◆ **Human Brain Mapping 34:1796–1810 (2013)** ◆

# Emotional Expressions in Voice and Music: Same Code, Same Effect?

**Nicolas Escoffier,[1] Jidan Zhong,[2] Annett Schirmer,[1,3] and Anqi Qiu[4,5,6]***

[1]*Department of Psychology, National University of Singapore, Singapore, Singapore*
[2]*NUS Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore, Singapore*
[3]*Duke/NUS Graduate Medical School, Singapore, Singapore*
[4]*Department of Bioengineering, National University of Singapore, Singapore, Singapore*
[5]*Clinical Imaging Research Centre, National University of Singapore, Singapore, Singapore*
[6]*Singapore Institute for Clinical Sciences, the Agency for Science, Technology and Research, Singapore, Singapore*

**Abstract:** Scholars have documented similarities in the way voice and music convey emotions. By using functional magnetic resonance imaging (fMRI) we explored whether these similarities imply overlapping processing substrates. We asked participants to trace changes in either the emotion or pitch of vocalizations and music using a joystick. Compared to music, vocalizations more strongly activated superior and middle temporal cortex, cuneus, and precuneus. However, despite these differences, overlapping rather than differing regions emerged when comparing emotion with pitch tracing for music and vocalizations, respectively. Relative to pitch tracing, emotion tracing activated medial superior frontal and anterior cingulate cortex regardless of stimulus type. Additionally, we observed emotion specific effects in primary and secondary auditory cortex as well as in medial frontal cortex that were comparable for voice and music. Together these results indicate that similar mechanisms support emotional inferences from vocalizations and music and that these mechanisms tap on a general system involved in social cognition. *Hum Brain Mapp 34:1796–1810, 2013.* © 2012 Wiley Periodicals, Inc.

**Key words:** prosody; theory of mind; ACC; MPFC; fMRI

## INTRODUCTION

Many scholars proposed parallels between the emotional expressions in music and those in the human voice. For example, Jean-Jacques Rousseau held that "melody, by imitating the inflections of the voice, expresses complaints, cries of sadness or of joy, threats, and moans; all the vocal signs of the passions are within its scope" (1781/1998, p. 322). The capacity for music to express such passions was linked to its acoustic substrate, which was thought to mirror the voice. According to observations by Herbert Spencer in the 19th century (1857) as well as more recent acoustic measurements [Juslin, 2000; Scherer, 1986], music utilizes the same acoustic features for expression as does the human voice [Curtis and Bharucha, 2010; Ilie and Thompson, 2006; Juslin and Laukka, 2003]. Thus, music can imitate vocal expression of emotions.

Given this shared mode of expression, we asked whether also the perception and brain representations of emotions in music and vocalizations are similar. We

---

Additional Supporting Information may be found in the online version of the article.

*Correspondence to: Anqi Qiu, Department of Bioengineering, National University of Singapore, 9 Engineering Drive 1, Block EA #03-12, Singapore 117576, Singapore.
E-mail: bieqa@nus.edu.sg

Received for publication 25 November 2010; Revised 15 November 2011; Accepted 5 December 2011

DOI: 10.1002/hbm.22029

Published online 16 April 2012 in Wiley Online Library (wileyonlinelibrary.com).

© 2012 Wiley Periodicals, Inc.

presumed such similarity at three processing levels. First, we predicted similarity in the way vocal and musical acoustic information is represented in the brain. Second, we thought that similarity may exist in the way emotions are inferred from the acoustic representation of vocalizations and music. Finally, we speculated that voice and music compare in the way emotional meaning or content is represented in the brain. In the following, we will review research that speaks to these processing levels and justify our hypotheses for the present study.

### Perceptual Representations of Voice and Music

The question of how voice and music are represented in the brain has a long research tradition. Before the advent of neuroimaging, inferences were made primarily based on neurological case studies. These suggested a speech network in left fronto-temporal regions with right hemisphere analogues supporting speech independent vocal representations [for review see Schirmer and Kotz, 2006]. Moreover, evidence from individuals who lost their musical abilities due to brain insult suggested a role of the right hemisphere in the representation of music [e.g., Confavreux et al., 1992].

Subsequent neuroimaging research largely confirmed these observations while providing additional insights [for reviews see Schirmer and Kotz, 2006; Warren, 2008]. For example, the passive listening to a mix of speech and nonverbal human vocalizations has been contrasted with the passive listening to a mix of other sounds. This revealed activation differences in the bilateral superior temporal sulcus (STS) and adjoining gyri [Belin et al., 2000, 2002] suggesting that these regions support voice sensitive or perhaps voice specific processing mechanisms. Research that compared the processing of neutrally spoken words, pseudo-words and/or reversed words with the processing of duration and intensity matched synthesized tones or instrumental notes [Binder et al., 2000; Tervaniemi et al., 2006] supported this notion. It suggested that temporal lobe representations of the voice differ even from the representations of its acoustic approximation, music. However, as individual tones may not accurately elicit musical processing, inferences about differences between music and voice representations in the brain may be premature.

### Emotional Evaluation Mechanisms for Voice and Music

A second level of similarity in the processing of emotions from voice and music may exist for the mechanisms that support explicit emotion judgments. For vocalization, neuroimaging research again implicated frontal and temporal regions. For example, emotion judgments, when compared to linguistic or acoustic judgments, were found to activate orbital and inferior frontal gyri as well as superior temporal regions lateralized to the right hemisphere [for a review see Schirmer and Kotz, 2006; Wildgruber et al., 2006]. Additionally, some studies suggested an involvement of medial frontal regions [Bach et al., 2008; Fecteau et al., 2007; Peelen et al., 2010; Wildgruber et al., 2002].

To the best of our knowledge, there are currently no published neuroimaging studies that investigated emotion evaluation mechanisms in music by comparing the evaluation of musical emotion expression with an evaluation of other musical attributes. Thus, the jury is still out as to whether and in what way music evaluation mechanisms compare to the ones that support vocal evaluation.

### Representation of Emotional Content for Voice and Music

Finally, voice and music processing may compare in how emotional content is represented in the brain. Notably, inquiry in the representations of emotional content has proved more difficult for vocal and music stimuli than for the thoroughly studied facial emotional expressions. Activations are often surprisingly small and unreliable [Wildgruber et al., 2006]. Yet, some general trends can be inferred from studies that used salient emotional exclamations (e.g., screams, weeps) or other means to increase experimental sensitivity.

In vocal processing studies, emotional content representations have been explored by contrasting emotional and neutral vocalizations. This revealed activation differences in previously identified voice sensitive regions. Vocal emotional expressions were found to activate STS and surrounding gyri more strongly than neutral vocalizations [Bach et al., 2008; Ethofer et al.; 2006; Fecteau et al., 2007; Grandjean et al., 2005; Phillips et al., 1998]. Moreover, a recent study employing a multi-voxel pattern analysis technique [Peelen et al., 2010] revealed emotion category specific spatial patterns of neuronal activity in the middle temporal gyrus reaching into the STS (BA 37) as well as in the medial prefrontal cortex (BA 10).

Research on music processing also probed the neuronal representations of emotional content. For example, Khalfa et al. [2005] compared the processing of sad and happy classical music. They found that minor (sad) and major (happy) excerpts activated the cingulate cortex and dorsolateral frontal regions, with major (happy) excerpts additionally activating premotor cortex. The involvement of cingulate and frontal regions has been supported by other studies contrasting minor and major melodies [Green et al., 2008] or happy and sad classical pieces [Mitterschiffthaler et al., 2007].

Although some of these music activations partially overlap with those reported for vocalizations, methodological differences make inferences about processing similarities difficult. First, voice and music research focused on different aspects of emotions. Voice research typically contrasted emotional with neutral vocalizations, whereas music research typically contrasted positive with negative emotion states. Second, while the stimuli used in voice experiments were highly representative of the vocalizations humans hear in everyday life, the same did not hold for the stimuli in music experiments [Janata, 2009]. These tended to be abstract, of short duration, or from genres of limited interest to the young listeners typically recruited for experiments (i.e., classical western music). Therefore, further research is needed to enable a better comparison of voice and music processing.

### The Present Study

As apparent from the above review, we still know relatively little about the processing of emotions in music and in what way this processing may be similar or different from the processing of emotions in the voice. With the present study, we sought to further this knowledge. To this end, we presented happy, neutral, and sad vocal and musical excerpts to participants who processed the excerpts' emotional content either explicitly or implicitly. In the explicit task, participants judged changes in emotional valence, whereas in the implicit task, participants judged pitch changes.

Given previous difficulties in the identification of emotional content effects, we hoped to increase experimental sensitivity in the following ways. First, we selected musical excerpts from popular genres that better reflected common, everyday musical experiences than excerpts selected in past research. Second, vocal and musical excerpts were of longer duration than in most previous studies and matched for basic acoustic properties. Third, unlike prior studies that used stimuli belonging to discrete emotion categories, we used stimuli for which emotions varied on a continuum from neutral to negative or neutral to positive. These variations were intended to mimic dynamic emotional changes present in real-life intercourse. Fourth, instead of making a categorical judgment after stimulus offset, participants in the present study used a joystick to assess emotions or pitch continuously, while listening to the sounds. This was expected to more effectively engage emotion evaluation systems [Zaki et al., 2009]. Finally, we used the mean rating results from each participant, rather than pre-experimental ratings, to select a subset of participant-specific sounds for statistical analysis.

To identify overlap and divergence of voice and music processing regions, we contrasted voice against rest and music against rest. On the basis of previous research we predicted that this would reveal activations along the bilateral STS for the voice contrast. We expected similar activations for the music contrast, if music evoked acoustic representations that are shared with the voice. To characterize the emotion evaluation of voice and music, we contrasted the explicit emotion task with the implicit task. On the basis of existing evidence from the voice literature, we expected lateral and/or medial frontal and temporal activations for vocalizations and potentially overlapping regions for music. Finally, we attempted to specify similarities in the representation of emotional content from vocalizations and music by contrasting emotional with neutral stimuli and happy with sad stimuli both in a whole brain analysis and in functionally defined regions of interest. Given that psychophysiological consequences [e.g., Ellis and Thayer, 2010; Schirmer and Escoffier, 2010] are similar for emotions conveyed vocally and musically, we expected their representation in the brain to be similar also.

## METHODS

### Participants

Twenty participants were invited to this study. The data from four participants were discarded due to technical problems in the recording of participant behavior. The 16 remaining participants (seven females) were right-handed and 18–26 years old (mean 21.7, SD 1.9). None of the participants reported auditory or neurological impairments. All participants gave written informed consent and were financially compensated for their time.

### Material

The music stimuli for the present fMRI experiment were obtained in two steps. First, we selected 139 nonvocal musical excerpts from popular music genres (i.e., pop, jazz, and classical music). These excerpts conveyed happy, neutral, or sad emotional states and were unfamiliar to the average listener. All excerpts were edited to be between 10 and 30 s long. In a second step, the excerpts were subjected to an emotion rating completed by 40 listeners, who did not participate in the main experiment. Listeners rated perceived happiness, perceived sadness, emotional clarity (i.e., how easily the emotion could be identified) and familiarity of each excerpt on a 10-point scale. We selected 12 happy, six neutral, and 12 sad excerpts with the lowest familiarity and highest clarity ratings to be included as final stimuli in the fMRI experiment. The average familiarity of the selected excerpts was rated as 4.5 ($SD = 1.1$) and did not significantly differ across the emotion categories ($P > 0.36$ for all pairwise comparisons). Participants were asked to give the maximum familiarity rating only if they could identify the excerpt. None of the excerpts selected could be identified. The mean rating scores for the selected excerpts as well as their average duration are presented in Table I (see Supporting Information for individual stimulus ratings).

The vocal stimuli for the present fMRI experiment were obtained by randomly swapping the vowels of 30 proper English sentences to render them meaningless. A professionally trained female vocalist then produced all sentences with a very happy, happy, neutral, sad and very sad intonation. This allowed us to counterbalance sentences

**TABLE I. Stimulus ratings (1–10 scales) and average durations for the voice and music excerpts**

| | Ratings | | | | | |
|---|---|---|---|---|---|---|
| Emotion | Perceived happiness | Perceived sadness | Emotional clarity | Familiarity | Duration (s) | Average pitch (Hz) |
| Voice | | | | | | |
| Happy | 6.1 (0.7) | 2.3 (0.7) | 6.6 (1.0) | – | 16 (2.3) | 182 (15) |
| Neutral | 3.7 (1.0) | 3.7 (1.0) | 5 (0.9) | – | 16.7 (2.0) | 160 (24) |
| Sad | 2.3 (1.0) | 6.2 (1.1) | 6.9 (1.0) | – | 19.6 (3.2) | 160 (16) |
| Music | | | | | | |
| Happy | 8.7 (1.2) | 2.8 (0.6) | 8.6 (0.9) | 4.5 (1.3) | 16.1 (2.5) | 141 (46) |
| Neutral | 5.2 (1.2) | 5.0 (1.1) | 6.4 (0.4) | 5.0 (0.7) | 17.1 (2.0) | 199 (103) |
| Sad | 3.4 (0.7) | 7.7 (0.9) | 8.2 (0.7) | 4.4 (1.0) | 18.8 (3.4) | 231 (95) |

Standard deviations are indicated in parentheses.

across vocal-emotional conditions such that a given participant heard a given sentence only once, while across subjects each sentence occurred in each vocal-emotional condition. The obtained 150 recordings were subjected to an emotion rating completed by 20 listeners, who did not participate in the main experiment. Participants rated perceived happiness, perceived sadness, and emotional clarity of each excerpt on a 10-point scale. Their rating scores and average durations are presented in Table I (see Supporting Information for individual stimulus ratings). Please note that this and the music rating were conducted primarily to test the suitability of each stimulus set for emotion research. Ratings for a given stimulus are relative in that they always depend on the stimuli/context in which they are rated. As such a direct comparison of the music and the voice ratings is not particularly meaningful. More relevant here is a comparison of ratings from listeners exposed to both music and voice. Such a comparison was conducted in the main experiment and revealed that the differences in perceived emotions between the experimental emotion conditions were comparable for music and voice. Music and voice stimuli were presented in random order in the fMRI experiment. Each participant listened to the 30 musical excerpts and to 30 sentences (each comprising 12 happy, six neutral, and 12 sad stimuli). All excerpts were presented twice resulting in 120 trials.

Both music and voice stimuli were at standard CD quality (44,100 Hz, 16 bits). To control for loudness differences all the stimuli were normalized to the same average root mean square amplitude. Furthermore, we ensured that music and voice compared in average pitch [$t(29.6) = 1.20$, $P = 0.24$; $\text{mean}_{\text{music}} = 188$ Hz; $\text{mean}_{\text{voice}} = 169$ Hz] as estimated with a harmonic power spectrum technique [Schroeder, 1968] and duration [$t(42.7) = 0.33$, $P = 0.74$; $\text{mean}_{\text{music}} = 17.3$ s; $\text{mean}_{\text{voice}} = 17.4$ s]. Please note that although the sounds used here were matched for basic acoustic parameters, they were nevertheless acoustically different. Such differences cannot be perfectly controlled as equating two sounds acoustically would make them identical and render their comparison meaningless. Therefore, in the present as in previously published work [e.g. Belin et al., 2000] differences in brain activation between music and voice could reflect both higher order functional as well as basic acoustic differences. One may nevertheless make inferences about higher order functional commonalities in the processing of music and voice by considering two points. First, regions outside auditory cortex likely reflect higher order processes. Second, within-modality contrasts should remove basic acoustic processes both different and common to music and voice. A conjunction analysis of such contrasts should thus inform about higher order functional similarities inside or outside auditory regions without the confounding effect of acoustic differences.

## Procedure

One day before the fMRI experiment, participants underwent a 15 min practice session to familiarize themselves with the response device and the tasks judging excerpts not presented during the main experiment. On the day of the experiment, participants were lying in the scanner with eyes closed wearing linear frequency response earplugs, which protected against the background scanner noise without altering the frequency content of the stimulus. Above these earplugs, participants wore MRI-compatible headphones (SereneSound system, Resonance Technology).

Figure 1 gives a graphical illustration of our procedure. Music and voice stimuli were presented in a pseudo-random order and with an inter-stimulus interval of 3 s. While listening, participants performed one of two tasks. One task required them to track the pitch of voice and music stimuli (i.e., pitch task), whereas the other task required them to track emotional expression (i.e., emotion task). They accomplished both tasks by operating a joystick placed beside their body and held by half the participants with the right hand and by the remaining participants with the left hand. The center position of the joystick was the reference. In the pitch task, this reference corresponded to the pitch perceived at stimulus onset. Half the participants were instructed to move the joystick

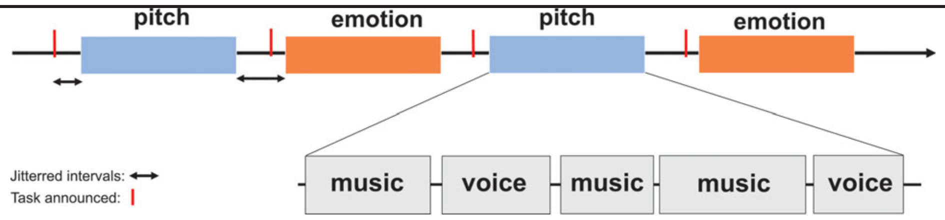

**Figure 1.**

FMRI experimental paradigm. The pitch and emotion tasks were presented in alternating blocks separated by a 25 s break. The music and voice stimuli within each block were presented in pseudorandom order. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

leftwards for increases in pitch and rightwards for decreases in pitch. The remaining participants performed the opposite motion. In the emotion task, joystick center position corresponded to a neutral expression. Half the participants were asked to move the stick leftwards if they felt the stimulus to become sad and rightwards if the felt the stimulus to become happy. The remaining participants performed the opposite motion. Tasks were presented in alternating blocks each comprising five stimuli. The task was announced by a synthesized female voice saying "pitch" or "emotion." Half the participants started with the pitch task, whereas the remaining participants started with the emotion task.

## Participant-Specific Selection of Stimuli to be Included in Data Analysis

The joystick movements in the emotion task were used to derive an average emotion rating for each participant and stimulus.[1] From these ratings, the six most positive, the six most neutral, and the six most negative musical excerpts and vocalizations were selected. Only these stimuli were considered experimental stimuli in the data analysis.

## Image Acquisition

Anatomical and functional MR images were acquired using a 3 Tesla Siemens Magnetom Trio system (Siemens, Erlangen, Germany). A high-resolution T1-weighted Magnetization Prepared Rapid Gradient Recalled Echo (MPRAGE) was applied to generate the anatomical image. Volumes consisted of 192 axial slices of 1 mm thickness with no gap [voxel size = 1 × 1 × 1; TR = 2,530 ms; TE = 1.64 ms; flip angle = 7°; field of view (FOV) = 256 × 256 mm; acquisition matrix = 256 × 256]. Functional images were acquired using single-shot echo-planar imaging (EPI) with 36 ascending 3.3 mm (no gap) axial slices parallel to the AC-PC plane (voxel size = 3 × 3 × 3.3; TR = 2 s, echo time = 30 ms; flip angle = 90°; FOV = 192 mm; matrix = 64 × 64). In total, 1337 EPI volumes were collected in one run. For better co-registration between the anatomical and functional images, 36-slices low-resolution anatomical images with the same slice position and orientation as the functional images were acquired with 3.609 mm thickness (no gap; FOV = 192 × 192 mm; matrix = 256 × 256).

## Data Analysis

Because the cortex is a convoluted sheet with sulco-gyral folding patterns, functionally distinct regions may be close to each other in a volume space but geometrically distant when measured along the cortex. This geometric property of the cortex is well preserved in a cortical surface model. Indeed, it has been shown that surface-based registration is superior to volume-based registration in fMRI group analysis [Anticevic et al., 2008]. Thus our data was both spatially normalized to and analyzed on a cortical surface model.

### *Image segmentation and cortical surface generation*

Anatomical MR images were first transformed to the Montreal Neurological Institute (MNI) space by affine registration using FLIRT from the FSL library [Jenkinson et al., 2002]. The algorithm maximized the correlation ratio between participant's image and the MNI average template (ICBM152) and images were resampled using a trilinear interpolation. All subsequent preprocessing steps were performed in MNI space. Every voxel in the anatomical images was first labeled as cerebrospinal fluid (CSF), gray matter (GM), white mater (WM), lateral ventricles, or subcortical structures (hippocampus, amygdala, caudate, putamen, globus pallidus, and thalamus) using a Markov-Random field model [Fischl et al., 2002]. An inner surface was then defined at the boundary of GM and WM and propagated to the boundary of GM and CSF to form an outer surface [Dale et al., 1999]. Both image labeling and surface generation were conducted using Freesurfer (http://surfer.nmr.mgh.harvard.edu). A fiducial surface, defined as a surface equidistant to the inner and outer surfaces, was then generated. To reduce noise features and small changes in shapes such as small spikes, this surface was smoothed by changing the location of each vertex toward the barycenter of its first neighbors. The smoothed fiducial surface was used as the geometric representation of the cortical surface for the analysis and figures. To facilitate localization of activation on the surface, we further divided the left and right template surfaces into thirty one regions based on anatomical definitions in Fischl et al. [2004] and Desikan et al. [2006].

### *Cortical surface registration*

To allow group analysis, we employed a multi-manifold large deformation diffeomorphic metric mapping (MM-LDDMM) algorithm to align the participants' cortical fiducial surfaces to a template generated from one participant in the study [Zhong and Qiu, 2010]. The MM-LDDMM algorithm utilizes different cortical representations (fiducial surface and gyral curves) and seeks an optimal diffeomorphic transformation (one-to-one, reversible, smooth, with topological preservation) to map one fiducial surface to another. The algorithm improves the quality of local alignment by incorporating the geometry of five reliable gyral curves, and controls the quality of both global and regional alignment by incorporating the geometry of the fiducial surface. The five gyral curves included the precentral gyrus, post-central gyrus, superior temporal gyrus, middle temporal gyrus and the inferior temporal gyrus. They were chosen because they are consistently present and are readily identifiable on the cortex around the regions of interest for our study. They were semi-automatically extracted from the fiducial surface using dynamic programming [Ratnanather, 2003; Zhong and Qiu, 2010] and manually labeled. The accuracy of the MM-LDDMM algorithm alignment of the fiducial surfaces was evaluated by comparing it with the diffeomorphic curve mapping and the diffeomorphic surface mapping [Zhong and Qiu, 2010; Zhong et al., 2010].

### *Projection of functional data to surface*

The functional volumes were first preprocessed using SPM5 (Wellcome Trust Centre for Neuroimaging). To this end, the temporal offsets between slices were corrected using a sinc interpolation. Data was then corrected for motion artifacts and a temporal high pass filter with a cut-off frequency of 1/190 Hz was used for baseline correction of the signal. Grand mean scaling was applied for global normalization of the data to remove the scanner gain. Intensity at each voxel was scaled by the average intensity derived from all voxels of all fMRI volumes and then multiplied by 100. The functional data was then transformed to MNI space and projected on each participant's fiducial surface in MNI space. For each participant, the low resolution anatomical image was rigidly aligned to the MNI-registered high resolution image. The functional volumes were then rigidly aligned to the low resolution anatomical image now in MNI space. Because the fiducial surface was generated from the MNI-registered high resolution image, both functional volumes and fiducial surface were in MNI space prior to the functional data projection. The projection required a reduction of dimensionality: the assignment of the functional data in the 3D volume to locations on the fiducial surface. The mapping of a voxel in the functional volumes to a vertex on the fiducial surface was performed in two steps. In the first step, we defined a set of voxels associated with each vertex on the fiducial surface as all GM voxels within 2 mm of the vertex. In the second step, we averaged the functional data over this set of voxels and assigned it to the vertex [Qiu et al., 2006b]. As the coregistration of anatomical and functional volumes was visually controlled, the procedure was conducted under the assumption of no misregistration. For fMRI group analysis, we then transferred individual functional data on the template surface derived from the MM-LDDMM transformation. Finally, the functional data was spatially smoothed on the template surface using orthonormal bases of the Laplace-Beltrami operator [equivalent to Gaussian smoothing with a kernel of 12 mm FWHM; Chung et al., 2005; Qiu et al., 2006a].

### *Surface-based statistical analysis: First level*

We analyzed our data in a general linear model framework [Friston et al., 1995], fitting a linear model at each location on the template surface. The analysis was performed in MATLAB using code from SPM5 (Wellcome Trust Centre for Neuroimaging). The first level model included regressors modeling all combinations of material (voice, music), task (pitch, emotion) and emotional content (happy, neutral, sad). This resulted in 12 experimental regressors. Additional regressors modeled the cues for rating pitch or emotion, the happy and sad voice and music stimuli not included for analysis, six movement parameters (translation and rotation for $x$, $y$, $z$ axis, respectively) and an intercept. For each regressor, events were modeled as short stimulus blocks, time-locked to stimulus onset and convolved with the synthetic hemodynamic response function provided by SPM5 [Friston et al., 1998].

### *Surface-based statistical analysis: Second level*

To examine commonalities and differences between conditions, the relevant individual participant contrasts were computed and subjected to second-level random effect analyses [Holmes and Friston, 1998]. First, commonalities between music and voice irrespective of task were identified using a conjunction of the two contrasts music minus resting baseline and voice minus resting baseline. Activation differences were explored using the contrast music minus voice. Second, overlaps between the neural correlates of emotion evaluation in music and voice were determined by computing the contrasts emotion task minus pitch task for both materials separately and computing their conjunction. Potential differences in emotion evaluation regions between music and voice were explored using an overall interaction contrast (i.e., [emotion music−pitch music]−[emotion voice−pitch voice]). Finally, we performed a whole brain analysis to investige the effect of specific emotional content. We contrasted happy and sad with neutral stimuli, and happy with sad stimuli, for both music and voice. The effect of emotional content was further investigated using an ROI analysis described further below.

All differences between conditions were tested using one sample t-tests implemented in a GLM model. Conjunctions were tested using the conjunction null hypothesis test [Nichols et al., 2005] after contrasts were entered in a second level ANOVA model. Because statistical comparisons were performed at 142,310 cortical locations, we controlled for multiple comparison at the cluster level using random field theory [Andrade et al., 2001; Worsley et al., 1996]. We set the vertex-level threshold at $P < 0.001$ uncorrected and a cluster size threshold at 112 $\text{mm}^2$, corresponding to a corrected significance level of $P < 0.05$. For vizualization, the t-statistical maps were rendered on the surface template using an uncorrected significance level of $P < 0.001$.

### Region of interest analysis

The effect of specific emotional content on regions identified by the whole-brain analysis was explored using functional regions of interest (ROIs). The average BOLD signal time course of each ROI resulting from the whole-brain analysis was extracted and the average signal was computed in a time-window time-locked to the onsets of each condition. The average signal during rest periods was then subtracted and the results scaled by this value and multiplied by 100. The resulting percent signal changes with reference to the resting baseline were then analyzed using ANOVAs with Material (voice, music), Task (emotion, pitch) and Emotional Content (happy, sad, neutral) as repeated measures factors. An ROI factor was added when relevant. It should be noted that some of the effects tested in the ANOVAs were not fully independent from the ROI selection procedure. We considered these only if they interacted with the effect of Emotional Content.

# RESULTS

## Behavioral Results

The continuous rating procedure used here differs from classical approaches and one might ask whether participants were indeed able to rate emotions using a joystick. To investigate this question we computed the correlation between the emotion ratings derived from joystick movements and the initial stimuli ratings derived using the classical, categorical procedure. The ratings from both procedures were highly correlated for both voice stimuli [$r = 0.83$, $t(148) = 17.8$, $P < 0.001$] and music stimuli [$r = 0.92$, $t(28) = 12.3$, $P < 0.001$]. This shows that participants were able to rate emotions using the joystick and that their mean rating values corresponded in accuracy to the accuracy of traditional categorical ratings.

To further assess perceived emotion from voice and music, the joystick ratings were analyzed using an ANOVA with Material (voice, music) and Emotional Content (happy, sad, neutral) as repeated measures factors. A significant Emotional Content effect [$F(2, 30) = 14$, $P < 0.01$] revealed that neutral stimuli were rated as less negative than sad stimuli [$t(30) = 9.2$, $P < 0.001$] and as less positive than happy stimuli [$t(30) = 8.3$, $P < 0.001$]. Additionally, a significant Material effect [$F(1, 15) = 50.4$, $P < 0.001$] indicated that participants perceived vocalizations to be more positive than music. Importantly, Emotional Content and Material failed to interact [$F(2, 30) = 1.6$, $P = 0.21$] suggesting that the Emotional Content effect was comparable for voice and music.

## Neuroimaging Results

### Music and voice processing

Regions involved in the processing of music and vocalizations were investigated by contrasting each stimulus type to the resting baseline and subjecting the resulting contrasts to a conjunction analysis. This analysis revealed activity in parts of Heschl's Gyrus (HG, BA 41) and the planum temporale (PT, BA 42) bilaterally (Fig. 2, Table II). Regions involved in the preferential processing of a stimulus type were investigated by contrasting music and vocalizations. For music, this revealed activity in the left PT and the right HG with activation along its medial border reaching to the first transverse sulcus at the anterior border of HG (Fig. 3, Table II). For vocalizations, we found activity in the superior and middle temporal gyrus (BA 21/22) and the cuneus (BA 19) in the left hemisphere. In the right hemisphere, activations showed in the superior temporal gyrus (BA 22), the temporal pole (BA 38) and the precuneus (BA 7, Fig. 4, Table II).

### Emotion evaluation

Regions implicated in the evaluation of both musical and vocal emotional expressions were identified by contrasting the emotion task with the pitch task for both music and vocalizations. The resulting contrasts were subjected to a conjunction analysis to determine regions of overlap (Fig. 5, Table II). This analysis revealed an activation cluster in the left and right medial superior frontal gyrus (BA 10), with the right cluster extending from the posterior part of the superior frontal gyrus (BA 10) to the rostral anterior cingulate gyrus (BA 32). In these clusters, activity was greater for the emotion as compared to the pitch task. No brain regions were identified in which activity was greater for the pitch as compared to the emotion task. Regions activated preferentially for music or voice during emotion evaluation were explored in an interaction analysis (see methods). This analysis revealed activity in the left intraparietal gyrus, which was further explored in the ROI analysis reported below.

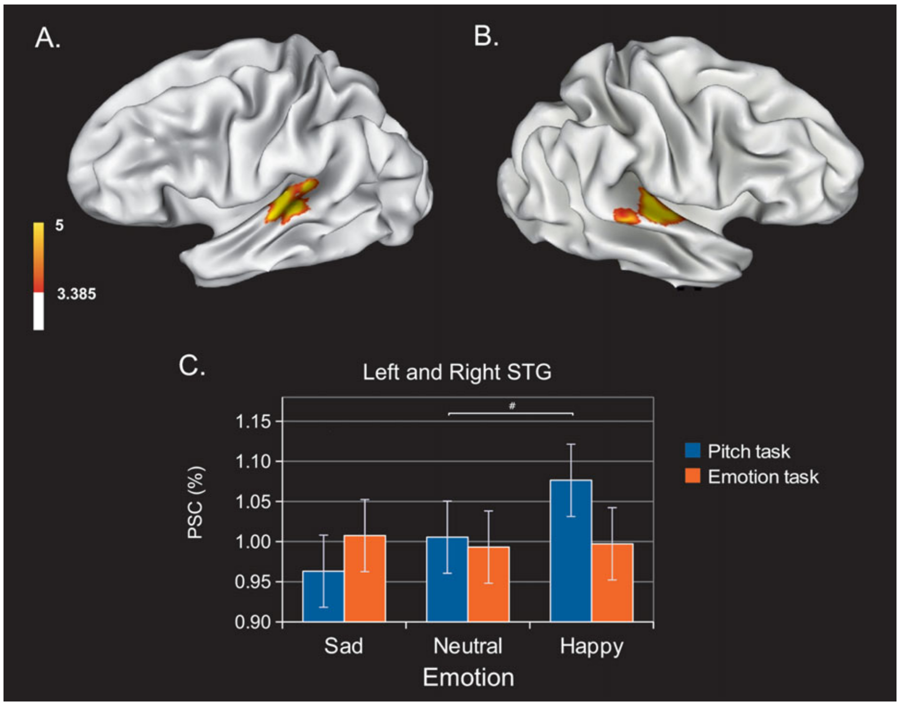

**Figure 2.**

Regions jointly activated by music and vocalizations when compared to rest in the left (**A**) and right (**B**) hemisphere. Across these regions, we observed an effect of emotional content in the pitch task but not in the emotion task (**C**). Error bars represent within-participant standard error of the mean (Loftus and Masson, 1994). The dash (#) sign represents a tendential effect.

### *Emotional content representation*

We investigated the effect of emotional content both in a whole-brain analysis and using a functional ROI approach. The whole brain analysis contrasted the emotional (happy and sad) with the neutral stimuli and the happy with the sad stimuli for both music and voice. These and the opposite contrasts were non-significant.

**TABLE II. Peak activations corrected for multiple comparisons**

| | | | MNI coordinates | | | | |
|---|---|---|---|---|---|---|---|
| Anatomical location | BA | Hemisphere | *x* | *y* | *z* | Z score | Area (mm$^2$) |
| Voice >Rest ~ Music > Rest | | | | | | | |
| Heschl's Gyrus | 41/42 | L | −60 | −23 | 15 | 5.46 | 1136 |
| Heschl's Gyrus/Temporal Plane | 41/42 | R | 60 | −19 | 12 | 4.95 | 841 |
| Voice > Music | | | | | | | |
| Superior/Middle Temporal Gyrus | 21/22 | L | −54 | −16 | 12 | 5.32 | 2224 |
| Temporal Pole | 38 | R | 54 | 14 | −24 | 3.66 | 132 |
| Superior Temporal Gyrus | 22 | R | 64 | −4 | −7 | 4.42 | 1149 |
| Cuneus | 19 | L | −1 | −80 | 37 | 4.26 | 196 |
| Precuneus | 7 | R | 9 | −69 | 45 | 4.22 | 316 |
| Music > Voice | | | | | | | |
| Planum Temporale | 42 | L | −51 | −35 | 13 | 4.16 | 144 |
| Heschl's Gyrus | 41 | R | 43 | −13 | −2 | 5.74 | 282 |
| Emotion Task > Pitch Task (Music ~ Voice) | | | | | | | |
| Medial superior frontal gyrus | 10 | L | −4 | 57 | 8 | 5.12 | 126 |
| Anterior cingulate gyrus | 32 | R | 11 | 45 | 0 | 4.36 | 109 |

[1] fMRI data analysis using dynamic changes around the average rating as a regressor revealed no significant results.

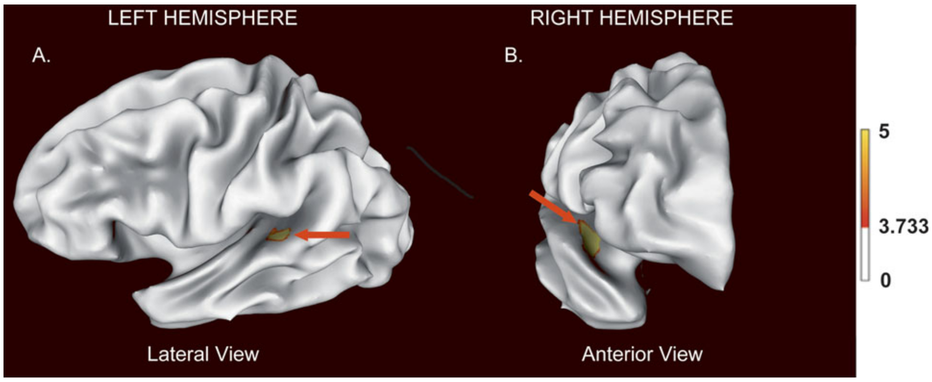

**Figure 3.**

Regions with greater activity when listening to music as compared to vocalizations. The statistical t-map was thresholded at an uncorrected vertex-level threshold of $P < 0.001$, and clusters smaller than the cluster size threshold (112 mm$^2$) were excluded. Significant effects were found in (**A**) left planum temporale and (**B**) the right first transverse sulcus and part of HG. The red arrows indicate the positions of the two activations.

ROI analyses were performed to investigate the effect of emotional content in the regions identified in the previous analyses. First, common and different activations for voice and music as described in the Music and voice processing section were subjected to separate ANOVAs with ROI (regions identified in the whole brain analysis), Task (emotion, pitch), Material (music, voice), and Emotional content (sad, neutral, happy) as repeated measures factors. For

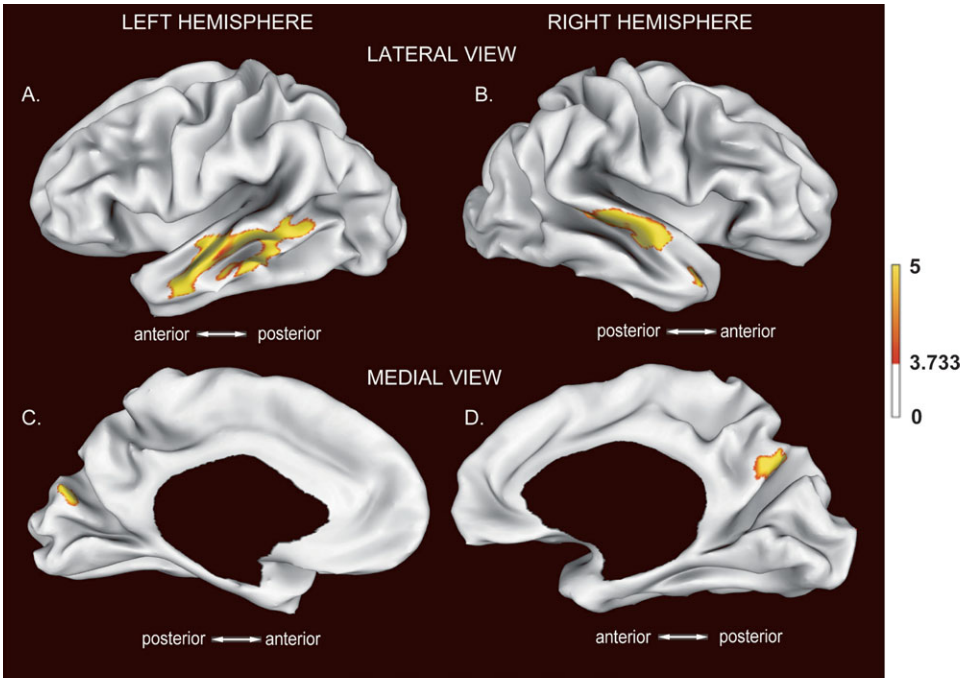

**Figure 4.**

Regions with greater activity when listening to vocalizations as compared to music. The statistical t-map was thresholded at an uncorrected vertex-level threshold of $P < 0.001$ and clusters smaller than the cluster size threshold (112 mm$^2$) were excluded. Significant differences were found in (**A**) the left ventral superior temporal gyrus (BA 22), dorsal middle temporal gyrus (BA 21), (**B**) right ventral superior temporal gyrus (BA22), temporal pole (BA 38), (**C**) left cuneus (BA 19), and (**D**) right precuneus (BA 7).

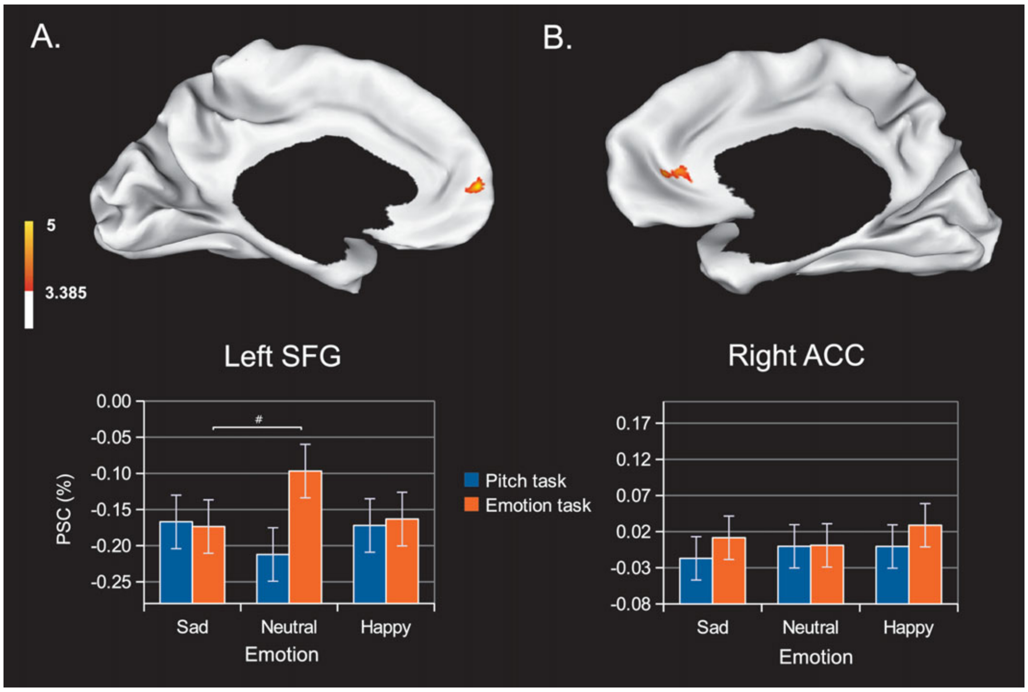

**Figure 5.**

Regions with greater activity during the emotion as compared to the pitch judgment task in the conjunction of music and voice. The statistical t map was thresholded at an uncorrected vertex-level threshold of $P < 0.001$, and clusters smaller than the cluster size threshold (100 mm$^2$) were excluded. Significant effects were found in (**A**) lateral superior frontal gyrus and (**B**) right anterior cingulate gyrus. Error bars represent within-participant standard error of the mean (Loftus and Masson, 1994). The dash (#) sign represents tendential effects. PSC, percent signal change.

activations common to voice and music, the Task by Emotional Content interaction was significant [$F(2, 30) = 3.70$, $P < 0.05$, Fig. 2]. Follow-up analysis for the pitch task revealed tendentially greater activity for happy as compared to sad [$F(1, 15) = 3.7$, $P = 0.07$] and neutral [$F(1, 15) = 3.2$, $P = 0.09$] stimuli. Notably, these observations were made in two thirds of the participants (11 out of 16). The difference between neutral and sad stimuli was non-significant ($P > 0.38$). No effects were observed in the follow-up analysis for the emotion task (all $Ps > 0.7$). For activations differing between voice and music, there were no significant effects involving Emotional Content ($Ps > 0.13$).

Second, we investigated the effect of Emotional Content on the frontal emotion evaluation regions identified in the whole brain analysis described in the section Emotion evaluation. For activations identified in this analysis an ANOVA with ROI (L mSFG, R mSFG/ACC), Task (emotion, pitch), Material (music, voice), and Emotional Content (sad, neutral, happy) as repeated measures factors revealed an interaction between ROI, Task and Emotional Content [$F(4, 60) = 3.71$, $P < 0.05$]. Follow-up analysis of the Task by Emotional Content interaction in the left medial SFG ROI [$F(2, 30) = 3.20$, $P = 0.05$, Fig. 5] revealed for the Emotion task a tendentially higher activity for neutral as compared to sad stimuli [$F(1, 15) = 3.9$, $P = 0.07$]. These observations were again made in 11 of the 16 participants. The difference between neutral and happy stimuli was non-significant [$F(1, 15) = 0.04$, $P = 0.84$]. No effects were observed in the follow-up analysis for the pitch task (all $Ps > 0.11$). There were no significant effects of Emotional Content in the right mSFG/ACC and the intraparietal region ($Ps > 0.1$).

## DISCUSSION

The aim of the present study was to compare the neuronal substrates that mediate emotional music and voice processing. The three different aspects of this processing that were of interest here are discussed separately below.

### Voice and Music Processing in the Temporal Lobe

While surprisingly complex information is already represented at the level of the brainstem and the diencephalon [Strait et al., 2009; Tervaniemi et al., 2006; von Kriegstein et al., 2008], most aspects of sound recognition rely on cortical representations and thus input from HG. Past research suggests that primary auditory cortex in HG is differently specialized in the left and right hemisphere [Schönwiesner et al., 2005; Zatorre and Belin, 2001]. Slow spectral changes perceived as changes in melody or pitch are favored by the right hemisphere, which presumably has a large temporal integration window suited for these changes. In contrast, fast spectral changes, as those representing consonants in speech, are favored by the left hemisphere, which has a relatively smaller integration window. Researchers believe that this specialization explains the different lateralization patterns for music and speech. A greater significance of slow spectral changes in music plays to the right hemisphere, whereas a greater significance of fast spectral change in speech linked to articulation plays to the left hemisphere [Zatorre et al., 2002].

The present results accord with this. Compared to the spoken, nonsensical utterances used here, music more strongly activated the right HG. That music and vocalizations activated the left HG to the same extent likely reflects the nature of the vocal stimuli (i.e., non-meaningful speech) and the absence of a linguistic task [Gandour et al., 2003; Imaizumi et al., 1998; Shtyrov et al., 2005]. Both music and voice activated the bilateral PT, which forms an extension of HG in the superior temporal gyrus. Surprisingly this activation was more left lateralized for music than for vocalizations. Again, this may be explained by the nature of the present stimuli. While music stimuli were multi-instrumental comprising several spatial sources, vocalizations originated from only one source. Thus, music may have been more likely to recruit auditory scene analysis processes supported by the left temporal plane [Deike et al., 2004; Micheyl et al., 2007].

Importantly, temporal regions more strongly activated when contrasting voice with music listening were the bilateral STG, the left MTG, and the right temporal pole. These activations map onto previous reports of voice sensitivity in the temporal lobe [Belin et al., 2000]. Moreover, together with the activations observed in the occipital and parietal lobe, they suggest that even when contrasted with melodically and harmonically rich music, vocalizations are more likely to recruit multi-sensory [Brosch et al., 2009; Kawase et al., 2005] and attention processes [Brosch et al., 2008; Sander at al., 2005]. A reason for this may be that vocalizations play a more important role in social interactions, which critically determine what humans attend to and act on [Pittam, 1994].

## Emotion Evaluation in the Frontal Lobe

Although vocalizations may be special in some respect, they share basic properties with music. In particular, the expression of emotions has been shown to be similar between the two [Juslin and Laukka, 2003] and for both, well accounted for by a current emotion model [Scherer, 2003; Zentner et al., 2009]. Evidence for culturally universal recognition of vocal [Chua and Schirmer, 2011; Pell and Skorup, 2008] and musical emotions [Fritz et al., 2009] furthermore suggests that emotion recognition for both types of sounds is largely innate. Thus, one would expect the underlying perceptual processes to be comparable. The present results support this idea. Contrasting the emotion task with the pitch task revealed no different but common activations for music and vocalizations. Specifically, both activated an area in the medial SFG bilaterally reaching into the anterior cingulate.

These activations compare to those reported by Peelen and colleagues [2010] who compared emotional processing of voice, body movements, and facial expressions. As such the present results corroborate the notion that the medial prefrontal cortex supports supramodal emotional representations. Additionally, they indicate that such representations go beyond basic nonverbal cues but include secondary means of emotional expression such as music. Interestingly, the structures identified here have also been implicated in mentalizing or theory of mind (ToM). For example, the processing of short stories that require listeners to infer the mental state of story agents activates the medial SFG relative to stories about physical events that do not require mentalizing [Fletcher et al., 1995; Gallagher et al., 2000; Gobbini et al., 2007]. These and other findings are reviewed by Carrington and Bailey [2009] who highlight the medial SFG as the most consistently activated structure related to ToM. Moreover, this structure appears to be of particular significance for a subset of ToM tasks related to empathy. Evaluating the emotional state of others in a real life situation (e.g., from audiovisual recording) activated this region [Zaki et al., 2009].

Interestingly, the medial SFG rarely emerges as a prominent structure in the music and voice processing literature. For voice, the SFG has been reported occasionally, but not frequently enough to feature in current models [Schirmer and Kotz, 2006; Wildgruber et al., 2006]. The reason for this may lie in the nature of the stimuli and tasks that were typically used. Stimuli were typically very clear with respect to their emotional value – that is they were selected as best representatives for a given emotion. Moreover, participants were required to assign stimuli to two or three categories (e.g., emotional, neutral). Notably, one study that deviated from this procedure reported ACC activity that partially overlaps with the medial SFG activity observed here [Wildgruber et al., 2002]. In this study, participants listened to two sentence pairs with subtle acoustic differences and were asked to indicate which of the two was more expressive. Thus, relative to other work, the task was more challenging and encouraged continuous emotion evaluation likely more suitable to elicit activity in the medial SFG. In line with this, the present study found activity in the left medial SFG at peak when the average rating of sounds was neutral—suggesting that the more ambiguous sounds recruited this region more strongly.

For music, medial frontal regions are traditionally not considered central processing hubs [e.g., Levitin and Tirovolas, 2009; Peretz and Zatorre, 2004]. For example, in the neuropsychological literature on emotion in music those regions have, to the best of our knowledge, never been investigated. In light of the present findings, this could explain why this literature never identified a general deficit in the perception of emotions from music [Gosselin et al., 2005, 2006, 2007, 2011; Peretz et al., 1998]. In the neuroimaging literature, medial frontal regions have recently been shown to respond to familiar and pleasant musical stimuli [Janata, 2009b; Plailly et al., 2007] and recent theoretical proposals highlight their importance [Janata, 2009a; Livingstone and Thompson, 2009]. For example, Janata [2009a] argues that empathic and self-referential processes supported by these regions critically contribute to the appreciation of music in everyday listening contexts. That, nevertheless, music emotion researchers failed to consistently observe medial frontal activations likely reflects the fact that one of their important functions, the evaluation of perceived emotions, has hitherto not been probed with ecologically valid musical stimuli.

## Brain Representations of Emotional Content

A final goal of the present study was to identify regions recruited by specific emotional content. To this end, we conducted both a whole brain and a more sensitive ROI analysis [Saxe et al., 2006]. Despite all the measures taken to increase experimental sensitivity, the whole brain analysis was non-significant. This maps onto existing work, which in the large majority reports negative results [Wildgruber et al., 2006]. Those that found emotion-specific activations used extreme emotion exclamations [Fecteau et al., 2007], resorted to less conservative statistical thresholds [Khalfa et al., 2005], restricted the analysis to specific regions, used more sensitive multivoxel pattern analysis [Ethofer et al., 2009] and/or used a larger number of participants [Ethofer et al., 2006].

Overall, this suggests that the areas representing specific emotional content for music and voice vary across and possibly within individuals making it difficult to identify consistent effects in a whole-brain analysis. Inter- and intra-individual variations in the localization of emotion-specific areas can be linked to music and voice having different relevance for different individuals in different contexts. For instance, brain responses to emotional voices have been shown to depend on social orientation [Schirmer et al., 2004], gender [Schirmer et al., 2008], and musical expertise [Strait et al., 2009]. Aside from differences in personal relevance, variation in brain activation could arise from differences in the contexts in which representations were acquired and the actions they were linked to.

Nevertheless, one may expect auditory emotional processing to have a common denominator across individuals. To identify this denominator, we complemented the whole brain analysis with a more sensitive ROI analysis guided by functional activations observed when contrasting stimulus type or task. Here, we found an emotion effect in the bilateral HG and PT as well as in the medial frontal cortex. In the bilateral HG and PT, activity tended to be greater for happy as compared to neutral and sad sounds. This effect is likely related to the greater arousal of happy relative to neutral and sad stimuli, which is conveyed by differences in tempo, intensity and pitch [Banse and Scherer, 1996]. In the medial frontal cortex, neutral sounds tended to elicit greater activity than emotional sounds. This effect is likely related to emotion evaluation difficulty, which was greatest for the neutral sounds that comprised few clear emotion signals [Leitman et al., 2010].

While the former effect in HG and PT was present during implicit but not explicit emotion processing, the latter effect in the medial SFG was present during explicit but not implicit emotion processing. Differential emotion content effects as a function of task have been reported previously [Critchley et al., 2000; Hariri et al., 2000, 2003]. For example, Critchley and colleagues found greater activity to emotional as compared to neutral facial expressions in primary visual cortex and amygdala during an implicit but not during an explicit task. In contrast, the explicit task produced greater activity to emotional as compared to neutral facial expressions in the fusiform gyrus and middle temporal gyrus among others. In line with this, the present study suggests that emotion content effects differ depending on whether emotions are evaluated implicitly or explicitly. The former seems to elicit changes in primary sensory and emotion structures, whereas the latter seems more likely to elicit changes in structures that support the explicit evaluation of social information.

## CONCLUSIONS

Together, the present results extend existing work by showing that music and voice recruit specific, yet partially overlapping, networks. Moreover, greater activations beyond PT for vocalizations relative to music confirm the notion of vocalizations being particularly important for the human auditory system. Nevertheless, the mechanisms by which listeners evaluate and represent emotions coded in voice and music are comparable. Judging emotions in both voice and music activates the medial SFG suggesting a common reliance on processes involved in social cognition. Moreover, emotion specific but not stimulus specific representations emerge in low-level perceptual systems during implicit processing and in high-level evaluative systems during explicit processing. Although the range of emotions expressed by music can extend beyond that of the voice and vice versa [Juslin and Laukka, 2003; Zentner et al., 2009], our findings imply basic overlap and support the notion that music may mirror the voice [Rousseau, 1781, 1992; Spencer, 1857].

## ACKNOWLEDGMENTS

The authors thank Pierce Hale and Amerie Baeg for their help with stimulus selection, editing, and rating. This research was supported by the NUS young investigator award R-581-000-066-101 (AS), A*STAR SERC 0921570130 (AS), A*STAR SERC 082-101-0025 (AQ), A*STAR SICS-09/1/1/001 (AQ), the Young Investigator Award at National University of Singapore (NUSYIA FY10 P07) (AQ), and National University of Singapore MOE AcRF Tier 1 (AQ).

## REFERENCES

Andrade A, Kherif F, Mangin JF, Worsley KJ, Paradis AL, Simon O, Dehaene S, Le Bihan D, Poline JB (2001): Detection of fMRI activation using cortical surface mapping. Hum Brain Mapp 12:79–93.

Anticevic A, Dierker DL, Gillespie SK, Repovs G, Csernansky JG, Van Essen DC, Barch DM (2008): Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia. Neuroimage 41: 835–848.

Bach DR, Grandjean D, Sander D, Herdener M, Strik WK, Seifritz E (2008): The effect of appraisal level on processing of emotional prosody in meaningless speech. Neuroimage 42:919–927.

Banse R, Scherer K (1996): Acoustic profiles in vocal emotion expression. J Pers Soc Psychol 70:614–636.

Belin P, Zatorre R, Lafaille P, Ahad P, Pike B (2000): Voice-selective areas in human auditory cortex. Nature 403:309–312.

Belin P, Zatorre RJ, Ahad P (2002): Human temporal-lobe response to vocal sounds. Cogn Brain Res 13:17–26.

Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET (2000): Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Brosch T, Sander D, Pourtois G, Scherer KR (2008): Beyond fear: Rapid spatial orienting toward positive emotional stimuli. Psychol Sci 19:362–370.

Brosch T, Grandjean D, Sander D, Scherer KR (2009): Cross-modal emotional attention: Emotional voices modulate early stages of visual processing. J Cogn Neurosci 21:1670–1679.

Carrington SJ, Bailey AJ (2009): Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum Brain Mapp 30:2313–2335.

Chua SM, Schirmer A (2011): Perceiving verbal and vocal emotions in a second language. Cog Emotion 25:1376–1392.

Chung MK, Robbins SM, Dalton KM, Davidson RJ, Alexander AL, Evans AC (2005): Cortical thickness analysis in autism with heat kernel smoothing. Neuroimage 25:1256–1265.

Critchley H, Daly E, Phillips M, Brammer M, Bullmore E, Williams S, Van Amelsvoort T, Robertson D, David A, Murphy D (2000): Explicit and implicit neural mechanisms for processing of social information from facial expressions: A functional magnetic resonance imaging study. Hum Brain Mapp 9: 93–105.

Confavreux C, Croisile B, Garassus P, Aimard G, Trillet M (1992): Progressive amusia and aprosody. Arch Neurol 49:971–976.

Curtis ME, Bharucha JJ (2010): The minor third communicates sadness in speech, mirroring its use in music. Emotion 10:335–348.

Dale AM, Fischl B, Sereno MI (1999): Cortical surface-based analysis I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

Deike S, Gaschler-Markefski B, Brechmann A, Scheich H (2004): Auditory stream segregation relying on timbre involves left auditory cortex. Neuroreport 15:1511–1514.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006): An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31:968–980.

Ellis RJ, Thayer JF (2010): Music and autonomic nervous system (dys)function. Music Percept 27:317–326.

Ethofer T, Anders S, Wiethoff S, Erb M, Herbert C, Saur R, Grodd W, Wildgruber D (2006): Effects of prosodic emotional intensity on activation of associative auditory cortex. Neuroreport 17:249–253.

Fecteau S, Belin P, Joanette Y, Armony JL (2007): Amygdala responses to nonlinguistic emotional vocalizations. Neuroimage 36:480–487.

Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002): Whole brain segmentation automated labeling of neuroanatomical structures in the human brain. Neuron 33:341–355.

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004): Automatically parcellating the human cerebral cortex. Cereb Cortex 14:11–22.

Fletcher PC, Happé F, Frith U, Baker SC, Dolan RJ, Frackowiak RSJ, Frith CD (1995): Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. Cognition 57:109–128.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995): Statistical parametric maps in functional imaging: A general linear approach. Hum Brain Mapp 2:189–210.

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

Fritz T, Jentschke S, Gosselin N, Sammler D, Peretz I, Turner R, Friederici AD, Koelsch S (2009): Universal recognition of three basic emotions in music. Curr Biol 19:573–576.

Gallagher HL, Happe F, Brunswick N, Fletcher PC, Frith U, Frith CD (2000): Reading the mind in cartoons and stories: An fMRI study of theory of mind in verbal and nonverbal tasks. Neuropsychologia 38:11–21.

Gandour J, Wong D, Dzemidzic M, Lowe M, Tong Y, Li X (2003): A cross-linguistic fMRI study of perception of intonation and emotion in Chinese. Hum Brain Mapp 18:149–157.

Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007): Two takes on the social brain: A comparison of theory of mind tasks. J Cogn Neurosci 19:1803–1814.

Gosselin N, Peretz I, Noulhiane M, Hasboun D, Beckett C, Baulac M, Samson S (2005): Impaired recognition of scary music following unilateral temporal lobe excision. Brain 128:628–640.

Gosselin N, Samson S, Adolphs R, Noulhiane M, Roy M, Hasboun D, Baulac M, Peretz I (2006): Emotional responses to unpleasant music correlates with damage to the parahippocampal cortex. Brain 129:2585–2592.

Gosselin N, Peretz I, Johnsen E, Adolphs R (2007): Amygdala damage impairs emotion recognition from music. Neuropsychologia 45:236–244.

Gosselin N, Peretz I, Hasboun D, Baulac M, Samson S (2011): Impaired recognition of musical emotions and facial expressions following anteromedial temporal lobe excision. Cortex 47:1116–1125.

Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005): The voices of wrath: Brain responses to angry prosody in meaningless speech. Nat Neurosci 8:145–146.

Green A, Baerentsen K, Stodkilde-Jorgensen H, Wallentin M, Roepstorff A, Vuust P (2008): Music in minor activates limbic structures: A relationship with dissonance? Neuroreport 19:711–715.

Hariri AR, Bookheimer SY, Mazziotta JC (2000): Modulating emotional responses: Effects of a neocortical network on the limbic system. Neuroreport 11:43–48.

Hariri AR, Mattay VS, Tessitore A, Fera F, Weinberger DR (2003): Neocortical modulation of the amygdala response to fearful stimuli. Biol Psychiat 53:494–501.

Holmes AP, Friston KJ (1998): Generalisability, random effects and population inference. Neuroimage 7:S754.

Ilie G, Thompson WF (2006): A comparison of acoustic cues in music and speech for three dimensions of affect. Music Percept 23:319–329.

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.

Janata P (2005): Brain networks that track musical structure. Ann NY Acad Sci 1060:111–124.

Janata P (2009a): Music and the self. In: Haas R, Brandes V, editors. Music that Works: Contributions of Biology, Neurophysiology, Psychology, Sociology, Medicine and Musicology. New York, NY: Springer. pp131–141.

Janata P (2009b): The neural architecture of music-evoked autobiographical memories. Cereb Cortex 19:2579–2594.

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 17:825–841.

Juslin PN (2000): Cue utilization in communication of emotion in music performance: Relating performance to perception. J Exp Psychol Human 26:1797–1813.

Juslin PN, Laukka P (2003): Communication of emotions in vocal expression and music performance: Different channels, same code? Psychol Bull 129:770–814.

Kawase T, Yamaguchi K, Ogawa T, Suzuki K, Suzuki M, Itoh M, Kobayashi T, Fujii T (2005): Recruitment of fusiform face area associated with listening to degraded speech sounds in auditory-visual speech perception: A PET study. Neurosci Lett 382:254–258.

Khalfa S, Schon D, Anton JL, Liégeois-Chauvel C (2005): Brain regions involved in the recognition of happiness and sadness in music. Neuroreport 16:1981–1984.

Leitman DI, Wolf DH, Ragland JD, Laukka P, Loughead J, Valdez JN, Javitt DC, Turetsky BI, Gur RC (2010): It's not what you say, but how you say it: A reciprocal temporo-frontal network for affective prosody. Front Hum Neurosci 26:4–19.

Levitin DJ, Tirovolas AK (2009): Current advances in the cognitive neuroscience of music. Ann NY Acad Sci 1156:211–231.

Livingstone S, Thompson W (2009): The emergence of music from theory of mind. Music Sci (Special issue 2009–2010):83–115.

Loftus GR, Masson ME (1994): Using confidence intervals in within-participant designs. Psychon B Rev 1:476–490.

Micheyl C, Carlyon R, Gutschalk A, Melcher J, Oxenham A, Rauschecker J, Tian B, Wilson CE (2007): The role of auditory cortex in the formation of auditory streams. Hear Res 229:116–131.

Mitterschiffthaler M, Fu C, Dalton J, Andrew C, Williams S (2007): A functional MRI study of happy and sad affective states induced by classical music. Hum Brain Mapp 28:1150–1162.

Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005): Valid conjunction inference with the minimum statistic. Neuroimage 25:653–660.

Peelen MV, Atkinson AP, Vuilleumier P (2010): Supramodal representations of perceived emotions in the human brain. J Neurosci 30:10127–10134.

Pell MD, Skorup V (2008): Implicit processing of emotional prosody in a foreign versus native language. Speech Commun 50:519–530.

Peretz I, Gagnon L, Bouchard B (1998): Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. Cognition 68:111–141.

Peretz I, Zatorre RJ (2004): Brain organization for music processing. Annu Rev Psychol 56:89–114.

Phillips ML, Young AW, Scott SK, Calder AJ, Andrew C, Giampietro V, Williams C, Bullmore ET, Brammer M, Gray JA (1998): Neural responses to facial and vocal expressions of fear and disgust. P Roy Soc B-Biol Sci 265:1809–1817.

Pittam J (1994): Voice in Social Interaction: An Interdisciplinary Approach. Thousand Oaks, CA: Sage Publications, Inc.

Plailly J, Tillmann B, Royet J-P (2007): The feeling of familiarity of music and odors: The same neural signature? Cereb Cortex 17:2650–2658.

Qiu A, Bitouk D, Miller MI (2006a): Smooth functional and structural maps on the neocortex via orthonormal bases of the Laplace–Beltrami operator. IEEE T Med Imaging 25:1296–1306.

Qiu A, Rosenau BJ, Greenberg AS, Hurdal MK, Barta P, Yantis S, Miller MI (2006b): Estimating linear cortical magnification in human primary visual cortex via dynamic programming. Neuroimage 31:125–138.

Ratnanather JT, Barta PE, Honeycutt NA, Lee N, Morris HM, Dziorny AC, Hurdal MK, Pearlson GD, Miller MI (2003): Dynamic programming generation of boundaries of local coordinatized submanifolds in the neocortex: Application to the planum temporale. Neuroimage 20:359–377.

Rousseau JJ (1992): Essay on the origin of languages. In: Scott J, editor. Essay on the Origin of Languages and Writings Related to Music. Hanover: University Press of New England.

Sander D, Grandjean D, Pourtois G, Schwartz S, Seghier M, Scherer KR, Vuilleumier P (2005): Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. Neuroimage 28:848–858.

Saxe R, Brett M, Kanwisher N (2006): Divide and conquer: A defense of functional localizers. Neuroimage 30:1088–1096.

Scherer KR (1986): Vocal affect expression: A review and a model for future research. Psychol Bull 99:143–165.

Scherer KR (2003): Vocal communication of emotion: A review of research paradigms. Speech Commun 40:227–256.

Schirmer A, Escoffier N (2010): Emotional MMN: Anxiety and heart rate correlate with the ERP signature for auditory change detection. Clin Neurophys 121:53–59.

Schirmer A, Escoffier N, Zysset S, Koester D, Striano T, Friederici A (2008): When vocal processing gets emotional: On the role of social orientation in relevance detection by the human amygdala. Neuroimage 40:1402–1410.

Schirmer A, Kotz S (2006): Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. Trends Cogn Sci 10:24–30.

Schirmer A, Zysset S, Kotz S, von Cramon Y (2004): Gender differences in the activation of inferior frontal cortex during emotional speech perception. Neuroimage 21:1114–1123.

Schönwiesner M, Rübsamen R, von Cramon Y (2005): Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci 22:1521–1528.

Schroeder MR (1968): Period histogram and product spectrum: New methods for fundamental-frequency measurement. J Acoust Soc Am 43:829–834.

Shtyrov Y, Pihko E, Pulvermüller F (2005): Determinants of dominance: Is language laterality explained by physical or linguistic features of speech? Neuroimage 27:37–47.

Spencer H (1857): The origin and function of music. Fraser's Magazine 56:396–408.

Strait DL, Kraus N, Skoe E, Ashley R (2009): Musical experience and neural efficiency—Effects of training on subcortical processing of vocal expressions of emotion. Eur J Neurosci 29:661–668.

Tervaniemi M, Szameitat AJ, Kruck S, Schroger E, Alter K, De Baene W, Friederici AD (2006): From air oscillations to music and speech: Functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. J Neurosci 26:8647–8652.

von Kriegstein K, Patterson RD, Griffiths T (2008): Task-dependent modulation of medial geniculate body is behaviorally relevant for speech recognition. Curr Biol 18:1855–1859.

Warren J (2008): How does the brain process music? Clin Med 8:32–36.

Wildgruber D, Pihan H, Ackermann H, Erb M, Grodd W (2002): Dynamic brain activation during processing of emotional intonation: Influence of acoustic parameters, emotional valence, and sex. Neuroimage 15:856–869.

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T (2006): Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog Brain Res 156:249–268.

Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996): A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 4:58–73.

Zaki J, Weber J, Bolger N, Ochsner K (2009): The neural bases of empathic accuracy. Proc Natl Acad Sci USA 106:11382–11387.

Zatorre R, Belin P (2001): Spectral and temporal processing in human auditory cortex. Cereb Cortex 11:946–953.

Zatorre R, Belin P, Penhune V (2002): Structure and function of auditory cortex: Music and speech. Trends Cogn Sci 6:37–46.

Zentner M, Grandjean D, Scherer K (2008): Emotions evoked by the sound of music: Characterization, classification, and measurement. Emotion 8:494–521.

Zhong J, Qiu A (2010): Multi-manifold diffeomorphic metric mapping for aligning cortical hemispheric surfaces. Neuroimage 49:355–365.

Zhong J, Phua DY, Qiu A (2010): Quantitative evaluation of LDDMM, FreeSurfer, and CARET for cortical surface mapping. Neuroimage 52:131–141.