HHS Public Access
Author manuscript
Neuroimage. Author manuscript; available in PMC 2021 April 01.

What you say versus how you say it: Comparing sentence comprehension and emotional prosody processing using fMRI

Anna Seydell-Greenwald, **Catherine E. Chambers**, **Katrina Ferrara**, **Elissa L. Newport**
Center for Brain Plasticity and Recovery, Georgetown University, Building D, Suite 145, 4000 Reservoir Road NW, Washington, DC 20057, USA

Abstract

While language processing is often described as lateralized to the left hemisphere (LH), the processing of emotion carried by vocal intonation is typically attributed to the right hemisphere (RH) and more specifically, to areas mirroring the LH language areas. However, the evidence base for this hypothesis is inconsistent, with some studies supporting right-lateralization but others favoring bilateral involvement in emotional prosody processing.

Here we compared fMRI activations for an emotional prosody task with those for a sentence comprehension task in 20 neurologically healthy adults, quantifying lateralization using a lateralization index. We observed right-lateralized frontotemporal activations for emotional prosody that roughly mirrored the left-lateralized activations for sentence comprehension. In addition, emotional prosody also evoked bilateral activation in pars orbitalis (BA47), amygdala, and anterior insula.

These findings are consistent with the idea that analysis of the auditory speech signal is split between the hemispheres, possibly according to their preferred temporal resolution, with the left preferentially encoding phonetic and the right encoding prosodic information. Once processed, emotional prosody information is fed to domain-general emotion processing areas and integrated with semantic information, resulting in additional bilateral activations.

Graphical Abstract

1. Introduction

More than a century of evidence has demonstrated that the left hemisphere of the human brain is the "eloquent" or "language-dominant" one in the vast majority of humans (Broca, 1861; Wernicke, 1874; Wada & Rasmussen, 1960; Geschwind & Levitsky, 1968; Pascual-Leone et al., 1991; Binder et al., 1995; Just et al., 1996; Bookheimer et al., 1997). However, understanding of verbal communication involves not only the comprehension of words and sentences. Variations in pitch, stress, timbre, and tempo can distinguish between a statement and a question (linguistic prosody) or reveal information about the speaker's emotional state and intent (emotional prosody). It has been suggested that the processing of emotional or affective prosody is lateralized to the right hemisphere (RH; Heilman et al., 1975; Tucker et al., 1977; Ross & Mesulam, 1979; Borod et al., 2002), and moreover that its functional-anatomical organization mirrors that of the left-hemisphere (LH) language system, with comprehension deficits arising from posterior temporal lesions and production deficits arising from inferior frontal lesions (Ross & Mesulam, 1979; Ross, 1981; Gorelick & Ross, 1987; Ross & Monnot, 2008). More recently, dual-stream models of speech processing (Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Fridriksson et al., 2016) have refined this frontal (production) and temporal (comprehension) distinction by proposing multiple substages of speech processing along a ventral (sound-to-meaning) and a dorsal (sound-to-articulation) stream that involve the superior and middle portions of the temporal lobe and posterior frontal and temporo-parietal cortex, respectively. Importantly, there is evidence that a similar dual-stream architecture in the right hemisphere (RH) may underlie prosody processing (Sammler et al., 2015). However, compared to the overwhelming support for left-lateralization of word- and sentence comprehension, the evidence for right-lateralization of emotional prosody processing and its analogous organization to the LH language system[1] is relatively sparse and inconclusive.

---

[1]To be clear, we do not wish to imply that "language" as a whole is confined to the LH. There is ample evidence of RH activation during language tasks that evoke left-lateralized activation, and there are aspects of language (like prosody) that are thought to be predominantly supported by the RH. We use the expression "LH language system/network/activation/areas" merely to denote those LH areas typically and consistently shown in meta-analyses and literature reviews of language processing (e.g., Price, 2010; Friederici, 2011; Hagoort, 2014).

Some lesion studies have concluded that RH lesions impair patients' ability to comprehend and/or produce emotional prosody more than LH lesions (Heilman et al., 1975; Tucker et al., 1977; Bowers et al., 1987; Blonder et al., 1991; Schmitt et al., 1997; Borod et al., 1998; Kucharska-Pietura et al., 2003; Shamay-Tsoory et al., 2004), but others have found no difference between LH and RH lesion effects (Schlanger et al., 1976; House et al., 1987; Twist et al., 1991; Van Lancker & Sidtis, 1992; Pell & Baum, 1997; Pell, 2006). Similarly, studies mimicking lesions with transcranial magnetic stimulation (TMS) have only sometimes suggested RH dominance for emotional prosody (Alba-Ferrara et al., 2012; but see Hoekert et al., 2010; Jacob et al., 2014). A recent meta-analysis concluded that lesions to either hemisphere had detrimental effects on emotional prosody perception, with those from RH lesions being somewhat greater (Witteman et al., 2011).

Functional neuroimaging studies of emotional prosody also paint an unclear picture regarding lateralization and localization of the emotional prosody system. While some report activation predominantly in the RH (George et al., 1996; Buchanan et al., 2000; Mitchell et al., 2003; Wiethoff et al., 2008; Wildgruber et al., 2005), others report bilateral activations (Imaizumi et al., 1997; Kotz et al., 2003; Grandjean et al., 2005; Ethofer et al., 2006, 2012; Beaucousin et al., 2007; Mitchell & Ross, 2008; Alba-Ferrara et al., 2011; Hervé et al., 2012). Moreover, activations are often quite localized, with some studies finding only temporal (Ethofer et al., 2006; Wiethoff et al., 2008; Mothes-Lasch et al., 2011) and others finding only frontal activations (George et al., 1996). Two quantitative meta-analyses (Witteman et al., 2012; Belyk & Brown, 2014) concluded that activations across studies suggest bilateral inferior frontal, but possibly right-lateralized superior temporal involvement in emotional prosody processing. Witteman et al. (2012) speculate that the right-lateralization in temporal cortex may reflect the RH's preference for prosody-relevant acoustic features rather than prosody processing per se. This notion is also reflected in recent multi-step models of emotional prosody processing (Schirmer & Kotz, 2006; Wildgruber et al., 2006; 2009). According to these models, extraction of prosody-relevant features happens predominantly in right temporal cortex, because its longer temporal receptive fields are better suited for processing dynamic pitch than its LH counterpart (Zatorre, 2001; Poeppel, 2003; Boemio et al., 2005). However, following feature extraction, prosody information is then passed on to bilateral inferior frontal cortex.

Claims from functional neuroimaging studies regarding right-lateralization of emotional prosody processing are usually based on the observation that significant activation appears exclusively (or predominantly) in the RH. A more rigorous and direct test of lateralization would be the lateralization index (LI), which quantitatively compares LH and RH activation and is commonly used for establishing lateralization or language activation. However, this measure is not commonly used in emotional prosody research. Similarly, the notion that the emotional prosody system is a RH mirror image of the classic LH language system gains its support from independent observations of RH and LH lesion effects, specifically, from the fact that in both cases, frontal lesions result in production deficits whereas temporal lesions result in comprehension deficits. However, given that brain lesions often affect large areas, this is rather coarse evidence for concluding that the two systems are mirror images of one another. A side-by-side comparison of "language" and "emotional prosody" activation in the same participants would provide a more spatially specific look at this "mirror hypothesis".

To fill these gaps, we performed fMRI on twenty neurologically healthy, right-handed young adults, interleaving conditions with varying demands on the emotional prosody and sentence comprehension systems in a within-subject design and computing LIs to quantify lateralization. During the active emotional prosody condition (Emot), participants heard content-neutral sentences spoken in happy, sad, or angry tone and identified the speaker's emotion. During the sentence comprehension condition (Neut), participants heard the same sentences spoken in neutral tone and identified their semantic content (food, gift, or trip). To control for low-level acoustic differences between emotional and neutral prosody stimuli, we included reversed speech conditions (EmotRev and NeutRev), during which the same stimuli were played in reverse. This rendered the sentences entirely incomprehensible but retained some of the key acoustic properties thought to carry the emotion signal, such as mean and variability of the fundamental frequency (Fairbanks & Pronovost, 1939; Fairbanks, 1940; Lieberman & Michaels, 1962; Williams & Stevens, 1972; Van Lancker & Sidtis, 1992; Banse & Scherer, 1996). A visual-motor control condition, in which no auditory stimuli were presented, served as the baseline.

This within-subject design allows us to test the lateralization hypothesis and the mirror hypothesis in a more direct and quantitative manner than has been done before. Our specific research questions and predictions were: (1) Is emotional prosody processing activation right-lateralized? If so, LIs for emotional prosody (contrasting conditions Emot and Neut) should be significantly below zero. (2) Is lateralization driven by the higher amplitude and pitch modulations characteristic of emotional compared to neutral speech? If so, LIs should be different when these differences are part of the contrast (Emot – Neut) than when they are removed by subtracting the difference between reverse speech conditions (EmotRev – NeutRev) from this contrast. (3) Is it as right-lateralized as sentence processing activation is left-lateralized? If so, absolute LIs should not differ between sentence comprehension (contrasting conditions Neut and NeutRev) and emotional prosody. (4) Are the two mirror images of each other? If so, the same areas that show positive LIs for sentence comprehension should show negative LIs for emotional prosody.

2. Materials and methods

2.1. Participants

Twenty neurologically healthy young adults (12 women, ages ranging from 18 to 31, M = 21.75, SD = 3.5 years) from the Georgetown University community volunteered to participate in the study for moderate financial compensation and gave informed consent to study procedures approved by the Georgetown University Institutional Review Board. All were native speakers of American English; none grew up with significant non-English language influences or was fluent in a second language prior to the age of 12.

2.2. Stimuli and design

2.2.1. Conditions and stimuli—Five experimental conditions (Figure 1A) were presented in counterbalanced order across three five-minute functional MRI runs. In the visuomotor control condition (VisMot), participants were instructed to fixate a dot at screen center and push a button held in their right hand whenever a "thumbs up" symbol appeared,

but to withhold button pushes if the "thumbs up" symbol was crossed out. This condition, which matched the other conditions with respect to the timing of visual stimulation and button pushes, served as the baseline.

In the emotional forward speech condition (Emot), participants heard an emotional female voice utter content-neutral sentences with happy, sad, or angry intonation and then pushed a button if the line drawing shown after the utterance matched the speaker's emotion. (A full list of sentences, along with the results of a validation study which confirmed that the intended emotion was indeed conveyed by each emotional rendering, can be found in supplementary materials S1.) The mapping of line drawing to emotion (a sun for "happy", tear drops for "sad", and a boxing glove for "angry") was trained prior to scanning and was understood without difficulty by all participants. We chose object symbols rather than emotional faces or words to avoid evoking brain activation associated with facial emotion recognition or with reading. Sentence content was chosen such that it could plausibly be spoken with any of the emotions (e.g., "Grandpa took me to the library"). During task training, participants heard an example sentence in all four versions (neutral, happy, sad, angry) to illustrate the need for basing their emotion judgments on prosody rather than sentence content.

In the neutral forward speech condition (Neut), participants heard the same female voice utter the same sentences in a neutral tone and pushed a button if the line drawing matched the content of the sentence. Again, participants were familiarized with the meaning of the line drawings (a plate and utensils for "food", a wrapped gift box for "gift", and a minivan with suitcases on the roof for "trip") prior to the scan.

In the emotional reversed speech (EmotRev) and neutral reversed speech (NeutRev) conditions, participants heard the same utterances as in the Emot and Neut conditions, respectively, but made unintelligible by playing the sound file in reverse. These served as acoustic controls for the forward conditions. Following each utterance, a line drawing of a loudspeaker appeared with a question mark superimposed, prompting participants to push a button if they had heard a beep at the end of the utterance. Beeps were played on 50% of trials, just as there were 50% "matches" in the Emot and Neut conditions and 50% "thumbs-up" in the visuomotor control condition.

2.2.2.   Stimulus generation and presentation—Auditory stimuli were recorded in stereo by a native speaker of American English and edited in Audacity ® (Version 2.1.1.) to remove background noise, normalize peak volume, and equate duration. Prior to editing, the sound clips differed moderately in duration. If they were between 2600 and 2800 ms, the duration was not changed; files with durations outside this window were adjusted to within this window using Audacity's "Change Tempo" option, which alters the speed of the file without affecting the pitch. None of the original recordings were more than 500 ms longer or shorter than the desired duration, resulting in only moderate "speeding up" and "slowing down" of stimuli during editing. The resulting sound files were screened by two independent listeners to ensure that they were easy to comprehend and subjectively sounded like "normal" speech (rather than unnaturally fast or slow). Reversed versions of the sound files were created using Audacity's "Reverse" effect. The "beep" played at the end of 50% of

reversed speech trials was generated with Audacity's "Tone" function as a 440 Hz sine wave of 100 ms, of which the initial and final 20 ms were edited with the "Fade In" and "Fade Out" effects, respectively, to avoid sudden onsets. The maximum amplitude of the beep was set to 80% of the maximum amplitude of the speech stimuli.

During scanning, auditory stimuli were presented over Sensimetrics insert earphones. To ensure the best possible listening conditions despite the scanner noise, participants wore Bilsom ear defenders over the ear phones, and the sound level was adjusted to the maximum level comfortable for the participant, using an Echo AudioFire sound card and SLA1 amplifier.

Visual stimuli were simple line drawings presented in black on a gray background to minimize eye strain. An Epson PowerLite 5000 projector was used to cast the images onto a screen mounted at the back of the scanner bore, which the participants could see via a slanted mirror mounted on the head coil.

2.2.3.   Trial timing—Conditions were presented in blocks of 24 seconds, each block comprising 6 trials. Each trial began with presentation of the fixation spot at screen center (Figure 1B), which coincided with the onset of the auditory stimulus in all except the visuomotor baseline condition. The decision screen was presented 3000 ms after trial onset and stayed on the screen for 1000 ms, resulting in a total trial duration of 4000 ms. For conditions EmotRev and NeutRev, the beep to which participants were to respond by pushing a button appeared simultaneously with the decision screen, ensuring that in all conditions, participants had to wait for the decision screen to determine the correct response. Participants were familiarized with task and trial timing prior to scanning, and reaction times during task training showed that the response window of 1000 ms was sufficiently long not to cause "time out" errors (which would be indistinguishable from intentional withholding of a button push).

2.2.4.   Block timing—Conditions were presented in a block design, with block order counterbalanced across three 5-minute functional runs. Each run comprised eleven 24-second blocks, each block was preceded by an instruction screen shown for 3 seconds, and the final block was followed by a "Thank you" screen. All runs had a visuomotor control block at the beginning, in the middle, and at the end, and one block of each of the 4 other conditions in the first and in the second half. The order of the conditions was counterbalanced across runs and participants. The time course for an example run is illustrated in Figure 1C.

2.3.   Imaging procedures

2.3.1.   MRI setup—Functional MRI data were acquired on a research-dedicated 3 Tesla Siemens Trio Tim scanner using a 12-channel birdcage head coil. Participants lay on their backs holding a Cedrus fiber optic button box in their right hand, with their legs supported by a wedge pillow and a blanket provided for warmth if requested. Foam cushions were used to stabilize the position of the head inside the headcoil and to minimize motion. Stimulus presentation and response collection were coordinated using E-Prime software (version 2.0).

2.3.1. Imaging sequences—Following a 1-minute "scout" scan (a low-resolution used to align slices for subsequent scans to be parallel to the AC-PC plane), we acquired three functional MRI runs (T2*-weighted) using echo-planar imaging with the following parameters: whole-brain coverage in 50 horizontal slices of 2.8 mm thickness, acquired in descending order with a distance factor of 7% between slices, matrix 64×64, effective voxel size 3×3×3 mm^3, repetition time (TR) of 3 seconds, echo time (TE) 30 ms, flip angle 90 degrees. Each run consisted of 100 volumes (of which the first 2 were discarded during preprocessing to allow for T1 saturation) and lasted 5 minutes.

At the end of the imaging session, we acquired a high-resolution anatomical scan (T1-weighted) while participants either rested or watched part of a nature documentary. The parameters for the MPRAGE were: whole-brain coverage in 176 sagittal slices, matrix 256×256, voxel size 1×1×1 mm^3, TR = 2530 ms, TE = 3.5 ms, inversion time = 1100 ms, flip angle = 7 degrees. This scan took 8 minutes to complete.

2.4. Imaging Analysis

Neuroimaging data were analyzed using BrainVoyager 20.6 (BVQX 3.6) and Matlab (R2017a) in combination with the BVQX toolbox.

2.4.1. Preprocessing—Preprocessing for anatomical data was performed using BrainVoyager's built-in functions for automatic inhomogeneity correction and brain extraction, followed by manual identification of landmarks and transformation into Talairach space using 9-parameter affine transformation.

Functional data from each run were separately co-registered with the participant's anatomical data in native space using 9-parameter gradient-based alignment and underwent slice-time correction, linear trend removal, and 3D motion correction to the first volume of the run using rigid-body transformation. We then transformed the functional data into Talairach space using the 9-parameter affine transformation determined for the anatomical data. Finally, the data were smoothed in 3D with a Gaussian kernel of 8 mm full width at half maximum.

2.4.2. GLM fitting—All functional runs from all participants were combined and fitted with a mixed-effects general linear model (fixed effects at the single-participant level and random effects at the group level). In addition to nuisance predictors capturing run- and participant-specific effects as well as motion-related effects (derived from the z-transformed 6-parameter output of the 3D motion estimation), the model had 5 stimulation-related predictors: one for each of the auditory conditions described above (Emot, Neut, EmotRev, NeutRev), and one to capture signal changes common to instruction periods. (The latter is not of interest for our analyses but serves to capture variance that would otherwise add noise.) These boxcar predictors were convolved with a standard hemodynamic response function (two gamma, peak at 5 seconds, undershoot peak at 15 seconds) prior to inclusion in the model. The visuomotor control condition periods were used as the model's baseline and not explicitly modeled. As part of the GLM estimation process, voxel time courses were normalized using percent signal change transformation and corrected for serial autocorrelations (AR2).

2.4.3.   Whole-brain analyses and thresholding of contrast maps—From the GLM results, we generated contrast maps comparing the beta values for the different conditions to each other for each voxel. The resulting t-maps were thresholded at $p < 0.001$ (single-voxel threshold). For each map, we then used the Cluster-Level Statistical Threshold Estimator plugin, which employs Monte-Carlo simulations to determine what cluster sizes would be expected in a random dataset of comparable size and smoothness. Following 1000 simulations, we set the cluster size threshold to retain only clusters whose size was unlikely to occur by chance ($k < 0.05$).

2.4.4.   Contrasts of interest—Our primary contrast of interest was [Emot – Neut], which identifies those voxels with a significant preference for emotional over neutral forward speech. We will refer to these voxels collectively as the "emotional prosody network." Based on the literature, activations associated with emotional prosody were expected predominantly in the right frontal and temporal cortex. Importantly, [Emot – Neut] activation could be driven by a number of factors. (1) Acoustic differences: auditory stimulation is more salient and variable in Emot. (2) Emotion content, which is present only in Emot[2]. (3) Attention direction: the task specifically directs the listener's attention to the speaker's emotion in Emot, but to sentence content in Neut. (4) Integration of prosodic with semantic information, and potential interactions between the two[2]. One might argue that it is unsatisfying not to know which of these factors contribute to any given voxel's activation. However, the advantage of this "big" contrasts is that it should capture the whole emotional prosody network, including stimulus- as well as task-driven effects.

To remove effects of acoustic differences (such as mean and variability of pitch) between Emot and Neut, we can estimate the effect of those differences on activation using contrast [EmotRev – NeutRev] and subtract the result from contrast [Emot – Neut]. Thus, we used contrast [(Emot – Neut) – (EmotRev – NeutRev)] to identify voxels that show an effect of emotional prosody processing [Emot – Neut] that is not explained by differences in these low-level acoustic features. Note that this is a very strict control, because the reversed emotional speech stimuli preserved some acoustic features relevant to emotional prosody and thus carried a recognizable, if subjectively degraded, emotion signal. (See supplementary materials S1 for the results of a validation study demonstrating above-chance performance for identifying emotion in the reverse speech stimuli.) While the experimental task in the reversed speech conditions did not require participants to judge emotion, emotional prosody processing might still have taken place in these conditions to the extent that it is automatic and can be supported by the somewhat degraded prosody signal in the reversed speech stimuli. Because of this, subtracting the difference between the emotional and neutral reversed speech conditions removes not only the effects of differences in pitch and amplitude modulations, but likely also some of the activation associated with automatic emotion processing. At the same time, it is important to note that [EmotRev – NeutRev]

---

[2]While our sentences were intended to be content-neutral, some participants may have perceived some sentences as having emotional semantic content. This would mean that condition Neut was not entirely free of emotion content, and that on some trials, participant might have perceived a conflict between semantic and prosodic emotion (e.g., if "Dad always gives me books for my birthday" is interpreted as sad in semantic content, but presented with happy prosody.) However, this likely happened only on few trials. Moreover, participants were aware that each sentence could be presented with any (happy, sad, angry, or neutral) prosody and thus knew not to interpret semantic content with regard to emotion.

does not fully capture stimulus-driven emotional prosody processing, because the natural features of prosodic speech (e.g., certain pitch contours) are distorted in the reverse speech stimuli and thus not present to the same degree as in EmotRev as in Emot. Thus, while [(Emot – Neut) – (EmotRev – NeutRev)] controls for basic acoustic differences (such as mean and variability of pitch) between emotional and neutral speech, it does not control for the presence of intact and recognizable prosodic features. In summary, contrast [(Emot – Neut) – (EmotRev – NeutRev)] captures predominantly task-driven effects (e.g., the fact that attention is on emotional prosody), but also activations associated with the extraction, identification, and interpretation of fully intact emotional prosody.

Lastly, contrast [Neut – NeutRev] identifies voxels with a significant preference for neutral forward over reversed speech. Like [Emot – Neut], [Neut – NeutRev] is a big contrast, as Neut exceeds NeutRev in several aspects, including the presence of semantic content, syntactic structure, lexical and morpheme recognition, etc. We thus refer to it as capturing the "sentence comprehension network" and, based on the literature, expect to find it lateralized to the left hemisphere, particularly superior temporal and inferior frontal cortex. This network has been studied extensively and was primarily of interest for comparison with the emotional prosody network.

2.4.5.   Region-of-interest (ROI) creation—To generate ROIs for comparing left- and right-hemisphere activations and computing lateralization indices, we binarized the whole-brain activation maps for all contrasts, flipped them across the midline, and took the union of the original and flipped maps, thus combining LH and RH activations into a bilaterally symmetric map for each contrast. We then took the union of these symmetrized contrast maps for all contrasts of interest ([Neut – NeutRev], [Emot – Neut], or [Emot – Neut]-[EmotRev-NeutRev]). The resulting bilaterally symmetric "union ROI" constrains the analysis to voxels of interest (i.e., voxels that show a significant effect for any of the contrasts of interest, and their homotopic counterparts) but is not biased towards either hemisphere or to any particular contrast.

To allow for a more detailed, regional look at activation and lateralization, we manually placed spherical ROIs (volume 257 mm^3) inside the union ROI. Sphere center locations were chosen so as to cover different anatomical regions of interest (e.g., amygdala, relevant Brodmann areas) within the larger ROI, without regard to peak activations.

2.4.6.   Lateralization indices—To quantify lateralization of activations for the contrasts of interest, we used the lateralization index (LI), which contrasts left and right activation and divides the result by the total activation:

$$LI = (\text{left activation} - \text{right activation})/(\text{left activation} + \text{right activation}).$$

Lateralization indices near −1 indicate right-lateralization, LIs near 1 indicate left-lateralization, and LIs around 0 indicate roughly equal activation in the two hemispheres. Importantly, what counts as "active" on either side is determined by the statistical threshold applied to the activation map. If the threshold is extremely lenient, most voxels on both sides (including noise) count as "active", whereas at strict thresholds, only a few voxels with

extreme activation values (potentially outliers) will figure into the computation and pull the LI into the direction of their hemisphere. To obtain a robust, threshold-independent estimate of the LI, we followed an approach introduced by Wilke & Schmithorst (2006). Each activation map of interest was evaluated at 20 different single-voxel thresholds, covering the range from t = 0.1 (very lenient) to the maximum t-value in the map in equal steps. At each threshold, the LI was determined using a bootstrapping procedure, which involved resampling the map voxels inside the ROI 100 times for each side, creating 100×100 LI estimates by comparing the sum of t-values of all suprathreshold voxels in a sample using the equation above for all left-right sample pairings, and then computing a trimmed mean (using only the central 50%) of the resulting 10,000 LI estimates for that threshold. In a final step, we computed a weighted average of the LIs across thresholds, using the t-threshold value as the weight such that LI estimates obtained at lower thresholds were allowed to influence the result at lower weights than LI estimates obtained at higher thresholds. To ensure that outliers did not drive the results, LIs were only computed if there was at least one cluster of at least 5 suprathreshold voxels on each side.

3. Results

3.1. Behavioral

Performance in the in-scanner tasks was overall high, and no participant scored lower than 89% correct in any condition. Average accuracy and reaction times for the different conditions are shown in Table 1.

Statistical comparisons (using paired t-tests on average accuracy and reaction times for each participant in each condition) revealed no significant differences between the conditions directly compared in our fMRI contrasts. The emotion and content identification tasks compared in fMRI contrast [Emot – Neut] did not differ from each other regarding accuracy (t(19) = 1.45, p = 0.164) or reaction time (t(19) = 1.90, p = 0.073), nor did reverse speech conditions EmotRev and NeutRev (t(19) = 1.03, p = 0.316 for accuracy and t(19) = 1.75, p = 0.096 for reaction time). There was also no significant difference between conditions Neut and NeutRev, on which the sentence comprehension contrast is built (t(19) = 1.50, p = 0.150 for accuracy and t(19) = 1.26, p = 0.223 for reaction time). However, accuracy was significantly higher in condition EmotRev than in condition Emot (t(19) = 4.20, p = 0.0005), although reaction times again did not differ significantly (t(19) = 1.50, p = 0.150).

3.2. Imaging

3.2.1. Whole-brain analyses—Figure 2 illustrates group-level activations for the three contrasts of interest. Detailed information about activation clusters' size and location are provided in Table 2. As expected, the sentence comprehension contrast [Neut – NeutRev] (Figure 2A) was associated with activations in left superior and middle temporal and inferior frontal cortex, angular gyrus (BA 39), and fusiform gyrus including the visual word form area (Cohen et al., 2000; Dehaene et al., 2002; McCandliss et al., 2003), which can be modulated by attention to speech stimuli even in the absence of visual stimulation (Yoncheva et al., 2010). This pattern of activation corresponds well with the classic left

hemisphere language network as identified similarly in many other functional neuroimaging studies (for reviews, see Vigneau et al., 2006; Price, 2012).

The Emotion recognition contrast [Emot – Neut] (Figure 2B) revealed fronto-temporal activations in the right hemisphere, including but not limited to areas roughly mirroring those activated by sentence comprehension. Like the sentence comprehension contrast, it activated superior and middle temporal, as well as inferior frontal cortex. In addition, it also evoked activations along the left superior temporal gyrus, as well as in the inferiormost portion of left inferior frontal cortex (pars orbitalis, Brodmann area 47), and extended bilaterally into the anterior insula and the amygdala. Interestingly, as best seen on the surface maps, the temporal cortex activations for [Emot – Neut] were more superior than those for [Neut – NeutRev]. The closer proximity to primary auditory cortex suggests that these activations may have partly been driven by the fact that the stimuli in condition Emot contain stronger pitch and amplitude modulations and are thus better at driving auditory cortex. Indeed, activation differences between the emotional and neutral reversed speech conditions (contrast [EmotRev – NeutRev]) were found predominantly in right, but also left superior temporal cortex (see Supplementary Materials S2).

Not surprisingly, then, subtracting the influence of these low-level acoustic differences from the emotional prosody contrast (see [Emot – Neut] – [EmotRev – NeutRev], Figure 2C) removed activations near primary auditory cortex bilaterally as well as left lateral temporal activations. However, activations in right auditory association areas (BA 22, 21) and inferior frontal cortex (BA 47 bilaterally, BA 45 and 46 on the right) remained. This suggests that their involvement in emotional prosody was not driven solely by low-level acoustic differences between emotional and neutral stimuli. Rather, we interpret these activations as showing task-related effects associated with attending to, identifying and interpreting emotional prosody cues. For example, the activation along right STS is in line with prior research demonstrating that auditory areas experience an attentional boost when attention is directed to their preferred features (e.g., Seydell-Greenwald et al., 2014). Similarly, left inferior frontal areas associated with sentence comprehension are activated more strongly during an active task requiring a comprehension-based decision compared to passive listening (e.g., Vannest et al., 2009). An analogous effect may be driving the right inferior frontal activations observed here. Overall, the areas of activation for emotional prosody observed here are in good agreement with the literature (e.g., Witteman et al., 2012).

To visually explore whether activation maps for sentence comprehension and emotional prosody were mirror images of each other, we flipped the sentence comprehension map across the midline and overlaid the right-hemisphere views of all maps (supplementary Figure S4). Common areas of activation were found along the length of the middle temporal gyrus (BAs 22, 21, and 38) and the inferior frontal gyrus (BAs 44, 45, and 6). However, there were considerable areas of non-overlap, with only emotional prosody significantly activating pars orbitalis, anterior insula, and amygdala, and only sentence comprehension activating fusiform gyrus and angular gyrus. However, the fact that activation is significant for one contrast (or one hemisphere) but not the other does not indicate a significant difference between the contrasts (or hemispheres; de Hollander et al., 2014). Thus, we next

took a more rigorous quantitative look at lateralization by comparing LH and RH activation for the different contrasts using an ROI-based approach.

3.2.2. Lateralization indices—Lateralization indices (Figure 3) were computed for a bilaterally symmetric "union ROI" (see section 2.4.5) comprising all voxels that showed a significant effect for any of the contrasts of interest in either hemisphere along with their homotopic counterparts. Even though this ROI includes all voxels (regardless of hemisphere) that showed activation for any contrast and was not biased towards any particular contrast, the expected lateralization pattern emerged. The sentence comprehension contrast ([Neut – NeutRev]) showed significant left-lateralization at group level (LI = 0.76, test against zero: $t(19) = 22.68$, $p < 0.001$). In contrast, both emotional prosody contrasts showed significant right-lateralization ([Emot – Neut], LI = −0.49, test against zero: $t(19) = −6.62$, $p < 0.001$; [Emot – Neut] – [EmotRev – NeutRev], LI = −0.38, test against zero: $t(19) = −5.29$, $p < 0.001$).

To directly compare strength of lateralization between the contrasts, we performed paired t-tests on the absolute LIs for all contrast pairs. Sentence comprehension was significantly more left-lateralized than either emotional prosody contrast was right-lateralized ([Neut – NeutRev] vs. [Emot – Neut], $t(19) = 4.02$, $p < 0.001$; [Neut – NeutRev] vs. [Emot – Neut] – [EmotRev – NeutRev], $t(19) = 5.31$, $p < 0.001$). Emotional prosody was somewhat more right-lateralized when acoustic feature differences were allowed to contribute to the effect than when they were not ([Emot – Neut] vs. [Emot – Neut] – [EmotRev – NeutRev], $t(19) = 1.88$, $p = 0.038$). However, this difference did not survive correction for multiple comparisons (the Bonferroni-corrected error level for the above 6 tests is $p < 0.008$).

We also repeated the lateralization analyses for an ROI encompassing the whole hemispheres (see supplementary Figure S5). While the LIs were overall less extreme due to the inclusion of voxels not involved in either sentence comprehension or emotional prosody processing, the overall pattern remained the same. There was significant group-level lateralization to the left for sentence processing ([Neut – NeutRev], LI = 0.64, test against zero: $t(19) = 11.58$, $p < 0.001$) and to the right for both emotional prosody contrasts ([Emot – Neut], LI = −0.45, test against zero: $t(19) = −6.51$, $p < 0.001$; [Emot – Neut] – [EmotRev – NeutRev], LI = −0.31, test against zero: $t(19) = −4.29$, $p < 0.001$). Sentence comprehension again was more left-lateralized than either emotional prosody contrast was right-lateralized ([Neut – NeutRev] vs. [Emot – Neut], $t(19) = 2.18$, $p = 0.042$; [Neut – NeutRev] vs. [Emot – Neut] – [EmotRev – NeutRev], $t(19) = 3.69$, $p = 0.002$), and emotional prosody was somewhat more right-lateralized when acoustic feature differences were allowed to contribute to the effect than when they were not ([Emot – Neut] vs. [Emot – Neut] – [EmotRev – NeutRev], $t(19) = 2.48$, $p = 0.011$), although in both cases, the p-values exceeded the Bonferroni-corrected error level of $p < 0.008$.

3.2.3. Region of interest analyses—The analysis across such large ROIs could mask local differences in lateralization. For example, right-lateralization of emotional prosody could be limited to the superior temporal lobe, as suggested by Witteman et al. (2012). Thus, we placed spherical ROIs at different anatomical locations of interest inside the union ROI

(see section 2.4.5. and Figure 4A) to capture local differences in areas of interest for either contrast.

Sphere ROIs in regions activated by both sentence comprehension *and* emotional prosody (BAs 22, 21, 45, 44, and 6, labeled in black in Figure 4B) coincided with the "classic" fronto-temporal network thought to support language function. In excellent agreement with the mirror hypothesis, they all showed significant left-lateralization for sentence comprehension and significant right-lateralization for emotional prosody. None showed a significant difference in right-lateralization between the two emotional prosody contrasts, indicating that right-lateralization for emotional prosody was not driven by acoustic differences between emotional and neutral stimuli. Left-lateralization for sentence comprehension tended to be stronger than right-lateralization for emotional prosody in all sphere ROIs, although the difference was not significant at Bonferroni-corrected error levels.

Sphere ROIs in regions that showed significant whole-brain activation *only* for sentence comprehension (FG and BA 39, labeled in blue in Figure 4B) also showed significant left-lateralization for sentence comprehension, which significantly exceeded right-lateralization for emotional prosody (not surprisingly, given that these areas did not respond to emotional prosody). In contrast, sphere ROIs in regions that showed significant whole-brain activation *only* for emotional prosody (amygdala, anterior insula, and BA 47, labeled in orange in Figure 4B) showed roughly equal activation in both hemispheres, with the exception of primary auditory cortex (BA 41), which was not significantly activated by the sentence comprehension contrast on either side, but showed weakly right-lateralized activation for emotional prosody, regardless of whether acoustic feature differences were included or removed from the contrast.

4. Summary and Discussion

4.1. Summary

Overall, our results answer the four questions raised at the end of the Introduction. As expected, sentence processing was associated with strongly left-lateralized frontotemporal activations. In support of the lateralization hypothesis, emotional prosody processing was associated with right-lateralized frontotemporal activations (1), even when the effect of stronger pitch- and amplitude modulation was controlled for (2). However, right-lateralization for emotional prosody processing was weaker than left-lateralization for sentence processing (3), and there was significant LH activation for emotional prosody. In support of the mirror hypothesis, the RH emotional prosody activation encompassed many of the same frontotemporal brain areas as the LH sentence processing activation (4). However, there was also considerable non-overlap. Bilateral pars orbitalis (BA 47), amygdala, and anterior insula were activated only for emotional prosody, and the sentence-comprehension activations in right cerebellum, left fusiform gyrus, and left inferior parietal cortex did not have a contralateral counterpart for emotional prosody processing. We next discuss our key findings with respect to previous research and models/hypotheses.

4.2. Emotional prosody activation was right-lateralized and not solely driven by the larger amplitude and pitch modulations of the emotional stimuli

As reviewed in the Introduction, the literature is divided regarding lateralization of emotional prosody processing. While early models based on lesion studies favored the idea of a right-lateralized frontotemporal emotional prosody network mirroring the LH language network, meta-analyses of functional neuroimaging studies have found little evidence of right-lateralization in inferior frontal cortex, and right-lateralization in temporal cortex is sometimes attributed to the acoustic features of the emotional prosody stimuli (e.g., Wiethoff et al., 2008).

The present data confirm a preference of right temporal cortex for emotional compared to neutral prosody. Contrasting reverse versions of emotional and neutral stimuli led to right-lateralized auditory activation (see supplementary Figure S2), which suggests some preference of right auditory cortex for those characteristics of the emotional prosody stimuli that are preserved in the reverse emotional speech stimuli but not present to the same degree in the reverse neutral stimuli (e.g., higher pitch- and amplitude modulations at the suprasegmental level). This observation is consistent with hypotheses suggesting that the auditory cortices of the two hemispheres preferentially process different aspects of the speech signal (acoustic lateralization hypothesis, Wildgruber et al., 2006), with the RH dominant for analyzing pitch (Van Lancker & Sidtis, 1992; Zatorre, 2001) or aspects of the speech signal that require integration over larger temporal windows (Poeppel, 2003; Boemio et al., 2005).

Interestingly, stronger amplitude and pitch modulations were by no means the sole driver of right-lateralization for emotional prosody processing, as right-lateralization remained when their effect was subtracted from the forward speech emotional prosody contrast (see right-lateralization for both emotional prosody contrasts in BAs 41, 22, and 21 in Figure 4). This suggests that right-lateralization is also driven by differences between conditions (Emot – Neut) that were not captured by contrast (EmotRev – NeutRev). These are, in no particular order: the presence of fully intact prosodic features associated with specific emotional interpretations (e.g., certain pitch contours that are degraded when playing the emotional stimuli in reverse), task-driven effects (e.g., the fact that attention was directed to emotional prosody in condition Emot, but not EmotRev), and potential interactions between emotional prosody and semantic content (e.g., if a participant interpreted the sentence "Dad cooked potroast for dinner last night" as something to be angry about, but the sentence was presented with happy prosody).

Importantly, right-lateralization was not equally strong across all areas activated by the emotional prosody contrasts. We next turn to discussing these interesting local variations, which have not been investigated before at this level of detail.

4.3. Lateralization of prosody activation showed strong local variations, with strong right-lateralization limited to areas mirroring the LH sentence processing activation

As can be seen in Figure 4B, of the ROIs with significant emotional prosody activation, only those that were homotopic to LH sentence comprehension areas (BAs 22, 21, 44, 45, and 6)

showed strong right-lateralization. This complementary lateralization, which is predicted by the mirror hypothesis, may have been driven in part by the non-independence of the contrasts used to identify activations for sentence comprehension and emotional prosody: the contrasts of interest, [Neut – NeutRev] and [Emot – Neut], each contain [Neut], but with with opposite signs. Thus, strong LH activation for neutral forward speech would drive LIs for sentence comprehension towards left-lateralization and those for emotional prosody processing towards right-lateralization. However, the fact that we observed strong RH activation for [Emot – Neut] demonstrates that right-lateralization for emotional prosody was not driven solely by LH activation for [Neut], but to a significant degree by RH activation for [Emot].

In contrast, the emotional prosody ROIs that were not homotopic to LH sentence comprehension areas (BA 47, amygdala, and anterior insula) were activated bilaterally, and there were no contralateral emotional prosody activations mirroring LH sentence comprehension activations in right cerebellum, left fusiform gyrus, and left inferior parietal cortex. This, along with the observation that right-lateralization for emotional prosody was lower than left-lateralization for sentence comprehension in almost all ROIs, shows that a strong version of the mirror hypothesis cannot be maintained. However, our data are compatible with a qualified version of the mirror hypothesis that is also in line with more recent models of emotional prosody processing.

Multi-step models of prosody processing (Schirmer & Kotz, 2006; Wildgruber et al., 2009; Brück et al., 2011; Frühholz et al., 2016) have moved away from regarding emotional prosody processing as purely RH-based. Rather, they posit several sub-stages of emotional prosody processing that differ with respect to hemispheric lateralization. Specifically, after the speech signal reaches bilateral auditory cortex, its processing is divided between the hemispheres according to acoustic features, with the RH preferentially decoding the suprasegmental speech parameters carrying prosodic information (step 1). The results of this analysis are passed on to auditory association areas in right anterior and posterior temporal cortex (step 2) for identification (i.e., comparison with familiar auditory objects, such as pitch contours associated with specific emotions) and integration with other (e.g., visual) cues. Following identification, prosody-derived emotion information is then passed on to bilateral inferior frontal cortex for evaluation and integration with semantic information (step 3).

Reminiscent of the mirror hypothesis and in line with the ROI analysis reported here, these models propose right-lateralization for emotional prosody in superior temporal cortex as well as inferior frontal cortex (more specifically, BAs 44 and 45). However, Wildgruber et al. also predict bilateral frontal activations in BA 47 (Wildgruber et al., 2006), which is exactly what we observed here. In addition, both models recognize that emotional prosody processing does not occur in isolation, but happens in the larger context of general emotional appraisal, which also involves areas not expected to mirror the sentence comprehension system. For example, the amygdala, which was activated here by emotional prosody processing, is also responsive to emotional faces (Adolphs et al., 1994; Calder, 1996) and non-verbal emotional vocalizations (Morris et al., 1999; Fecteau et al., 2007). In this context it is important to note that vocal emotion processing in general is not right-lateralized (e.g.,

Warren et al., 2006; Scott et al., 2009; Bestelmeyer et al., 2014; see Frühholz et al., 2016, for a review). Rather, right-lateralization seems to be most pronounced when parameters carrying emotional prosody information have to be disentangled from parameters that could potentially carry semantic information (even if the stimulus material is not comprehensible). In other words, it seems that the mirror hypothesis holds to the extent that there is a division of labor between the hemispheres regarding "core decoding", i.e., which parameters of the speech signal they preferentially process. However, it does not capture the full complexity of the sentence comprehension and emotional prosody networks.

4.5. Caveats

As with any functional neuroimaging study, an important caveat is that activation of a brain area during a certain task should not be taken as an indication that this brain area is necessary for performing the task. However, the results of the few artificial lesion studies published to date are consistent with our neuroimaging results: Alba-Ferrara et al. (2012) found that only TMS to the right, but not to the left posterior superior temporal cortex prolonged reaction times in a prosody judgment task, in agreement with the right-lateralized activation observed here. Hoekert et al. (2010) found that rTMS over both the right and the left inferior frontal gyrus slowed emotion judgments, with no differences between the two sides, which is consistent with our data if the stimulation affected Brodmann area 47, where prosody activation was strongly bilateral.

Another caveat is that lateralization for emotion may be modulated by valence, with a RH advantage for negative and a LH advantage for positive emotions (Davidson, 1984; Sackeim et al., 1982; Ahern & Schwartz, 1985; Canli et al., 1998). Because two thirds of our stimuli were of negative valence (angry and sad vs. happy utterances), valence effects might have contributed to stronger RH activation. However, there is little evidence for a valence effect in emotional prosody processing, with several studies failing to find one (e.g., Borod et al., 1998; Wildgruber et al., 2002, 2005; Kotz et al., 2003; Kucharska-Pietura et al., 2003; Wildgruber et al., 2005; see also Murphy et al., 2003, for a meta-analysis) and others producing weak and inconsistent effects (Buchanan et al., 2000; Kotz et al., 2006; Wittfoth et al., 2010). It is thus highly unlikely that the strong right-lateralization observed here was driven solely by a valence effect.

Lastly, our experimental conditions were not perfectly equivalent with regards to task demands. While performance was overall high and did not differ significantly between the conditions compared in our key fMRI contrasts (Emot vs. Neut, and Neut vs. NeutRev), this could partly reflect a ceiling effect or inadequate power/violations of assumptions of the statistical tests used and does not exclude the possibility that differences in task difficulty contributed to the observed activations. In addition, the forward speech conditions were associated with a matching task (does the visual symbol match the auditory information) whereas the reverse speech control conditions were associated with a simple detection task (was there a beep). This difference in task type (matching vs. detection) could potentially have contributed to activations observed for the sentence comprehension contrast (Neut – NeutRev), but not to activations for the emotional prosody contrasts (Emot – Neut) and

(EmotRev – NeutRev) because the conditions put in direct contrast have equivalent task demands.

5. Conclusion

This study used a within-subject design to directly juxtapose fMRI activation and quantify its lateralization for sentence comprehension and emotional prosody processing. Our results show that emotional prosody processing and sentence comprehension activate homotopic portions of superior temporal and inferior frontal cortex, suggesting a division of labor between the hemispheres when analyzing a speech signal that carries both semantic and prosodic information. However, right-lateralization of emotional prosody processing was not as strongly lateralized as left-lateralization of sentence comprehension in these areas, and each function also activated unique areas not activated by the other. We conclude that the classic "mirror hypothesis" has merit as a short-hand for describing hemispheric preferences, although it does not capture the full complexity of linguistic and paralinguistic processing, which involves additional, non-homotopic areas.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Adolphs R, Tranel D, Damasio H, Damasio A (1994) Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. Nature 372:669–672. [PubMed: 7990957]

Ahern GL, Schwartz GE (1985) Differential lateralization for positive and negative emotion in the human brain: EEG spectral analysis. Neuropsychologia 23:745–755. [PubMed: 4080136]

Alba-Ferrara L, Ellison A, Mitchell RLC (2012) Decoding emotional prosody: Resolving differences in functional neuroanatomy from fMRI and lesion studies using TMS. Brain Stimulation 5:347–353. [PubMed: 21824835]

Alba-Ferrara L, Hausmann M, Mitchell RL, Weis S (2011) The neural correlates of emotional prosody comprehension: disentangling simple from complex emotion. PLoS ONE 6(12): e28701 Doi:10.1371/journal.pone.0028701. [PubMed: 22174872]

Banse R, Scherer K (1996) Acoustic profiles in vocal emotion expression. J Personality 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. Cerebral Cortex 17:339–352. [PubMed: 16525130]

Belyk M, Brown S (2014) Perception of affective an linguistic prosody: an ALE meta-analysis of neuroimaging studies. Soc Cogn Affect Neurosci 9:1395–1403. [PubMed: 23934416]

Bestelmeyer PEG, Maurage P, Rouger J, Latinus M, Belin P (2014) Adaptation to vocal expressions reveals multistep perception of auditory emotion. J Neurosci 34:8098–8105. [PubMed: 24920615]

Binder JR, et al. (1995) Lateralized human brain language systems demonstrated by task subtraction functional magnetic resonance imaging. Arch Neurol 52:593–601. [PubMed: 7763208]

Blonder L, Bowers D, Heilman K (1991) The role of the right hemisphere in emotional communication. Brain 114:1115–1127. [PubMed: 2065243]

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395. [PubMed: 15723061]

Bookheimer SY, et al. (1997) A direct comparison of PET activation and electrocortical stimulation mapping for language localization. Neurology 48(4):1056–1065. [PubMed: 9109900]

Borod JC et al. (1998) Right hemisphere emotional perception: evidence across multiple channels. Neuropsychology 12:446–458. [PubMed: 9673999]

Borod JC, Bloom RL, Brickman AM, Nahutina L, Curko EA (2002) Emotional processing deficits in individuals with unilateral brain damage. Appl Neuropsychol 9:23–36. [PubMed: 12173747]

Bowers D, Coslett H, Bauer R, Speedie L, Heilman K (1987) Comprehension of emotional prosody following unilateral hemispheric lesions: Processing defect versus distraction defect. Neuropsychologia 25:317–328. [PubMed: 3601038]

Broca PM (1861) Remarques sur le siége de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). Bull d l Soc Anatom 6:330–357.

Brück C, Kreifelts B, Wildgruber D (2011) Emotional voices in context: A neurobiological model of multimodal affective information processing. Phys Life Rev 8:383–403. [PubMed: 22035772]

Buchanan TW, et al. (2000) Recognition of emotional prosody and verbal components of spoken language: an fMRI study. Cogn Brain Res 9:227–238.

Calder AJ (1996) Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. Cogn Neuropsychol 13:699–745.

Canli T, Desmonds JE, Zhao Z, Glover G, Gabrieli JDE (1998) Hemispheric asymmetry for emotional stimuli detected with fMRI. NeuroReport 9:3233–3239. [PubMed: 9831457]

Cohen L, et al. (2000) The visual word form area: spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123:291–307. [PubMed: 10648437]

Davidson RJ (1984) Affect, cognition, and hemispheric specialization Pages 320–365 in Izard CE, Kagan J, & Zajonc R (Eds): Emotion, cognition, and behavior. New York: Cambridge University Press.

Dehaene S, Le Clec'H G, Poline JP, Le Bihan D, Cohen L (2002) The visual word form area: a prelexical representation of visual words in the fusiform gyrus. NeuroReport 13:321–325. [PubMed: 11930131]

de Hollander G, Wagenmakers EJ, Waldorp L, Forstmann B (2014) An antidote to the imager's fallacy, or how to identify brain areas that are in limbo. PLoS ONE 9(12): e115700. doi:10.1371/journal.pone.0115700 [PubMed: 25546581]

Ethofer T, et al. (2006) Effects of prosodic emotional intensity on activation of associative auditory cortex. NeuroReport 17:249–253. [PubMed: 16462592]

Ethofer et al. (2012) Emotional voice areas: Anatomic location, functional properties, and structural connections revealed by combined fMRI/DTI. Cereb Cortex 22:191–200. [PubMed: 21625012]

Fairbanks G (1940) Recent experimental investigations of vocal pitch in speech. J Acoust Soc Am 11:457–466.

Fairbanks G, Pronovost W (1939) An experimental study of the pitch characteristics of the voice during the expression of emotions. Speech Monographs 6:87–104.

Fecteau S, Belin P, Joanette Y, Armony JL (2007) Amygdala responses to nonlinguistic emotional vocalizations. NeuroImage 36:480–487. [PubMed: 17442593]

Fridriksson J, Yourganov G, Bonilha L, Basilakos A, Den Ouden DB, Rorden C (2016) Revealing the dual streams of speech processing. Proc Natl Acad Sci USA 113:15108–15113. [PubMed: 27956600]

Friederici A (2011) The brain basis of language processing: from structure to function. Physiol Rev 91:1357–1392. [PubMed: 22013214]

Frühholz S, Trost W, Kotz SA (2016) The sounds of emotions – Towards a unifying neural network perspective of affective sound processing. Neurosci Biobehav Rev 68:96–110. [PubMed: 27189782]

George M, et al. (1996) Understanding emotional prosody activates right hemisphere regions. Arch Neurol 53:665–670. [PubMed: 8929174]

Geschwind N, Levitsky W (1968) Human brain: left-right asymmetries in temporal speech region. Science 161(3837):186–187. [PubMed: 5657070]

Gorelick PB, Ross ED (1987) The aprosodias: further functional-anatomical evidence for the organization of affective language in the right hemisphere. J Neurol Neurosurg Psychiatry 50:553–560. [PubMed: 2438386]

Grandjean D, et al. (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci 8:145–146. [PubMed: 15665880]

Hagoort P (2014) Nodes and networks in the neural architecture for language: Broca's region and beyond. Curr Opin Neurobiol 28:136–141. [PubMed: 25062474]

Heilman KM, Scholer R, Watson RT (1975) Auditory affective agnosia. Disturbed comprehension of affective speech. J Neurol Neurosurg Psychiatry 38:69–72. [PubMed: 1117301]

Hervé PY, Razafimandimby A, Vigneau M, Mazoyer B, Tzourio-Mazoyer N (2012) Disentangling the brain networks supporting affective speech comprehension. NeuroImage 61:1255–1267. [PubMed: 22507230]

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. [PubMed: 17431404]

Hoekert M, Vingerhoets G, Aleman A (2010) Results of a pilot study on the involvement of bilateral inferior frontal gyri in emotional prosody perception: an rTMS study. BMC Neuroscience 11:93 http://www.biomedcentral.com/1471-2202/11/93. [PubMed: 20698964]

House A, Rowe D, Standen PJ (1987) Affective prosody in the reading voice of stroke patients. J Neurol Neurosurg Psychiatry 50:910–912. [PubMed: 3625214]

Imaizumi S, et al. (1997) Vocal identification of speaker and emotion activates different brain regions. NeuroReport 8:2809–2812. [PubMed: 9295122]

Jacob H, Brück C, Plewnia C, Wildgruber D (2014) Cerebral processing of prosodic emotional signals: Evaluation of a network model using rTMS. PLoS ONE 9(8):e105509 Doi:10.1371/journal.pone.0105509. [PubMed: 25171220]

Just MA, Carpenter PA, Keller TA, Eddy WF, Thulborn KR (1996). Brain activation modulated by sentence comprehension. Science 274(5284):114–116. [PubMed: 8810246]

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. [PubMed: 12972367]

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. [PubMed: 17015086]

Kucharska-Pietura K, Phillips ML, Gernand W, David AS (2003) Perception of emotions from faces and voices following unilateral brain damage. Neuropsychologia 41:1082–1090. [PubMed: 12667543]

Lieberman P, Michaels SB (1962) Some aspects of fundamental frequency and amplitude as related to the emotional content of speech. J Acoust Soc Am 34:922–927.

McCandliss BD, Cohen L, Dehaene S (2003) The visual word form area: expertise for reading in the fusiform gyrus. Trends Cogn Sci 7:293–299. [PubMed: 12860187]

Mitchell RLC, Elliott R, Barry M, Cruttenden A, Woodruff PWR (2003) The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. Neuropsychologia 41:1410–1421. [PubMed: 12757912]

Mitchell RLC, Ross ED (2008) fMRI evidence for the effect of verbal complexity on lateralization of the neural response associated with decoding prosodic emotion. Neuropsychologia 46:2880–2887. [PubMed: 18635241]

Morris JS, Scott SK, Dolan RJ (1999) Saying it with feeling: neural responses to emotional vocalizations. Neuropsychologia 37:1155–1163. [PubMed: 10509837]

Mothes-Lasch M, Mentzel HJ, Miltner WHR, Straube T (2011) Visual attention modulates brain activation to angry voices. J Neurosci 31:9594–9598. [PubMed: 21715624]

Murphy FC, Nimmo-Smith I, Lawrence AD (2003) Functional neuroanatomy of emotions: A meta-analysis. Cognitive, Affective, & Behav Neurosci 3:207–233.

Pascual-Leone A, Gates JR, Dhuna A (1991) Induction of speech arrest and counting errors with rapid-rate transcranial magnetic stimulation. Neurology 41:697–702. [PubMed: 2027485]

Pell MD, Baum SR (1997) The ability to perceive and comprehend intonation in linguistic and affective contexts by brain-damaged adults. Brain Lang 57:80–99. [PubMed: 9126408]

Pell MD (2006) Cerebral mechanisms for understanding emotional prosody in speech. Brai Lang 96:221–234.

Poeppel D (2003) The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. Speech Comm 41:245–255.

Price CJ (2010) The anatomy of language: a review of 100 fMRI studies published in 2009. Ann NY Acad Sci 1191:62–88. [PubMed: 20392276]

Price CJ (2012) A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. NeuroImage 62:816–847. [PubMed: 22584224]

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724. [PubMed: 19471271]

Ross ED, Mesulam MM (1979) Dominant language functions of the right hemisphere? Prosody and emotional gesturing. Arch Neurol 36:144–148. [PubMed: 435134]

Ross ED, Monnot M (2008) Neurology of affective prosody and its functional-anatomic organization in the right hemisphere. Brain Lang 104: 51–74. [PubMed: 17537499]

Ross ED (1981) The Aprosodias. Functional-anatomic organization of the affective components of language in the right hemisphere. Arch Neurol 38:561–569. [PubMed: 7271534]

Sackeim HA, et al., (1982) Hemispheric asymmetry in the expression of positive and negative emotions. Arch Neurol 1982:210–218.

Sammler D, Grosbras MH, Anwander A, Bestelmeyer PEG, Belin P (2015) Dorsal and ventral pathways for prosody. Curr Biol 25:3079–3085. [PubMed: 26549262]

Schirmer A, Kotz SA (2006) Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends in Cognitive Sciences 10:24–30. [PubMed: 16321562]

Schlanger BB, Schlanger P, Gerstman LJ (1976) The perception of emotionally toned sentences by right hemisphere-damaged and aphasic subjects. Brain Lang 3:396–403. [PubMed: 949593]

Schmitt JJ, Hartje W, Willmes K (1997) Hemispheric asymmetry in the recognition of emotional attitude conveyed by facial expression, prosody and propositional speech. Cortex 33:65–81. [PubMed: 9088722]

Scott SK, Sauter D, McGettigan C (2009) Brain mechanisms for processing perceived emotional vocalizations in humans Chapter 5.5 in Brudzynski SM (ed): Handbook of Mammalian Vocalization. Oxford: Academic Press.

Seydell-Greenwald A, Greenberg AS, Rauschecker JP (2014) Are you listening? Brain activation associated with sustained nonspatial auditory attention in the presence and absence of stimulation. Hum Brain Map 35: 2233–2252.

Shamay-Tsoory SG, Tomer R, Goldsher D, Berger BD, Aharon-Peretz J (2004) Impairment in cognitive and affective empathy in patients with brain lesions: anatomical and cognitive correlates. J Clin Exp Neuropsychol 26:1113–1127. [PubMed: 15590464]

Tucker DM, Watson RT, Heilman KM (1977) Discrimination and evocation of affectively intoned speech in patients with right parietal disease. Neurology 27:947–950. [PubMed: 561908]

Twist DJ, Squires NK, Spielholz NI, Silverglide R (1991) Event-related potentials in disorders of prosodic and semantic linguistic processing 1. Neuropsychiatry, Neuropsychology, and Behavioral Neurology 4:281–304.

Van Lancker D, Sidtis JJ (1992) The identification of affective-prosodic stimuli by left- and right-hemisphere-damaged subjects: All errors are not created equal. J Speech Hearing Sci 35:963–970.

Vannest J, et al. (2009) Comparison of fMRI data from passive listening and active-response story processing tasks in children. J Magn Reson Imaging 29:971–976. [PubMed: 19306445]

Vigneau M, et al. (2006) Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. NeuroImage 30:1414–1432. [PubMed: 16413796]

Wada J, Rasmussen T (1960) Intracarotid injection of sodium amytal for the lateralization of cerebral speech dominance. J Neurosurg 17:266–282.

Warren JE, et al. (2006) Positive emotions preferentially engage and auditory-motor "mirror" system. J Neurosci 26:13067–13075. [PubMed: 17167096]

Wernicke C (1874) Der aphasische Symptomencomplex - eine psychologische Studie auf anatomischer Basis. (Cohn & Weigert, Breslau).

Wiethoff S, et al. (2008) Cerebral processing of emotional prosody – influence of acoustic parameters and arousal. NeuroImage 39:885–893. [PubMed: 17964813]

Wildgruber D, Pihan H, Ackermann H, Erb M, Grodd W (2002) Dynamic brain activation during processing of emotional intonation: influence of acoustic parameters, emotional valence, and sex. NeuroImage 15:856–869. [PubMed: 11906226]

Wildgruber D, Ackermann H, Kreifelts B, Ethofer T (2006) Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog Brain Res 156:249–268 [PubMed: 17015084]

Wildgruber D, et al. (2005) Identification of emotional intonation evaluated by fMRI. NeuroImage 24:1233–1241. [PubMed: 15670701]

Wildgruber D, Ethofer T, Grandjean D, Kreifelts B (2009) A cerebral network model of speech prosody comprehension. Int J Speech-Lang Pathol 11:277–281.

Wilke M, Schmithorst VJ (2006) A combined bootstrap/histogram analysis approach for computing a lateralization index from neuroimaging data. NeuroImage 33:522–530. [PubMed: 16938470]

Williams CE, Stevens KN (1972) Emotions and speech: Some acoustical correlates. J Acoust Soc Am 52:1238–1250. [PubMed: 4638039]

Witteman J, van Heuven VJP, Schiller NO (2012) Hearing feelings: A quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. Neuropsychologia 50:2752–2763. [PubMed: 22841991]

Witteman J, van Ijzendoorn MH, van de Velde D, van Heuven VJJP, Schiller NO (2011) The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. Neuropsychologia 49:3722–3738. [PubMed: 21964199]

Wittfoth M, Schröder C, Schardt DM, Dengler R, Heinze HJ, Kotz SA (2010) On emotional conflict: interference resolution of happy and angry prosody reveals valence-specific effects. Cereb Cortex 20:383–392. [PubMed: 19505993]

Yoncheva YN, Zevon JD, Maurer U, McCandliss BD (2010) Auditory selective attention to speech modulates activity in the visual word form area. Cereb Cortex 20:622–632. [PubMed: 19571269]

Zatorre RJ (2001) Neural specializations for tonal processing. Ann NY Acad Sci 930:193–210. [PubMed: 11458830]

Highlights

- Within-subject comparison of emotional prosody and sentence comprehension fMRI

- Emotional prosody judgments evoke right-lateralized frontotemporal fMRI activations

- Right-lateralization is not driven exclusively by acoustic feature differences

- Sentence comprehension evokes left-lateralized frontotemporal fMRI activations

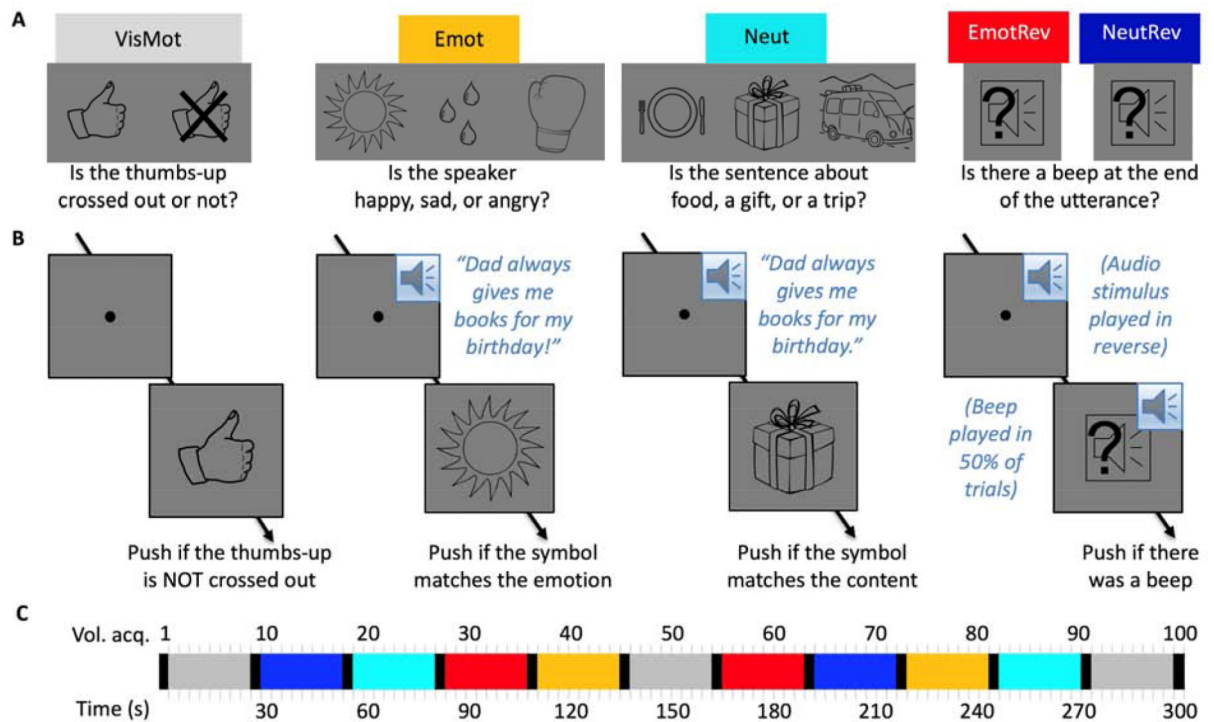- Emotional prosody and sentence processing activations partly mirror one another

Figure 1.
Stimuli and Design. (A) Experimental conditions and associated instruction screens. (B) Example trials. (C) Example time course of a 5-minute functional MRI run. Instruction periods are shown in black, all other colors as labelled in (A). Three runs were presented in counterbalanced order across participants.
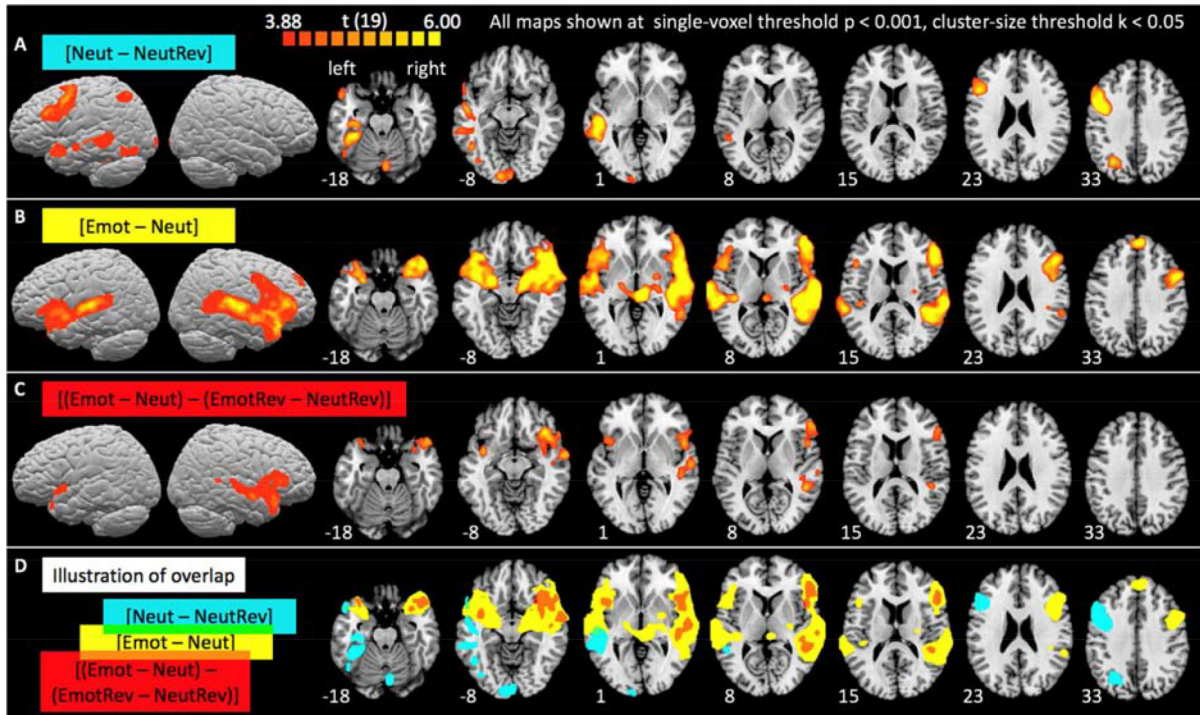
Figure 2.

Whole-brain activation maps for sentence comprehension (A) and for emotional prosody processing including (B) and excluding (C) the influence of acoustic differences between emotional and neutral stimuli. Group-level activations are projected onto the surface and superimposed on horizontal slices of the Colin 25 template brain in Talairach space. Slices were chosen to best show activations and are shown in neurological convention (left hemisphere on the left). Numbers indicate the z-Talairach coordinate of the depicted slices. Detailed information about activation clusters' locations are provided in Table 2. (D) Overlap of the binarized activation maps from A (blue), B (yellow), and C (red).
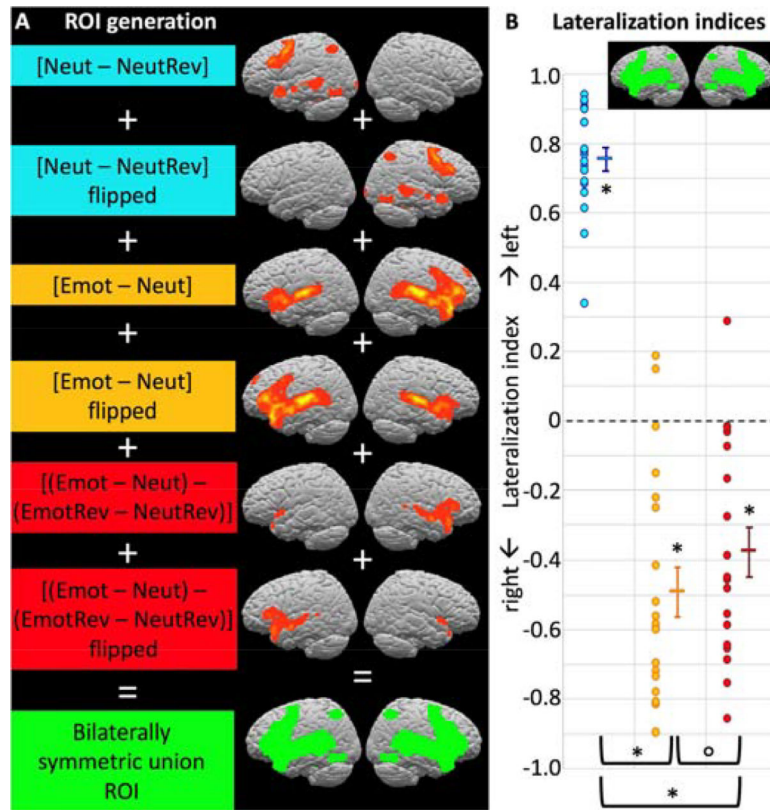
Figure 3.

Lateralization indices for the three contrasts of interest. (A) A bilaterally symmetric ROI not biased towards either hemisphere or contrast was generated from the union of all contrasts' activations and their left-right reversed mirror images (see Section 2.4.5. for details on ROI generation). Blue, yellow, and red dots represent individual participants' LIs for contrasts [Neut – NeutRev], [Emot – Neut], and ([Emot – Neut] – [EmotRev – NeutRev]), respectively; horizontal bars represent the group means, with error bars representing the standard error of the mean. As expected, sentence comprehension was significantly left-lateralized for all individual participants and at group level, whereas emotional prosody processing was right-lateralized at group level and for most individual participants. Left-lateralization for sentence comprehension was significantly stronger than right-lateralization for emotional prosody processing, and emotional prosody processing was somewhat more right-lateralized when acoustic differences were allowed to contribute (see Section 3.2.1. for statistical details.) Supplementary Figure S5 shows LIs in an anatomical, whole-hemisphere ROI.
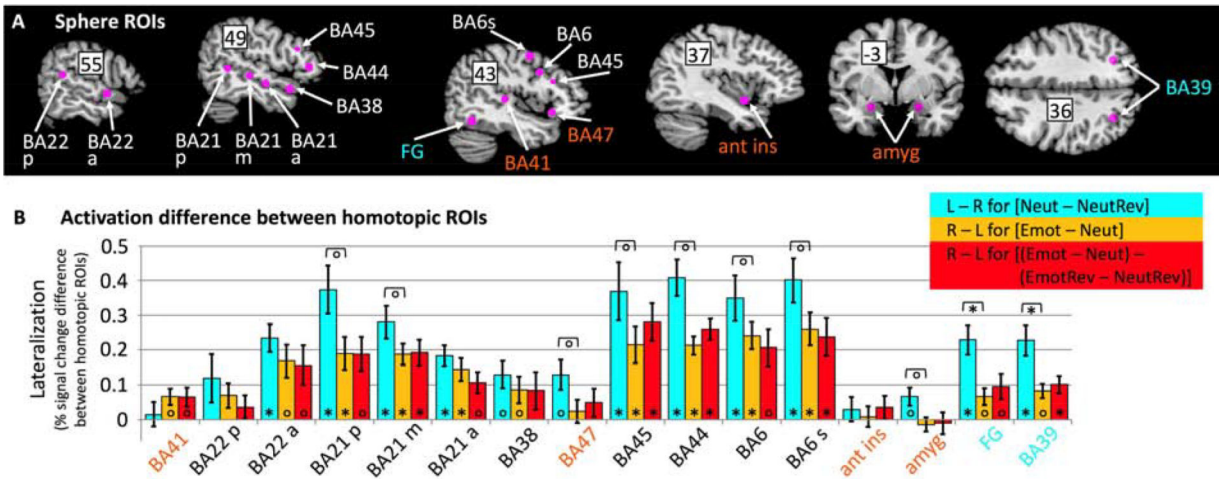
Figure 4.

Sphere ROI analyses comparing lateralization strength for sentence comprehension (blue) and emotional prosody processing including (yellow) and excluding (red) the influence of acoustic differences between emotional and neutral stimuli. (A) Spherical ROIs along with their name labels (BA = Brodmann area; p = posterior, m = mid, a = anterior, s = superior; FG = fusiform gyrus; amyg = amygdala, ant ins = anterior insula) superimposed on brain slices selected to best show the ROIs. Numbers superimposed on the slice indicate the relevant Talairach coordinate (x for the sagittal slices, y for the coronal, z for the horizontal slice.) ROIs activated only by sentence comprehension are labeled in blue; ROIs activated only by emotional prosody are labeled in orange; all other ROIs were activated by both sentence comprehension and emotional prosody (albeit in different hemispheres). (B) Strength of lateralization, as quantified by subtracting percent signal change in the homotopic sphere ROI from percent signal change in the ROI of the hemisphere thought to be dominant for the contrasts of interest. Positive values indicate stronger activation on the side thought to be dominant for the contrast, i.e. on the left for sentence comprehension contrast [Neut – NeutRev] (blue), and on the right for emotional prosody contrasts [Emot – Neut] (yellow) and [(Emot – Neut) – (EmotRev – NeutRev)] (red). Error bars indicate standard error of the mean across participants. Significance indicators at the base of bars indicate significant lateralization; significance indicators underneath brackets indicate differences in lateralization strength between sentence comprehension and emotional prosody. Stars indicate significance at Bonferroni-corrected error levels (p < 0.0006, correcting for 80 tests), circles indicate significance at uncorrected p < 0.05.

Table 1.

In-scanner performance

	Mean (SD) percent correct	Mean (SD) reaction time (ms)
Vis	98.80 (1.88)	519 (83)
Emot	96.53 (2.76)	489 (72)
Neut	97.29 (2.86)	501 (82)
EmotRev	98.89 (2.04)	501 (88)
NeutRev	98.06 (3.18)	493 (79)

Table 2.

Activation peaks in the whole-brain analysis

Contrast of interest	Cluster location	Peak Talairach coordinates x,y,z *	Peak t-value
Neut – NeutRev	Left superior and middle temporal cortex, BAs 22, 21, 37	−48, −28, 1	7.63
	Left fusiform gyrus, BAs 37, 36	−33, −37, −17	7.77
	Left anterior temporal cortex, BA 38	−48, 11, −14	6.32
	Left inferior frontal cortex, BAs 6, 9, 44 (extending into 46 and 45)	−45, 5, 34	11.92
	Left inferior parietal cortex, BAs 39, 7	−27, −61, 37	6.22
	Right cerebellum	6, −76, −29	6.09
Emot – Neut	Left superior/middle temporal cortex, BAs 41, 42, 22, 21, 38	−57, −37, 11	9.28
	Left anterior insula	−33, 5, −11	9.90
	Left amygdala	−20, −6, −10	8.19
	Left inferior frontal cortex, BAs 47, 45, 44	−42, 17, −5	7.28
	Midline frontal cortex, BA 9	3, 47, 37	8.22
	Right thalamus	6, −27, 1	8.92
	Right amygdala	21, −7, −8	10.09
	Right anterior insula (no separate peak, but bridging activation between other peaks)		
	Right superior/middle temporal cortex, BAs 41, 42, 22, 21, 37, 38	63, −16, 7	12.53
	Right inferior/middle frontal cortex, BAs 47, 45, 44, 46, 6	48, 35, 10	10.83
(Emot – Neut) – (EmotRev – NeutRev)	Left anterior insula, extending into anterior temporal BA 38	−35, 2, −11	5.38
	Left inferior frontal cortex, BA 47	−42, 15, −3	4.75
	Right amygdala	24, 2, −14	5.47
	Right anterior insula (no separate peak, but bridging activation between other peaks)		
	Right superior/middle temporal cortex, BAs 22, 21, 38	45, −43, 7	6.23
	Right inferior frontal cortex, BAs 47, 45, 46	33, 23, −11	6.50

*
When activations merged into one big cluster at p < 0.001, we determined these peaks by increasing the threshold until the clusters separated, allowing identification of local maxima.