Differential influences of emotion, task, and novelty on brain regions underlying the processing of speech melody

ETHOFER, Thomas, et al.

Abstract
We investigated the functional characteristics of brain regions implicated in processing of speech melody by presenting words spoken in either neutral or angry prosody during a functional magnetic resonance imaging experiment using a factorial habituation design. Subjects judged either affective prosody or word class for these vocal stimuli, which could be heard for either the first, second, or third time. Voice-sensitive temporal cortices, as well as the amygdala, insula, and mediodorsal thalami, reacted stronger to angry than to neutral prosody. These stimulus-driven effects were not influenced by the task, suggesting that these brain structures are automatically engaged during processing of emotional information in the voice and operate relatively independent of cognitive demands. By contrast, the right middle temporal gyrus and the bilateral orbito-frontal cortices (OFC) responded stronger during emotion than word classification, but were also sensitive to anger expressed by the voices, suggesting that some perceptual aspects of prosody are also encoded within these regions subserving explicit processing of vocal [...]
Differential Influences of Emotion, Task, and Novelty on Brain Regions Underlying the Processing of Speech Melody

Thomas Ethofer\textsuperscript{1,2}, Benjamin Kreifelts\textsuperscript{1}, Sarah Wiethoff\textsuperscript{1}, Jonathan Wolf\textsuperscript{1}, Wolfgang Grodd\textsuperscript{1}, Patrik Vuilleumier\textsuperscript{2}, and Dirk Wildgruber\textsuperscript{1}

Abstract

We investigated the functional characteristics of brain regions implicated in processing of speech melody by presenting words spoken in either neutral or angry prosody during a functional magnetic resonance imaging experiment using a factorial habituation design. Subjects judged either affective prosody or word class for these vocal stimuli, which could be heard for either the first, second, or third time. Voice-sensitive temporal cortices, as well as the amygdala, insula, and mediodorsal thalami, reacted stronger to angry than to neutral prosody. These stimulus-driven effects were not influenced by the task, suggesting that these brain structures are automatically engaged during processing of emotional information in the voice and operate relatively independent of cognitive demands. By contrast, the right middle temporal gyrus and the bilateral orbito-frontal cortices (OFC) responded stronger during emotion than word classification, but were also sensitive to anger expressed by the voices, suggesting that some perceptual aspects of prosody are also encoded within these regions subserving explicit processing of vocal emotion. The bilateral OFC showed a selective modulation by emotion and repetition, with particularly pronounced responses to angry prosody during the first presentation only, indicating a critical role of the OFC in detection of vocal information that is both novel and behaviorally relevant. These results converge with previous findings obtained for angry faces and suggest a general involvement of the OFC for recognition of anger irrespective of the sensory modality. Taken together, our study reveals that different aspects of voice stimuli and perceptual demands modulate distinct areas involved in the processing of emotional prosody.

INTRODUCTION

Modulation of speech melody (prosody) provides a powerful cue to express one’s affective state. Previous neuroimaging studies demonstrated that voice-sensitive regions (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000) in the middle part of the right superior temporal gyrus (STG) exhibit stronger responses to a variety of vocally expressed emotions than to neutral prosody (Ethofer et al., 2007; Ethofer, Anders, Wiethoff, et al., 2006; Grandjean et al., 2005). Another brain structure where differential responses to vocally expressed emotions were often reported is the amygdala (Quadflieg, Mohr, Mentzel, Miltner, & Straube, 2008; Fecteau, Belin, Joanette, & Armony, 2007; Ethofer, Anders, Erb, Droll, et al., 2006; Ethofer, Pourtois, & Wildgruber, 2006; Sander et al., 2005; Morris, Scott, & Dolan, 1999; Phillips et al., 1998), although its role in processing of vocal emotions is still controversial, as different studies found either increased (Quadflieg et al., 2008; Fecteau et al., 2007; Sander et al., 2005; Phillips et al., 1998), decreased (Morris et al., 1999), or no differential activity (Wildgruber et al., 2005) in amygdala responses to affective prosody or nonlinguistic vocalizations. In addition, it has been demonstrated that the orbito-frontal cortex (OFC) displays a particular sensitivity to angry expressions in faces (Blair, Morris, Frith, Perrett, & Dolan, 1999) but probably also in voices (Hornak et al., 2003). So far, however, the specific role of these different regions remains unclear, and the exact aspects of affective vocal information encoded in the amygdala, the OFC, and other related brain regions are still largely unknown.

Explicit evaluation of affective prosody has been linked to both right posterior middle temporal gyrus/superior temporal sulcus (MTG/STS; Ethofer, Anders, Erb, Herbert, et al., 2006; Wildgruber et al., 2005; Mitchell, Elliott, Barry, Cruttenden, & Woodruff, 2003) and the bilateral inferior frontal cortex/orbito-frontal cortex (IFG/OFC; Ethofer, Anders, Erb, Herbert, et al., 2006; Wildgruber et al., 2004; Wildgruber, Pihan, Ackermann, Erb, & Grodd, 2002; Imaizumi et al., 1997), as these brain regions show task-related brain activity with stronger responses during judgment of affective prosody than a...
variety of control tasks including judgment of emotional word content (Ethofer, Anders, Erb, Herbert, et al., 2006; Mitchell et al., 2003), phonetic identification (Wildgruber et al., 2005), evaluation of linguistic prosody (Wildgruber et al., 2004), and speaker identification (Imaizumi et al., 1997). The suggestion that these brain regions cooperate during voice processing is supported by functional magnetic resonance imaging (fMRI) data testing for their functional (Oehlser, Wise, Alex Dresner, & Scott, 2007) and effective (Ethofer, Anders, Erb, Herbert, et al., 2006) connectivity, as well as by results from diffusion tensor imaging demonstrating structural connectivity between these areas (Glasser & Rilling, 2008).

Besides the emotional information, novelty constitutes an important factor that determines the behavioral relevance of a stimulus as automatic attention to novel stimuli is crucial to detect possible significant changes in the environment (Sokolov, 1963). Repeated exposure to a certain stimulus leads to a faster and more efficient processing of that stimulus. This phenomenon is called perceptual priming and reflects an implicit form of memory (Tulving & Schacter, 1990). At the neural level, stimulus repetition leads to a decrease in the neural responses, an effect that is termed “repetition suppression” (for a review, see Grill-Spector, Henson, & Martin, 2006). This neural adaptation to repeated stimulus exposure is a robust effect and has been investigated both at the neural level in monkeys, using single-cell recordings (e.g., Sobotka & Ringo, 1996; Li, Miller, & Desimone, 1993), and at the hemodynamic level in humans, using fMRI (e.g., Grill-Spector et al., 1999; Buckner et al., 1995). Across various studies, this effect has been exploited to investigate different aspects of face processing (Henson & Mouchlianitis, 2007; Bunzeck, Schutze, & Duzel, 2006; Hasson, Nusbaum, & Small, 2006; Yi, Kelley, Marois, & Chun, 2006; Soon, Venkatraman, & Chee, 2003; Henson, Shallice, Gorno-Tempini, & Dolan, 2002; Henson, 2000) and voice processing (Dehaene-Lambertz et al., 2006; Hasson et al., 2006; Orfanidou, Marslen-Wilson, & Davis, 2006; Cohen, Jobert, Le Bihan, & Dehaene, 2004; Belin & Zatorre, 2003). Concerning emotional communication, Ishai, Pessoa, Bikle, and Ungerleider (2004) used an adaptation paradigm to examine whether habituation in face-sensitive regions (Haxby et al., 2001; Kanwisher, McDermott, & Chun, 1997) is modulated by emotion displayed in facial expressions. They found that face-sensitive regions exhibit a stronger habituation for emotional than for neutral faces. So far, similar adaptation paradigms have not been used to investigate processing of affective prosody.

In the present study, we investigated whether brain regions showing differential activity to distinctive stimulus features (i.e., angry vs. neutral prosody) are additionally modulated by task demands (i.e., when prosody is relevant vs. incidental); and vice versa, whether regions displaying task-dependent activity are also modulated by stimulus-driven factors. We chose anger as a representative and well-discriminated category of emotion expressed by prosody with strong social relevance. Based on previous neuroimaging studies, we hypothesized that the bilateral STG (Ethofer et al., 2007; Ethofer, Anders, Wiethoff, et al., 2006; Grandjean et al., 2005), the amygdala (Quadflieg et al., 2008; Sander et al., 2005; Morris et al., 1999; Phillips et al., 1998), and the OFC (Blair et al., 1999) should exhibit sensitivity to anger as expressed by prosody, whereas a distinct set of cortical regions comprising the right posterior MTG/STS (Ethofer, Anders, Erb, Herbert, et al., 2006; Wildgruber et al., 2005; Mitchell et al., 2005) and the bilateral IFG/OFC (Ethofer, Anders, Erb, Herbert, et al., 2006; Wildgruber et al., 2002, 2004; Imaizumi et al., 1997) was expected to show task-related brain activity.

Furthermore, we wanted to clarify whether task- and stimulus-driven components of the brain structures subserving prosody recognition are influenced by the novelty of the stimuli and whether habituation in these regions is modulated by task instructions or expressed emotion. To this end, we presented adjectives and nouns spoken in angry or neutral prosody during an event-related fMRI experiment and instructed our participants either to judge the emotion expressed in the voice or to classify the spoken word as adjective or noun. Each stimulus was presented three times. Thus, the main fMRI experiment conformed to a 3 × 2 × 2 factorial adaptation design with presentation (first, second, third), emotion (angry, neutral), and task (emotion classification, word classification) as separate factors. In addition, we also performed an fMRI localizer scan to define voice-sensitive brain regions (Belin et al., 2000) and to verify that stimulus-driven effects due to the expressed affective prosody occurred within these regions.

**METHODS**

**Subjects**

Twelve subjects (6 women, 6 men, mean age = 23.2 years) took part in the behavioral study and 24 subjects (12 women, 12 men, mean age = 26.3 years) were included in the fMRI study. All participants were right-handed native speakers of German and had no history of neurological or psychiatric diseases. Right-handedness was assessed using the Edinburgh Inventory (Oldfield, 1971). The study was approved by the Ethical Committee of the University of Tübingen and conducted according to the Declaration of Helsinki.

**Stimulus Material**

The stimulus set used in the present study comprised 16 adjectives and 16 nouns which were spoken by six professional actors (3 men, 3 women) in neutral or angry prosody (8 adjectives and 8 nouns, respectively). Stimulus examples are provided as supplemental material. These
stimuli were selected from a larger corpus of stimuli and rated in several prestudies (see Ethofer et al., 2007) to ensure that they are emotionally neutral and low arousing with respect to the word content, that they represent the emotional category intended by the actors, and to evaluate them with respect to emotional valence and arousal as expressed by prosody. The gender of the speakers was balanced over prosodic categories (angry, neutral) and word classes (adjectives, nouns). For each stimulus, the mean intensity ($I$) and variability (standard deviation [SD] of $I$, normalized to mean $I$), mean fundamental frequency ($F_0$), and variability of $F_0$ (SD of $F_0$, normalized to the mean $F_0$) were calculated from $I$ and $F_0$ contours determined for the whole stimulus duration using Praat software (version: 4.3.20, www.praat.org; Boersma, 2001). Since a previous fMRI study of our group revealed that mean $I$, number of syllables, and stimulus duration can influence activation within voice-sensitive regions (Wiethoff et al., 2008), stimulus selection aimed at minimizing differences in these parameters over prosodic categories (angry, neutral) and word classes (adjectives, nouns). Valence (ranging from 1 = highly negative to 9 = highly positive) and arousal (ranging from 1 = very calming to 9 = highly arousing) expressed by word content and affective prosody, number of syllables, and acoustic parameters of the stimuli are presented in Table 1. To verify that repeated presentation of these stimuli results in priming with shortening of reaction times and that the tasks to be used in the main fMRI experiment are comparable with respect to task difficulty, we evaluated our stimuli in a behavioral prestudy.

**Experimental Design**

In both the behavioral and the main fMRI experiment, stimuli were presented in an event-related design. To maximize behavioral priming and repetition suppression effects, we limited the maximal time interval between the repetitions by clustering the stimuli in eight segments consisting of 12 trials. Each segment contained two adjectives and two nouns (one spoken in angry and one in neutral prosody, respectively), which were presented three times. Stimulus presentation of the 12 trials within these segments and the order of the eight segments were fully randomized. Each trial started with a visual display of the judgment scale that informed the subject whether the next stimulus should be classified according to the affective prosody (negative, neutral) or word class (adjective, noun). This display was presented for 6 sec and the voice stimuli were presented 2 sec after onset of the scale. Within each segment, subjects had to classify half of the stimuli with respect to affective prosody and the other half according to the word class. The task did not change for the three presentations of each stimulus. Word classification was chosen as non-emotional control task because its task difficulty is close to that of emotion classification (see below results of the behavioral prestudy), whereas gender classification is usually much easier to perform and can result in biases due to different levels of cognitive effort (e.g., Bach et al., 2008). For the sake of statistical power and to have an equal number of response alternatives in the emotion classification and the word classification task, we restricted the present study to angry and neutral prosody. To reduce synchronization of scanner noise and acoustic stimulation, trial onsets were jittered relative to scan onset in steps of 500 msec and the intertrial interval ranged from 8.85 to 12.85 sec. Subjects were instructed that they should classify the stimuli as accurately and quickly as possible, and that they could report their decision during stimulus presentation.

**Table 1.** Valence and Arousal as Expressed by Word Content and Affective Prosody, Number of Syllables, and Acoustic Parameters of the Stimuli

<table>
<thead>
<tr>
<th></th>
<th>Valence (Word Content)</th>
<th>Arousal (Word Content)</th>
<th>Valence (Prosody)</th>
<th>Arousal (Prosody)</th>
<th>Number of Syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjectives (AP)</td>
<td>4.9 ± 0.5</td>
<td>3.2 ± 0.4</td>
<td>2.9 ± 0.5</td>
<td>5.9 ± 0.5</td>
<td>2.8 ± 0.5</td>
</tr>
<tr>
<td>Adjectives (NP)</td>
<td>5.1 ± 0.5</td>
<td>3.3 ± 0.3</td>
<td>4.8 ± 0.2</td>
<td>2.7 ± 0.4</td>
<td>2.5 ± 1.2</td>
</tr>
<tr>
<td>Nouns (AP)</td>
<td>5.1 ± 0.4</td>
<td>2.0 ± 0.5</td>
<td>3.0 ± 0.6</td>
<td>6.0 ± 1.2</td>
<td>3.1 ± 1.0</td>
</tr>
<tr>
<td>Nouns (NP)</td>
<td>5.2 ± 0.4</td>
<td>1.8 ± 0.5</td>
<td>5.1 ± 0.3</td>
<td>3.0 ± 0.4</td>
<td>3.0 ± 1.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean $I$ [a.u.]</th>
<th>SD of $I$ [%]</th>
<th>Mean $F_0$ [Hz]</th>
<th>%SD of $F_0$ [%]</th>
<th>Duration [sec]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjectives (AP)</td>
<td>78.6 ± 0.3</td>
<td>10.9 ± 3.3</td>
<td>154.7 ± 39.7</td>
<td>15.3 ± 6.9</td>
<td>0.84 ± 0.35</td>
</tr>
<tr>
<td>Adjectives (NP)</td>
<td>78.6 ± 0.3</td>
<td>12.7 ± 2.1</td>
<td>143.0 ± 42.6</td>
<td>8.9 ± 4.0</td>
<td>0.82 ± 0.25</td>
</tr>
<tr>
<td>Nouns (AP)</td>
<td>78.5 ± 0.3</td>
<td>10.4 ± 2.3</td>
<td>164.8 ± 47.8</td>
<td>18.3 ± 12.2</td>
<td>0.86 ± 0.31</td>
</tr>
<tr>
<td>Nouns (NP)</td>
<td>78.6 ± 0.3</td>
<td>10.9 ± 3.0</td>
<td>150.1 ± 46.6</td>
<td>12.8 ± 6.3</td>
<td>0.81 ± 0.32</td>
</tr>
</tbody>
</table>

All values represent mean ± standard deviation.

AP = angry prosody; NP = neutral prosody; a.u. = arbitrary units; Hz = Hertz.
In the fMRI localizer scan, a block design with 32 auditory stimulation blocks (8 sec duration) and 16 silent blocks (8 sec duration) was employed. The distribution of the silent blocks was pseudorandomized and presentation of auditory stimulation blocks was fully randomized. Auditory stimulation blocks included 16 blocks with human voices (e.g., speech, sighs, laughs), 8 blocks with animal sounds (cries of various animals), and 8 blocks with modern human environmental sounds (e.g., doors, bells, telephones, cars). (For a detailed description of the stimulus material, see Belin et al., 2000.) Subjects were instructed to close their eyes and to listen to the stimuli. In both fMRI sessions, stimuli were presented binaurally via magnetic resonance-compatiable headphones with piezoelectric signal transmission (Jancke, Wustenberg, Scheich, & Heinze, 2002).

**Analysis of Behavioral Data**

Accuracy rates during the two tasks were compared using two-tailed paired t tests. Reaction times were submitted to a three-factorial ANOVA with repetition (first, second, third), emotion (angry, prosody), and task (emotion vs. word classification) as within-subject factors. All values are given in mean ± standard error of the mean.

**Image Acquisition**

Structural and functional imaging data were acquired using a 1.5-T whole-body scanner (Siemens AVANTO, Erlangen, Germany). A magnetization prepared rapid acquisition gradient-echo sequence was employed to acquire high-resolution (1 × 1 × 1 mm³) T1-weighted structural images (TR = 1980 msec, TE = 3.09 msec, TI = 1100 msec). Functional images were obtained using a multislice echo-planar imaging (EPI) sequence (26 axial slices acquired in descending order, slice thickness 4 mm + 1 mm gap, TR = 2.17 sec, TE = 40 msec, field of view [FOV] = 192 × 192 mm², 64 × 64 matrix, flip angle = 90°, bandwidth 1532 Hz/Px). EPI time series consisted of 490 images for the main experiment and 255 images for the localizer scan. For off-line correction of EPI image distortions, a static fieldmap (36 slices acquired in descending order, slice thickness = 3 mm + 1 mm gap, TR = 487 msec, TE (1) = 5.28 msec, TE (2) = 10.04 msec, FOV = 192 × 192 mm², 64 × 64 matrix) was acquired prior to the functional measurements.

**Data Analysis**

For both the main experiment and the localizer scan, the first five fMRI volumes were discarded from further analysis to exclude measurements that preceded T1 equilibrium. Functional images were analyzed using statistical parametric mapping software (SPM5, Wellcome Department of Imaging Neuroscience, London, UK). For both sessions, preprocessing steps comprised realignment to the first volume of the time series, unwarping by use of a static field map (Andersson, Hutton, Ashburner, Turner, & Friston, 2001), normalization into MNI space (Montreal Neurological Institute, Collins, Neelin, Peters, & Evans, 1994, resampled voxel size: 3 × 3 × 3 mm³) and spatial smoothing with an isotropic Gaussian filter (10 mm full width at half maximum). Statistical analysis relied on a general linear model (Friston et al., 1994). For the main experiment, separate regressors were defined for each of the 12 conditions of the 3 × 2 × 2 factorial design using a stick function convolved with the hemodynamic response function. Events were time-locked to stimulus onset. For the localizer scan, three regressors were defined for presentation of human voices, animal voices, and modern human environmental sounds using a boxcar function with a length of 8 sec that were convolved with the hemodynamic response function. To remove low-frequency components, a high-pass filter with a cut-off frequency of 1/128 Hz was used. Serial autocorrelations were accounted for by modeling the error term as a first-order autoregressive process with a coefficient of 0.2 (Friston et al., 2002) plus a white noise component (Purdon & Weisskoff, 1998).

The following contrast images were calculated for the main experiment:

1. To examine brain regions showing stimulus-driven effects, we compared activation to angry prosody relative to neutral prosody (anger > neutral).
2. To investigate brain regions showing task-driven effects, we compared activation during judgment of affective prosody relative to judgment of word class (emotion > word class). To exclude that task-related differences in brain activation can be solely explained by differences in task difficulty, parameter estimates were extracted from the most significantly activated voxel of every brain region with stronger responses during emotion- versus word-classification task, and then submitted to a regression analysis with reaction time as independent variable. The regression residuals obtained from these analyses were subsequently tested in separate paired t tests to investigate whether task-related brain activity was still significant after neutralization of possible contributions of task difficulty.
3. To identify brain regions showing repetition suppression effects, we compared activation to the first versus second presentation and to the second versus the third presentation of voice stimuli (first > second and second > third, respectively).
4. To identify brain regions where repetition suppression is modulated by affective prosody, we computed the interaction between repetition and emotion. To maximize sensitivity, we used the respective responses of the first and third presentation for this analysis ([angry prosody, first presentation] > [angry prosody, third presentation]) > ([neutral prosody, first presentation] > [neutral prosody, third presentation]).
To investigate brain regions where repetition suppression is modulated by task demands, we computed the interaction between repetition and task. Again, to maximize sensitivity, we used the first and third presentation for this analysis ([judgment of emotion, first presentation > judgment of emotion, third presentation] > [judgment of word class, first presentation > judgment of word class, third presentation]).

Statistical inference was based on second-level random effects analyses. Activations are reported descriptively at a height threshold of $p < .005$ and an extent threshold of $k > 8$ voxels. Significance was assessed at the cluster level with an extent threshold of $p < .05$ (corresponding to a minimal cluster size of 180 voxels), corrected for multiple comparisons across the whole brain. For the activations within the amygdala and the OFC, the search volume was restricted to these brain regions as defined by the automatic anatomical labeling toolbox (Tzourio-Mazoyer et al., 2002) and correction for multiple comparisons was carried out using small volume correction (Worsley et al., 1996).

To examine whether brain regions showing task-driven or stimulus-driven effects are modulated by repeated presentation, parameter estimates from these regions were submitted to repeated measures three-factorial ANOVAs with repetition (first, second, third), emotion (angry, prosody), and task (judgment of emotion > judgment of word class) as within-subject factors.

To investigate whether stimulus-driven effects due to affective prosody are located within voice-sensitive regions, brain activation maps during perception of human voices were contrasted with brain activation maps during perception of animal sounds and environmental sounds. Statistical parametric $t$ maps obtained at the single subject level were obtained at a height threshold of $p < .005$ (uncorrected) and probability maps showing a differential activation at this threshold for (i) at least 50% and (ii) at least 75% of the subjects were calculated across the whole brain.

**RESULTS**

**Behavioral Data**

In the behavioral prestudy, participants identified stimuli during the word classification task with a slightly higher accuracy (92.4 ± 1.8%) than during the emotion classification task (90.1 ± 2.1%). This difference in accuracy was not significant. A three-factorial ANOVA with repetition (first, second, third), emotion (angry, neutral), and task (emotional classification, word classification) as within-subject factors and reaction time as dependent variable yielded a significant main effect of repetition [$F(2, 10) = 29.6, p < .001$]. This was due to an acceleration of reaction times between the first (1.30 ± 0.11 sec) and the second presentations [1.03 ± 0.09 sec, paired $t(11) = 7.8, p < .001$, one-tailed; see Figure 1A]. No such acceleration occurred between the second and the third presentations (1.03 ± 0.10 sec). No main effect of task or emotion was found. A significant interaction between emotion and task [$F(1, 11) = 17.3, p < .01$] was due to faster reaction times for angrily than for neutrally spoken words during the emotion classification task, whereas the reverse pattern with longer reaction times for angry than for neutral prosody was observed during the word classification task. No other first- or second-order interaction was significant.

In the fMRI study, participants judged the stimuli with similar accuracy rates as those in the behavioral prestudy. However, the difference in accuracy rates between the word classification task (92.8 ± 0.7%) and the emotional classification task (89.0 ± 1.6%) was now slightly more pronounced and reached significance [paired $t(23) = 2.6, p < .05$, two-tailed]. A similar pattern of reaction times was found in the fMRI study as in the behavioral prestudy (see Figure 1B). The three-factorial ANOVA on reaction times yielded a significant main effect of repetition [$F(2, 22) = 105.6, p < .001$]: Again, there was an acceleration between the first (1.51 ± 0.07 sec) and the second presentations [1.19 ± 0.06 sec, paired $t(23) = 12.0, p < .001$, one-tailed], but also between the second and third presentations [1.13 ±
ANOVA on responses in these regions revealed a significant main effect of repetition for the bilateral IFG/OFC $[F(2, 22) = 24.3, p < .001$ and $F(2, 22) = 8.5, p < .01$ for the right and left side, respectively] and for the right posterior MTG/STS $[F(2, 22) = 20.9, p < .001]$. Moreover, a significant main effect of emotion was also found for the bilateral IFG/OFC $F(1, 23) = 18.8, p < .001$ and $F(1, 23) = 8.6, p < .01$ for the right and left side, respectively] and the right posterior MTG $[F(1, 23) = 23.2, p < .001]$. The interaction between emotion and repetition was marginally significant for the bilateral IFG/OFC $[F(2, 22) = 3.4, p = .051$ and $F(2, 22) = 3.3, p = .06$ for the right and left side, respectively]. No other first- or second-order interaction was found to be significant in brain regions showing task-related differences in activation.

Comparison of event-related responses during first and second presentations revealed widespread activations in cortical areas including the bilateral STG, the middle and inferior frontal cortex, the parietal and occipital cortex, as well as subcortical structures, such as the thalamus and the caudate nucleus (see Figure 4A, brain areas marked in red and yellow). Regions that showed additionally significant habituation of their hemodynamic responses between the second and third presentations were located in the bilateral inferior frontal and superior temporal cortex (see Figure 4A, brain areas marked in blue and light blue).

A whole-brain analysis revealed that the right and left OFC were the only brain areas showing an interaction between emotion and repetition. This interaction was due to stronger repetition decreases for angry than for neutral prosody (see Figure 4B and Table 2). A similar whole-brain analysis on interactions between task and repetition did not yield any suprathreshold clusters.

**DISCUSSION**

The present neuroimaging study allowed us to disentangle the differential contributions of emotion, task instruction, and stimulus repetition (or novelty) to cerebral responses evoked by affective prosody. Brain imaging results converged with behavioral data to indicate differential effects of stimulus and task factors on voice processing.

**Behavioral Results**

We found reliable evidence for priming with acceleration of reaction times between the first and the second presentations during the behavioral prestudy. This acceleration of the subjects’ responses did not interact with task instructions or emotion expressed by prosody, in agreement with previous behavioral findings for facial expressions (Ishai et al., 2004). Moreover, similar hit rates and reaction times for the two tasks indicated a similar degree of difficulty. However, there was an interaction between emotion and task, reflecting the fact that subjects responded faster to words spoken in angry than in neutral prosody during emotion classification, whereas...
angry prosody resulted in longer reaction times during word classification. A possible explanation for this pattern might be that, in some cases, one can already infer the stimulus’ emotional intention from the first few syllables, whereas the decision that a stimulus is neutral requires listening to the whole word before excluding any emotionality. On the other hand, slower reaction times for angry than for neutral prosody during word classification might reflect an involuntary shift of the subjects’ attention to the emotional information in the voice (Grandjean, Sander, Lucas, Scherer, & Vuilleumier, 2007; Wambacq, Shea-Miller, & Abubakr, 2004).
Table 2. Brain Regions Showing a Main Effect of Emotion, Main Effect of Task, or an Interaction between Emotion and Repetition

<table>
<thead>
<tr>
<th>Anatomical Definition</th>
<th>MNI Coordinates</th>
<th>Z Score</th>
<th>Cluster Size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Main Effect of Emotion (Angry &gt; Neutral)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right STG</td>
<td>57 30 -3</td>
<td>5.53</td>
<td>593*</td>
</tr>
<tr>
<td>Left STG</td>
<td>-57 -15 -3</td>
<td>4.74</td>
<td>387*</td>
</tr>
<tr>
<td>Right IFG/OFC</td>
<td>48 30 -9</td>
<td>3.82</td>
<td>310*</td>
</tr>
<tr>
<td>Left IFG/OFC</td>
<td>-48 27 -9</td>
<td>4.83</td>
<td>267*</td>
</tr>
<tr>
<td>Left Insula</td>
<td>-24 15 -18</td>
<td>3.85</td>
<td>107</td>
</tr>
<tr>
<td>Right Insula</td>
<td>30 6 -18</td>
<td>3.72</td>
<td>70</td>
</tr>
<tr>
<td>Mediodorsal Thalamus</td>
<td>6 3 6</td>
<td>3.46</td>
<td>74</td>
</tr>
<tr>
<td>Left amygdala</td>
<td>-18 -9 -15</td>
<td>3.87</td>
<td>43**</td>
</tr>
<tr>
<td>Right amygdala</td>
<td>18 -9 -18</td>
<td>2.90</td>
<td>9**</td>
</tr>
<tr>
<td><strong>Main Effect of Task (Emotional Classification &gt; Word Classification)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left IFG/OFC</td>
<td>-45 36 -9</td>
<td>5.66</td>
<td>357*</td>
</tr>
<tr>
<td>Right IFG/OFC</td>
<td>45 39 -6</td>
<td>4.20</td>
<td>275*</td>
</tr>
<tr>
<td>Right DMPFC</td>
<td>12 27 48</td>
<td>4.53</td>
<td>173</td>
</tr>
<tr>
<td>Right posterior MTG/STS</td>
<td>51 -33 3</td>
<td>3.09</td>
<td>17</td>
</tr>
<tr>
<td><strong>Interaction between Emotion and Repetition ((Angry First &gt; Angry Third) &gt; (Neutral First &gt; Neutral Third))</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right OFC</td>
<td>54 30 -9</td>
<td>3.93</td>
<td>69***</td>
</tr>
<tr>
<td>Left OFC</td>
<td>-42 30 -15</td>
<td>3.04</td>
<td>31***</td>
</tr>
</tbody>
</table>

Height threshold: \( p < .005 \) (uncorrected), extent threshold \( k > 8 \) voxels.

IFG = inferior frontal gyrus; OFC = orbito-frontal cortex; DMPFC = dorsomedial prefrontal cortex; STG = superior temporal gyrus; MTG = middle temporal gyrus.

\*\( p < .05 \) (corrected at cluster level for multiple comparisons across the whole brain).

\*\*\( p < .05 \) (corrected at cluster level for multiple comparisons within the amygdala).

\*\*\*\( p < .05 \) (corrected at cluster level for multiple comparisons with the OFC).

In general, the same pattern of behavioral responses and reliable priming effects was observed during fMRI scanning. Again, the facilitation of responses with repetition did not interact with task instructions or vocally expressed emotion. However, the difference in hit rates and reaction times between the two tasks was slightly more pronounced in the fMRI study, suggesting that the scanner noise might have impeded judgment of affective prosody more than judgment of the word class. Furthermore, the interaction between emotion and task failed to reach significance in the fMRI study, which was mostly due to a lack of prolongation of reaction times to angrily spoken words during word classification. Importantly, however, hit rates for emotion discrimination were similar for the behavioral and the fMRI study (90.1% and 89.0%, respectively), indicating that this was not due to a general incomprehensibility of affective prosody in the presence of scanner noise. We speculate that the higher level of attention the subjects had to maintain during scanning to successfully complete the task made them less susceptible to involuntary shifts of attention induced by affective prosody.

Brain Areas Showing Stimulus-driven Activity

Brain regions responding stronger to angry than to neutral prosody were situated bilaterally in the middle part of the STG, IFG/OFC, amygdala, insula, and mediodorsal thalamus. Stimulus-driven effects in the bilateral mid-STG strongly overlapped with sensitivity to human voices relative to animal voices or modern environmental sounds. These combined results confirm that a selective enhancement of neural responses by affective prosody occurs within voice-sensitive regions of the associative auditory cortex (Belin et al., 2000). The mid-STG has previously been reported to show stronger responses to a variety of emotional prosodic categories (Ethofer et al., 2007; Ethofer, Anders, Wiethoff, et al., 2006; Grandjean et al., 2005). As in previous studies, this enhancement was particularly prominent in the right hemisphere. However, in the present study, the left mid-STG also showed significant stimulus-driven effects that survived correction for multiple comparisons across the whole brain, suggesting that associative auditory cortices of both hemispheres are involved in decoding affective information from the voice, with the right-hemispheric areas showing only a relative superiority. This interpretation is supported by results from studies of patients with unilateral brain lesions demonstrating that left-hemispheric brain lesions produce impairments midway between healthy controls and patients with right-hemispheric lesions during comprehension of affective prosody (Kucharska-Pietura, Phillips, Gernand, & David, 2005).

Remarkably, our results reveal that brain activity in the bilateral mid-STG was not influenced by task demands. In keeping with previous results (Ethofer, Anders, Wiethoff, et al., 2006), this insensitivity to task instructions accords with the suggestion that the mid-STG constitutes a relatively early stage for the perceptual analysis of voices, whose activity is mainly driven by stimulus features irrespective of whether the subjects’ attention is focused on affective prosody or some other aspect of the stimuli (Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Brain activity in the bilateral mid-STG showed clear repetition suppression effects, however, there was no effect of task instructions or emotion on habituation in this brain region.
Other brain regions that showed a similar response pattern, with clear sensitivity to stimulus-driven factors but insensitivity to task demands, and repetition suppression effects unaffected by either task or emotion, included the bilateral amygdala, insula, and mediodorsal thalami. Neuroimaging evidence for an emotional modulation of voice processing in these brain regions is less established than for the mid-STG. In particular, for the amygdala, conflicting results have been reported with either increased (Phillips et al., 1998) or decreased (Morris et al., 1999) responses to fearful vocalizations (e.g., screams). However, the present study is in line with previous studies reporting stronger amygdala responses to angry relative to neutral prosody in the amygdala (Quadflieg et al., 2008; Sander et al., 2005) and the anterior insula (Quadflieg et al., 2008). Furthermore, increased activity to affective prosody within the mediodorsal thalamus for a variety of emotional categories including anger has been recently reported (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007), suggesting that some enhancement of neural activity might already occur at early processing levels downstream to the auditory cortex.

The limited temporal resolution of fMRI impedes direct inference on the precise temporal distribution of brain activity. However, based on their shared functional properties including stimulus-driven brain activity, insensitivity to task demands, and marked repetition suppression without influences of task or emotion, we speculate that the mediodorsal thalamus, the amygdala, the insula, and the mid-STG constitute an early processing stage during recognition of vocal emotions, which responds to anger prosody in a largely automatic manner.

Brain Areas Showing Task-related Activity

Brain regions showing stronger responses during the emotional classification than during the word classification task were located in the right posterior MTG, in the bilateral IFG/OFC, and in the right medial prefrontal
cortex. Task-related activation within the right posterior MTG is consistent with results from previous fMRI studies, in which participants made explicit judgments on the emotional category (Wildgruber et al., 2005; Mitchell et al., 2003) or emotional valence (Ethofer, Anders, Erb, Herbert, et al., 2006) of auditory stimuli. It has been suggested (Wildgruber et al., 2006) that activation in the right posterior MTG might reflect effortful extraction of suprasegmental features (Poeppel et al., 2004; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Zatorre, 2001) that are characteristic for emotional information in the voice.

Enhanced responses within the bilateral IFG/OFC during explicit categorization of affective prosody is also consistent with previous neuroimaging results (Ethofer, Anders, Erb, Herbert, et al., 2006; Wildgruber et al., 2004; Imaizumi et al., 1997), and with lesion studies demonstrating that damage to either the right or left IFG/OFC can impair comprehension of affective prosody (Hornak et al., 2003). It has been suggested that the involvement of the prefrontal cortex during processing of emotional prosody might reflect engagement of working memory components because activity in the lateral prefrontal cortex increases with working memory load (Mitchell, 2007). However, this result was not specific for emotional prosody because it occurred in a similar manner during lexico-semantic processing. A recent fMRI study demonstrated activation of an auditory–motor mirror system in the premotor cortex during passive listening to various affective vocalizations (Warren et al., 2006). No differential task- or stimulus-dependent responses were found in this brain region in the present study. Suppression of spontaneous orofacial gestures due to the explicit categorization task may offer a possible explanation for this discrepancy.

To our knowledge, no imaging study so far has indicated that the right medial prefrontal cortex is implicated in active judgment of affective prosody. Although this effect is therefore only reported descriptively here, it converges with the view that medial prefrontal regions are critically involved in more subjective appraisal of emotional significance of sensory events, as already found in the visual domain (Posner et al., 2008; Vuilleumier, Armony, & Dolan, 2003; Lane, Fink, Chau, & Dolan, 1997). Reaction times obtained for the emotion classification task were slightly longer than those for the word
classification task, raising the question of whether task-related effects might potentially be biased by differences in task difficulty. However, by using additional regression analyses, we were able to remove all the variance correlating with reaction time, and thus, to demonstrate that, for all regions showing these differences, the task-related activation was still significantly larger for emotion than for word classification. Therefore, it is unlikely that different levels of difficulty can account for the task-related activations found in our study.

In addition to task-related differences, the right posterior MTG/STS and the bilateral IFC/OFC also exhibited stimulus-driven effects, with stronger responses to angry than to neutral prosody. Thus, these results lend support to the notion that neural activity within these cortical areas is not exclusively driven by task demands, but is also modulated by stimulus properties. Furthermore, these findings are in agreement with previous reports on the right MTG/STS and the IFC/OFC, showing differential responses to vocally expressed happiness and anger (Johnstone, van Reekum, Oakes, & Davidson, 2006). The presence of stimulus-driven effects within task-dependent brain areas, albeit at a lower level of significance than in the mid-STG, suggests that enhanced neural responses to affective prosody spread through the neural system beyond early cortical areas engaged by voice analysis, and that emotional features may also boost activity in higher-order cortical areas underlying explicit judgment of prosody.

**General Effects of Stimulus Repetition and Specific Interactions between Repetition and Affective Prosody**

The most significant decrease of hemodynamic responses across the three stimulus presentations was found within the left inferior frontal cortex, a region that has been reported to show repetition suppression effects for various types of stimuli including faces/scenes (Bunzeck et al., 2006), objects (Vuilleumier, Schwartz, Duhoux, Dolan, & Driver, 2005), visually presented words (Meister, Buelte, Sparing, & Boroojerdi, 2007), spoken words/pseudowords (Orfanidou et al., 2006), and spoken sentences (Dehaene-Lambertz et al., 2006; Hasson et al., 2006). A significant habituation of brain responses across the three presentations was also found within bilateral superior temporal cortices in keeping with studies reporting repetition suppression effects in the auditory cortex for repeated exposure to acoustic stimuli (Dehaene-Lambertz et al., 2006; Hasson et al., 2006; Saad et al., 2006). However, comparison of brain responses to the first and second stimulus presentations revealed widespread activations in cortical areas including bilateral sensory cortices in the temporal and occipital lobes, prefrontal and motor cortices, as well as subcortical structures. These findings strongly suggest that the phenomenon of repetition suppression is not restricted to sensory areas, but presumably occurs for many brain regions engaged by a certain stimulus in the context of a given task.

To disentangle habituation effects that are specific for processing of angry prosody, we presented our stimuli within a factorial design. Such factorial habituation designs represent an effective way of isolating habituation effects attributable to certain cognitive factors from other confounding components which might also show habituation (such as phonological, lexical, semantic, working memory, or motor components) as these confounding factors are eliminated in the interaction between repetition and the cognitive factor of interest. The only brain region which showed a specific habituation with an interaction between repetition and emotion as expressed by prosody was found in the bilateral OFC. This interaction was due to the fact that hemodynamic responses in the OFC strongly habituated for stimuli spoken in angry prosody, whereas habituation for neutral trials was much less pronounced. Thus, the OFC showed particularly strong responses to stimuli that were both novel and emotional, suggesting that this region may be essential for prompting the organism to evaluate and respond to stimuli with affective value when these have not been encountered before.

This functional property makes the OFC an ideal site for detection of socially relevant stimuli and adds to lesion studies suggesting that the OFC is critical for normal social functioning (Beer, John, Scabini, & Knight, 2006), especially for correct recognition and appropriate reaction to the emotion of anger (Blair & Cipolotti, 2000). Furthermore, these novel results on vocally expressed anger are in line with a previous neuroimaging study (Blair et al., 1999) showing enhanced responsiveness of the OFC to angry relative to neutral faces. Taken together, these data point to a more general role of the OFC for the detection of negative emotional information irrespective of modality—a hypothesis that is anatomically well supported by the fact that it receives highly processed afferents from all sensory cortices (for a review, see Rolls, 2004) and received recent support from an fMRI study evidencing differential responses of this region in response to valence information expressed by word content (Alia-Klein et al., 2007).

In summary, our results shed new lights on several functional properties of the brain regions implicated in the processing of affective prosody. The bilateral amygdala, insula, mediiodorsal thalami, as well as voice-sensitive regions in the mid-STG, demonstrated stimulus-driven effects with stronger responses to angry than to neutral prosody. These stimulus-driven effects were not influenced by task demands, suggesting that these brain regions represent a processing stage for identification of emotional cues in the voice that operates relatively independent of task-related control. In contrast to this, the right MTG/STS and the bilateral IFG/OFC showed significant task-related increases, but also
responded stronger to angry than to neutral prosody, indicating that some features from early processing levels are also encoded in higher-order cortices that subserve explicit recognition of affective prosody. Habituation of responses due to repeated stimulus presentation was observed in widespread cortical and subcortical regions, demonstrating that repetition suppression is not limited to perceptual stages but involves a distributed system recruited by processing of emotional voices. Importantly, the OFC was the only area that showed an emotion-specific habituation with enhanced responses during initial presentations of angry prosody that strongly decreased across repetitions. This unique pattern with pronounced responses to stimuli that were both novel and emotional points to this region as a key structure for appraisal of socially relevant information and, together with previous findings for angry facial expressions, suggests a major role for the OFC in processing of anger across different sensory modalities. Increased OFC activations to vocal anger in patients with social phobia (Quadflieg et al., 2008) indicate that this might be of clinical relevance for diagnosis and, possibly, also monitoring of treatment of patients with social disorders.

Acknowledgments
This study was supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 550-B10). Reprint requests should be sent to Thomas Ethofer, Laboratory for Behavioral Neurology & Imaging of Cognition, Department of Neurosciences and Clinic of Neurology, University Medical Center of Geneva, 1 Rue Michel-Servet, 1211 Geneva, Switzerland, or via e-mail: Thomas.Ethofer@medecine.unige.ch.

REFERENCES


This article has been cited by:


2. Benjamin Kreifelts, Thomas Ethofer, Elisabeth Huberle, Wolfgang Grodd, Dirk Wildgruber. 2009. Association of trait emotional intelligence and individual fMRI-activation patterns during the perception of social signals from voice and face. *Human Brain Mapping* NA-NA. [CrossRef]