their own behaviours using well-recognized communication modes.

The expression of affective information supports functional, adaptive behaviour by allowing a robot to convey how external events influence its internal state so that humans (and other robots) can interpret and predict its intentions and behaviors [3]. For example, a robot nurse assistant, deployed in the dynamic environment of a hospital, should be able to greet patients, appear happy when informing them of good results and express sorrow or encouraging emotions, when the test results are not satisfying. If this robot does not behave according to social norms, the interaction can quickly turn unpleasant and unnerving.

Animators and artists have been trying to design robots that actuate behavior that can convey affective information regarding their internal state and intentions for decades (see the Disney classic, "The Illusion of Life"[4]). In social HRI, there is still no established protocol and no common guidelines on how to design affective expressions for robots. Different research teams used facial expression [3], speech (i.e., voice level, pitch) [5], body posture, orientation and motion (i.e., acceleration, curvature) [6], head motion[7], gaze direction, sound and colour as either the primary method of expression or to provide expression redundancy on robots with different platforms (see [8] for a comprehensive review). Previous findings have shown that robot emotions can be identified only with the use of body motion, without speech or facial expressions [7] and that distinctive patterns of body movements are associated with specific emotions [9].

Despite the previous findings, the contribution of each modality to the perception of robot affective states has not been clearly established yet. More precisely, it is not currently possible to predict the effect that a specific modality has on the perception of expressions actuated by a robot. For example, if a robot is expressing fear with a facial expression, could the communication become stronger if the robot also used a withdrawing body movement and a non-verbal whimpering sound? Additionally, if the robot's face is not visible, could then non-facial and non-verbal affective expressions provide additional cues to continue the social interaction?

Our work addresses a small yet impactful gap in our understanding of human-robot social interaction. Lack of knowledge regarding how each modality actually conveys affective information to the user about the robot's internal state and intention makes it difficult to create robots that actuate synchronized affective behaviour corresponding to the situational context of the interaction. In many HRI studies, affective expressions are hand-designed by trained artists or animators by “blending” primitive movements of emotions. Too often these designs rely on a gut “feeling” that the expression is right, rather than a systematic understanding of human affective behavior. Social HRI would benefit from a set of standard guidelines on how to use and combine expressive cues into accurately perceived affective robot expressions, grounded on theories from human social and affective interaction and validated in empirical HRI studies.

In this position paper, we propose the creation of guidelines on how to use and combine facial, bodily and vocal cues into accurately recognized affective expressions that convey the affective states and intentions of robots operating in social HRI scenarios. In this position paper, we propose a series of empirical HRI studies to evaluate the contribution of each modality on the accurate perception of a robot´s affective state, within the context of social interaction. We consider the following contribtuions of the proposed research to the field : (1) we will create and validate of a pool of affective expressions using facial expression, body expression and vocal expression and (2) we will establish the affective contribution of each modality in the accurate recognition of a robot’s affective states and intentions. Based on these findings we will suggest how to create accurately perceived affective expressions for social HRI scenarios, using use one, two or more modalities.

II. Expression and perception of affect in cognitive science and HRI

In humans, affect expression occurs through combinations of visual and auditory communication channels such facial, vocal and bodily expressions. A number of studies in psychology, cognitive science, and HRI investigate how these cues contribute individually and how they interact within the human sensory system to perceive affective information in human or robotic counterparts.

A. Facial, Bodily and Vocal expression

Facial expression: A fair amount is known and accepted about affective facial expressions, such as some ways in which they are conveyed and recognized and how to code them based on the well-established coding FACS system [10]. The modeling of facial expressions on robots is not an easy task due to the mechanical limitations of the robots’ faces. Researchers using highly expressive robots often rely on the FACS and map the corresponding joints in the robot's face to the FACS descriptors. Numerous studies suggest that the recognition rates for facial expressions are substantially high for basic emotions (e.g., [3], [11]).

Bodily expression: Although initial studies suggested that the human body does not function as an additional source of information in the communication of affect, more recent research has shown that human body language plays an important role in effectively communicating certain emotions either combined with facial expressions as well as on its own [12]–[14]. A human study investigating the recognition of the basic emotions of anger, fear, happiness and sadness, conveyed only through body language, found recognition rates greater than 85 % for all the emotions [15]. De Gelder [16] postulates that body expressions may provide more information than the face when discriminating between fear and anger or fear and happiness. Another study [17], found that body posture was the influencing factor over the recognized emotion when observers were presented with incongruent combinations of facial expressions and posture or movement.

In HRI, it has been shown that affective body language can be interpreted accurately without facial or vocal cues (e.g., [7], [9]). Beck and colleagues [7] investigated the effect of varying a robot’s head position on the interpretation of static emotional key poses. They discovered that even small cues, like a different head position, can change the perception of a pose significantly. Moving the head down lead to decreased arousal, valence and stance whereas moving the head up increased these dimensions. However, this study focused on emotional expression through static postures, rather than movement. Our work considers both form and movement since studies indicate that both are useful and important for the perception of affect from body expressions[12].

Vocal expression: Speech prosody can reflect affect through changes in pitch, intensity, rate, timing and voice quality. With the aim of synthesizing emotional speech that can add more naturalness to human-robot interaction, Brezeal [18] adapted several correlates of human speech to a synthesizer, allowing a robot to speak in either an angry, calm, disgusted, fearful, happy, sad, or surprised manner and studied how well human subjects perceived the intended affect. Aly and Tapus [19] also propose a design for vocal patterns corresponding to a subset of emotions.

B. Complementarity of modalities

The correlation between facial expression, body motion and speech has been intensively investigated in psychology and cognitive science. Findings suggest the complementarity of certain modalities so that the perception of affective information can be ameliorated when two or more channels are considered at the same time. Several HRI efforts were driven towards synchronizing expressive modalities. Costa et al., [20] showed that gestures are a valuable addition to the recognition of facial expressions of robots. Le at al. [21] synthesized expressive body gestures with speech. The findings of Salem et al. [22] suggest that a robot is evaluated more positively when gestures are displayed along with speech. A recent study by Aly and Tapus [19] validated the role of multimodality in increasing the clearness of emotional content. Their findings prove the role of facial expressions in enhancing the expressiveness of the robot behavior and the role of the generated gestures in recognizing target emotions.

To the best of our knowledge, there is no comprehensive comparison that investigates how the modalities of facial expression, body expression, and vocal expression, contribute to the attribution of particular internal affective states in naturalistic social HRI scenarios. The majority of HRI studies in affect expression focus either on the validation of single modalities in isolation (e.g., [7], [23]) or the validation of multimodal expressions at large without consideration of how specific modalities contribute and interact towards the attribution of particular internal affective states (e.g.,[9]). This is especially important in the context of social HRI scenarios, because designs based on inappropriately combined cues may result in misinterpretation of the robot’s affective state and intention. In contrast, we also evaluate combinations of two and three modalities, to establish the added value of each modality.

III. Transfering Affect and Intentions through Facial, Bodily and Vocal expressions

In this section, we present our methodological approach towards the creation of guidelines on how to use and combine expressive cues into accurately recognized affective expressions. Our work is grounded in theories of human affect and social interaction, as well as empirical findings from HRI studies which indicate that there are several reliable features of facial expression, body expression and vocal expression that can be manipulated to function as social cues conveying affective information, both in humans and in robots.

Our first goal is to evaluate the perception of single-modality expressions, to establish whether each of the three modalities can offer a basis for intuitive-affective interaction between humans and robots. Based on literature descriptions of affective state expressions that consider form and movement we will design of a pool of affective expressions for a robotic platform, using facial expression, body expression and vocal expression that clearly describe three affective states: happy, sad, surprise. The neutral state is included as a baseline. Table 1 presents our tentative design space for bodily expressions, based on the work of de Meijer [14], Coulson [13] and Kleinsmith et al. [12]. We will systematically compare head motion sequences (straight, backward bend, forward bend), arm motion sequences (parallel to the body, vertical extension, parallel extension) and upper body motion sequences (straight torso, backward chest bend, forward chest bend). Affective vocal expressions will be synthesized using a commercial TTS engine, to add relevant prosodic cues to free of emotional context sentences.

In our first experiment, we will validate the design of the affective expressions. We will ask a wide audience of human participants to watch each expression and label it with the corresponding affective state. At this stage, no contextual information will be provided, to guarantee that the expressions are suitable for a number of situational contexts and interaction scenarios. With this study, we will be able to perform a detailed analysis of the perception of the features of each expression (e.g., head, arms) and compare the findings with our hypotheses and the psychological research they are based on. The findings will allow us to identify flaws in our design and systematically revise our expressions for the three modalities.

Once we have established accurately recognized single-modality expressions for all the affective states, we will combine the expressions, to test the contribution of each modality and the complementarity of modalities on the accurate recognition of robot affective states. Our second experiment will be based on a 3x4 within-subject design with two independent variables: the combination of modalities (one modality, two modalities, and three modalities) and the affective state (happy, sad, angry, neutral). In a Wizard-of-Oz setup, participants will watch videos together with a robot, while an experimenter is controlling the robot's expressions to convey different affective states in response to the video content. We will evaluate one dependent variable: correctly/incorrectly recognized affective state. Furthermore, we will sample the participant´s Heart Rate Variability (HRV) and Galvanic Skin Response (GSR), with the purpose of understanding to which extent the psychophysiological signals disclose information related with the robot´s expressions.

Without the facial expressions, or the ability to move, the robot will initially rely only on verbal prosody cues to convey its internal affective state. Based on the abovementioned findings that suggest the complementarity of modalities, the face and body are expected to offer additional levels of expressiveness to specific affective states.

Table 1: Body expression design based on literature descriptions

	Body Expression		
Target Emotion	**Head**	**Upper body**	**Arms**
Sadness	***Forward head bend***	***Forward chest bent***	***Arms at side of trunk***
Happiness	***Backward head bent***	***Straight trunk***	***Vertical and lateral extension***
Surprise	***Backward head bent***	***Backward chest bent***	***Vertical extension***
Neutral	***Straight head***	***Straight trunk***	***Arms at side of trunk***

IV. Looking ahead: Questions and parting thoughts

Naturalistic human-robot interactions require that robots actuate synchronized affective behaviour corresponding to the situational context of the interaction. Expressing affect in robots is equivalent to creating the “illusion of life” in robots: making people think and feel that the mechanical being they see in front of them actually has a persona and feelings [4]. To accomplish this kind of interaction requires understanding of the human perception system based on theories from psychology and cognitive science, and more empirical effort in the area of social HRI, so as to validate the influence of each modality on the accurate perception of affective expressions. Our work on affect expression is still in its initial phases, but it gives rise to some critical questions that we look forward to discussing with the HRI, psychology and cognitive science communities and investigate in our future empirical work:

- *Do humans base their perceptions of affect on one modality more than another? That is, does either the facial expression, body or vocal expression dominate in perceptions of the affective expression, or are they all essential?*
- *Does combining visual and audio cues result in enhanced recognition of specific affective states? If so, which combinations of facial expression, body or vocal expression are best for each state?*
- *How important is it that facial, body and vocal expressions are consistent? For example, what do humans perceive if a robot has a happy face with a concerned voice?*

V. CONCLUSION

The main objective of our research is the creation of rules and guidelines to combine facial, bodily and vocal expressions into accurately recognized affective expressions for robots that engage in naturalistic social interactions with humans. In this paper, we propose a series of empirical studies which will contribute towards the definition of a pool of candidate expressions that consist of one or multiple modalities (facial expression, body expression, and vocal expression) to express happy, sad and surprised states in a robotic platform. Since these affective states are mapped onto different quadrants of the valence-arousal space, we expect our findings to generalize to other affective states. With regards to generalizability to other robot embodiments which may not be able to perform expressions through all the three modalities, we aim to offer a set of expressions with one modality, two modalities, and three modalities rather than pick only one best expression for each affective state.

Besides the empirically validated affective expressions using facial expression body expression and vocal expression, the novel contribution of our work is deepening the understanding of how to express well recognized affective states through these three modalities. We expect that our findings, which will take the form of mappings between modalities and the affective states, will contribute towards the definition of design guidelines which HRI researchers could readily employ to design expressive robots for naturalistic social HRI scenarios.

REFERENCES

[1] J. A. Russell, “Core affect and the psychological construction of emotion.,” Psychol. Rev., vol. 110, no. 1, pp. 145–172, 2003.

[2] P. Baxter, “Cognitive Architectures for Social Human-Robot Interaction,” in Proceedings of HRI ’16 The Eleventh ACM/IEEE International Conference on Human Robot Interaction, 2016, pp. 579–580.

[3] C. Breazeal, “Emotion and sociable humanoid robots,” Int. J. Hum. Comput. Stud., vol. 59, no. 1–2, pp. 119–155, Jul. 2003.

[4] T. Ribeiro and A. Paiva, “The illusion of robotic life,” in Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI ’12, 2012, p. 383.

[5] C. Breazeal, “Emotive qualities in robot speech,” in Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180), 2001, vol. 3, pp. 1388–1394.

[6] J. Mumm and B. Mutlu, “Human-robot proxemics,” in Proceedings of the 6th international conference on Human-robot interaction - HRI ’11, 2011, p. 331.

[7] A. Beck, L. Canamero, and K. A. Bard, “Towards an Affect Space for robots to display emotional body language,” in 19th International Symposium in Robot and Human Interactive Communication, 2010, pp. 464–469.

[8] N. Mavridis, “A review of verbal and non-verbal human–robot interactive communication,” Rob. Auton. Syst., vol. 63, pp. 22–35, Jan. 2015.

[9] D. McColl and G. Nejat, “Recognizing Emotional Body Language Displayed by a Human-like Social Robot,” Int. J. Soc. Robot., vol. 6, no. 2, pp. 261–280, Apr. 2014.

[10] P. Ekman and D. Keltner, “Universal facial expressions of emotion,” Calif. Ment. Heal. Res. Dig., vol. 8, no. 4, pp. 151–158, 1970.

[11] N. Mirnig, E. Strasser, A. Weiss, B. Kühnlenz, D. Wollherr, and M. Tscheligi, “Can You Read My Face?,” Int. J. Soc. Robot., vol. 7, no. 1, pp. 63–76, Nov. 2014.

[12] A. Kleinsmith and N. Bianchi-Berthouze, “Affective Body Expression Perception and Recognition: A Survey,” IEEE Trans. Affect. Comput., vol. 4, no. 1, pp. 15–33, Jan. 2013.

[13] M. Coulson, “Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence,” J. Nonverbal Behav., vol. 28, no. 2, pp. 117–139, 2004.

[14] M. de Meijer, “The contribution of general features of body movement to the attribution of emotions,” J. Nonverbal Behav., vol. 13, no. 4, pp. 247–268, Dec. 1989.

[15] B. de Gelder and J. Van den Stock, “The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions.,” Front. Psychol., vol. 2, p. 181, 2011.

[16] B. de Gelder, “Towards the neurobiology of emotional body language,” Nat. Rev. Neurosci., vol. 7, no. 3, pp. 242–249, Mar. 2006.

[17] J. Van den Stock, R. Righart, and B. de Gelder, “Body expressions influence recognition of emotions in the face and voice.,” Emotion, vol. 7, no. 3, pp. 487–494, Aug. 2007.

[18] C. Breazeal and L. Aryananda, “Recognition of Affective Communicative Intent in Robot-Directed Speech,” Auton. Robots, vol. 12, no. 1, pp. 83–104, 2000.

[19] A. Aly and A. Tapus, “Multimodal Adapted Robot Behavior Synthesis within a Narrative Human-Robot Interaction,” in Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on, 2015, pp. 2986–2993.

[20] S. Costa, F. Soares, and C. Santos, “Facial Expressions and Gestures to Convey Emotions with a Humanoid Robot,” in Social Robotics, vol. 8239, Springer International Publishing, 2013, pp. 542–551.

[21] Q. A. Le, J. Huang, and C. Pelachaud, “A Common Gesture and Speech Production Framework for Virtual and Physical Agents.”

[22] M. Salem, K. Rohlfing, S. Kopp, and F. Joublin, “A friendly gesture: Investigating the effect of multimodal robot behavior in human-robot interaction,” in 2011 RO-MAN, 2011, pp. 247–252.

[23] M. Haring, N. Bee, and E. Andre, “Creation and Evaluation of emotion expression with body movement, sound and eye color for humanoid robots,” in 2011 RO-MAN, 2011, pp. 204–209.