Predicting Vocal Emotion Expressions from the Human Brain

Sonja A. Kotz,[1]* **Christian Kalberlah,**[2,3] **Jörg Bahlmann,**[4,5]
Angela D. Friederici,[4,6] **and John-D. Haynes**[2,3,6]

[1]*Minerva Research Group "Neurocognition of Rhythm in Communication", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[2]*Max Planck Fellow Group "Attention and Awareness", Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[3]*Bernstein Center for Computational Neuroscience, Charité – Universitätsmedizin Berlin, Berlin, Germany*
[4]*Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[5]*Helen Wills Neuroscience Institute, University of California, Berkeley, California*
[6]*Graduate School of Mind and Brain, Humboldt-University, Berlin, Germany*

◆━━━━━━━━━━━━━━━━━━━━━━━━━━◆

Abstract: Speech is an important carrier of emotional information. However, little is known about how different vocal emotion expressions are recognized in a receiver's brain. We used multivariate pattern analysis of functional magnetic resonance imaging data to investigate to which degree distinct vocal emotion expressions are represented in the receiver's local brain activity patterns. Specific vocal emotion expressions are encoded in a right fronto-operculo-temporal network involving temporal regions known to subserve suprasegmental acoustic processes and a fronto-opercular region known to support emotional evaluation, and, moreover, in left temporo-cerebellar regions covering sequential processes. The right inferior frontal region, in particular, was found to differentiate distinct emotional expressions. The present analysis reveals vocal emotion to be encoded in a shared cortical network reflected by distinct brain activity patterns. These results shed new light on theoretical and empirical controversies about the perception of distinct vocal emotion expressions at the level of large-scale human brain signals. *Hum Brain Mapp 34:1971–1981, 2013.* © 2012 Wiley Periodicals, Inc.

Key words: emotion; vocal expressions; prosody; multivariate pattern analysis; fMRI

◆━━━━━━━━━━━━━━━━━━━━━━━━━━◆

INTRODUCTION

The affective state of a person during social interaction can be transferred by vocal expressions that consist of verbal or nonverbal information, the latter termed prosody [Scherer, 1995]. Although previous neuroimaging studies have shown activation differences between the perception of emotional and neutral vocal expressions in a fronto-temporo-striatal brain network [Beaucousin et al., 2007; Grandjean et al., 2005; Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006], it remains unclear to which degree discrete emotional vocal expressions are processed in these brain regions. In light of theories on discrete facial and vocal "basic emotions" [e.g., Banse and Scherer, 1996; Ekman, 1992], regulated by distinct yet overlapping brain circuitries [e.g., Calder et al., 2001; Davidson and Irwin, 1999; LeDoux, 2000], any fine-grained patterning of specific emotions being communicated could reconcile current theoretical and empirical discrepancies. By evaluating the results of conventional activation-based functional magnetic resonance imaging (fMRI) and multivariate pattern analysis, we aim to close this gap.

Early behavioral work on emotional vocal [Pittam and Scherer, 1993] and facial expressions [Ekman and Friesen, 1976] has shown that some emotion categories are easily confused, whereas others are well recognized. "Surprise," a brief and involuntary response to an unexpected event, is often ambiguous. It may be immediately followed by expressions of "joy" or "fear" [Ekman and Friesen, 1976]. Recognition of "surprise" expressions may thus rely on the situational context [Davidson, 2000]. However, emotions such as "anger" or "fear" are acoustically [Banse and Scherer, 1996] and visually [Pittam and Scherer, 1993] distinct and, thus, are more easily recognized even though they may be culturally constrained. Furthermore, conventional fMRI results reveal overall increases in activation for vocal emotion compared with neutral expressions in a fronto-temporo-striatal network [e.g., Ethofer et al., 2006; Grandjean et al., 2005; Kotz et al., 2003; Wildgruber et al., 2005]. However, to date, there is no differentiation between specific emotional vocal expressions at the level of large-scale human brain signals. Recent applications of multivariate pattern analysis have revealed that information about speech and speaker identity [Formisano et al., 2008] can be decoded from voice-sensitive areas of the auditory cortex (AC). Only one study investigated the encoding of emotion expressions in fine-grained patterns of fMRI signals, but restricted the analysis to voice-sensitive regions of AC [Ethofer et al., 2009].

Here, we use a "searchlight" decoding approach [Haynes et al., 2007; Kriegeskorte et al., 2006] to investigate in an unbiased fashion across the brain whether local activity patterns in other brain regions than the AC encode vocal emotion expressions. We applied fMRI at 3 T and trained a classifier to recognize characteristic signatures of specific vocal emotion expressions. This assesses whether it is possible to read out a person's percept of distinct vocal emotion expressions from their brain activity alone and reveals where this information is neurally and locally encoded in the brain.

METHODS

Participants

Twenty right-handed participants took part in the fMRI study (10 females, mean age = 25.2 ± 2.5 standard deviation, SD). They were native speakers of German and had normal or corrected-to-normal vision and normal hearing. None of the participants had a history of neurological, major medical, or psychiatric disorder. The participants gave informed consent before the experiment.

Stimuli and Design

Five different emotion expressions were used (angry, happy, sad, surprised, and neutral). Five different consonant-vowel syllable pairs (/baba/, /dada/, /fafa/, /lala/, and /tata/) were recorded from five different female speaker of German. Thus, 25 different items (five syllable pairs × five speakers) per emotion expression were generated, i.e., the stimulus material consisted of 125 items in total. The voice material was digitized at a 16-bit/44.1 kHz sampling rate, and the amplitudes were normalized across speakers (with Cool Edit Version 2000) at 75 dB. Results from the acoustical analyses can be found in Table I.

Design and fMRI Measurement

The fMRI session of each participant was divided into six functional runs. Each run comprised of 25 blocks. One

TABLE I. Results of the acoustic analysis of the RAW stimulus material

Condition	Mean duration (s)	SD	Mean intensity (dB)	SD	Mean F_0 (Hz)	SD
Angry	0.53	0.08	71.83	3.42	238.78	26.55
Happy	0.59	0.07	64.71	4.59	290.85	25.93
Neutral	0.64	0.10	65.18	8.06	196.57	9.07
Sad	0.64	0.13	59.60	6.12	224.10	17.18
Surprised	0.54	0.07	67.78	4.14	305.53	28.01

Mean values and standard deviations (SD) for duration (in seconds), intensity (in decibel), and fundamental frequency (in Hz) of each vocal emotion expression.
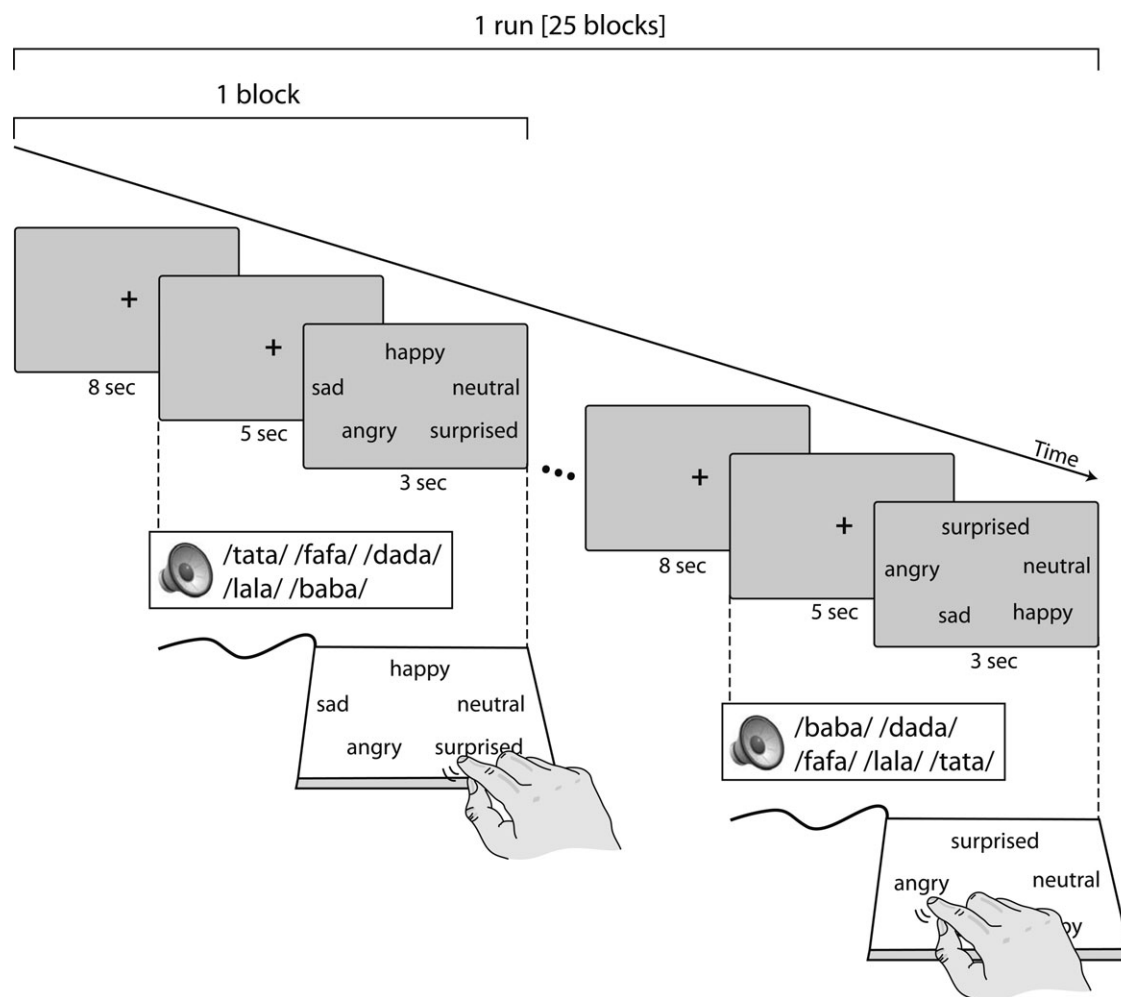
Figure 1.

Schematic trial sequence describing a run, the timing of one of 25 blocks within each run, and the response alternation in the emotional categorization task in which participants decided via button press which emotional category a syllable pair belonged to.

block contained five items from one emotion condition. Syllable pairs and voices were varied in a block: each of the five syllable pairs and each of the five voices were presented in random order per block. In one run, each of the five emotion categories was presented five times (25 blocks). Thus, all 125 items of the stimulus material occurred in each functional run without repetition. This randomization was performed at the level of individual runs, i.e., across both, functional runs and individual participants. At the end of each block, participants categorized the emotional valence of the perceived syllables. The categorization was accomplished with a response-mapping screen and a touchpad on which participants pointed with the right index finger. At the response-mapping screen boxes with the words expressing the five emotions (angry, happy, sad, surprised, and neutral) were displayed in a "pentagon manner" (equal distance between neighboring

words, displayed in the center of the screen, see Fig. 1). As can be seen in Figure 1, the functional runs consisted of 25 repetitions of the following three segments: (1) 8 s of rest (fixation cross), (2) 5 s of auditory stimulation, and (3) 3 s for the behavioral response.

To exclude a possible confound of a covert motor response preparation, the position of the labels for the five emotion categories was randomized from trial to trial. Thus, the motor responses used to indicate the same emotion category differed across blocks. Participants used the touchpad to point at the boxes with the emotion words on the screen. The selection of an emotion expression was indicated by a color change of the box (from black to red). Before the actual fMRI session, participants practiced the use of the touchpad in the scanner for ∼ 8 min. Each syllable pair was presented with a stimulus onset asynchrony of 1 s; thereafter, participants had 3-s time for a categorization.

Imaging was performed on a 3 T scanner (Medspec 30/100/Bruker, Ettlingen). Stabilization cushions were used to reduce head motion. An MDEFT (data matrix = 256 × 256, time of repetition (TR) = 1.3 s, and echo time (TE) = 7.4 ms) and an EPI-T1 (TE = 14 ms and TR = 3 s) sequence were used. Functional MRI volumes (18 slices, TR = 2 s, TE = 30 ms, and field of view (FOV) = 19.2 cm) at an isotropic resolution of 3 × 3 × 3 mm^3 covering prefrontal, temporal, and parts of the parietal and occipital cortex were acquired. Slices of anatomical and functional images were positioned parallel to AC-PC. Six functional runs of 208 volumes were collected for each participant.

Univariate Analysis

Data processing was performed using SPM2 (http://www.fil.ion.ucl.ac.uk/spm). The first five images of each run were deleted to avoid magnetic saturation effects. After standard motion correction of the functional EPI volumes, a rigid body coregistration of the EPI volumes to a within-session whole-brain EPI (42 slices, 3 × 3 × 3 m^3 voxel size, TR = 3 s, TE = 30 ms, and FOV = 19.2 cm) was performed without reslicing the functional volumes. Then, the transformation matrix for the normalization of the whole-brain EPI to the standard MNI EPI template was calculated. This matrix was finally applied to the coregistered functional EPI volumes, and the volumes were resliced for further analysis.

The first analysis was designed to identify brain regions, in which activity was significantly increased during the presentation of the vocal emotion expressions. The analysis was performed with a general linear model (GLM) as implemented in SPM2. Before GLM analysis, images were spatially smoothed with a 7-mm full-width-at-half-mean (FWHM) isotropic Gaussian Kernel to account for anatomical variability across participants and to satisfy the assumptions of the Gaussian random field theory [Worsley et al., 1996]. In the statistical model, boxcar regressors for every emotion condition (a total of 5 s of auditory stimulation) were convolved with a canonical hemodynamic response function. Baseline contrast images (i.e., contrasts for each emotion category against baseline across functional runs, data not shown here) as well as differential contrast images against neutral condition were generated for each participant (i.e., four contrast images per participant across functional runs: angry > neutral; happy > neutral; sad > neutral; and surprise > neutral, see Fig. 3). The single-subject differential contrast images were entered into a second-level random effects analysis for each of the contrasts. The group analysis consisted of a one-sample t-test across the contrast images for all participants to determine whether observed differences between conditions were significantly larger than zero.

## Searchlight Mapping and Pattern Classification Analysis

In the absence of statistically significant activation differences in the univariate analysis, information about distinct vocal emotion expression may still be encoded by fine-grained differences between local cortical activity patterns. This has been shown for other cognitive domains when using a multivariate pattern classification analysis [Haynes and Rees, 2005, 2006; Haynes et al., 2007; Kalberlah et al., 2011; Kamitani and Tong, 2005; Kriegeskorte et al., 2006; Polyn et al., 2005; Soon et al., 2008]. We, therefore, also conducted a multivariate pattern classification analysis with the goal to identify brain regions, which contain distinct information about the type of vocal emotion expression from which a stimulus was chosen. We used a searchlight decoding analysis [Haynes et al., 2007; Kriegeskorte et al., 2006] that assesses the decodable information contained in each local spherical cluster of voxels, separately for each position in the brain. For each position in the brain (i.e., for all in-mask voxels as obtained by conventional SPM GLM analysis), we defined a small spherical cluster of voxels with a radius of 9 mm. For this subset of voxels, we extracted the fMRI signal separately for each emotion and each run. Then, we trained a linear support-vector-classifier (LIBSVM implementation, http://www.csie.ntu.edu.tw/~cjlin/libsvm/) to decode the individual vocal emotion expressions using the data from all but one run and tested this classifier on the data of the remaining run. In a cross validation this was repeated run times. The mean prediction accuracy directly yields a measure for the emotion-specific information contained in local activity patterns within a searchlight centered at that specific position in the brain. The analysis was then repeated for each position in the brain yielding a three-dimensional map of local decoding accuracy. In a next step, these decoding accuracy maps were spatially normalized to a template brain (Montreal Neurological Institute EPI template), spatially smoothed with a 6-mm FWHM Gaussian kernel, and then entered into a second level t-test to identify brain regions from where prediction was significantly above chance level across all participants.

RESULTS

Behavioral Results

A one-way ANOVA with the factor emotion (angry, happy, sad, surprise, and neutral) was conducted to investigate whether participant's categorization of the four emotion and neutral expressions significantly differed from each other. Overall, the ability of the participants to determine the correct emotional expression was very high: 81.4% of the stimuli were judged correctly (SD = 31.6). However, these values differed between the emotion categories as confirmed by the main effect of emotion [$F(4,76) = 10.46$, $P < 0.001$ (Greenhouse-Geisser corrected)]. Participants showed the best performance in categorizing angry (88.3%) and surprise (90.8%) expressions. The categories sad (82.5%), happy (73.2%), and neutral (72.1%) received significantly lower categorization values. Paired t-tests corroborate these results: categorization values of angry expressions were significantly
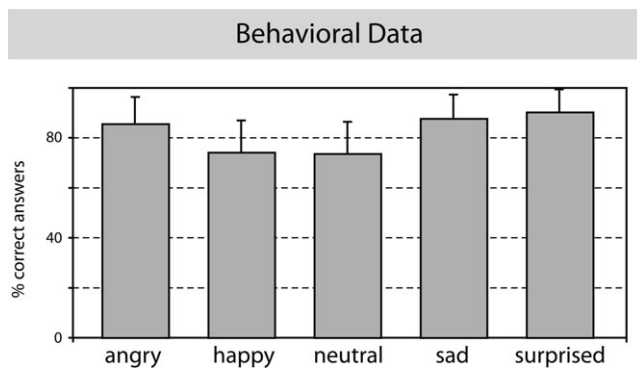
Behavioral Data

Figure 2.

Behavioral response pattern in % correct for each emotional category.

different from happy ($t(19) = 14.7$, $P < 0.001$), sad ($t(19) = 3.3$, $P < 0.001$), and neutral expressions ($t(19) = 16.7$, $P < 0.001$). Furthermore, categorization values of surprise expressions significantly differed from those of happy ($t(19) = 30.2$, $P < 0.001$), sad ($t(19) = 5.5$, $P < 0.001$), and neutral expressions ($t(19 = 29.2$, $P < 0.001$). There was no difference between surprised and angry expressions ($t(19) = 1.5$, ns) and happy and neutral expressions ($t(19) = 0.4$, ns). Moreover, the surprise expressions differed from happy ($t(19) = 30.2$, $P < 0.001$) and from neutral expressions ($t(19) = 29.2$, $P < 0.001$; see Fig. 2).

Univariate Analysis

The overall contrast (all vocal expressions versus baseline, data not shown) confirmed a well-known network activation including prefrontal, temporal, occipital, and subcortical brain regions. However, when conducting direct contrasts between the categories (angry > neutral, happy > neutral, sad > neutral, and surprise > neutral), there was no activation that survived the FDR-corrected threshold of $P < 0.01$.

Thus, with conventional activation-based measures no differences between emotions are discernable. At a relaxed criterion ($P < 0.001$, uncorrected) the comparison between each vocal emotion expression and the neutral expression revealed differences in several prefrontal and subcortical brain regions (see Fig. 3): (1) angry > neutral in the right putamen and the bilateral superior middle frontal gyrus, (2) happy > neutral in the left inferior frontal gyrus (IFG), (3) sad > neutral in the bilateral caudate and the left anterior cingulate cortex [spreading into the superior middle frontal gyrus (MFG)], and (4) surprised > neutral in the bilateral hippocampi, the left middle temporal gyrus (MTG), and the anterior cingulate cortex.

Multivariate Pattern Analysis

In the absence of overall statistically significant activation differences between the individual emotions, informa-

tion about the vocal emotion expression may still be encoded by fine-grained differences between local cortical activity patterns. This has been shown for other cognitive domains such as perception [Haynes and Rees, 2005, 2006; Kamitani and Tong, 2005; Kriegeskorte et al., 2006], memory [Polyn et al., 2005], intentions [Haynes et al., 2007; Soon et al., 2008], and attention [Kalberlah et al., 2011]. We, therefore, proceeded with a multivariate pattern classification analysis to identify brain regions that contain distinct information about the type of emotion expression from which a stimulus was chosen. Classification accuracy of different types of emotion expressions was significantly above chance in the following right hemisphere regions: posterior superior temporal gyrus (pSTG) extending to the anterior superior temporal sulcus (aSTS), the cingulum, the anterior insula and the adjacent frontal operculum (IFO), the IFG, and the MFG. Encoding in the left hemisphere was above chance in the anterior and posterior portion of the posterior MTG (pMTG) and the posterior cerebellum (pCE; see Fig. 4, Table II).

Then, in a closer look, we asked whether specific vocal emotion expressions could be classified more reliably than other expressions. Thus, for the informative brain areas, we compared the accuracy with which one vocal emotional expression could be distinguished from all
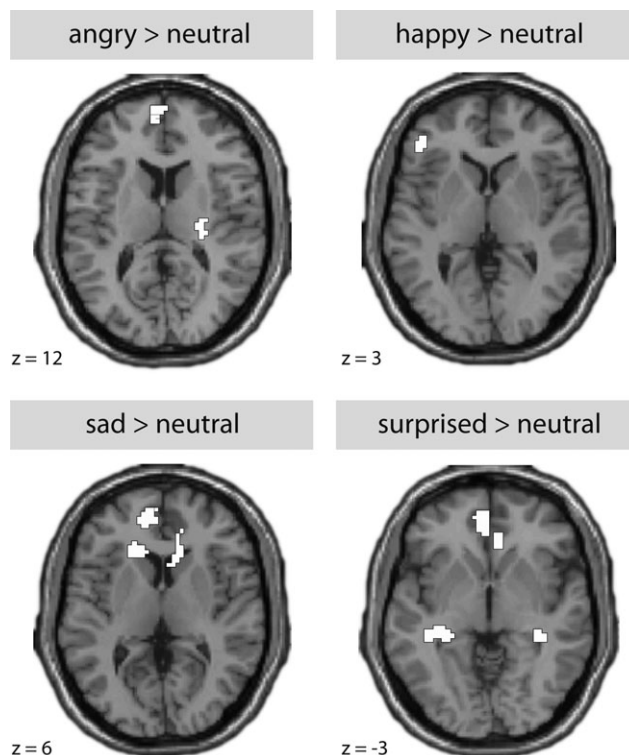
Figure 3.

The results of the GLM analysis are displayed in axial views (uncorrected: $P < 0.0001$) contrasting each emotional against neutral vocal expressions.
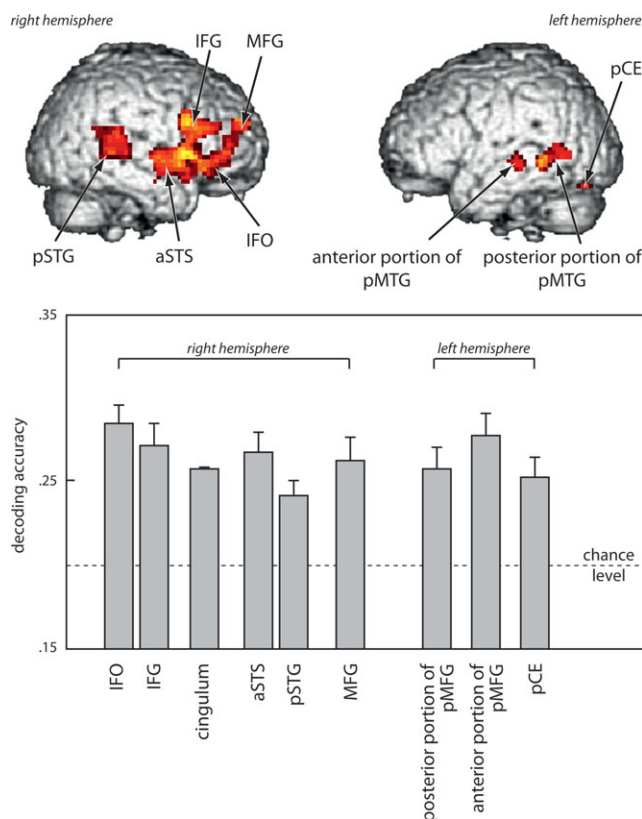
Figure 4.

Three-dimensional rendering of regions where decoding accuracies were significantly above chance (FDR corrected, $P < 0.01$, 10 voxel threshold, see also Table II). Decoding accuracies at peak voxels are plotted for each reported brain area ± standard error across participants. Abbreviations: IFO, anterior insula/frontal operculum complex; IFG, inferior frontal gyrus; aSTS, anterior superior temporal sulcus; pSTG, posterior superior temporal gyrus; MFG, middle frontal gyrus; pMFG, posterior middle frontal gyrus; pCE, posterior cerebellum.

others (pairwise classification), respectively, for each single condition. These classifications were calculated for single searchlights centered at the peak voxels of the investigated areas. Importantly, whole-brain pairwise classifications for all emotions showed accuracies significantly above chance level at all investigated locations ($P < 0.01$, FDR corrected across brain). These results confirm that the information contained in the identified brain areas is sufficient to distinguish single vocal emotion expressions. Nevertheless, when we compare the classification performances with the mean performance of all pairwise classifications, we find a significantly enhanced encoding for surprise expressions in the right IFG and reduced encoding of anger in the IFO ($P < 0.05$). This is visualized by the polar plot "fingerprints" shown in Figure 5. To allow for a direct comparison between areas, the mean classification accuracy for all five pairwise classifications was set to 1 for each area.

DISCUSSION

Our results of the present multivariate pattern classification confirm that vocal emotion expressions materialize from the common encoding of suprasegmental information along the right posterior to anterior STG/STS extending via the right IFO complex into frontal cortex. Furthermore, affective information was also decoded from left hemispheric activity patterns in regions known to support the processing of syllable sequences. The present findings corroborate but critically extend previously reported fMRI evidence [Ethofer et al., 2009; Grandjean et al., 2005; Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006] on vocal emotion expressions as outlined below.

Encoding of Suprasegmental Information in the Right STS and STG

The presence of information about vocal emotion expressions in the right pSTG and aSTS supports the

TABLE II. Results of the decoding analysis

Label	Peak voxel (MNI space)	z-score	Decoding accuracy (chance level at 0.2)	Extend (voxel)
Right hemisphere				
IFO	51/9/0	5.14	0.285 ± 0.011	739
IFG	54/18/21	4.90	0.272 ± 0.013	
Cingulum	6/21/27	4.69	0.258 ± 0.001	54
aSTS	51/–36/15	4.64	0.268 ± 0.012	174
pSTG	54/–51/21	4.22	0.242 ± 0.009	
MFG	30/54/24	4.16	0.263 ± 0.014	127
Left hemisphere				
Posterior portion of pMTG	–48/–48/–3	4.61	0.258 ± 0.013	111
Anterior portion of pMTG	–63/–27/–6	4.27	0.278 ± 0.013	33
pCE	–18/–78/–21	4.04	0.253 ± 0.012	29

Five-class decoding of vocal emotion expressions (angry/happy/sad/surprised/neutral), FDR corrected, $P < 0.01$, 10 voxel extent threshold (see also Fig. 4).
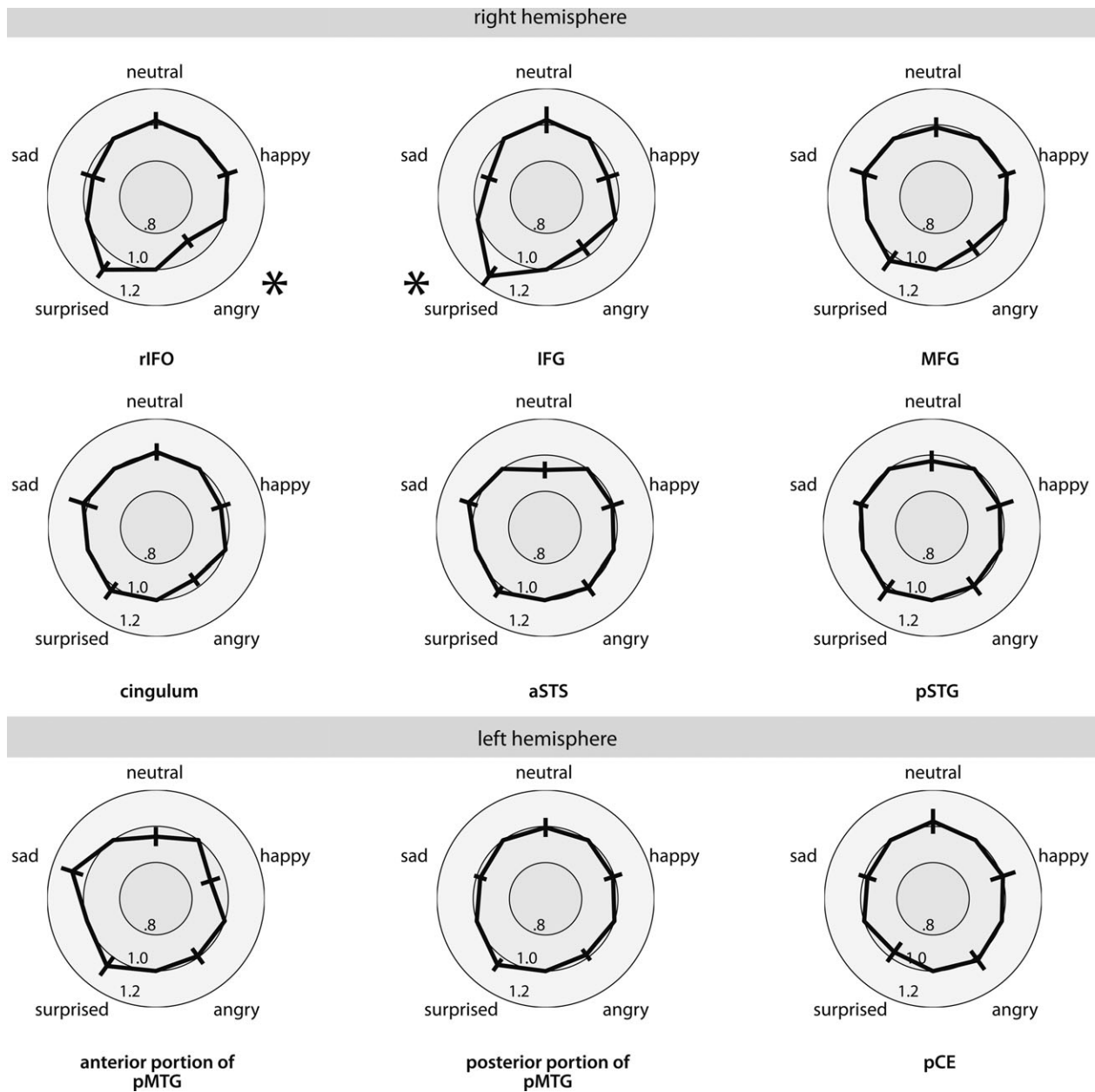
Figure 5.

Relative prediction accuracies of the single conditions versus all other conditions (pairwise classification) referring to mean accuracy across all pairwise classifications. Prediction accuracies were calculated for searchlights centered at the peak voxel of each region observed in five-class decoding. *Indicates a significant difference compared with mean accuracy (*P* < 0.05).

functional relevance of these areas in the encoding of suprasegmental (i.e., the intonation contour of nonverbal expressions) information as a salient and acoustically complex emotional event [Grandjean et al., 2005; Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006]. In line with previous data on voice-specific activation along the right STS [Formisano et al., 2008; but see bilateral decoding for vowels], the encoding of suprasegmental properties and emotional saliency may reflect acoustic parameters that are both emotion and voice specific. For example, activation of the right middle STS is linked to acoustic properties of the voice [Kriegstein and Giraud, 2004] or emotional enhancement of the voice [Grandjean et al., 2005], whereas the right aSTS/STG has been tied to

paralinguistic aspects of voice processing [Belin et al., 2004], voice recognition [Kriegstein and Giraud, 2004], and voice prototypicality [Lattner et al., 2005]. Also, the right pSTS/STG responds to voice familiarity [Kriegstein and Giraud, 2004] and spectral properties of the voice [Lattner et al., 2005]. Thus, previous neuroimaging evidence on emotional vocalization and intonation [Ethofer et al., 2009; Griffiths et al., 1997; Kotz et al., 2003, 2006; Sander and Scheich, 2005; Schirmer and Kotz, 2006; Wildgruber et al., 2006] as well as pattern encoding [Ethofer et al., 2009] may reflect voice-specific representation of nonverbal emotional speech sound sequences [see Belin et al., 2004 for an alternative view]. Furthermore, the current encoding findings also fit lesion and nonhuman primate data. For instance, lesion studies have implicated the right posterior perisylvian cortex in emotional vocalization [e.g., Adolphs et al., 2001; Borod et al., 2002; Heilman et al., 1984]. Reporting vocal primate data, Remedios et al. [2009] discussed the option of a "vocalization-related pathway" that includes the lateral belt [Tian et al., 2001], the STS [Russ et al., 2008], the aSTS [Wang, 2000], the temporal polar region [Petkov et al., 2008], the anterior insula [Remedios et al., 2009], and the ventrolateral prefrontal cortex [Romanski and Goldman-Rakic, 2002; Romanski et al., 2005; Russ et al., 2008] in con-specific vocalizations. Considering the combined evidence on human and nonhuman emotional and nonemotional vocalization, the encoding approach supports the notion that the posterior–anterior extension of the right STG/STS may be differentially involved in voice-specific representation that can be modulated by the affective state of a speaker.

Encoding of Vocal Emotion Expressions in the Right Anterior Insula, Frontal Operculum, and IFG

The anterior insula and the adjacent operculum [IFO; see Jabbi et al., 2007] have recently received much attention in social cognition research. The bilateral IFO responds to visual negative and positive gustatory emotions [Jabbi et al., 2007; Wicker et al., 2003] and transforms "observed" to "experienced" emotion [Craig, 2002; Critchley, 2005; Gallese et al., 2004; Singer, 2006]. However, little is known about an auditory "emotional" equivalent. Some evidence relates the right anterior insula not only to "sympathetic" arousal [Craig, 2005], emotionally filtered speech [Kotz et al., 2003], and human laughing and crying [Sander and Scheich, 2005] but also to sound sequences [Griffiths et al., 1997; Rauschecker, 2005], slow click rates [Ackermann et al., 2001], and vocalization in singing [Perry et al., 1999; Riecker et al., 2000]. Bilateral activation of the fronto-opercular complex and the insula is reported for slow prosodic modulations [filtered speech; Meyer et al., 2002] and pitch perception [Wong et al., 2004; Zatorre et al., 1994]. These low level but complex acoustic phenomena may extend from rhythmic representation

[Griffiths et al., 1997] to affective states [Craig, 2004] and nicely complement data on vocal communicative sound representation in the primate insular cortex [Remedios et al., 2009]. The very fact that this brain signature emerges together with patterns in the right temporal lobe also goes nicely hand in hand with primate evidence on con-specific vocalization [Remedios et al., 2009; Russ et al., 2008; Tian et al., 2001; Wang, 2000]. Remedios et al. [2009] speculated that the selective response of insula neurons to con-specific vocalization could reflect the identification of individual vocalizations and consequently a behavioral reaction. This concept is also discussed in human literature [Bamiou et al., 2003; Hickok and Poeppel, 2007] and is supported by clinical evidence on auditory agnosia [e.g., Engelien et al., 1995; Griffiths et al., 1997] and stroke [Cancelliere and Kertesz, 1990; Habib et al., 1995]. Although it is clear that the right anterior insula/operculum together with the right STG/STS play a specific role in the recognition of human vocal emotion expressions, it is less clear how much this pattern relates to the transfer of perceived to experienced emotion [Craig, 2002; Critchley, 2005; Gallese et al., 2004; Singer, 2006] or whether this pattern is a necessary phylogenetic precursor for response behavior as speculated in the primate [Remedios et al., 2009] and the human [Hickok and Poeppel, 2007] literature.

In particular, the latter explanation raises the issue of amygdala involvement that plays a critical role in automatic emotional response behavior across sensory domains [e.g., Fecteau et al., 2007; Grandjean et al., 2005; Koelsch et al., 2006; Olsson and Phelps, 2007; Scott et al., 1997; Wildgruber et al., 2006]. One reason why we do not find encoding in the amygdala may be related to the use of an explicit emotional categorization task, as explicit tasks seem to result in deactivation of the amygdala [Adolphs, 2002; Critchley et al., 2000; Morris et al., 1999]. Contrary, activation in the right inferior frontal gyrus has been reported to enhance during explicit processing of emotional vocalization such as the encoding of emotional evaluation and/or labeling [Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006]. This is the pattern we find in this study. Most importantly and novel in the current results, encoding in the right IFG is specifically enhanced for vocal expressions of "surprise." Surprise as one of the "basic emotions" that can be acoustically and situationally ambiguous [Davidson, 2000; Ekman and Friesen, 1976] in particular, enhances activation in the right IFG as evidenced in our data. Chandrasekhar et al. [2008] defined the activation pattern observed for "surprise" as a reflection of outcome uncertainty that engages lateral orbitofrontal as well as attention-related areas such as the cingulum and precuneus [Crottaz-Herbette and Menon, 2006; Kanske and Kotz, 2011; Small et al., 2003]. Although not statistically significant, the encoding of vocal "surprise" expressions was slightly enhanced in the right cingulum and right MFG possibly reflecting the fact that emotional categorization of "surprise" expressions may entail more attention to alternative outcomes than other vocal emotion expressions.

Left Hemisphere Contributions to Encoding of Vocal Emotion Expressions

Multivariate pattern analysis proved sensitive to encode emotionally intoned nonsense syllable pairs void of lexical-semantic information. If the emotional stimuli had been meaningful words, the integration of vocal emotion expression and lexical meaning should have resulted in left and right IFG encoding [Schirmer and Kotz, 2006; Wildgruber et al., 2006]. The observed pattern in the left posterior and middle MTG may speak for the segmental encoding of the syllable pairs [Desai et al., 2008; Liebenthal et al., 2005]. Interestingly, this left hemisphere pattern goes hand in hand with encoding in the left posterior cerebellum. There are numerous recent reports that cerebellar damage leads to linguistic, cognitive, and affective disturbances [e.g., Baillieux et al., 2008; Fusar-Poli et al., 2009; Imaizumi et al., 1997; Paquier and Marien, 2005; Schmahmann, 2004] and that emotional face processing [Fusar-Poli et al., 2009] and vocal processing [Imaizumi et al., 1997; Wildgruber et al., 2005] recruit the cerebellum (bilaterally and unilaterally). According to Stoodley and Schmahmann [2010], activation of posterior cerebellar lobules in concert with association cortices is consistently seen in studies utilizing emotionally salient stimuli. Thus, the current decoding pattern most likely reflects the sequencing of emotional speech in form of nonsense syllable pairs.

CONCLUSIONS

The application of multivariate pattern analysis convincingly demonstrates that the processing of vocal emotion expressions is supported by a right frontotemporal network [Ethofer et al., 2009; Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006], which extends to (i) the IFO complex, a social resonance structure in primates and humans, (ii) the right IFG with distinct finger printing for certain emotion expressions, and (iii) the left cerebro-cerebellar regions known to be involved in syllable sequencing. The results clearly evidence that the categorization of vocal emotion expressions relies on a broad network that allows decomposing complex emotional expressions [Davidson, 2000; Davidson and Irwin, 1999; Pittam and Scherer, 1993]. Critically though, the current decoding approach opens up an exciting and challenging venue to explore social communication by extending the rather minimal current stimulus set to emotional utterances or combined emotional utterances and dynamic facial expressions in unilateral and bilateral communicative settings (see, for example, Peelen et al., 2010].

ACKNOWLEDGMENTS

REFERENCES

Ackermann H, Riecker A, Mathiak K, Erb M, Grodd W, Wildgruber D (2001): Rate-dependent activation of a prefrontal-insular-cerebellar network during passive listening to trains of click stimuli: An fMRI study. Neuroreport 18:4087–4092.

Adolphs R (2002): Neural systems for recognizing emotion. Curr Opin Neurobiol 12:169–177.

Adolphs R, Tranel D, Damasio H (2001): Emotion recognition from faces and prosody following temporal lobectomy. Neuropsychology 15:396–404.

Baillieux H, De Smet HJ, Paquier PF, De Deyn PP, Mariën P (2008): Cerebellar neurocognition: Insights into the bottom of the brain. Clin Neurol Neurosurg 110:763–773.

Bamiou DE, Musiek FE, Luxon LM (2003): The insula (Island of Reil) and its role in auditory processing. Literature review. Brain Res Brain Res Rev 42:143–154.

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

Beaucousin V, Lacheret A, Turbelin MR, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007): FMRI study of emotional speech comprehension. Cereb Cortex 17:339–352.

Belin P, Fecteau S, Bédard C (2004): Thinking the voice: Neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Borod JC, Bloom RL, Brickman AM, Nakhutina L, Curko EA (2002): Emotion processing deficits in individuals with unilateral brain damage. Appl Neuropsychol 9:23–36.

Calder AJ, Lawrence AD, Young AW (2001): Neuropsychology of fear and loathing. Nat Neurosci Rev 2:352–363.

Cancelliere AE, Kertesz A (1990): Lesion localization in acquired deficits of emotional expression and comprehension. Brain Cogn 13:133–147.

Chandrasekhar PV, Capra CM, Moore S, Noussair C, Berns GS (2008): Neurobiological regret and rejoice functions for aversive outcomes. Neuroimage 39:1472–1484.

Craig AD (2002): How do you feel? Interoception: The sense of the physiological condition of the body. Nat Rev Neurosci 3:655–666.

Craig AD (2004): Human feelings: Why are some more aware than others? Trends Cogn Sci 8:239–241.

Craig AD (2005): Forebrain emotional asymmetry: A neuroanatomical basis? Trends Cogn Sci 9:566–571.

Critchley HD (2005): Neural mechanisms of autonomic, affective, and cognitive integration. J Comp Neurol 493:154–166.

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.

Crottaz-Herbette S, Menon V (2006): Where and when the anterior cingulate cortex modulates attentional response: Combined fMRI and ERP evidence. J Cogn Neurosci 18:766–780.

Davidson RJ (2000): Affective style, psychopathology, and resilience: Brain mechanisms and plasticity. Am Psychol 55:1196–1214.

Davidson R, Irwin W.1999. The functional neuroanatomy of emotion and affective style. Trends Cogn Sci 3:11–21.

Desai R, Liebenthal E, Waldron E, Binder JF (2008): Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci 20:1174–1188.

Ekman P (1992): An argument for basic emotions. Cogn Emot 6:169–200.

Ekman P, Friesen WV (1976): Measuring facial movement. Environ Psychol Nonverb Behav 1:56–75.

Engelien A, Silbersweig D, Stern E, Huber W, Döring W, Frith C, Frackowiak RS (1995): The functional anatomy of recovers from auditory agnosia. A PET study of sound categorization in a neurological patient and normal controls. Brain 118:1395–1409.

Ethofer T, Anders S, Erb M, Herbert C, Wiethoff S, Kissler J, Grodd W, Wildgruber D (2006): Cerebral pathways in processing of affective prosody: A dynamic causal modeling study. Neuroimage 30:580–587.

Ethofer T, Van De Ville D, Scherer K, Vuilleumier P (2009): Decoding of emotional information in voice-sensitive cortices. Curr Biol 19:1028–1033.

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

Formisano E, De Martino YF, Bonte M, Goebel R (2008): "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322:970–973.

Fusar-Poli P, Placentino A, Carletti F, Landi P, Allen P, Surguladze S, Benedetti F, Abbamonte M, Gasparotti R, Barale F, Perez J, McGuire P, Politi P (2009): Functional atlas of emotional faces processing: A voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. J Psychiatry Neurosci 34:418–432.

Gallese V, Keysers C, Rizzolatti G (2004): A unifying view of the basis of social cognition. Trends Cogn Sci 8:396–403.

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.

Griffiths TD, Rees A, Witton C, Cross PM, Shakir RA, Green GG (1997): Spatial and temporal auditory processing deficits following right hemisphere infarction. A psychophysical study. Brain 120:785–794.

Habib M, Daquin G, Milandre L, Royere ML, Rey M, Lanteri A, Salamon G, Khalil R (1995): Mutism and auditory agnosia due to bilateral insular damage--role of the insula in human communication. Neuropsychologia 33:327–339.

Haynes JD, Rees G (2005): Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15:1301–1307.

Haynes JD, Rees G (2006): Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523–534.

Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007): Reading hidden intentions in the human brain. Curr Biol 20:323–328.

Heilman KM, Bowers D, Speedie L, Coslett HB (1984): Comprehension of affective and nonaffective prosody. Neurology 34:917–921.

Hickok G, Poeppel D (2007): The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Imaizumi S, Mori K, Kiritani S, Kawashima R, Sugiura M, Fukuda H, Itoh K, Kato T, Nakamura A, Hatano K, Kojima S, Nakamura K (1997): Vocal identification of speaker and emotion activates different brain regions. Neuroreport 8:2809–2812.

Jabbi M, Swart M, Keysers C (2007): Empathy for positive and negative emotions in the gustatory cortex. Neuroimage 34:1744–1753.

Kalberlah C, Chen Y, Heinzle J, Haynes JD (2011): Beyond topographic representation: Decoding visuospatial attention from local activity patterns in the human frontal cortex. Int J Imag Syst Tech 21:201–210.

Kamitani Y, Tong F (2005): Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kanske P, Kotz SA (2011): Emotion speeds up conflict resolution: A new role for the ventral anterior cingulate cortex? Cereb Cortex 21:911–919.

Koelsch S, Fritz T, von Cramon DY, Müller K, Friederici AD (2006): Investigating emotion with music: An fMRI study. Hum Brain Mapp 27:239–250.

Kotz SA, Meyer M, Alter K, Besson M, von Cramon DY, Friederici AD (2003): On the lateralization of emotional prosody: An event-related functional MR investigation. Brain Lang 86:366–376.

Kotz SA, Meyer M, Paulmann S (2006): Lateralization of emotional prosody in the brain: An overview and synopsis on the impact of study design. Prog Brain Res 156:285–294.

Kriegeskorte N, Goebel R, Bandettini P (2006): Information-based functional brain mapping. Proc Natl Acad Sci USA 103:3863–3868.

Kriegstein KV, Giraud AL (2004): Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22:948–955.

Lattner S, Meyer ME, Friederici AD (2005): Voice perception: Sex, pitch, and the right hemisphere. Hum Brain Mapp 24:11–20.

LeDoux J (2000): Emotion circuits in the brain. Annu Rev Neurosci 23:155–184.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005): Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631.

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002): FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum Brain Mapp 17:73–88.

Morris JS, Scott SK, Dolan RJ (1999): Saying it with feeling: Neural responses to emotional vocalizations. Neuropsychologia 37:1155–1163.

Olsson A, Phelps EA (2007): Social learning of fear. Nat Neurosci 10:1095–1102.

Paquier PF, Marien P (2005): A synthesis of the role of the cerebellum in cognition. Aphasiology 19:3–19.

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

Perry DW, Zatorre RJ, Petrides M, Alivisatos B, Meyer E, Evans AC (1999): Localization of cerebral activity during simple singing. Neuroreport 10:3979–3984.

Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK (2008): A voice region in the monkey brain. Nat Neurosci 11:367–374.

Pittam J, Scherer KR (1993): Vocal expression and communication of emotion. In: Lewis M, Haviland JM, editors. Handbook of Emotion, 3rd ed. New York: Guilford Press. pp 185–197.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005): Category-specific cortical activity precedes retrieval during memory search. Science 310:1963–1966.

Rauschecker JP (2005): Neural encoding and retrieval of sound sequences. Ann N Y Acad Sci 1060:125–135.

Remedios R, Logothetis NK, Kayser C (2009): An auditory region in the primate insular cortex responding preferentially to vocal communication sounds. J Neurosci 29:1034–1045.

Riecker A, Ackermann H, Wildgruber D, Dogil G, Grodd W (2000): Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. Neuroreport 11:1997–2000.

Romanski LM, Goldman-Rakic PS (2002): An auditory domain in primate prefrontal cortex. Nat Neurosci 5:15–16.

Romanski LM, Averbeck BB, Diltz M (2005): Neural representation of vocalizations in the primate ventrolateral prefrontal cortex. J Neurophysiol 93:734–747.

Russ BE, Ackelson AL, Baker AE, Cohen YE (2008): Coding of auditory-stimulus identity in the auditory non-spatial processing stream. J Neurophysiol 99:87–95.

Sander K, Scheich H (2005): Left auditory cortex and amygdala, but right insula dominance for human laughing and crying. J Cogn Neurosci 17:1519–1531.

Scherer KR (1995): Expression of emotion in voice and music. J Voice 9:235–248.

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

Schmahmann JD (2004): Disorders of the cerebellum: Ataxia, dysmetria of thought, and the cerebellar cognitive affective syndrome. J Neuropsychiatry Clin Neurosci 16:367–378.

Scott SK, Young AW, Calder AJ, Hellawell DJ, Aggleton JP, Johnson M (1997): Impaired auditory recognition of fear and anger following bilateral amygdala lesions. Nature 385:254–257.

Singer T (2006): The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. Neurosci Biobehav Rev 30:855–863.

Small DM, Gitelman MD, Nobre AC, Parish TB, Mesulam MM (2003): The posterior cingulate and medial prefrontal cortex mediate the anticipatory allocation of spatial attention. Neuroimage 18:633–641.

Soon CS, Brass M, Heinze HJ, Haynes JD (2008): Unconscious determinants of free decisions in the human brain. Nat Neurosci 11:543–545.

Stoodley CJ, Schmahmann JD (2010): Evidence for topographic organization in the cerebellum of motor control versus cognitive and affective processing. Cortex 46:831–844.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001): Functional specialization in rhesus monkey auditory cortex. Science 292:290–293.

Wang X (2000): On cortical encoding of vocal communication sounds in primates. Proc Natl Acad Sci USA 97:11843–11849.

Wicker B, Keysers C, Plailly J, Royet JP, Gallese V, Rizzolatti G (2003): Both of us disgusted in My insula: The common neural basis of seeing and feeling disgust. Neuron 40:655–664.

Wildgruber D, Riecker A, Hertrich I, Erb M, Grodd W, Ethofer T, Ackermann H (2005): Identification of emotional intonation evaluated by fMRI. Neuroimage 24:1233–1241.

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

Wong PC, Parsons LM, Martinez M, Diehl RL (2004): The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. J Neurosci 24:9153–9160.

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.

Zatorre RJ, Evans AC, Meyer E (1994): Neural mechanisms underlying melodic perception and memory for pitch. J Neurosci 14:1908–1919.