Human brain spots emotion in non humanoid robots

Stéphanie Dubal, Aurélie Foucher, Roland Jouvent, and Jacqueline Nadel
CNRS USR 3246, GH Pitié-Salpêtrière, Paris, France

The computation by which our brain elaborates fast responses to emotional expressions is currently an active field of brain studies. Previous studies have focused on stimuli taken from everyday life. Here, we investigated event-related potentials in response to happy vs neutral stimuli of human and non-humanoid robots. At the behavioural level, emotion shortened reaction times similarly for robotic and human stimuli. Early P1 wave was enhanced in response to happy compared to neutral expressions for robotic as well as for human stimuli, suggesting that emotion from robots is encoded as early as human emotion expression. Congruent with their lower faceness properties compared to human stimuli, robots elicited a later and lower N170 component than human stimuli. These findings challenge the claim that robots need to present an anthropomorphic aspect to interact with humans. Taken together, such results suggest that the early brain processing of emotional expressions is not bounded to human-like arrangements embodying emotion.

Keywords: emotion; ERP; N170; P1; non humanoid robots

INTRODUCTION

The computation by which our brain elaborates fast responses to emotional expressions is currently an active field of brain studies. Until now, studies have focused on stimuli taken from everyday life—namely facial expressions of emotion, emoticons (also called smileys) created to convey a human state of emotion and instantly taken as such, landscapes or social pictures likely to generate recalls of past emotional events, etc. Less is known, however, concerning brain responses to expressive stimuli devoid of humanity and that have never been seen before. Non-humanoid robots can help providing such kind of information.

Robots are a potential scaffold for neuroscientific thought, to put it in Ghazanfar and Turesson’s words (Ghazanfar and Turesson, 2008). This comment intends to announce a turn in the exploration of the brain by neuroscientists. More and more is known about the fact that the Mirror Neuron System responds anthropomorphically to robotic movements, when the mechanical motion is closely matched with the biological one (Gazzola et al., 2007; Oberman et al., 2007). Interference between action and observation of a biological motion performed by a humanoid robot suggests similar action effect of robotic and human motor perception (Chaminade et al., 2005), although an artificial motion may require more attention (Chaminade et al., 2007). Electromyographic studies have found that we respond by similar automatic imitation to the photograph of a robot opening or closing hand and to the photograph of a human hand (Chaminade et al., 2005; Press et al., 2005; Bird et al., 2007; Oberman et al., 2007). In sum, the human brain tends to respond similarly to human compared to robotic motion. But, what about emotion, so deeply anchored in our social brain (Brothers, 1990; Allison et al., 2000)? Does the human brain respond similarly to emotion expressed by human and robot?

To explore this question, we compared event-related potentials (ERPs) evoked by human expressions of emotion and by emotion-like patterns embodied in non-humanoid robots that should prevent the patterns from being captured as conveying mental states of emotion and recalling emotional reactions already experienced.

Recent brain findings may feed assumptions about whether we can react similarly to robotic and human emotional stimuli. Perceptual processing of facial emotional expressions engages several visual regions involved in constructing perceptual representations of faces as well as activations of subcortical structures such as the amygdala (Adolphs, 2002; Vuilleumier and Pourtois, 2007). Functional MRI studies found greater activations in the occipital and temporal cortex and in the amygdala in response to expressive faces as compared to neutral faces (Breiter et al., 1996; Vuilleumier et al., 2001). Electroencephalographic studies have localized temporally this amplification of activity for emotional content at about 100 ms post-stimulus, with occipital components of the ERPs amplitude being higher for emotional than neutral faces (Pizzagalli et al., 1999; Batty and Taylor, 2003; Pourtois et al., 2005). In non-human primates, intracell recordings found temporal lobe face-selective neurons differentiating emotion vs neutral
expression at a latency of 90 ms, and differentiating monkey identity about 70 ms later (Sugase et al., 1999). These delays meet P1 and N170 visual components latencies in human EEG studies, with the P1 wave being modulated by emotion, and the N170 wave being mainly modulated by facial configuration, and originated in higher-level visual areas selective of face recognition (Bentin et al., 1996; George et al., 1996; Pizzagalli et al., 1999; Batt y and Taylor, 2003; Rossion and Jacques 2008).

The emotional modulation of the P1 has been described in the literature in response to human (Batty and Taylor, 2003) and schematic emotional faces (Eger et al., 2003) and also in response to emotionally arousing stimuli such as body expressions (Meeren et al., 2005), words (Ortigue et al., 2004) or complex visual scenes (Carretie et al., 2004; Alorda et al., 2007). Such perceptual enhancement has been found in response to positive as well as negative eliciting stimuli (Batty and Taylor, 2003; Brosch et al., 2008). The P1 visual component has a field distribution compatible with activation of occipital, visual brain areas, and may be due to several striate and extrastriate generators within the posterior visual cortex (Di Russo et al., 2003; Vanni et al., 2004; Joliot et al., 2009). The posterior activation is likely to be modulated by the amygdala, acting remotely on sensory cortices to amplify visual cortical activity (Amaral et al., 2003), and may as well involve frontal sources (Carretie et al., 2003). Such participation would be related to the top-down regulation of attention designed to enhance perceptual processing of emotionally relevant stimuli. P1 amplitude enhancement in response to emotional stimuli compared to neutral stimuli may reflect a sensory gain for the processing of emotional stimuli. This is congruent with results from fMRI studies that reported emotional amplification in several visual regions of early perceptual processing (e.g. Lang et al., 1998).

Now suppose that we are facing a non-humanoid robot displaying emotional and neutral patterns. We wonder whether the encoding of emotion from such a machine-like set-up may occur as early as P1: will we find a P1 enhancement for the emotional patterns, notwithstanding the fact that the set-up is a complex combination of emotional features and typical components of mechanical devices?

To investigate this question, we created pictures of robotic patterns presenting two contrasted properties: some of the robotic features are similar to inner features of emotion in the human face (mouth, eyes, brows) but a great number of other features are familiar components of machine set-ups (cables, nails, threads, metallic pieces). In such a case, we do not know what the prominent perceptual saliency should be. If we reason by analogy we could consider that the answer is already given in studies having used schematic faces (Eger et al., 2003; Krombholz et al., 2007). Those studies have shown similar brain responses to the emotional expressions of real faces compared to schematic faces. However, our robots are not analogous to schematic faces, at least regarding the criterion of complexity. In schematic faces, only the main invariant of an expression has been selected so as to provide both simpler and prototypical patterns. Reversely, our designs are complex mixtures of a prominent feature of emotion and familiar components of mechanical devices. Schematic faces show much less variance than human faces (Kolassa et al., 2009) or robots. It is not given for granted that the invariant of the robotic emotional expression will be extracted from the whole. Those contrasted properties of our designs allow us to explore whether the brain can spot emotion in complex non-humanoid stimuli.

**METHOD**

**Participants**

Fifteen right-handed subjects took part in the experiment (seven women and eight men, mean age 22.1 ± 3.28). All had normal or corrected to normal vision. All participants provided informed consent to take part in the experiment that had been previously approved by the local ethical committee.

**Stimuli and design**

Eight photos of four different robotic set-ups displaying happy or neutral expressions and eight photos from four different human faces displaying happy or neutral expressions were selected (Figure 1). The eight robotic expressions were created following the Facial Action Coding System (FACS) criteria. Each emotional expression was validated by a FACS-certified researcher and a FACS expert. The robotic set-ups from which the pictures were drawn were all inspired by Feelix (Canamero and Fredslund, 2001; Canamer o and Gaus sier, 2005) and designed by Gau ssi er and Canet (Canamero and Gaus si er, 2005; Nadel et al., 2006). They were composed of a robotic architecture linked to a computer. The eyes, eyebrows, eyelids and mouth were moved by 12 servomechanisms connected to a 12-channel serial servo-controller with independent variable speed. An in-house software was used to generate the emotional patterns and to command the different servo-motors accordingly. At a pre-experimental stage, robotic expressions were recognized by 20 adults as happy or neutral with a mean percent score of 95% for happy and 84% for neutral (Nadel et al., 2006).

Concerning human pictures, six were drawn from Ekman and Friesen’s classical set of prototypical facial expressions and two were posed by an actor exhibiting the same facial action units as those described in the FACS (Ekman and Friesen, 1976). All 16 pictures were black and white. They were equalized for global luminance. Pictures were presented in the centre of the computer screen and subtended a vertical visual angle of 7°.

**Behavioural data recording and analysis**

Participants were asked to discriminate if pictures displayed a neutral or an emotional expression and to press a choice
button by either the left or the right hand in a counterbalanced order. Subjects were asked to react as quickly as possible although avoiding incorrect responses. The participants completed a total of 256 trials including a practice block of 16 trials, followed by four blocks of 60 experimental trials, lasting about 20 min. Each block consisted of random combinations of four robotic and four human displays of happy and neutral expressions. Each trial began with the presentation at the centre of the screen of a fixation cross for 100 ms, followed by a picture stimulus for 500 ms. Interval between trial varied randomly between 1000 and 2400 ms.

Accuracy, by means of percentage of correct identification of emotional expression, and response time were analysed separately by a 2 × 2 repeated measures analysis of variance (ANOVA), with media (robot, human) and emotion (happy, neutral) as the repeated measures.

Electrophysiological recording and analysis
Electroencephalograms (EEGs) were acquired from 62 Ag/AgCl electrodes positioned according to the extended 10–20 system and referenced to the nose. Impedances were kept below 5 kΩ. The horizontal electrooculogram (EOG) was acquired using a bipolar pair of electrodes positioned at the external canthi. The vertical EOG was monitored with supra and infraorbital electrodes. The EEG and EOG signals were sampled at 500 Hz, filtered online with a bandpass of 0.1–100 Hz and offline with a low pass at 30 Hz using a zero-phase-shift digital filter. Eye blink was removed offline and all signals with amplitudes exceeding ±100 μV in any given epoch were discarded. Data were epoched from −100 to 500 ms after stimulus onset, corrected for baseline over the 100 ms window pre-stimulus, transformed to a common average reference and averaged for each individual in each of the four experimental conditions. Peak amplitudes were analysed at the 19 most parietal–occipital electrode sites for the N170 (P7/8, P5/6, P3/4, P1/2, Pz, PO3/4, PO5/6, PO7/8, POZ, O1/2 and Oz) and the 10 most posterior parietal–occipital for the P1 (PO3/4, PO5/6, PO7/8, POZ, O1/2 and Oz). Peak latency was measured for the P1 by locating the most positive deflection from the stimulus onset within the latency window of 80–120 ms. N170 was identified in grand average waveforms as the maximum voltage area within the latency windows of 130–190 ms. P1 Peak amplitude and latency were analysed separately by a 2 × 2 × 10 repeated measures ANOVA, with media (robot, human), emotion (happy/neutral) and electrodes (10 levels) as intrasubject factors. For the N170, an identical ANOVA was run with the factor electrode involving 19 levels. A Greenhouse–Geisser correction was applied to P-values associated with multiple degrees of freedom repeated measures. Paired t-tests were used for 2 × 2 comparisons.

RESULTS
Behavioural effects of emotion
Participants correctly discriminated 94% of the expressions as emotional or neutral, irrespective of the robotic vs human media support, $F(1,14) = 0.001$, $P = 0.97$, or emotion, $F(1,14) = 2.35$, $P = 0.15$. Mean reaction times (Figure 2) were not significantly affected by media, $F(1,14) = 3.43$, $P = 0.087$, while there was a significant effect of emotion, $F(1,14) = 10.34$, $P < 0.001$. Responses to happy displays were shorter (692 ± 187 ms) than responses to neutral displays (743 ± 180 ms) independent of media effects, $F(1,14) = 2.67$, $P = 0.52$.

Early emotion effects on P1 component
P1 amplitude (Figure 3) was higher for happy than for neutral expressions, $F(1,14) = 25.68$, $P < 0.001$. Mean amplitude was, respectively, 7.7 ± 3.36 μV for happy and 6.17 ± 3.25 μV for neutral. Separate analysis within each media showed that
P1 emotion effect was significant for human \(F(1,14) = 16.65, P < 0.001\) as well as for robots \(F(1,14) = 15.82, P < 0.001\). P1 amplitude was also affected by human vs robotic media, \(F(1,14) = 4.68, P < 0.05\), with P1 responses being greater for robotic expressions (7.54 ± 3.1 µV) than for human expressions (6.33 ± 3.6 µV). There was no interaction between emotion and media, \(F(1,14) = 2.34, P = 0.14\). P1 latency was shorter in response to human than to robotic expressions, \(F(1,14) = 9.018, P < 0.01\). The P1 peak culminated parieto–occipitally at 97 ± 2.2 ms in response to human expressions, and at 101 ± 2.5 ms in response to robotic expressions. P1 latency did not vary with emotion, \(F(1,14) = 2.29, P = 0.15\), nor did it interact with media, \(F(1,14) = 0.07, P = 0.93\).

Media effect on N170 component

Media affected N170 latency and amplitude (Figure 4), with N170 latency being earlier for human than for robotic media, \(F(1,14) = 13.81, P < .01\). N170 culminated temporo–occipitally at 154 ± 3.5 ms for human expressions and at 162 ± 3.8 ms for robotic expressions.

There was a significant media effect on N170 amplitude, \(F(1,14) = 56.33, P < 0.0001\). N170 was higher for human expressions (7.1 ± 3.7 µV) than for robotic ones (4.26 ± 3.5 µV).

There was no effect of emotion on N170 latency, \(F(1,14) = 0.09, P = 0.75\), or amplitude, \(F(1,14) = 1.51, P = 0.225\), and no interaction between expression and media for latency, \(F(1,14) = 1.15, P = 0.45\), or amplitude, \(F(1,14) = 1.59, P = 0.23\).

An electrode effect on N170 latency, \(F(18,252) = 3.73, P < 0.001\), \(\varepsilon = 0.006\), and amplitude, \(F(18,252) = 4.95\),
DISCUSSION
We compared ERP and behavioural responses to emotion elicited by robotic and human displays. The temporal dynamics of early brain electrical responses showed two successive visual components: an occipital P1 component amplitude that was modulated by emotion in response to robotic stimuli as well as to human stimuli and a temporo-occipital N170 component differentiating brain responses to robotic vs human stimuli.

These results are congruent with previous findings concerning human emotional expression. An enhancement of the P1 for emotional content has been shown in emotional detection tasks (Pizzagalli et al., 1999; Fichtenholtz et al., 2007) and also in studies exploring the impact of emotion on attentional orientation (Pourtois et al., 2005; Brosch et al., 2008). Such findings suggest that P1 enhancement may reflect an increased attention towards motivationally relevant stimuli during early brain processing.

Within the perspective of natural selective attention, perceptual encoding is proposed to be in part directed by underlying motivational systems of approach and avoidance (Lang et al., 1997; Schupp et al., 2006): Certain kinds of stimuli trigger selective attention due to their biological meaning. Emotional cues would guide visual attention and receive privileged processing that would in turn enhance the saliency of emotional stimuli (Adolphs, 2004; Rey et al., 2010). This hypothesis is supported by EEG and fMRI findings of higher perceptual activations in visual regions for emotional stimuli compared to neutral ones, when using human faces or face representations such as schematic faces (Vuilleumier et al., 2001; Eger et al., 2003; Britton et al., 2008). In this context, recent findings point out that the P1 would reflect a more general form of early attentional orienting based on other stimulus features such as salience (Taylor, 2002; Briggs and Martin, 2008). For example, P1 amplitude is enhanced in response to images of upright faces compared to inverted faces (Taylor, 2002).

This interpretation goes well along the fact that at the level of perceptual encoding affective properties may be confounded with physical properties of the stimulus (Keil et al., 2007). The initial sensory response in the visual cortex is sensitive to simple features associated with emotionality (Pourtois et al., 2005; Keil et al., 2007). The fact that simple emotional features are likely to modulate P1 component may explain why happy robotic patterns elicited a higher P1 than neutral robotic ones. Indeed, the difference between the neutral and happy robotic displays was the lip pattern.

The detection of happy faces is especially dependent on the mouth region. Without the mouth region, detection times to happy faces are longer (Calvo et al., 2008) and the lower section of the face including the mouth is sufficient to accurately recognize happiness (Calder et al., 2000). The mouth region provides the most important diagnostic feature for the recognition or identification of happy facial expressions (Adolphs, 2002; Leppanen and Hietanen, 2007; Schyns et al., 2007). Our findings suggest that an upward lip pattern is not only sufficient for a face to be classified as happy but also sufficient to pick up and process selectively the emotional feature among a mixed combination of emotional and non-emotional components, at the perceptual level of information processing as soon as 90 ms poststimulus. In our robots, the upward lip pattern may be the salient element rather than the metallic piece beneath the mouth or the cables suspended behind the eye.

Behavioural results are congruent with such a facilitated detection of happy stimuli. Happy stimuli elicited quicker reaction times than neutral stimuli, for robot as well as for human stimuli. Such facilitation is in line with the literature...
of behavioural studies using human emotional faces as stimuli. Concerning positive emotional stimuli in particular, studies using dot-probe tasks have found greater attentiveness for happy faces (Brosch et al., 2008; Unkelbach et al., 2008). Happy faces including schematic ones are identified more accurately and more quickly than any other faces (Leppanen and Hietanen, 2004; Palermo and Coltheart, 2004; Juth et al., 2005). The absence of a faster detection of happy faces in a few studies (Ohman et al., 2001; Juth et al., 2005; Schubo et al., 2006) accounted for a saliency effect, as measured by Calvo et al. (2008).

An increased saliency in the mouth region of happy robots may explain the faster categorization of happy robots as emotional. Stimulus categorization can actually be quickened by salient features that would be detected without requiring a complete visual analysis of the stimulus (Fabre-Thorpe et al., 2001), leading to shortened reaction times to detect target cues. Indeed, rapid recognition of happy faces may be achieved by using the smile as a shortcut for categorizing facial expressions (Leppanen and Hietanen, 2007).

Detecting emotionally significant stimuli in the environment might be an obligatory task of the organism. Emotional detection does apparently not habituate as a function of passive exposure, and exists in passive situations as well as categorization tasks (Schupp et al., 2006; Bradley, 2009). Early posterior cortical activity in the P1 temporal window has been shown to be enhanced by emotional cues in tasks where emotional cues were either target (instructed attention) or non-target cues (motivated attention). For example, Schupp et al. (2007) found an early emotional occipital effect while some picture categories served as target category as well as they were passively displayed in separate runs.

Cortical indices of selective attention due to implicit emotional and explicit task-defined significance do not differ at the level of perceptual encoding. Same conclusion comes from fMRI studies: as presumed by Bradley et al. (2003), the attentional engagement prompted by motivationally relevant stimuli involve the same neural circuits as those that modulate attentional resource allocation more broadly, e.g. by task or instructional requirements. To summarize, P1 amplitude may be enhanced by both voluntary allocation of attentional resources and the automatic capture of motivationally relevant stimuli. Our findings show that such perceptual enhancement for emotional stimuli extends to complex patterns mixing emotional features and typical components of mechanical devices.

Our robotic results for P1 and N170 are consistent with differences between human and non-human processing such as those evidenced in Moser’s et al. (2007) fMRI study. Emotional computerized avatar faces elicited an amygdala activation altogether with less activity in several visual regions selective of face processing, compared to human faces (Moser et al., 2007). N170 reflects high-level visual processing of face, namely the structural encoding (Bentin et al., 1996). This component is typically higher in response to faces, including schematic ones (Sagiv and Bentin, 2001), than to any other category of objects, such as animals or cars (Carmel and Bentin, 2002; George et al., 1996). Accordingly, robotic pictures elicited a smaller and delayed N170 compared to human pictures. As far as the face configuration of a stimulus accounts for the N170 amplitude, a smaller N170 is coherent with the properties of our robotic displays: they had no contour, no nose, no chin, no cheeks, and the emotional features, though prototypical ones, were made of or combined with metallic pieces and cables (Nadel et al., 2006).

Physical characteristics of the non-humanoid designs may also explain why robotic stimuli elicited a higher and delayed P1 compared to human stimuli. P1 differences between robot and human stimuli may involve low-level features. P1 amplitude has indeed also been found to be sensitive to low-level features including complexity (Bradley et al., 2007).

As early as it is to human expressions, the human visual system is sensitive to expressive displays of complex non-humanoid mechanical set-ups that are not visually encoded strictly like faces, as the N170 shows. Thus, our data do not support the hypothesis that humans are predisposed to react emotionally to facial stimuli by biologically given affect programs (Dimberg et al., 1997). It is, however, compatible with another evolutionary perspective proposing that our modern brain still obeys to ‘ancestral priorities’ (New et al., 2007). It is argued that human attention evolved to reliably develop certain category-specific selection criteria including some semantic categories based on prioritization. Emotional patterns should be part of those semantic categories producing an attention bias, insofar as they are major informative signals for survival and for social relationships. Once selected, such patterns might have come to be progressively disembodied from faces and become a signal of their own. A key function of emotion is to communicate simplified but high impact information (Arbib and Fellous, 2004). The P1 response to non-humanoid robots suggests that high impact information can be drawn from mixed displays where emotional invariant has to be selected among other salient physical features.

Two main consequences of such a prioritization deserve emphasis. First, it could be of interest for roboticists to have in mind that human-like entities might not be the unique way to develop social robots: if our brain responds similarly to non anthropomorphic emotional signals, is there a need to design expressive agents that share our morphology? Are anthropomorphic signals of emotion a necessary ingredient for Robot based therapy of emotional impairments?

Secondly, our results can lead to speculate about the evolution of the modern brain regarding emotion. Following the Darwinian thesis (Darwin, 1872), a cognitive evolutionary perspective describes the continued heavy use of facial expressions by modern humans even while possessing a tremendously powerful language, (as) a vestige of the early adaptation of
Homo (Donald, 1991). Instead, this vestige may be thought of as an easy vehicle for a flexible modern brain capable to process ‘undressed emotion’.

**FUNDING**

This work was funded by a EU Program Feelix Growing (IST-045169).

**REFERENCES**


