Article:

Fritz, T., Mueller, K., Guha, A. et al. (4 more authors) (2018) Human behavioural discrimination of human, chimpanzee and macaque affective vocalisations is reflected by the neural response in the superior temporal sulcus. Neuropsychologia, 111. pp. 145-150.

https://doi.org/10.1016/j.neuropsychologia.2018.01.026

HUMAN BEHAVIOURAL DISCRIMINATION OF HUMAN, CHIMPANZEE AND MACAQUE AFFECTIVE VOCALISATIONS IS REFLECTED BY THE NEURAL RESPONSE IN THE SUPERIOR TEMPORAL SULCUS

Thomas Fritz [1,4], Karsten Mueller[1], Anika Guha[1], Andre Gouws[2], Liat Levita[3],

Timothy J. Andrews*[2] & Katie E. Slocombe*[2]

[1] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Science, Leipzig, Germany

[2] Department of Psychology, University of York, UK

[3] Department of Psychology, University of Sheffield, UK

4 **Institute for Psychoacoustics and Electronic Music (IPEM), University of Ghent, Belgium**

***Joint senior author (equal contributions)**

Corresponding author: Timothy Andrews (tim.andrews@york.ac.uk)

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.

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 & 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, Molnár, & Miklósi, 2006) and the likely eliciting context (Pongrácz, Molnár, Miklósi, & Csányi, 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, Laakso, Aulanko & Leinonen, 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., 2008). 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. (2008) 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.

METHODS

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.

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, Fillion-Bilodeau, & Gosselin, 2008). 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 non-dominant hand. Participants were asked to wash their hands with water and dry them gently before the electrodes were attached. For analysis, a smoothing baseline removal and a low-pass filter (1 Hz) were used. SCR was defined as an increase of more than 0.02 microsiemens of the skin conductance level, occurring between 1-5 s after presentation of each stimulus. For each stimulus, we calculated the proportion of 15 participants who registered a SCR. Four stimuli used in the main study (chimpanzee food calls (positive)) were not included in this experiment and so average values from the rest of the chimpanzee food call stimuli were used for these four stimuli.

Behavioural Experiment:

Six stimulus conditions were employed during behavioural assessment of human ability to discriminate emotional valence across species: human-positive, human-negative, chimpanzee-positive, chimpanzee-negative, macaque-positive, macaque-negative. Participants were asked to rate each stimulus on perceived affective valence (from extremely negative to extremely positive). After hearing each sound participants were presented with a visual analogue scale and asked to rate each sound by pressing a number between one (very negative) and eight (very positive) on the keyboard. All stimuli were presented binaurally through headphones in a random order using E-PRIME v1.2.

fMRI

Stimulus conditions used in the fMRI experiment were identical to those used in the behavioural experiment. Stimuli from all conditions were presented in an optimized pseudo-random order for event-related fMRI designs, determined by OPTSEQ2 (http://surfer.nmr.mgh.harvard.edu/optseq). Auditory stimuli were delivered to participants via piezoelectric headphones (Magnetic Resonance Systems Ltd.) with Presentation v0.71 (http://www.neurobs.com). In order to maintain participants' attention to the acoustic properties of the stimuli, participants were tasked with pressing a button upon presentation of a target amplitude-matched car-horn sound (car_miata.wav; www.partnersinrhyme.com), which was randomly distributed 12 times throughout the entire duration of the experiment (20 minutes).

Scanning was performed on a 3T system (GE Healthcare Signa HD) at the York Neuroimaging Centre. In order to ensure maximal brain coverage, 38 contiguous para-axial slices (Resolution 2.25x2.25x3 mm) were acquired per TR using a gradient-echo EPI pulse sequence (TR=3 s, TE=33 ms, flip angle=90°, FOV=288mm, matrix= 128x128). A total of 423 volumes were acquired in each functional run (the first three volumes were discarded to allow for T1 saturation). After the functional scanning, T1-weighted anatomical images were obtained for each participant (1x1x1 mm resolution, FOV=290mm).

Univariate analyses of the fMRI data were performed with FEAT (http://www.fmrib.ox.ac.uk/fsl). In all scans the initial 9 s of data were removed to reduce the effects of magnetic stimulation. Motion correction (MCFLIRT, FSL49) was applied followed by temporal high-pass filtering (Gaussian-weighted least-squared straight line fittings, sigma=15 s). Spatial smoothing (Gaussian) was applied at 5 mm FWHM. Parameter estimates were generated for each condition by regressing the hemodynamic response of each voxel against a box-car regressor convolved with a single-gamma HRF. The frequency, arousal (behavioural) and arousal (GSR) measures were also added as regressors. A ‘species’ contrast was run to determine if there were any regions that showed progressive response to vocalizations from different species (human [1] > chimpanzee [0] > macaque[-1]). This enabled us to identify voxels that showed a bigger response to human vocalizations, an intermediate response to chimp vocalizations and a smaller response to macaque vocalizations. Next, individual participant data were entered into higher-level group analyses using a mixed-effects design (FLAME, FSL). Functional data were first co-registered to an in-plane FLAIR anatomical image then to a high-resolution T1-anatomical image, and finally onto the standard MNI brain (ICBM152). Z-statistic (Gaussianised T) statistic images were generated using an uncorrected threshold of $p<0.001$ (Worsley et al, 1992).

RESULTS

Properties of the stimuli

Acoustical analysis

Statistical comparisons between acoustical parameters of the different categories of vocalisations were performed using two-way ANOVAs with species (human/chimp/macaque) and valence (positive/negative) as factors. No significant interaction or main effects were found for mean RMS amplitude, indicating stimuli were effectively equalized in amplitude. The duration of sounds varied across species (*F*(2,24) = 10.19, *p* = .010, partial eta squared = .46). Bonferroni-corrected post-hoc tests revealed that macaque sounds were significantly shorter than chimpanzee (*p* = .001) and human sounds (p = .004), but that human and chimpanzee sounds were similar in length (p = 1.000; Table 1). No effect of valence (*F*(1,24) = 0.22, *p* = .645, partial eta squared = .01) and no interaction between valence and species was found (*F*(2,24) = 0.09, *p* = .918, partial eta squared = .01).

There was a main effect of species on the frequency of sounds (*F*(2,24) = 7.70, *p* = .003, partial eta squared = 0.391). Bonferroni-corrected post-hoc tests showed that macaque sounds were significantly higher in frequency than human sounds (*p* = .002; Table 1), but macaque and chimpanzee (*p* = .170) and chimpanzee and human (*p* = .200) sounds were similar in frequency. Overall, negative sounds had significantly higher peak frequencies than positive sounds (*F*(1,24) = 11.64, *p* = .002, partial eta squared = .327; Table 1) and a significant interaction between valence and species was found (*F*(2,24) = 6.62, *p* = .005, partial eta squared = .355). Independent samples t-tests at the level of each species revealed that whilst positive and negative sounds had similar peak frequencies for human (*t* (10.55) = 0.61, *p* = .557) and chimpanzee (*t*(16) = 0.39, *p* = .705) sounds, macaque negative sounds had significantly higher peak frequencies than positive sounds (*t*(16) = 4.21, *p* = .001; Bonferroni-corrected alpha level= .017).

Arousal ratings

A 3 x 2 repeated measures ANOVA with within-subjects factors of species (human, chimpanzee, macaque) and valence of vocalisation (positive, negative) was performed on the behavioural arousal rating data. There were significant main effects of species (*F*(2, 18) = 8.19, *p* = .003), and valence (*F*(1, 9) = 8.87, *p* = .015), which was qualified by a significant interaction between valence and species *(F*(2, 18) = 12.74, *p* < .001). At the level of positively-valenced stimuli, species had a significant effect on arousal ratings (F(2,18) = 6.16, p = .009), with bonferroni corrected pair-wise comparisons revealing that human stimuli were rated as higher arousal (mean =4.54 , sd = 1.61) than chimpanzee stimuli (mean = 3.47, sd =1.05 ; p = .040), there was a trend for human stimuli to be rated as higher arousal than macaque stimuli to (mean = 3.44, sd = 1.39, p = .098), whilst chimpanzee and macaque vocalisations were not rated as significantly different (p = 1.000). At the level of negatively-valenced stimuli, species had a significant effect on arousal ratings (F(2,18) = 10.55, p = .001), with bonferroni corrected pair-wise comparisons revealing that human stimuli were rated as higher arousal (mean =5,12 , sd = 1.73) than macaque stimuli (mean = 3.29, sd =1.61 ; p = .012), but not chimpanzee stimuli (mean= 4.50, sd = 1.26; p = .031). Human and chimpanzee vocalisations were not rated as significantly differently (p = .326).

We then considered the effect of valence at each level of the factor species. Post-hoc pairwise comparisons (with bonferroni corrected alpha level of .017) found that arousal ratings of positive human vocalisations (*M*=5.12, *SD*=1.73) were not significantly different from negative human vocalisations (*M* = 4.54, *SD* = 1.61, *t*(9) = 2.58, *p* = .030). In contrast, arousal ratings for positive chimpanzee vocalisations (*M* = 3.47, *SD* = 1.05) were significantly lower than ratings for negative chimpanzee vocalisations (*M* = 4.50, *SD* = 1.26, *t*(9) = 4.19, *p* =.002). However, ratings for positive macaque vocalisations (*M* = 3.29, *SD* = 1.61) did not differ significantly from negative macaque vocalisations (*M* = 3.44, *SD* = 1.39, *t*(9) = 0.98, *p* = .351).

Galvanic Skin Response

A 3 x 2 between subjects ANOVA with factors of species (human, chimpanzee, macaque) and valence of vocalisation (positive, negative) was performed on the GSR data. As the dependent variable for this analysis was the proportion of participants (N = 15) who showed an above threshold GSR response to each vocalisation, this analysis was run at the level of stimulus (N = 9 stimuli within each category (e.g. human positive)). There were no significant main effects of species (*F*(2,48) =0.98 , *p* =.383), or valence(*F*(1,48) = 0.04, *p* = .843), and no significant interaction between valence and species *(F*(2,48) = 0.82, *p* = .445).

Behavioural data

Figure 1 shows the mean perceived affective valence across participants. A 3 x 2 repeated measures ANOVA with within-subjects factors of species (human, chimpanzee, macaque) and valence of vocalisation (positive, negative) was performed on the post-scan valence rating data. Although there was no main effect of species (*F*(2,36) = 0.68, *p* = .512), this analysis revealed a significant main effect of valence(*F*(1, 18) = 196.09, *p* < .001), which was qualified by a significant interaction between valence and species *(F*(2, 36) = 65.31, *p* < .001). Post hoc pairwise comparisons (with bonferroni corrected alpha level of .017) found that valence ratings for positive human vocalisations (*M*=7.04, *SD*=0.56) were significantly higher than ratings for negative human vocalisations (*M* = 2.19, *SD* = 0.66, *t*(18) = -21.11, *p* < .001; Figure 1). Similarly, ratings for positive chimpanzee vocalisations (*M* = 5.07, *SD* = 0.81) were significantly higher than ratings for negative chimpanzee vocalisations (*M* = 3.74, *SD* = 0.92, *t* = -5.14, *p* < .001). However, ratings for positive macaque vocalisations (*M* = 4.35, *SD* = 0.75) did not differ significantly from negative macaque vocalisations (*M* = 4.62, *SD* = 0.84, *t*(18) = 1.29, *p* = .215).

fMRI data

A whole-brain analysis showed activity for the 'species' contrast (human>chimpanzee>macaque) in regions of the superior temporal gyrus (Figure 2 & Table 2). This activity was evident in both posterior (MNI - LH: -62, -18, -2; RH: 48, -36, 2) and anterior (MNI - LH: -48, -4, -32; RH: 52, -14, -16) regions of the superior temporal sulcus (STS) as well as the right planum temporale (MNI 58, -12, 4). This contrast also revealed activity in the left inferior frontal gyrus (MNI -50, 22, 0), medial frontal regions (MNI - LH: -6, 54, 34; RH: 10, 48, 38) and orbitofrontral regions (MNI -32, 28, -22). Finally, there was significant activation in the right hippocampal region (MNI 30, -28, -8). To determine whether the fMRI results might be explained by differences in the auditory stimuli, the peak frequency and arousal values (behavioural and GSR) of the stimuli were entered as additional covariates in the first-level analysis for all subjects. We did not find any regions that showed activity that covaried with these regressors.

DISCUSSION

The aim of this study was to investigate the behavioural and neural differences in the human ability to discriminate emotional valence from vocalisations produced by different primate species. We found that human participants were best at discriminating positive and negative emotional content from human vocalisations. To a lesser extent, they were able to discriminate positive and negative signals from chimpanzee vocalisations. However, they were not able to discriminate positive from negative macaque vocalisations. The neural correlates of these behavioural differences were found in regions of the superior temporal gyrus corresponding to differences in the response to affective vocalisations from different species (human>chimpanzee>macaque). These results mirror the phylogenetic distance between these species and show that the difference in the ability to discriminate emotional content in vocalisations is associated with a relatively early stage of auditory processing.

The behavioural results show that participants were able to correctly perceive the valence of human and chimpanzee vocalisations corresponding to the emotional context in which the vocalisations were recorded. In line with the findings of Belin et al. (2008), the participants were unable to accurately judge the valence of macaque vocalisations. The higher discrimination of human affective vocalizations presumably reflects our experience and familiarity with human sounds (Zajonc, 1968; Bornstein & D'Agostino, 1992). However, familiarity and experience cannot explain why unfamiliar chimpanzee vocalisations were correctly attributed to the emotional context in which the stimuli were recorded but the equally unfamiliar macaque vocalisations were not. Instead, the findings suggest an effect of phylogenetic distance within the primate order, modulating the capability to assess emotional context of primate vocalisations.

The idea of a phylogenetic influence on our ability to identify the emotional context of another species' animal vocalisation must relate to underlying neural commonalities with the species phylogenetically closer to us. This may be in terms of either (1) vocalisation production – we may have a repertoire of emotional acoustic expression closer to the chimpanzee than the macaque, which is why it may be easier to infer a situational context of the emotional vocalisation – or (2) vocalisation perception – we may be prone to emotional contagion in the acoustic domain through a mechanism shared with other great apes (Fritz & Koelsch, 2013; Fecteau et al., 2004; Heffner & Heffner, 1984) which, for example, may also give rise to emotional contagion in great ape laughter, chimpanzee pant hoots and human music.

Our results cannot be explained by low level acoustic properties of the stimuli, as the amplitudes of positive and negative sounds were effectively equalized and peak frequency was added as covariate, ensuring that any differences in stimuli peak frequencies were controlled for in the analysis. We also included measures of arousal determined behaviourally or using GSR as covariates. Activity in the STG did not covary with these measures. This does not mean that these regions are not sensitive to arousal. Previous studies have shown that high arousal vocalizations activate the STG more than low arousal vocalizations (Wiethoff et al. 2008, Ethofer et al., 2012). As all our stimuli were selected to show arousal, we may not have had sufficient variation in arousal values to detect to detect the differences previously reported.

The human STG has been shown to be intimately involved in the perception of human vocalisations (Belin et al., 2000; Wiethoff et al. 2008, Ethofer et al., 2012) and a number of studies that have shown that the human STG responds more to human vocalisations compared non-vocal sounds (Fecteau et al., 2004). The human STG also responds more to human vocalisations compared to the vocalisations of other species, suggesting a within-species bias (Fecteau et al., 2004; Joly et al., 2012). Equally, the role of the STG in the processing of species-specific vocalisation is demonstrated by lesion studies showing that damage to the STG in macaque monkeys impairs the ability to discriminate between species-specific vocalisations (Heffner & Heffner, 1984). A greater specificity of the STG responses for own-species vocalisations is consistent with studies suggesting that activation of the anatomically proximate Superior Temporal Sulcus (STS) may increase as a function of the communicative significance of stimuli (Redcay, 2008; Schultz et al., 2012). Consistent with this idea, the same stimuli elicited greater neural responses in the STS when listeners interpreted them as speech rather than non-speech (Möttönen et al., 2006; Dehaene-Lambertz et al., 2005).

We also found activity in the left inferior frontal gyrus to be differentially active to vocalizations from different species. This is consistent with a study by Buccino and colleagues (2004) in which they compared the neural response to visualizing silent human, monkey and dog vocalizations. They found silent vocalizations in humans and monkeys activated the left inferior frontal gyrus, but the observation of barking did not produce any activation in this region. These results suggest that the interpretation of actions made by other individuals may require activation of the motor system. If the movement is not compatible with the observer, this may affect processing of the action. Finally, we found activity in medial- and orbito-frontal regions associated with processing the valence of stimuli. This is consistent with Belin and colleagues (2008) who showed higher selectivity to human emotional vocalizations compared to macaque and cat vocalizations. Activity in these regions presumably reflects a higher level representation of the emotional properties of the stimulus.

In summary, both behavioural assessments and neural responses in the STG corresponded to the distance of the phylogenetic relationship of the vocalising species with humans. It is plausible that the mediation of neural and behavioural responses with phylogenetic distance between species corresponds to a greater acoustical similarity in emotional vocalisation production, so that vocalisations from a species phylogenetically closer to us may be perceived as more communicatively significant. For future research, it would be of interest to extend the current paradigm to include vocalisations of more primate species. It would be especially meaningful to investigate human behavioural and brain physiological measures in response to vocalisations from all great ape species.

Acknowledgments

We would like to thank Pascal Belin for his helpful advice on the experiment, and Marc Hauser for providing the rhesus macaque vocalisations used in this experiment. We are indebted to Anne Schneider and Po Ling Chen, who collected most of these data and contributed valuable arguments to the paper.

REFERENCES:

Ashburner J, Friston KJ. (2005). Unified segmentation. Neuroimage, 26(3):839-51.

Banse, R., & Scherer, K. (1996). Acoustic Profiles in Vocal Emotional Expression. Journal of Personality and Social Psychology, 70(3), 614-636.

Belin, P., Zatorre RJ, Lafaille P, Ahad P, Pike B. (2000). Voice-Selective areas in human auditory cortex. Nature, 403(6767), 309-312.

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. Behavior research methods, 40(2), 531-539.

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalisations. Proceedings of the Royal Society, 275(1634), 473-481.

Bornstein, R. F., D'Agostino, P. R. (1992). Stimulus Recognition and the Mere Exposure Effect. Journal of Personality and Social Psychology, 63(4): 545-52.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., ... & Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. Journal of cognitive neuroscience, 16(1), 114-126.

Coulson M (2004) Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. J Nonverbal Behav 28:117–139.

Darwin, C. (1872). The Expression of the Emotions in Man and Animals Chicago: University of Chicago Press.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. Neuroimage, 24, 21–33.

Ekman P, Friesen WV, Ancoli S (1980) Facial signs of emotional experience. J Pers Soc Psychol 39:1125–1134.

Fecteau, S., Armony, J. L., Joanette, Y., & Belin, P. (2004). Is Voice Processing Species-specific in Human Auditory Cortex? An FMRI Study." NeuroImage 23: 840-848.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R.. (1998). Event-related fMRI: characterizing differential responses. Neuroimage, 7(1):30-40.

Fritz, T., Jentschke, S., Gosselin, N., Sammler, D., Peretz, I., Turner, R., et al. (2009). Universal recognition of three basic emotions in music. Current Biology, 19(7), 573-576.

Fritz, T., & Koelsch, S. (2013). Acoustically mediated emotional contagion as an across-species homology underlying music processing. Evolution of Emotional Communication. Altenmueller, E., and Zimmerman, E., eds. Oxford University Press.

Gläscher J. (2009). Visualization of group inference data in functional neuroimaging. Neuroinformatics 7(1):73-82. doi: 10.1007/s12021-008-9042-x.

Heffner, H. E., Heffner, R. S. (1984). Temporal Lobe Lesions and Perception of Species-Specific Vocalisations by Macaques. Science, 226(4670), 75-76.

Joly, O., Pallier, C., Ramus, F., Pressnitzer, D., Vanduffel, W., & Orban, G. A. (2012). Processing of vocalisations in humans and monkeys: a comparative fMRI study. Neuroimage, 62(3), 1376-1389.

Koelsch, S., Fritz, T., von Cramon, D. Y., Müller, K., & Friederici, A. D. (2006). Investigating emotion with music: An fMRI study. Hum Brain Mapp, 27(3), 239-250.

Linnankoski, I., Laakso, M., Aulanko, R., & Leinonen, L. (1994). Recognition of emotions in macaque vocalisations by children and adults. Language and Communication, 14, 183-192.

McComb, K., Taylor, A. M., Wilson, C., & Charlton, B. D. (2009). The cry embedded within the purr. Current Biology, 19(13), R507-R508.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. American Naturalist, 855-869.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage, 30, 563–569.

Mueller, K., Mildner, T., Fritz, T., Lepsien, J., Schwarzbauer, C., Schroeter, M. L., & Moeller, H. E. (2011). Investigating brain response to music: a comparison of different fMRI acquisition schemes. Neuroimage, 54(1), 337-343.

Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., Baldoli, C., Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. Proceedings of the National Academy of Sciences, 107(10), 4758-4763.

Pongrácz, P., Molnár, C., & Miklósi, A. (2006). Acoustic parameters of dog barks carry emotional information for humans. Applied Animal Behaviour Science, 100(3-4), 228-240.

Pongrácz, P., Molnár, C., Miklósi, Á., & Csányi, V. (2005). Human listeners are able to classify dog (Canis familiaris) barks recorded in different situations. Journal of Comparative Psychology, 119, 136-144.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. Neuroscience & Biobehavioral Reviews, 32, 123–142.

Sauter DA, Scott SK (2007) More than one kind of happiness: Can we recognize vocal expressions of different positive states? Motiv Emot 31:192–199.

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross- cultural recognition of basic emotions through nonver- bal emotional vocalisations. Proceedings of the National Academy of Sciences, USA, 107, 2408–2412.

Scherer KR, Banse R, Wallbott HG (2001) Emotion inferences from vocal expression correlate across languages and cultures. J Cross Cult Psychol 32:76–92.

Schultz, S., Vouloumanos, A., Pelphrey, K 2012. The STS differentiates communicative and noncommunicative auditory signals. Journal of Cognitive Neuroscience, 24(5), 1224-1232.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. Journal of Personality and Social Psychology Monograph Supplement, 9(2), Pt. 2: 1-27.

Table 1 Mean amplitude*, duration (s) and peak frequency (Hz) of the vocalisations

	Amplitude (*sd*)	Duration (*sd*)	Frequency (*sd*)
Human positive	3422.86 (118.01)	1.44 (0.30)	921.38 (257.34)
Human negative	3521.41 (187.54)	1.49 (1.02)	994.72 (509.22)
Chimpanzee positive	3490.80 (142.83)	1.54 (0.17)	1293.17 (420.56)
Chimpanzee negative	3495.25 (105.13)	1.54 (0.25)	1519.14 (1038.93)
Macaque positive	3551.68 (75.08)	0.78 (0.38)	1123.07 (699.43)
Macaque negative	3501.98 (113.46)	0.92 (0.39)	2624.32 (807.94)

* the root mean square amplitude has dimensionless sample units.

Table 2 Regions of the brain that were more activated by human>chimpanzee>macaque emotional vocalizations.

		Voxels	Z_{max}	X	Y	Z
posterior STS	L	208	4.4	-62	-18	-2
	R	118	4.5	48	-36	2
anterior STS	L	6	3.2	-48	-4	-32
	R	131	3.9	52	-14	-16
IFG	L	48	3.7	-50	22	0
Medial prefrontal	L	16	3.4	-6	54	34
	R	36	3.7	10	48	38
Planum Temporale	R	21	3.6	58	-12	4
Hippocampus	R	5	3.4	30	-28	-8
Orbitofrontal	L	4	3.3	-32	28	-22

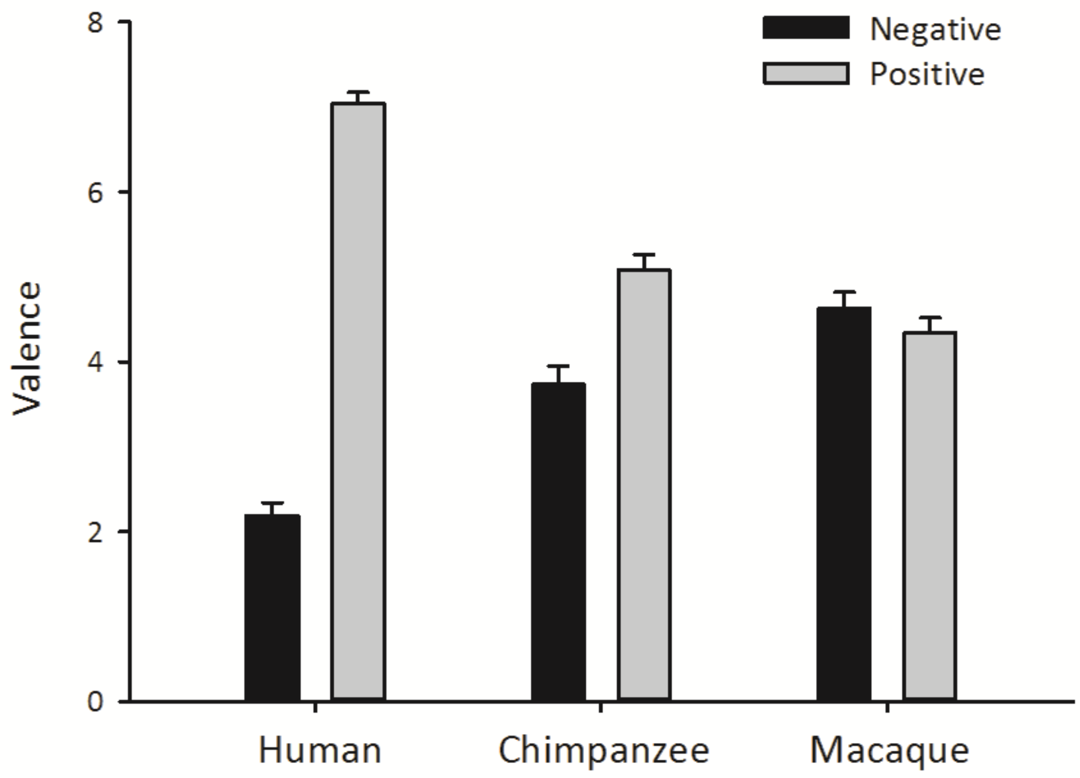

Figure 1 Mean post-scan behavioural valence ratings for human, chimpanzee and macaque (monkey) calls (N = 19 participants). Participants rated emotionally-valenced vocalisations ranging from 1 (very negative) to 8 (very positive). Error bars represent SEM.

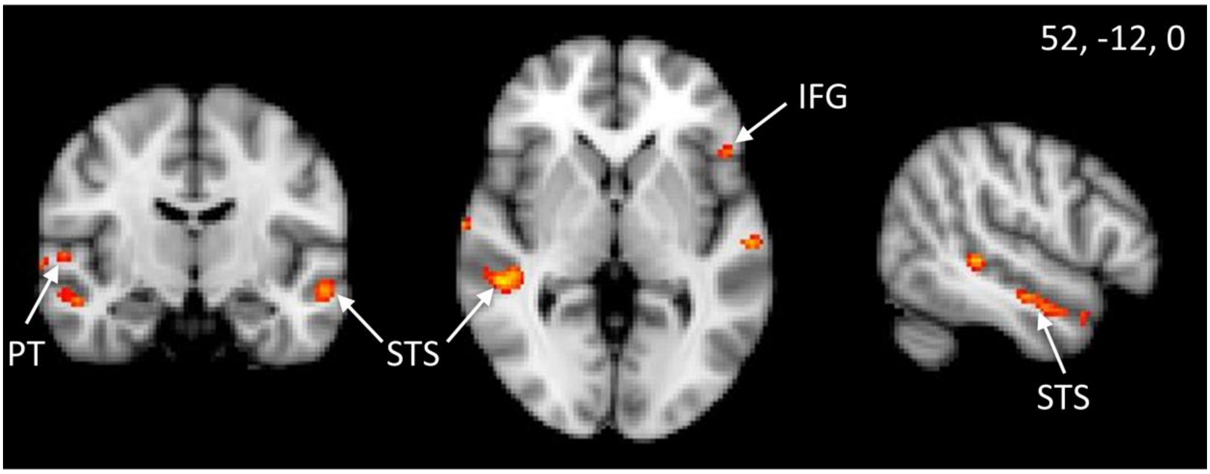

Figure 2 Coronal, axial and sagittal, slices showing significant brain activity differences for human > chimpanzee > macaque emotional vocalizations. The responses are based on the general linear model with peak frequency and arousal (measured behaviourally and using GSR) included as covariates. In line with the phylogenetic distance, an increase in brain response was evident across species in auditory regions of the superior temporal gyrus, such as the superior temporal sulcus (STS) and the right planum temporale (PT). Other activity was evident in the left inferior frontal gyrus (IFG). Coordinates are given in MNI space and the images are shown from a radiological display convention.