Assessment of human response to robot facial expressions through visual evoked potentials

Richard Craig, Ravi Vaidyanathan, Christopher James and Chris Melhuish

Abstract—The focus of this work is to investigate and quantify the ability of a humanoid ‘hybrid face’ robot to effectively convey emotion to a human observer by mapping their physiological (EEG) response to perceived emotional information. Specifically, we examine the event related response during two implicit emotion recognition experiments to determine the modulation of the face-specific N170 brain response component to robot facial expressions. EEG recordings were taken from a range of test subjects observing the BERT2 robot cycle through a range of facial emotions in each emotion recognition experiment. Results from both experiments demonstrate that the stimuli evoke the N170 component and that digital facial expressions with high correlations can be discriminated. Emotional expressions evoke a larger response relative to neutral stimuli, with negative evoking an increased amplitude and latency to positive emotions, and demonstrate that the response to robot facial expressions evoke similar brain activity to that of a human emotions. This study is the first of its nature to investigate and quantify the human physiological response to digital facial expressions as conveyed in real-time by a humanoid robot.

1. INTRODUCTION

The ability to correctly make, distinguish and recognize facial expressions is a central aspect in effective social interaction; it is now well-recognized in the robotic community that possession of similar capacity will enable robots to more effectively interact with human beings. In contrast to industrial robot automation or most Human-Computer Interaction (HCI), close Human Robot Interaction (HRI) is typified by the shared manipulation of objects or even direct contact (e.g. moving an infirm person). Pipe recently stated “to be accepted, (robots) will have to be somehow emotionally compatible with us in their human-robot interactions”, and asserted a human’s correctly placed trust in the robot is imperative [1]. For a robot to interact effectively with humans, it must make the human believe that it has beliefs, desires, life-like intentions; these interactions must be consistent to be able to exploit our natural human tendencies to anthropomorphise and respond socially to these cues from the robot [4][5]. The capacity for a robot to convey recognized facial expressions akin to human emotions has been a very popular topic in current robotic research.

Evolutionary views of human emotions have categorized them as adaptations for the rapid production and recognition of emotional cues; for example, fearful faces have been shown to elicit a rapidly-processed involuntary orienting of spatial attention towards their location [2]. Ekman and Friesen described and tested the assumption of “universality” in human interaction, asserting six primary emotions; happy, surprise, fear, anger, disgust and sadness [3]. Ekman’s facial action coding system (FACS) [6], postulates that all human expressions can be viewed as a combination of the basic expressions of happiness, sadness, anger, fear, disgust, and surprise. Depending on the configuration or the shape of ‘Action Units’ a 3-dimensional ‘affect space’ can be used to mathematically represent facial expression using axes; Arousal (i.e. high, neutral, or low), Valence (i.e. positive, neutral, or negative), and Stance (i.e. open, neutral, or closed) [8].

A. Neurological Basis for Facial Recognition

Psychological investigations into recognition of facial expressions have been complemented by neurological experiments utilising electroencephalography (EEG), magnetoencephalography (MEG), position electrode topography (PET) and functional magnetic resonance imaging (fMRI). Neurological findings suggest that recognizing emotion and facial expressions draws on multiple strategies through combinations of task specific neuron sources [13] from different brain regions such as the fusiform gyrus and temporal sulcus. The left and right hemispheres of the brain are differentially related to the processing of emotions with the perception and expression of facial emotion lateralized to a greater extent in the right hemisphere. The right hemisphere is superior for processing negative emotions, with the amygdala being linked to recognition of fearful and sad faces, the cingulate sulcus for happy and the orbital frontal recognising angry facial expressions.

Event Related Potentials (ERP) are described by the positive or negative voltage deflection occurring after the
onset of a specific stimulus. The face-specific N170 ERP component, the ERP response to visual facial stimulus, is generated by a structural encoding process sources at both early and later stage processing of facial configuration and analysis [13]. These analyses relate to the whole configuration of the stimulus rather than either head detection or emotional analysis of internal features in isolation. It has been shown that N170 activation is delayed when presented with human heads with no internal face structure and human faces with no external structure for context [15]. Recent studies have investigated different face categories (upright, inverted, eyeless/eyes-only, internal face/external head images) along with other stimuli such as hand cars and houses. The activation of face-selective neurons by full or part faces demonstrate a highly specific pattern of N170 adaptation for combinations of stimulus sensitive neurons, with contributions from eye-selective and orientation components.

Studies into facial identity (e.g. [12]) provide a number of perceptual paradigms such as stimulus orientation, feature displacement, distinctive effects and image negation, these techniques have helped our understanding how identity is coded. The most consistent result is the importance of configuration in face recognition. Configural information is a term used to describe the interrelationship between facial features such as the shape and position of the mouth in relation to the shape and position of the eyes / brow. Carey and Diamond [27] referred to two forms of configural information which they identified as first order (inter feature spatial information) and second order (featural information) relational properties.

Scalp-recorded, event-related potentials (ERPs) vary with the presented stimuli. Studies have found effects of emotional stimuli on the Late Positive Potentials (LLP) found with the 330 – 420 ms range at fronto-central electrode sites. The late components of the ERP would show larger positive potentials for both pleasant and unpleasant pictures, compared with neutral stimuli [26].

B. The BERT2 Hybrid face

The second generation Bristol Eulomotion Robotic Torso (BERT2) robot, currently being developed at the Bristol Robotics Laboratory (BRL) as part of the Cooperative Human Robot Interaction Systems (CHRIS) project, is a platform designed to facilitate the study of interactions between humans and social robots by through facial expressions for affective exchanges. BERT2 is a hybrid face robot, as it combines a digital face with a permanent exterior structure. It is designed to provide the flexibility of a digital countenance with some of the benefits of a fully actuated face. The aim is to evoke human-like reaction and trust of such robots, without the complexity in control or actuation of a full facial robot. The BERT2 digital facial expressions are implemented by a combination of basis facial expressions: happiness, sadness, anger, fear, surprise, tiredness, sternness, and disgust (Fig. 1) [7].

C. Human Physiological Response to Robotic Emotion

Until very recently there have been no investigations into the human physiological response to robot facial expressions and perception of intended emotion. The image stimuli conventionally used in brain studies have taken from everyday life (human faces, cars, houses); Dubal expanded this set to include images of robot faces displaying expressions [11]. This pioneering work was designed to investigate the human physiological response to a humanoid robot as a stimulus, and to evaluate the use of robots as test beds for future experiments to investigate human social cognition.

D. Synopsis of Investigation

The focus of this work is to investigate the ability of the BERT2 hybrid face to effectively convey emotion to a human observer by mapping the physiological response to perceived emotional information through modulation of one or more ERP components and the affect of presenting stimuli from a humanoid robot with increased contextual information (internal and external features) compared to via a computer monitor (internal features only). EEG recordings were taken from a range of test subjects observing the BERT2 robot cycle through a range of facial emotions. The facial expressions used by the digital face in this experiment are similar to those designed by Bazo [7], with an increase in the pupil dilation. EEG recordings were contrasted to existing results of responses to human facial expressions to quantify the capacity of the robot to convey emotion in a
manner similar to a human being. Section II of the paper outlines the materials and methods used for the experiments conducted in this work, including a summary of EEG protocols. Section III details the experimental results, while Section IV presents a discussion of the results and conclusions based on the work.

II. MATERIAL AND METHODS

A. Participants

Twelve healthy adult volunteers (three male and nine female) aged between 22 and 58 years (mean 33.67; SD 13.12) were included in the study. Their vision was normal or corrected-to-normal (4 subjects) and they reported no history of neurological disease.

With a mean 15 years of education, only three subjects had previous experience in research participation. The majority of subjects had little contact with robotics or ICT technology. All twelve participated in each experiment with lasted approximately two hours in total. Subjects were provided with an information sheet and asked to sign an informed consent form, which outlined the experiment along with their rights as participants.

B. Stimulus

The subject independent variable was emotion, stimuli were computer-generated facial expressions depicting the eight emotions shown in Figure 1, and a ‘Neutral’ expression Stimulus with a high affect space correlation coefficient was used to reduce perception bias. In addition to the six basic emotions suggested by Ekman, expressions for Stern and Tired were included as a pair with high correlation (r= 0.92) but low emotional valence. Subjects were asked to simply observe presented stimuli and were naive as to the specific research questions; this was an implicit emotion experiment.

C. Experimental Setup

Participants were seated comfortably within an isolated room, either 1m away from a monitor or robot and trained to maintain fixation on stimulus presented within a 95mm by 120mm area (Vertical visual angle 6.8º). A screen resolution of 640x480 was used for both the computer monitor and robot head. All participants were instructed to sit as quietly as possible throughout. The subject was sat facing the monitor or robot and asked to simply observe the stimuli, they were naive to the research questions; this was an implicit emotion task. After completing a P300 ERP testing spell to ensure suitable electrode placement and detection of the stimulus, two series of experiments were conducted to determine human response to structured facial expressions and the effect of presentation from with (Robot) or without (Monitor) context.

The protocol was to present five blocks of stimuli with a two minute break after each block. A block contained five series by the nine facial expressions with the order of presentation randomized by a Latin square for every subject and stimulus source. Using a computer monitor presented the internal features only, while using a humanoid robot provided additional configuration information.

Screen Stimulus. The digital expression when presented on To assist with attention, a small white fixation cross was displayed for a random time period up to 2s to stop habituation. Subjects were presented a stimulus for 1s with a blank inter-stimulus interval of 1s. The digital facial expressions were measured to ensure a consistent stimulus between the screen and robot.

Robot Stimulus. To investigate the human response to the intended emotion depicted from the digital countenance, the expression was presented via the BERT2 robot. This experiment varied the context of the facial expression by providing additional configural information.

Emotion Recognition. A post-task experiment was conducted to determine the recognition accuracy of presented stimuli. Facial expressions where presented for four seconds only, after which the participant was asked to choose a corresponding emotion (a forced choice paradigm). ‘Static’ presented the same fixed emotional expression as the previous tasks. ‘Animation’ presented the neutral expression before a transition to an emotional expression. ‘Realism’ presented the emotion expression including eye saccades and random blinks. ‘Both’ combined the animation transition and realism effects once emotion expression was presented.

Fig 2 shows a photograph of an experimental test session with a test subject (with their face blotted out) in front of the robot.

D. ERP Recording

Continuous Electroencephalogram (EEG) was recorded using a 16 channel 24bit g.tec USBAmp. The ear-referenced signals were sampled at 256Hz and an online bandpass-filter (0.10-30Hz) applied. Recordings were made using Ag-AgCl electrodes mounted according to the International 10-20 electrode placement[8][9] standard located; three prefrontal (Fp1,2,3), three fronto-central(F3,4,2), three centro-parietal(C3,4,2), two temporo-parietal(T7,8), five parietal(P3,4,7,8,2) and one occipital(O2). Electrode impedance was below 20 kΩ, used the left earlobe as a reference and Fpz as the common ground.

E. Data Analysis

Offline the EEG data was segmented into epochs to include a pre-stimulus of 100ms and continuing 800ms after
the stimulus onset. The mean of the 100ms pre-stimulus served as the baseline for normalization. Any epoch with amplitude above ±100μV was excluded from further analysis. A zero-phase shift digital low pass filter from 0.1-20 Hz was applied in addition to online system filtering. Electrodes Fp1 and Fp2 were used as rejection channels for EoG/EMG noise artifacts. Artifacts were rejected and devoid of humanity, had high correlation ratios to reduce any perceptual bias.

A distinctive N170 component with its typical lateral occipital-temporal distribution could be reliably identified in the grand-mean response epochs from all subjects, for all emotions.

III. RESULTS

Our interest was to investigate the ability of the digital facial expressions to effectively convey emotion to a human observer, the physiological response to perceived difference in emotional content and affect on the structural encoding of each expression, and the affect of increasing configural information by presenting stimuli from the computer screen and BERT2 humanoid robot. The digital faces, unfamiliar and devoid of humanity, had high correlation ratios to reduce any perceptual bias.

A distinctive N170 component with its typical lateral occipital-temporal distribution could be reliably identified in the grand-mean response epochs from all subjects, for all emotions.

The grand mean results per emotion location where used as dependent variables for analysis. Analysis of Variance (ANOVA) was used to determine the statistical significance of the facial expressions and the effect of context presentation. Emotional facial expressions evoked a larger ERP response than neutral, with the facial expression for Angry evoking greater amplitudes consistently between the two sources (Angry_Monitor ($F_{(1,259)}=3.03$, $p=0.083$), Angry_robot ($F_{(1,259)}=4.43$, $p=0.036$). Analysis revealed differences in the latency and amplitude of the N170 response between emotions (Fig.3) but these did not reach significance.

There is a modulation in the amplitude and latency in the physiological response to presented facial expression from the robot. A mean trend delay of 7ms was noted across presentation sources. ANOVA confirmed that physiological responses to the facial expression stimuli across the two presentation sources is statistically different ($F_{(1,17)}=11.73$, $p<0.01$). EEG topography of neutral and angry facial N170 responses shows the affect on results between sources. The Angry expressions evoked a larger N170 ERP response than neutral, but this relationship is muted in the results from the robot (fig.5).

The Late Positive Potentials for the emotional facial expressions presented by the computer follow the trend of an increase in amplitude for positive and negative expressions, though the results from the robot appear to return to the baseline with no difference between neutral or emotion expressions (fig.6).

A post-stage facial expression experiment was conducted to determine the recognition accuracy of presented stimuli (Table 1). It was conducted after the main experiments to ensure the participants were naive to the research questions and avoid any predisposition or psychological bias. The artificial facial expressions had been recognised by a previous trial of twelve [7] and an additional trial of ten...
participants with higher recognition accuracy. These trials comprised of students and a lower mean age and reduced standard deviation. The ‘Static’ presented the same fixed emotional expression as the previous tasks. ‘Animation’ presented the neutral expression before a transition to an emotional expression over a two second interpolation. ‘Realism’ presented the digital facial expression but included random eye saccades and blinking. ‘Both’ combined the animation transition and realism effects once emotion expression was presented.

IV. DISCUSSION

This study is the first of its nature to investigate and quantify the human physiological response to digital facial expressions from a humanoid robot. We examined the ability of the digital facial expressions to effectively convey emotion to a human observer and used EEG to record event related potentials to determine the perception of artificial emotion and the variance due to the context presentation of stimuli from a computer monitor and the BERT2 humanoid robot. Care was taken to ensure the participants were naive to the research questions as psychology provides an extensive literature regarding how context, manipulated by an emotionally-laden task, can bias information processing in a subsequent task on emotional faces[22].

A distinctive N170 component could be reliably identified in the grand-mean response epochs from all subjects, for all emotions, from both sources. The facial expressions used as the stimuli have high correlation but low information bandwidth, but are able to modulate the ERP response through the perceived face diagram of the digital facial expression. Our results are consistent with previous studies into the human response to conveyed emotion that both pleasant and unpleasant expressions evoked a larger N170 ERP response than neutral and that this difference activity is located on the right parietal brain area (Fig.5). These results are in line with other trials conducted with human images, though our results for the robot facial expressions are modulated with a lower amplitude and increased latency, congruent with a similar study [11]. The latency of the N170 face specific ERP increases along with a muted amplitude as the presented stimuli deviates from an ‘ideal human face’. This trend demonstrated by distorted or rotated human faces with or without contextual information [15][17] was extended by Dubal who found that robot expressions are encoded as early as human faces but evoke a later and muted response. This work confirms the trend by extending human physiological responses to facial expressions from physical robots.

Consistent with the results from studies, both pleasant or unpleasant expressions evoked a larger N170 response ERP than neutral and that this difference activity is located on the right lateral occipital-temporal area. Positive emotions were evoked earlier than negative, which follows current literature. Human perception of robot emotion is subject to increased latency and muted amplitude relative to an ideal human face indicating that a number of neuron clusters associated with internal features and head detection and may be engaged requiring additional time to acquire configural information. The human face has more than forty-four muscles that produce the numerous subtle expressions.

It has been demonstrated that structural faces devoid of human context, presented at the centre of the subject field of view, elicit a N170 physiological response that is modulated by expressions from the digital face. Emotional faces evoke stronger responses compared to the baseline neutral expression with the negative emotion (anger) evoking the larger. An average trend delay of ~7ms occurred in
physiological responses between the two sources of computer monitor and BERT2 hybrid face. It is speculated that this might be due to the increase in configural information or a delay as attention is refocused to the hybrid face. While a fixation cross was presented on the monitor, subjective comments from participants noted that focusing attention was easier and more natural for the hybrid face. It was noted that the human focus was ‘drawn automatically’ to the correct location as the moulded representation of ears helped to set the configuration parameters. Subject responses to the interaction with the BERT2 face were positive with people quickly accepting and becoming familiar with the digital facial expressions. One participant commented that like a human disfigurement, the focus of attention moves beyond physical characteristics.

While the physiological differences between the two sources may be put down to the perceptual salience of stimuli, a significant modulation of the P100 component would be expected along with a similar perceptual response to facial expression if luminance was a significant variable. Stimuli from the robot include additional contextual information (head size, shape, colour, configuration with internal features, etc) that provide similar perception stimuli, but evoke additional neuron clusters that allow particular internal neurological decision processes to reject the artificial face from later processing leading to reduced late positive potentials.

The post experiment recognition rates were lower than those found by Bazo[7], but this was predicted as the participants were drawn from a sample with a greater mean and age range. Subjective comments from the non-technical sample set demonstrated that overall they found it difficult to recognise some of the more subtle intended emotion from the digital facial expressions.

Studies into human preference or ‘likeability’ of robot faces and animation have highlighted the ‘uncanny valley’; a sudden negative drop in a linear trend relationship from non-human/distinctly artificial faces towards optimal human faces. This subjective measure has been supported by numerous studies, though with little investigation into the underlying cognitive process. Future work will seek to investigate the neurological response to an artificial faces capable of representing a range of appearances across the uncanny valley, from simple structural face diagrams (without context) to a highly rendered ‘optimum human’ stimulus. Further work into robot social cognition will investigate the quality and effectiveness of non verbal social cues in human robot interactions. There is a tremendous potential for mutually beneficial future research in this area between psychology, neuroscience and robotics. Reciprocal influences between social cognitive neuroscience and humanoid robotics promise a better understanding of social interactions that will ultimately lead to increasing the social acceptance of future robotic companions.

V. ACKNOWLEDGMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant EP/F018692/X/1 and by the European Commission under the Robotics and Cognitive Systems, ICT Project CHRIS (FP7-215805) and the authors gratefully acknowledge this support.

References