doi:10.1093/scan/nsl027 SCAN (2006) 1, 242–249

The voice of emotion: an FMRI study of neural responses to angry and happy vocal expressions

Tom Johnstone,[1] Carien M. van Reekum,[1,2] Terrence R. Oakes,[1] and Richard J. Davidson[1,2]
[1]Waisman Laboratory for Brain Imaging and Behavior, and [2]Psychology Department, University of Wisconsin-Madison, USA

The human voice is one of the principal conveyers of social and affective communication. Yet relatively little is known about the neural circuitry that supports the recognition of different vocally expressed emotions. We conducted an FMRI study to examine the brain responses to vocal expressions of anger and happiness, and to test whether specific brain regions showed preferential engagement in the processing of one emotion over the other. We also tested the extent to which simultaneously presented facial expressions of the same or different emotions would enhance brain responses, and to what degree such responses depend on attention towards the vocal expression. Forty healthy individuals were scanned while listening to vocal expressions of anger or happiness, while at the same time watching congruent or discrepant facial expressions. Happy voices elicited significantly more activation than angry voices in right anterior and posterior middle temporal gyrus (MTG), left posterior MTG and right inferior frontal gyrus. Furthermore, for the left MTG region, happy voices were related to higher activation only when paired with happy faces. Activation in the left insula, left amygdala and hippocampus, and rostral anterior cingulate cortex showed an effect of selectively attending to the vocal stimuli. Our results identify a network of regions implicated in the processing of vocal emotion, and suggest a particularly salient role for vocal expressions of happiness.

Keywords: emotion; prosody; FMRI; brain; happiness

The human voice is one of the principal conveyers of social and affective communication. From the earliest stages of development, infants response to affect-laden vocal expressions from their mothers (Fernald, 1989; Fernald and Morikawa, 1993). Vocal affect remains a primary channel of emotion expression during development (Shackman and Pollak, 2005) and throughout our lives, perhaps more so now than ever, given how much of our social interaction is carried out over the phone. Despite the fact that emotional vocal expressions are as ubiquitous as facial expressions in everyday life and are recognized across cultures at rates comparable to facial expressions (Scherer and Wallbott, 1994), vocal expressions of emotion have received far less attention from psychologists and cognitive neuroscientists than facial expressions. As a consequence, far less is known about the neural circuitry that underlies the perception of emotion in the voice.

Hughlings-Jackson (1915) observed that patients with severe linguistic deficits due to left hemisphere brain damage still had the ability to communicate emotions through the voice, and suggested that the right hemisphere might subserve such functions. Early neurological evidence of a right hemisphere specialization for affective speech comprehension (Tucker *et al.*, 1977) has subsequently been supported by a number of studies in which a deficit was observed in the perception of affective prosody in right-hemisphere-damaged listeners compared with left-hemisphere-damaged patients (Bowers *et al.*, 1987; Heilman *et al.*, 1984; Peper and Irle, 1997; Ross, 1981).

To date, only a small number of imaging studies of emotional prosody have been reported. Using positron emission tomography (PET), George *et al.* (1996) reported greater right prefrontal activation during processing of the emotional prosody than during processing of the emotional propositional content of spoken sentences. Pihan *et al.* (1997) reported a right hemisphere lateralization in DC components of the scalp electroencephalography (EEG) signal for the perception of both temporal (accented syllable duration) and frequency (F0 range) mediated emotional prosody. Imaizumi *et al.* (1998), in a study using magnetoencephalography (MEG), found evidence supporting the existence of prosody-specific right hemisphere processing, but also the involvement of certain left hemisphere centers in both linguistic and prosodic processing.

Most of the aforementioned studies focused on whether a right hemisphere lateralization exists for the processing of emotional prosody. Very few studies have attempted to pinpoint more specific neural circuits underlying affective voice perception. Evidence points to the right superior temporal cortex as being particularly involved in processing suprasegmental human vocal sounds (Belin *et al.*, 2000, 2002). Mitchell *et al.* (2003) found areas of posterior middle temporal gyrus (MTG) and superior temporal sulcus (STS) that activated more when attending to affective prosody as compared with semantic content of

Received 27 August 2006; Accepted 15 September 2006

The authors thank Ron Fisher, Michael Anderle and Kalthleen Ores-Walsh for assistance in data collection. This study was supported by NIMH grants MH069315, MH67167 and P50-MH069315.

Correspondence should be addressed to Tom Johnstone, Waisman Laboratory for Brain Imaging and Behavior, 1500 Highland Ave, Madison, WI 53705, USA. E-mail: itjohnstone@wisc.edu.

spoken words. Grandjean *et al.* (2005) and Sander *et al.* (2005) have reported FMRI data that revealed a region in STS that showed greater activation in response to angry speech as compared with neutral speech, and suggest that the STS might be the vocal analogue of the fusiform face processing area. In an FMRI study of five vocal emotions, Wildgruber *et al.* (2005) identified a right hemisphere network consisting of posterior STS, and dorsolateral and orbitobasal prefrontal cortex that showed selective activation during an emotion recognition task. Differential activations for the five emotions were not observed. In a recent FMRI study, Ethofer *et al.* (2006) identified regions in the right posterior MTG and STS and bilateral inferior/middle frontal gyrus that activated more when individuals identified affective prosody than when identifying the emotional content of the spoken words. No distinction was made between responses to the different expressed emotions studied.

Indeed, whereas brain regions that show heightened responses to specific emotional facial expressions have been identified (e.g. the amygdala for fear, insula for disgust), no such specificity has been found for vocal expressions of emotion. The question remains, therefore, of whether specific neural regions are more engaged in the processing of some emotions than others. In this study, we used FMRI to examine brain activation in response to vocal expressions of anger and happiness, two emotions that were chosen due to their similar acoustic characteristics (e.g. Banse and Scherer, 1996; Scherer *et al.*, 2003) as well as their relevance and regular occurrence in day-to-day social interaction. Based on the hypothesis that affiliative social vocal signals are prevalent throughout our lives and serve a fundamental purpose in social binding, from mother–infant interaction through all stages of development to adult communication, we hypothesized that emotional expressions of happiness would preferentially engage parts of the temporal cortex and inferior frontal regions previously shown to be involved in the processing of affective prosody.

A further question concerning the perception of vocal expressions of emotion is how directed attention towards or away from the expressed emotion affects the associated neural response. Given that emotional vocal expressions are commonly perceived in combination with facial expressions, we also examined the effect of selective attention to vocal emotion expressions when they are simultaneously presented with facial emotion expressions, and conversely how brain activation differs when attention is directed to the face rather than the voice. Pourtois *et al.* (2005) demonstrated an area in left MTG that showed heightened activation when congruent vocal expressions and facial expressions of happiness or fear were simultaneously presented, as compared with when only one expressive modality was presented. We wished to further examine which brain regions involved in the processing of vocal emotion show effects of selective attention to the vocal modality, and the extent to which such effects differ between emotions. In this study, we sought to address these questions by simultaneously presenting congruent or discrepant facial expressions, and instructing individuals to make emotion judgments based either on the face or voice. We hypothesized that multimodal areas in MTG would show greater responses to congruent than to discrepant pairs of vocal and facial emotions, particularly for expressed happiness.

METHODS

Participants

Forty right-handed subjects were recruited through local newspaper and pinup advertisements. All subjects provided informed consent, and all studies were performed in accordance with the policies of the UW-Madison's Human Subjects Committee. Participants were screened by phone with an MRI-Compatibility Form, the Edinburgh Handedness Survey and a Structured Clinical Interview for DSM-IV Axis I Disorders (SCID). Prior to the actual scanning session, subjects underwent a simulated scan in a mock scanner to acclimate them to the MRI scanner environment, and to train them in the performance of the experimental task. Subjects ranged in age from 18–50 years, with number and gender balanced within each decade: 18–29: *M* (8), *F*(6); 30–39: *M* (6), *F*(6); 40–50: *M* (8), *F*(6).

Experimental task

We used event-related FMRI to examine brain activation in response to angry and happy vocal expressions, while participants concurrently viewed either emotionally congruent or discrepant facial expressions, a task similar to that previously used to examine the crossmodal processing of fear expressions (Dolan *et al.*, 2001). Facial stimuli consisted of 16 greyscale images of posed expressions of anger and 16 expressions of happiness, half of them females, taken from the Karolinska Directed Emotional Faces set (Lundqvist *et al.* 1998). Vocal stimuli consisted of short phrases (dates and numbers) lasting on average 1 s, spoken with either angry or happy prosody, taken from the Emotional Prosody Speech and Transcripts dataset (Linguistic Data Consortium, Philadelphia, PA, USA, 2002). 16 angry expressions and 16 happy expressions, half of each spoken by female actors, were used. All expressions were normalized to the same mean signal amplitude.

Stimuli were presented for 1 s with an interstimulus interval of 15 s. Each of the four different types of stimulus pairs [angry voice+angry face (AA), angry voice + happy face (AH), happy voice + angry face (HA) and happy voice+happy face (HH)] were presented 20 times each in a pseudo-random order, across two scan runs. Participants performed a two-response (angry or happy) discrimination task. Half the participants were randomly selected to make their decision on the basis of the facial expression (Attentional Focus: 'face' condition), while the other half were instructed to make their decision on the basis of the vocal expression (Attentional Focus: 'voice' condition). Apart from the

instruction to attend to either facial or vocal stimuli, all participants performed the identical task, with an identical set of stimulus pairs presented to all subjects. Participants were instructed to press one button on a two-button response pad if the attended stimulus was an angry expression, and to press the other button if the attended stimulus was a happy expression. The matching of buttons to responses was counterbalanced across subjects within each response group. Participants were instructed to respond as quickly and accurately as possible. Stimuli were presented and responses were recorded using EPrime software.

Image acquisition

Images were acquired on a GE Signa 3.0 Tesla high-speed imaging device with quadrature head coil. Anatomical scans consisted of a high-resolution 3D T1-weighted inversion recovery fast gradient echo image (inversion time = 600 ms, 256×256 in-plane resolution, 240 mm FOV, 124×1.1 mm axial slices), and a T1-weighted spin echo coplanar image with the same slice position and orientation as the functional images (256×256 in-plane resolution, 240 mm FOV, 30×4 mm sagittal slices with a 1mm gap). Functional scans were acquired using a gradient echo EPI sequence (64×64 in-plane resolution, 240 mm FOV, TR/TE/Flip = 2000 ms/30 ms/90°, 30×4 mm interleaved sagittal slices with a 1mm interslice gap; 290 3D volumes per run).

We also collected skin conductance measures in response to stimuli for a subset of participants ($N = 24$, 13 from Attentional Focus: 'voice' and 11 from Attentional Focus: 'face'). Skin conductance was collected with a Coulbourn Instruments' (Allentown, PA, USA) skin conductance coupler with 8 mm Ag-AgCl electrodes placed on the tips of the index and third finger. The electrodes were filled with an NaCl paste (49.295 grams of unibase and 50.705 g of isotonic NaCl 0.9%). The electrode leads are shielded and the signal low-pass filtered to reduce RF interference from the scanner. The skin conductance signal was then digitized and sampled at 20 Hz using a National Instruments' (Austin, TX) DAQPad 6020E.

ANALYSIS

All analyses were carried out using AFNI (Cox, 1996), unless otherwise noted. Individual subject data were slice-time corrected to correct for temporal offsets in the acquisition of slices, and motion corrected to the functional image closest in time to the acquisition of high-resolution anatomical images to optimize alignment with anatomical images. Image distortion was corrected using estimated B0 field-maps to shift image pixels along the phase encoding direction in the spatial domain (Jezzard and Balaban, 1995). Individual participant data were then analyzed using a general linear model with the response to each condition (HH, HA, AH and AA) modeled with a unit-magnitude ideal hemodynamic response function convolved with a binary stimulus train corresponding to the respective stimulus onsets and offsets. Separate regressors were used to model correct and incorrect trials. Additional regressors based on estimated motion time courses were included as motion covariates to model variance due to residual head motion (Johnstone *et al.*, 2006). Slow baseline drifts were modeled with second-order Legendre polynomials.

To examine the main effect of expressed vocal emotion, we calculated the contrast between stimulus pairs containing happy and angry vocal stimuli $(HH + HA) - (AH + AA)$. Estimated contrasts were then converted to percent signal change by dividing each contrast estimate by the baseline signal value and multiplying by 100. Percent signal change contrast estimates were normalized to the 152-brain T1-weighted Montreal Neurological Institute (MNI) template with FLIRT software (Jenkinson 2002), using a three-stage registration procedure. First, functional volumes were registered to the T1 coplanar volumes using a three degree of freedom (DOF) rigid body registration. The T1 coplanar volumes were registered to the T1 high-resolution volumes using a six DOF rigid body registration. Finally, the T1 high-resolution volumes were registered to the MNI template using a 12 DOF linear transform. The three transformations were then combined and applied with a sinc interpolation to the percent signal change images from the individual-subject general linear modeling (GLM) analysis to achieve normalization of percent signal change contrast images to the MNI template.

These estimates were statistically analyzed voxelwise with a mixed-effects GLM, with Subject as a random factor nested within the between-subjects factor Attentional Focus ('voice' *vs* 'face'). Results were corrected for multiple comparisons using a combined voxelwise and cluster-size threshold, derived by Monte Carlo simulation based upon the whole brain gray matter search volume and an estimate of the data set spatial correlation based upon the GLM residual images. Because individual participants differed in the number of correct responses they made and their response times (RT), mean percentage correct and mean RT for each participant were included as covariates in the mixed model GLM, thus controlling for individual differences in overall task performance. We also investigated attention-modulated changes to the functional connectivity between activated brain regions by calculating inter-region correlations separately for each Attentional Focus condition.

Skin conductance responses were estimated using a GLM-based deconvolution, resulting in estimated second-by-second skin conductance responses for each condition, from which the mean of the response amplitude was estimated.

RESULTS

The data from six participants were dropped due to technical problems in recording subject responses from two participants and large, uncorrectable EPI distortion in four other participants. Thus the data analyzed were from

34 participants, 18 in the Attentional Focus: 'face' condition and 16 in the 'voice' condition. For all subsequent analyses, we present data for correct response trials only. Because the performance was ~80% correct across trials and conditions, there were not enough error trials for a reliable separate analysis of error–trial data.

Task performance

RT and percentage of correct trials were submitted to a multivariate repeated measures GLM with Attentional Focus ('voice' *vs* 'face') as a between-subjects factor, and Face Emotion ('happy' *vs* 'angry') and Voice Emotion ('happy' *vs* 'angry') as within-subjects factors. Means and standard errors for the performance measures are shown in Table 1.

Attentional Focus. There was a significant effect of Attentional Focus [$F(2, 31) = 3.74$, $P = 0.035$], which was due to RT being significantly longer in the 'voice' condition than in the 'face' condition [$F(1, 32) = 4.25$, $P = 0.047$], but no significant difference in the percentage correct [$F(1, 32) = 2.19$, $P = 0.148$].

Face and Voice Emotion. There was no effect of Face Emotion on the performance measures [$F(2, 31) < 1$]; Nor was the effect of Voice Emotion on the performance measures significant [$F(2, 31) = 2.88$, $P = 0.071$].

Interaction of face and voice emotion with attentional focus. There was a significant three-way interaction effect of Attentional Focus, Face Emotion and Voice Emotion on the performance measures [$F(2, 31) = 6.95$, $P = 0.003$]. This was due to RT being significantly shorter and percentage correct significantly higher for conditions in which the Face Emotion and Voice Emotion were congruent ($P = 0.004$ for each Face Emotion comparison within angry and happy voice) than when they were discrepant (angry voice: happy vs. angry face, $P = 0.007$; happy voice: happy *vs* angry face, $P = 0.003$), but only in the Attentional Focus: 'voice' condition.

Skin conductance. Skin conductance responses showed a main effect of Attentional Focus [$F(1, 22) = 4.58$, $P = 0.044$], with responses higher in the Attentional Focus: 'face' condition than in the 'voice' condition. There was a nonsignificant interaction between Attentional Focus and Vocal Emotion [$F(1, 22) = 3.90$, $P = 0.061$], which reflected marginally higher responses to angry voices than to happy voices in the attend-to-face condition ($P = 0.061$), but no such difference in the attend-to-voice condition ($P = 0.451$).

FMRI

Main effect of Vocal Emotion. Across both Attentional Focus conditions, there was significantly more activation in trials with happy voices than in trials with angry voices within MTG (BA21) in a right anterior, $F(1, 32) = 24.40$, $P = 0.000$, and right posterior region, $F(1, 32) = 16.47$, $P = 0.000$, and a left region, $F(1, 32) = 17.10$, $P = 0.000$, as well as a region in right inferior frontal gyrus, $F(1, 32) = 15.82$, $P = 0.000$. Furthermore, for the left MTG region, there was a significant interaction between Vocal Emotion and Facial Emotion, $F(1, 32) = 6.19$, $P = 0.018$, with happy voices being related to higher activation in this region relative to angry voices, but only when paired with happy faces ($P = 0.000$), and not when the faces were angry ($P = 0.204$; see Figure. 1). Higher levels of activation for angry voices than for happy voices were seen in right fusiform gyrus, $F(1, 32) = 11.38$, $P = 0.002$, and in the supplementary motor area, $F(1, 32) = 10.10$, $P = 0.003$.

Table 1 Reaction times (RT) and percentage correct for each condition.

	Attentional Focus: 'face'		Attentional Focus: 'voice'	
Voice Face	RT (ms)	% correct	RT (ms)	% correct
AA	562 (39)	85.8 (1.7)	668 (40)	87.5 (1.9)
HA	583 (35)	88.6 (2.2)	766 (38)	78.4 (2.3)
AH	561 (53)	84.7 (2.5)	698 (56)	80.0 (2.7)
HH	582 (34)	87.8 (1.9)	587 (36)	90.0 (2.0)

Numbers in parentheses are standard errors.

Interaction of Vocal Emotion and Attentional Focus. A number of brain regions showed a significant Vocal Emotion by Attentional Focus interaction. Activation in the left inferior parietal lobule, $F(1, 32) = 25.68$, $P = 0.000$, was greater in response to happy voices than to angry voices in the Attentional Focus: 'voice' condition ($P = 0.005$), but greater for angry voices than happy voices in the 'face' condition ($P = 0.000$). A similar pattern of activation was observed in left insula, $F(1, 32) = 19.84$, $P = 0.000$, happy > angry in 'voice' condition, $P = 0.000$ and angry > happy in face condition, $P = 0.041$, as well as in left amygdala and hippocampus, $F(1, 32) = 29.78$, $P = 0.000$, happy > angry in 'voice' condition, $P = 0.000$ and angry > happy in 'face' condition, $P = 0.006$, and in rostral anterior cingulate cortex (ACC), $F(1, 32) = 13.92$, $P = 0.001$, happy > angry in 'voice' condition, $P = 0.003$ and angry > happy in 'face' condition, $P = 0.045$; see Figure. 2. In right middle occipital gyrus, $F(1, 32) = 10.89$, $P = 0.002$, activation was greater to angry voices than to happy voices in the 'voice' condition ($P = 0.004$) with no significant difference in the 'face' condition ($P = 0.13$).

None of the regions showing significant main effects of Vocal Emotion or interactions with Attentional Focus showed significant correlations with either percentage correct or RT. A summary of clusters showing either a main effect of Vocal Emotion, or an interaction of Vocal Emotion and Attentional Focus is provided in Table 2.

Correlations between brain regions. We then assessed the degree to which the brain regions showing an Attentional Focus by Vocal Emotion interaction showed correlated happy–angry vocal emotion activation across participants, as a function of Attentional Focus. Such an analysis provides a measure of the degree to which Attentional

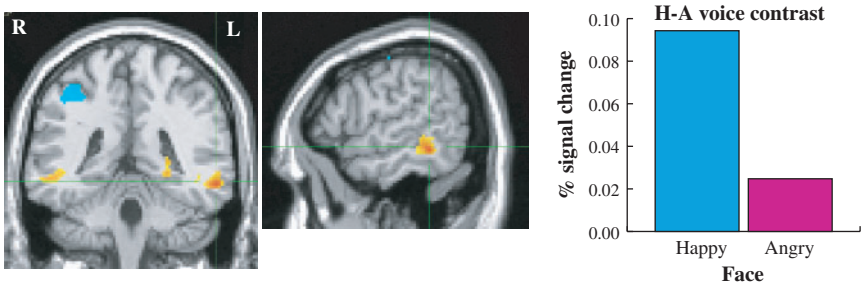

Fig. 1 Left MTG region that showed an interaction between Facial Emotion and Vocal Emotion. The happy–angry vocal emotion contrast was significantly greater when vocal expressions were accompanied by happy facial expressions than when accompanied by angry facial expressions. Images thresholded at $P < 0.05$ corrected for multiple comparisons. Cluster MNI coordinates: −58, −36, −9.

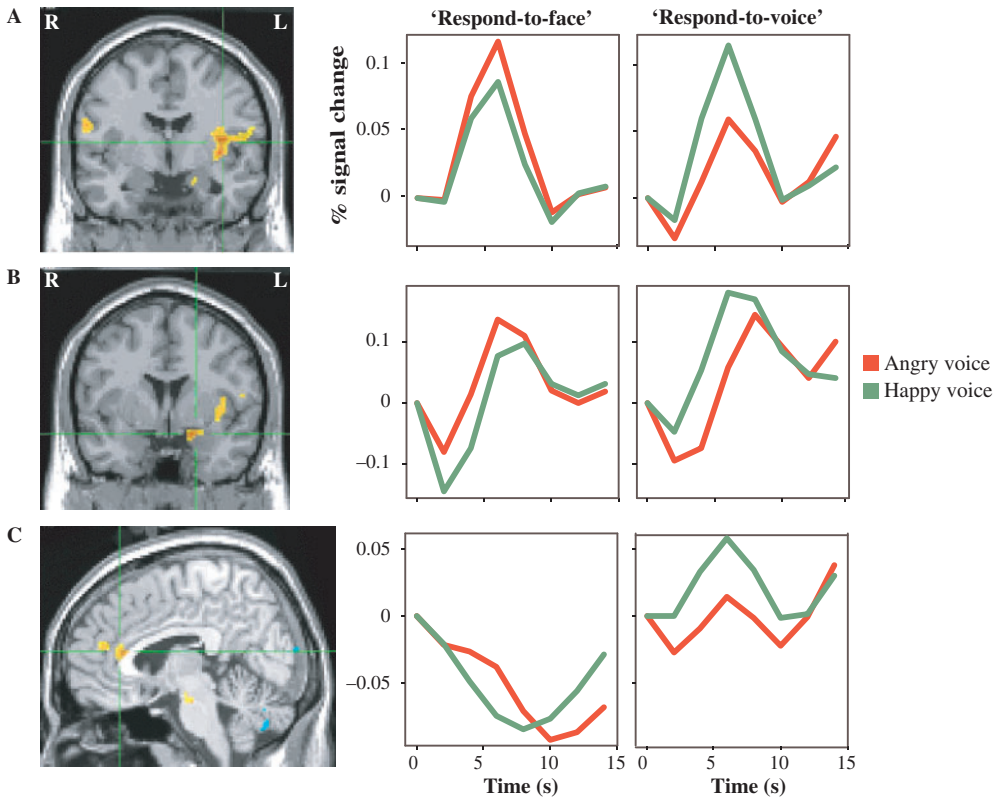

Fig. 2 Happy–angry vocal emotion contrast showed an interaction with Attentional Focus in left insula (**A**; MNI coordinates: −37, −6, 13), left amygdala (**B**; MNI coordinates: −17, −6, −17) and rostral ACC (**C**; MNI coordinates: −5, 29, 19). In all cases, activation was higher to happy voices than to angry voices in the attend-to-voice condition, but the reverse in the attend-to-face group. Images thresholded at $P < 0.05$ corrected.

Focus influenced how activated regions tended to co-activate across participants. Table 3 provides the correlations of the happy–angry Vocal Emotion contrast between the clusters that showed an Attentional Focus by Vocal Emotion interaction. In the attend-to-voice condition, left insula was positively correlated with left inferior parietal lobule, left amygdala/hippocampus and rostral ACC. Rostral ACC and left amygdala/hippocampus were positively correlated in the attend-to-voice condition. In the attend-to-face condition, rostral ACC was correlated positively with left amygdala/hippocampus, and the right visual cortex was positively correlated with the left inferior parietal lobule cluster.

DISCUSSION

In this study, we sought to identify neural regions that showed differential responses to vocal expressions of anger *vs* happiness. Based on the notion that affiliative social vocal signals are of particular salience and on previous research (Pourtois *et al.*, 2005) that showed areas of MTG to respond to happy voices, we hypothesized that emotional expressions of happiness would preferentially engage parts of the temporal cortex and inferior frontal regions previously shown to be involved in the processing of affective prosody.

We also sought to determine the extent to which explicitly attending to vocal emotional expressions *vs* attending to

Table 2 Clusters that showed either a main effect of Vocal Emotion or an interaction of Vocal Emotion with Attentional Focus (all clusters $P < 0.05$ corrected)

Brain region	BA	Hemi	Direction of effect	Size (mm^3)	*x*	*y*	*z*
H-A voice contrast							
MidTG	21	L	H > A	2624	−58	−36	−9
Anterior midTG	21	R	H > A	2464	49	−4	−24
InfFG	47	R	H > A	2320	37	40	0
SMA	6	Medial	A > H	1168	1	5	57
Fusiform gyrus	19	R	A > H	760	21	−68	−16
Posterior midTG		R	H > A	752	52	−38	−7
Interaction H-A voice by Attentional Focus							
Inferior parietal lobule	40	L	Face: A > H Voice: H > A	4840	−60	−40	31
Insula	13	L	Face: A > H Voice: H > A	4384	−37	−6	13
Amygdala/hippocampus		L	Face: A > H Voice: H > A	3416	−17	−6	−17
Middle occipital gyrus	18	R	Face: *n.s.* Voice: A > H	2168	35	−81	−2
Rostral ACC	24	Medial	Face: A > H Voice: H > A	1280	−5	29	19

MidTG, middle temporal gyrus; InfFG, inferior frontal gyrus; SMA, supplementary motor area; ACC, anterior cingulate cortex; L = left; R = right;. *n.s*, not significant.

Table 3 Correlation between brain regions showing an Attentional Focus x Voice Emotion interaction for the happy–angry Vocal Emotion contrast

		Left inferior parietal lobule	Left insula	Left amygdala hippocampus	Right middle occipital	Rostral ACC	
'attend-to-face'	Left inferior parietal lobule		**0.620** (**0.01**)	−0.043 (0.88)	0.134 (0.62)	0.356 (0.18)	'attend-to-voice'
	Left insula	0.305 (0.21)		**0.667** (**0.005**)	−0.040 (0.88)	**0.662** (**0.005**)	
	Left amygdala hippocampus	0.206 (0.41)	0.202 (0.42)		0.105 (0.70)	**0.493** (**0.05**)	
	Right middle occipital	**0.542** (**0.02**)	0.213 (0.40)	0.318 (0.20)		0.040 (0.88)	
	Rostral ACC	0.424 (0.08)	0.234 (0.35)	**0.520** (**0.03**)	0.195 (0.44)		

'attend-to-face' in left bottom part of the table—'attend-to-voice' in right top part of the table. Numbers in parentheses are *P*-values. Numbers in bold are significant at $P < 0.05$.

a different modality (faces) would impact the neural circuitry underlying the perception of happy and angry emotional expressions. Participants were presented with pairs of one vocal and one facial expression of either anger or happiness and instructed to make an angry *vs* happy decision based on either the vocal expressions or the facial expressions. Because all stimulus pairs were identical for all participants, we were able to examine differences in neural responses that were attributable to the differing attentional focus on either vocal or facial stimuli. Conversely, we were also able to identify those neural circuits involved in processing happy and angry vocal expressions that do not appear to be affected by changing one's attentional focus to one expression modality or another.

A striking finding of this study was the widespread network demonstrating greater activation to happy voices than to angry voices. Confirming our hypothesis, we observed higher activation to happy voices compared with angry voices in right anterior and posterior MTG, left MTG and right inferior frontal gyrus.

The MTG is known to be involved in the processing of complex auditory stimuli, including music, speech and emotional prosody (Ethofer *et al.*, 2006), particularly in the right hemisphere (Mitchell *et al.*, 2003). Further evidence has implicated the MTG in the processing of vocal expressions of fear and particularly happiness (Pourtois *et al.*, 2005). Although Pourtois *et al.* reported no significant difference in PET measures of regional cerebral blood flow (rCBF) for fear *vs* happy expressions, the rCBF was in fact marginally significantly greater (at $P = 0.1$ with $N = 8$ participants) for happy expressions than for fearful expressions. The results of this study coupled with those of Pourtois *et al.* suggest that happy expressions preferentially engage the middle temporal region. It is worth noting that this effect is unlikely merely due to happy voices being more arousing that angry voices, since skin conductance response amplitudes were no higher for happy vocal expressions compared with angry vocal expressions in the subset of participants for whom this measure was collected.

We also observed greater activation to happy voices than to angry voices in right inferior frontal gyrus. The inferior frontal gyrus has been associated with both more cognitive aspects of emotional judgment, as well as attaching reward value to stimuli. Consistent with current concepts of emotional speech perception (Schirmer and Kotz, 2006),

our data suggest that following acoustic differentiation in temporal cortices, information is transferred through connections to the inferior frontal regions for further elaboration and integration with cognitive and affective processes related to ongoing task planning and performance.

A further finding in this study was the combined effect of facial and vocal expressions of happiness in left MTG. Thus, activation was highest in left MTG to happy voices paired with happy faces, replicating and extending the finding of Pourtois *et al.* (2005) with happy and fearful bimodal expressions. The MTG has previously been implicated in the bimodal processing of facial and vocal information. Using event-related potentials (ERP) and source dipole modeling, Joassin *et al.* (2004) found a negative-going component in response to bimodal *vs* unimodal face and voice processing in MTG that they attributed to facial information influencing the processing of vocal information. The left MTG thus seems to be involved in the integration of cues from at least the visual and auditory modalities, and perhaps other modalities as well. This fits with current theory on the existence of a ventral auditory stream that serves to attach meaning to sound, a component of which is the posterior MTG, which combines auditory information with a wide range of information from other sensory and semantic brain regions (e.g. Hickok and Poeppel, 2004). Future research might examine to what extent activation in this region corresponds to individual ability to perceive and resolve mixed social signals, a subtle but important component of perceiving and understanding emotional expressions.

In contrast with brain regions showing a main effect of vocal emotion regardless of attentional focus, a network of brain regions including the left insula, left amygdala and hippocampus, and rostral ACC responded more to happy voices than to angry voices when attending to the voice, but showed either no difference or greater activation to angry voices than to happy voices when attending to the face. Moreover, functional connectivity between these regions was significant only for individuals attending to the voices. This result further supports the notion that vocal expressions of happiness are particularly salient social cues, engaging a network of brain regions involved in the perception and generation of emotional responses specifically when individuals attend to the voice.

The involvement of these brain regions in the processing of positively valenced expressions might seem at odds with much previous research on responses to valenced facial expressions, in which greater activation of amygdala, hippocampus and insula to negative stimuli than to positive stimuli has frequently been reported (see Adolphs, 2002). It should be noted, however, that evidence concerning the involvement of the human amygdala and insula in the processing of valenced sounds is less consistent than evidence of amygdala involvement in processing affective visual stimuli. Intact recognition of vocal expressions of fear have been observed in patients with bilateral lesions to the amygdala (Anderson and Phelps, 1998), indicating that this brain structure is possibly not as crucial for differentially processing valenced vocal expressions as it is for facial expressions. Some studies have reported reduced amygdala response to fearful vocal expressions relative to neutral expressions (Morris *et al.*, 1999). One of the primary functions of the amygdala is as a significance detector that alerts other parts of the brain to potentially salient stimuli (Davis and Whalen, 2001). In this sense, an increased response to happy expressions as observed in this study is consistent with happy vocal expressions as highly salient social signals.

Some studies have reported increased activation in insula in response to positive stimuli. Menon and Levitin (2005) demonstrated increased engagement of a network including nucleus accumbens, ventral tegmental area and the insula when individuals listened to music. The former are known to be important in reward processing, with the insula involved in generating reward-related autonomic responses. In an analysis of 66 patients with focal brain damage, Adolphs *et al.* (2002) found damage to the right insula (but not left, as found in this study) to be associated with low performance in recognizing emotional prosody. It has been suggested that one component of perceiving others' emotional expressions is the empathic generation of similar feelings in the listener, a function for which the insula is suited (Adolphs, 2001). The large response in insula when attending to happy voices in this study might reflect such an empathic response.

There are a number of limitations to the current study. Angry and happy vocal expressions were selected because both are high arousal emotions, and the sound recordings were normalized with respect to mean intensity. Nonetheless, it remains an open question as to whether the effects observed in this study reflect primarily the valence of the stimuli, the arousal of the stimuli or some other physical aspect of the stimuli. Angry and happy vocal expressions have grossly similar acoustic characteristics; both are characterized by high intensity and highly variable fundamental frequency, for example. Angry and happy vocal expressions do differ in more subtle ways, however. Future studies might use acoustic characteristics of the stimuli as covariates in the GLM analysis to control for effects due purely to the physical characteristics of the vocal expressions. Such an approach would also allow for a meaningful comparison between expressed emotions that differ markedly in their physical acoustic characteristics, such as expressions of sadness, boredom or neutral speech. More comprehensive measurements of autonomic responses to the stimuli would also help shed light on the relation between activation in different brain regions and perceptual *vs* response components.

In summary, we observed attention-independent activation to happy *vs* angry vocal expressions in a network including

inferior frontal and middle temporal cortices, and greater activation to happy *vs* angry vocal expressions in amygdala and insula regions when explicitly attending to these expressions. The results suggest that happy vocal expressions are particularly salient social signals that engage an extensive brain network, including sensory cortex, limbic and somatosensory regions, and prefrontal cortex, that underlies our ability to perceive and understand our social cohorts.

Conflict of Interest

None declared.

REFERENCES

Adolphs, R. (2002). Neural systems for recognizing emotion. *Current Opinion in Neurobiology*, *12*, 169–77.

Adolphs, R., Damasio, H., Tranel, D. (2002). Neural systems for recognition of emotional prosody: a 3-D lesion study. *Emotion*, *2*, 23–51.

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*, 231–9.

Anderson, A.K., Phelps, E.A. (1998). Intact recognition of vocal expressions of fear following bilateral lesions of the human amygdala. *Neuroreport*, *9*, 3607–13.

Banse, R., Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality & Social Psychology*, *70*, 614–36.

Belin, P., Zatorre, R.J., Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Brain Research. Cognitive Brain Research*, *13*, 17–26.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, *403*, 309–12.

Bowers, D., Coslett, H.B., Bauer, R.M., Speedie, L.J., Heilman, K.M. (1987). Comprehension of emotional prosody following unilateral hemispheric lesions: processing defect versus distraction defect. *Neuropsychologia*, *25*, 317–28.

Cox, R.W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, *29*, 162–73.

Davis, M., Whalen, P.J. (2001). The amygdala: Vigilance and emotion. *Molecular Psychiatry*, *6*, 13–34.

Dolan, R.J., Morris, J.S., de Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 10006–10.

Ethofer, T., Anders, S., Erb, M., et al. (2006). Cerebral pathways in processing of affective prosody: a dynamic causal modeling study. *NeuroImage*, *30*, 580–7.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: is the melody the message? *Child Development*, *60*, 1497–510.

Fernald, A., Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development*, *64*, 637–56.

George, M.S., Parekh, P.I., Rosinsky, N., et al. (1996). Understanding emotional prosody activates right hemisphere regions. *Archives of Neurology*, *53*, 665–70.

Grandjean, D., Sander, D., Pourtois, G., et al. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*, 145–6.

Heilman, K.M., Bowers, D., Speedie, L., Coslett, H.B. (1984). Comprehension of affective and nonaffective prosody. *Neurology*, *34*, 917–21.

Hickok, G., Poeppel, D. (2004). Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.

Hughlings-Jackson, J. (1915). On affectations of speech from diseases of the brain. *Brain*, *38*, 107–74.

Imaizumi, S., Mori, K., Kiritani, S., Hosoi, H., Tonoike, M. (1998). Task-dependent laterality for cue decoding during spoken language processing. *Neuroreport*, *9*, 899–903.

Jenkinson, M., Bannister, P., Brady, M., Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*, 825–41.

Jezzard, P., Balaban, R.S. (1995). Correction for geometric distortion. *Magnetic Resonance in Medicine*, *34*, 65–73.

Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M., Campanella, S. (2004). When audition alters vision: an event-related potential study of the cross-modal interactions between faces and voices. *Neuroscience Letters*, *369*, 132–7.

Johnstone, T., Ores Walsh, K.S., Greischar, L.L., et al. (2006). Motion correction and the use of motion covariates in multiple-subject fMRI analysis. *Human Brain Mapping*, *27*, 779–88.

Lundqvist, D., Flykt, A., Öhman, A. (1998). *The Karolinska Directed Emotional Faces,* KDEF [CD-ROM]. Stockholm: Karolinska Institute.

Menon, V., Levitin, D.J. (2005). The rewards of music listening: response and physiological connectivity of the mesolimbic system. *NeuroImage*, *28*, 175–84.

Mitchell, R.L., Elliott, R., Barry, M., Cruttenden, A., Woodruff, P.W. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia*, *41*, 1410–21.

Morris, J.S., Scott, S.K., Dolan, R.J. (1999). Saying it with feeling: neural responses to emotional vocalizations. *Neuropsychologia*, *37*, 1155–63.

Peper, M., Irle, E. (1997). Categorical and dimensional decoding of emotional intonations in patients with focal brain lesions. *Brain and Language*, *58*, 233–64.

Pihan, H., Altenmuller, E., Ackermann, H. (1997). The cortical processing of perceived emotion: a DC-potential study on affective speech prosody. *Neuroreport*, *8*, 623–7.

Pourtois, G., de Gelder, B., Bol, A., Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *41*, 49–59.

Ross, E.D. (1981). The aprosodias. functional-anatomic organization of the affective components of language in the right hemisphere. *Archives of Neurology*, *38*, 561–9.

Sander, D., Grandjean, D., Pourtois, G., et al. (2005). Emotion and attention interactions in social cognition: brain regions involved in processing anger prosody. *NeuroImage*, *28*, 848–58.

Scherer, K.R., Johnstone, T., Klasmeyer, G. (2003). Vocal expression of emotion. In: Davidson, R.J., Scherer, K.R., Goldsmith, H.H., editors. *Handbook of Affective Sciences*. New York: Oxford University Press, pp. 433–56.

Scherer, K.R., Wallbott, H.G. (1994). Evidence for universality and cultural variation of differential emotion response patterning [erratum appears in *Journal of Personality & Social Psychology* 1994, 55]. *Journal of Personality & Social Psychology*, *66*, 310–28.

Schirmer, A., Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*, 24–30.

Shackman, J.E., Pollak, S.D. (2005). Experiential influences on multimodal perception of emotion. *Child Development*, *76*, 1116–26.

Tucker, D.M., Watson, R.T., Heilman, K.M. (1977). Discrimination and evocation of affectively intoned speech in patients with right parietal disease. *Neurology*, *27*, 947–50.

Wildgruber, D., Riecker, A., Hertrich, I., et al. (2005). Identification of emotional intonation evaluated by fMRI. *NeuroImage*, *24*, 1233–41.