Human behavioural discrimination of human, chimpanzee and macaque affective vocalisations is reflected by the neural response in the superior temporal sulcus

Thomas Fritz[a,d], Karsten Mueller[a], Anika Guha[a], Andre Gouws[b], Liat Levita[c], Timothy J. Andrews[b,*,1], Katie E. Slocombe[b,1]

[a] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Science, Leipzig, Germany
[b] Department of Psychology, University of York, UK
[c] Department of Psychology, University of Sheffield, UK
[d] Institute for Psychoacoustics and Electronic Music (IPEM), University of Ghent, Belgium

ABSTRACT

Accurate perception of the emotional content of vocalisations is essential for successful social communication and interaction. However, it is not clear whether our ability to perceive emotional cues from vocal signals is specific to human signals, or can be applied to other species' vocalisations. Here, we address this issue by evaluating the perception and neural response to affective vocalisations from different primate species (humans, chimpanzees and macaques). We found that the ability of human participants to discriminate emotional valence varied as a function of phylogenetic distance between species. Participants were most accurate at discriminating the emotional valence of human vocalisations, followed by chimpanzee vocalisations. They were, however, unable to accurately discriminate the valence of macaque vocalisations. Next, we used fMRI to compare human brain responses to human, chimpanzee and macaque vocalisations. We found that regions in the superior temporal lobe that are closely associated with the perception of complex auditory signals, showed a graded response to affective vocalisations from different species with the largest response to human vocalisations, an intermediate response to chimpanzees, and the smallest response to macaques. Together, these results suggest that neural correlates of differences in the perception of different primate affective vocalisations are found in auditory regions of the human brain and correspond to the phylogenetic distances between the species.

1. Introduction

The ability of humans to convey their emotional state, or interpret the emotional state of others, is dependent on a range of complex social cues (Coulson, 2004; Sauter and Scott, 2007; Ekman et al., 1980). The ability to process emotional expression is thought to have evolved because of the advantages associated with understanding an organism's state and behavioural intentions both within and between species (Darwin, 1872). Vocalisations represent one key channel through which emotional state can be communicated. Cross-cultural studies in humans have shown that the recognition of basic emotions through non-verbal vocalisations is universal (Scherer et al., 2001; Sauter et al., 2010). Universality across cultures has also been found for the recognition of certain acoustic emotional expressions in both speech (Banse and Scherer, 1996) and music (Fritz et al., 2009). It is not clear, however,

whether the human ability to perceive emotional cues from vocal stimuli is species specific, or if it allows for accurate perception of emotional cues from other species.

Cross-species understanding of emotional signals may be possible due to the similarity of the acoustical signals used to express emotion across species. For example, positive and negative vocalisations in animals and birds are characterized by particular spectral and temporal structures; pure-toned, high-frequency calls are associated with fear, submission or affiliation, while harsh, low-frequency sounds tend to be aggressive or threatening (Morton, 1977). Indeed dog barks have been found to conform to these rules and humans, regardless of their experience of dogs can use these regularities to identify the emotional content of barks (Pongrácz et al., 2006) and the likely eliciting context (Pongrácz et al., 2005). Similarly, humans can accurately judge the pleasantness and urgency of cat purrs recorded in food solicitiation and

---

non-solicitiation contexts, independent of cat ownership status (McComb et al., 2009). Humans have also had some success in categorizing the context associated with vocal production in macaques (Linnankoski et al., 1994). However, when presented with a wide range of macaque and cat vocalisations, humans had difficulty determining whether these vocalisations derived from pleasant or unpleasant contexts (Belin et al., 2008a, b). This suggests that there may be limitations in the ability of humans to interpret the affective content of vocalisations of other animals.

The brain regions underpinning human responses to the emotional and communicative signals of other species are not well understood. Buccino et al. (2004) found that silent human speech movements and, to a lesser extent, macaque lip-smacking activated frontal and motor areas, whereas silent dog barking was not associated with any frontal activation of human cortex. Thus, it is possible that a different level of processing and understanding is associated with signals that activate a motor 'simulation' of the signal (only possible for signals similar to those in the receiver's repertoire). Belin et al. (2008a, b) found an interesting dissociation between human participants' failure to correctly rate the valence of cat and macaque vocalisations, and the differential responses of the orbitofrontal cortex to positive and negative vocal stimuli from all tested species (humans, cats and macaques).

Although previous studies have examined human behavioural and neural responses to affective vocalisations from a number of different species (e.g. macaques, dogs, cats), the evolutionary relationship between humans and these other species is highly varied and human ability to accurately interpret the affective content of signals from cats and dogs may have changed over the long period of domestication with these species. For example, it has been suggested that the structure of food solicitation vocalisations in cats are adapted to exploit human sensitivity to infant cries (McComb et al., 2009). In order to disentangle whether human ability to perceive emotion from vocal signals is species-specific or dependent on a phylogenetically shared system, comparisons between closely and distantly related non-domesticated species are required. Our study sought to address this issue by testing whether the degree to which humans can perceive the affective content of non-verbal vocalisations of other primates depends on the phylogenetic distance between species. We measured both the behavioural and neural response of humans to emotional vocalisations produced by different primate species: human, chimpanzee and macaque.

Chimpanzees are our closest relatives, with a common ancestor estimated at around 6 million years ago, whereas a common ancestor with macaques has been estimated at around 25 million years ago (Rhesus Macaque Genome Sequencing and Analysis Consortium, 2007). We sought to determine whether the degree of phylogenetic distance to other primates might play a role in human ability to process their emotional vocalisations. First, we compared human ability to behaviourally discriminate positive and negative vocalisations produced by humans, chimpanzees and macaques. Our prediction was that human ability to discriminate between differently valenced vocalisations should reflect phylogenetic distance between the species. Next, we compared the human neural responses to vocalisations produced by humans, chimpanzees and macaques. We predicted that regions involved in the auditory processing of emotional vocalisations as well as more amodal regions involved in processing emotional signals should show a graded response to vocalisations from different species, reflecting the phylogenetic distance between the species.

2. Methods

2.1. Participants

All data were collected at the University of York. Nineteen adults (all right-handed, fifteen females; mean age 25.9) participated in both the behavioural and fMRI experiments. One participant had to be excluded from the study because of movement artefacts during the fMRI task. Participants had normal hearing, no history of neurological or psychiatric conditions, and had no experience working with or studying non-human primates. All participants were recruited using the Psychology Experiment Booking System at the University of York and gave written informed consent. Ethical approval was obtained from the Department of Psychology and York Neuroimaging Centre Ethics Committees in accordance with the Declaration of Helsinki.

2.2. Stimuli

Auditory stimuli consisted of 54 vocalisations from affective contexts: 18 chimpanzee, 18 macaque, 18 human. For all species, stimuli consisted of 9 positively- and 9 negatively-valenced vocalisations. The human vocalisations were non-linguistic sounds with positive (laughs, pleasure) and negative (cries, fearful screams) valence, which were taken from the Montreal Affective Voices data set (Belin et al., 2008a, b). The chimpanzee and macaque vocalisations were classified into positive and negative valence based on the affective context of the recording.

The chimpanzee sounds were recorded in the field in the Budongo Forest (Uganda) and from captive chimpanzees at the Wolfgang Koehler Primate Research Centre in Leipzig (Germany) by Katie Slocombe. The chimpanzee positive vocalisations consisted of rough grunts given during feeding on high quality food. Negative vocalisations were screams given by victims of directed aggression and whimpers given by juveniles when separated from their mothers. The rhesus macaque vocalisations were acquired from semi-free range monkeys on Cayo Santiago and provided by Harvard University. The rhesus macaque positive vocalisations consisted of harmonic arches given whilst feeding on high quality food and gurneys given during affiliative interactions. Negative vocalisations were gekkers given in agonistic interactions and screams given by victims of aggression. Each stimulus consisted of between one and seven vocalisations. In order to assess if there were basic acoustic differences between different categories of stimuli (species; valence), acoustical analyses were performed on each stimulus using Raven Pro 1.3 (http://www.birds.cornell.edu/brp/raven/RavenVersions.html) and mean values for each stimulus were then calculated. For each call, the following measures were obtained: Mean Amplitude/Root Mean Squared (RMS) volume and mean peak frequency in Hz (this measure was determined by using spectral slices from the middle of each call). In addition, the duration of the stimulus was measured in seconds (from start of first call to end of last call).

In order to assess the arousal value of each stimulus, we obtained explicit ratings of the stimuli and implicit physiological responses to them from two new sets of participants who had not heard these sounds before. For the explicit ratings we asked 10 adult participants (5 females; mean age = 27.1 years) with normal hearing and no experience with non-human primates to rate each stimulus on perceived arousal (1–8 Likert scale from extremely negative (1) to extremely positive (8)). After hearing each sound through headphones, participants used a response sheet to provide their rating. For the autonomic measure of arousal we measured galvanic skin response in 15 adult participants (4 females, mean age = 22.8 years). These participants undertook the same valence rating experiment described below, whilst concurrently having their autonomic response to each stimulus measured. AMP-36 psychophysiological monitoring system (Biopac, Santa Barbara, CA) together with the AcqKnowledge software (Version 4.1, Biopac), was used to monitor the skin conductance response as it varied with the eccrine sweat gland activity. The computer running AcqKnowledge and the computer running E-prime 1.2 (Psychology Software Tools, Inc, Pittsburgh, PA) were interfaced allowing generation of digital TTL timestamps for each stimulus on the Biopac channel recording, so that stimuli presentations during the study were co-registered with Skin Conductance Response (SRC) record. The SCR was sampled at 200 Hz using disposable electrodermal gel electrodes (Biopac model EL507) attached to the distal phalanx of the pointer and middle fingers of the