NeuroImage 63 (2012) 137–147

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

On the spatial organization of sound processing in the human temporal lobe: A meta-analysis

Annett Schirmer [a,b,*], P. Mickle Fox [c], Didier Grandjean [d,e,**]

[a] National University of Singapore, Department of Psychology, Singapore
[b] Duke/NUS Graduate Medical School, Singapore
[c] Research Imaging Institute, University of Texas Health Science Center, San Antonio, TX, USA
[d] Neuroscience of Emotion and Affective Dynamics Lab, Department of Psychology, University of Geneva, Switzerland
[e] Swiss Center for Affective Sciences, University of Geneva, Switzerland

ARTICLE INFO

Article history:
Accepted 18 June 2012
Available online 23 June 2012

Keywords:
Voice
Grammar
Lateralization
Discourse
Language
Speech
Melody
Segmentation
fMRI
PET
Lesion

ABSTRACT

In analogy to visual object recognition, proposals have been made that auditory object recognition is organized by sound class (e.g., vocal/non-vocal, linguistic/non-linguistic) and linked to several pathways or processing streams with specific functions. To test these proposals, we analyzed temporal lobe activations from 297 neuroimaging studies on vocal, musical and environmental sound processing. We found that all sound classes elicited activations anteriorly, posteriorly and ventrally of primary auditory cortex. However, rather than being sound class (e.g., voice) or attribute (e.g., complexity) specific, these processing streams correlated with sound knowledge or experience. Specifically, an anterior stream seemed to support general, sound class independent sound recognition and discourse-level semantic processing. A posterior stream could be best explained as supporting the embodiment of sound associated actions and a ventral stream as supporting multimodal conceptual representations. Vocalizations and music engaged these streams evenly in the left and right hemispheres, whereas environmental sounds produced a left-lateralized pattern. Together, these results both challenge and confirm existing proposal of temporal lobe specialization. Moreover, they suggest that the temporal lobe maintains the neuroanatomical building blocks for an all-purpose sound comprehension system that, instead of being preset for a particular sound class, is shaped in interaction with an individual's sonic environment.

Introduction

Sound perception results from the analysis of air pressure waves originating from moving or vibrating objects in the environment. Given that the number of such objects is typically greater than one, the pressure waves that arrive at the ear are highly complex and require sophisticated processing mechanisms to enable sound comprehension. Research into these mechanisms revealed initial projection of sound information from the ear to subcortical centers and then onto the primary auditory cortex in Heschl's gyrus. From there, sound information enters a complex system extending anteriorly, posteriorly and ventrally into the temporal lobe. This system presumably comprises sound class and attribute specific processing modules.

Sound class specific processing has been postulated for the superior temporal gyrus and sulcus (STG/STS) (Belin et al., 2000; Fecteau et al., 2004; Schirmer and Kotz, 2006). Comparisons of human vocalizations and speech with non-vocal sounds revealed activations in the anterior and mid STG/STS that are stronger in the right than in the left hemisphere. Thus, some hold that anterior and mid STG/STS are specialized for the processing of auditory information from conspecifics (Belin, 2006; Petkov et al., 2009). A second traditional notion is that the left-hemisphere is specialized for the processing of speech or linguistically relevant sounds, whereas the right hemisphere is specialized for the processing of music. This notion arose from evidence that left hemisphere lesions (e.g., Wernicke's area in the posterior superior temporal cortex) disturb language functions (Wernicke, 1874), whereas right hemisphere lesions disturb musicality (Confavreux et al., 1992; Peretz et al., 1997).

The processing of sound attributes, such as what, where, and how information, has been linked to dedicated pathways extending from the auditory cortex (Belin and Zatorre, 2000; Rauschecker and Tian, 2000; Zatorre et al., 2002). One such pathway reaches along the STG/STS into the temporal pole. It was identified in electrophysiological recordings in non-human primates (Romanski et al., 1999; Tian et al.,

---

* Correspondence to: A. Schirmer, Department of Psychology, Faculty of Arts and Social Sciences, National University of Singapore, 9 Arts Link, Block AS4, Level 2, 117570, Singapore. Fax: +65 67731843.
** Correspondence to: D. Grandjean, Neuroscience of Emotion and Affective Dynamics Lab, Department of Psychology, University of Geneva, 40 bd du Pont d'Arve, CH-Geneva 1205, Switzerland. Fax: +41 223799844.
E-mail addresses: schirmer@nus.edu.sg (A. Schirmer), didier.grandjean@unige.ch (D. Grandjean).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.06.025

2001) and subsequently replicated in human functional neuroimaging studies (Alain et al., 2001; Arnott et al., 2005; De Santis et al., 2007; for a meta-analysis of human work see Arnott et al., 2004; DeWitt and Rauschecker, 2012). Because of its sensitivity to recognizable or intelligible sounds it appeared important for auditory "what" processing (Scott et al., 2000) and was consequentially termed the auditory "what" pathway. Moreover, because its anterior extension was found to correlate with acoustic complexity and combinatory processing demands, some proposed that it serves sound structure building relevant for the analysis of sound sequences such as words, spoken sentences or musical phrases (DeWitt and Rauschecker, 2012; Friederici, 2012; Lau et al., 2008).

A second pathway extends from auditory cortex posteriorly to the angular gyrus including temporal plane and planum polare. Given its role in sound localization (Rauschecker and Tian, 2000), it was termed the "where" pathway. However, other processes, apart from sound localization depend on posterior STG and thus other functions have been proposed. For example, lesions to Wernicke's area within the posterior STG result in difficulty to access word meaning (Wernicke, 1874) suggesting a role of this region in lexical and/or semantic processing. Accordingly, one proposal is that the posterior STG supports sound-to-meaning transformations (Lau et al., 2008). Another proposal is that posterior STG enables the analysis of spectral changes relevant for sound identification (Belin and Zatorre, 2000). Lastly, some researchers propose the posterior pathway to interface sound perception with motor and somatosensory representations in the frontal and parietal lobes and to support motor and sensory prediction during auditory perception and vocal production (Friederici, 2012; Rauschecker and Scott, 2009).

A third pathway extends from the auditory cortex ventrally into the middle temporal gyrus (MTG) and the temporo-parietal junction. This pathway was first postulated by Hickok and Poeppel (2000). In a literature review, these authors observed that brain lesion and neuroimaging work implies a left-lateralized mid/posterior MTG contribution to semantic processing. Thus, Hickok and Poeppel concluded that the mid/posterior MTG forms the extension of a ventral pathway, which supports the mapping of sound-based lexical representations on semantic knowledge stored elsewhere in the brain. This proposal received support from multiple subsequent studies. Moreover, recent reviews of this work and other work related to speech processing (Friederici, 2012; Price, 2010) highlighted the role of a ventral pathway in meaning analysis albeit reporting an overlap of this pathway with other language functions.

Taken together, existing work suggests a range of functional subdivisions within the temporal lobe. Although each of these divisions is founded on a supporting body of evidence, the latter often comes from specifically designed studies and may not fully consider other, potentially conflicting, work. Moreover, existing evidence is likely compromised by the many limitations of current neuroimaging research including underpowered designs, a prevalence of false positive results and a lack of direct replication (Yarkoni et al., 2010). Therefore, models of auditory processing and temporal lobe function need to also consider larger scale evidence as is afforded by a meta-analysis of published work.

Here, we set out to provide such evidence. In a comprehensive literature search, we identified 297 relevant studies that employed human vocalizations, music and inanimate environmental sounds and that found activations in the temporal lobe. Activations elicited to human vocalizations and music were classified into four levels of structural complexity ranging from low (e.g., syllables/tones) to high complexity (e.g., discourse/orchestral pieces). Using activation likelihood estimation (Eickhoff et al., 2009; Turkeltaub et al., 2002), we determined convergence in brain activation for the different sound classes and complexity levels. Moreover, we contrasted activations elicited by human vocalizations with those of music and environmental sound and we contrasted activations associated with different levels of structural complexity. Through these analyses, we hoped to identify voice, speech or music specialized regions and shed light on some of the proposed sound processing pathways.

Methods

Data collection

The positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies included in the present meta-analysis fulfilled the following criteria. They were published in English, conducted in healthy, young or middle-aged adults, used auditory stimuli, and reported whole brain contrasts with peak activations in the temporal lobe. Studies presenting auditory information concurrently with information in other modalities were excluded, unless the authors analyzed auditory processing separately.

The relevant studies were identified in two steps. First, we searched the BrainMap database (http://www.brainmap.org/; Laird et al., 2005) for studies involving auditory stimulation that elicited activity in the temporal lobe. Second, we conducted a PubMed search limited to articles published before August 31 2010. The keywords (*intonation OR prosody OR prosodic OR vocal OR voice OR speech OR vocalization OR vocalisation OR vocalizations OR vocalisations*) were used for identifying articles on the perception of human vocalizations. The keywords (*animal OR primate OR animals OR primates OR mammal OR mammals*) *AND* (*vocalization OR vocalisation OR call OR sound OR sounds OR vocal OR vocalizations OR vocalisations*) were used for identifying articles on the perception of non-human animal vocalizations. The keywords *(music OR musical)* and *(sounds OR tones OR sound OR tone)* were used for identifying studies on the perception of music and other sounds, respectively. To restrict our search to fMRI or PET studies, we always combined these keywords with the following search terms: (*fMRI OR functional magnetic resonance imaging OR functional neuroimaging OR PET OR positron emission tomography*).

Data set

As the number of studies on animal vocalizations was small (16), animal vocalizations were excluded from this meta-analysis. Likewise excluded were comparisons that involved mixed sound categories (e.g., that combined the presentation of animal vocalizations and environmental sounds) and artificially modified vocalizations that no longer sounded as coming from a human (e.g., noise vocoded speech). The final data set comprised 297 studies, 52 using PET, 244 using fMRI, and one using both PET and fMRI (see Supplementary materials). As the majority of peak activations (71.5%) were specified in Talairach space and space conversions are not without problems, we aimed at reducing noise by converting the smaller number of peak activations specified in MNI (Montreal Neurological Institute) space to Talairach space. Space conversion was achieved using the brainmap toolbox (http://brainmap.org/icbm2tal/). All activations were positive.

For analysis purposes, we specified three factors and sorted all contrasts according to these factors. The first factor was called *Sound Class* and comprised three levels: (1) human vocalizations, (2) music, and (3) environmental sounds. The level human vocalizations included intelligible spoken syllables, words, sentences, discourse as well as purely vocal sounds (e.g., humming or affective bursts). Studies providing activations for the latter kind of stimuli were very small (5). Thus, they were not analyzed separately but simply folded into the main analysis of human vocalizations. This analysis and an analysis excluding these few activations revealed comparable results. The level music included individual tones, simple melodies, harmonies and orchestral pieces. The level environmental sounds included noise and sounds produced by natural events or objects in the environment.

The second factor was called *Contrast Category* and comprised four levels: (1) stimulus vs rest, (2) stimulus vs different stimulus from the same sound class (e.g., spoken sentences vs. spoken words), (3) stimulus vs stimulus from a different sound class (e.g., vocalizations vs music), and (4) stimulus in one task vs the same stimulus in another task (e.g., vocalizations during speaker sex vs phoneme categorization).

The third factor was called *Structural Complexity*. Its levels differed as a function of Sound Class. For human vocalizations, four complexity levels were specified as (1) phonemes/syllables, (2) words, (3) sentences and (4) discourse. For music, four complexity levels were specified as (1) individual tones, (2) individual melodies, (3) melodies in a harmonic context and (4) multi-instrumental pieces. The complexity levels for environmental sounds were not as differentiated as the complexity levels for human vocalizations and music. For a comparison between environmental sounds and human vocalizations, we roughly equated environmental sounds with the complexity level 2 and thus used only spoken words for this comparison. The number of studies for each sound class, contrast and complexity is presented in Table 1.

Data analysis

Given the number of contrasts available for the different factors and factor levels, three main comparisons were deemed feasible and subjected to statistical analysis: (1) human vocalizations vs music, (2) human vocalizations vs environmental sounds, and (3) human vocalizations of different *Structural Complexity* (phonemes/syllables vs words, words vs sentences, sentences vs discourse). The analysis of these comparisons was done using two different approaches. First, we followed the traditional approach taken by previous meta-analyses (e.g., Kim, 2011; Lamm et al., 2011; Sabatinelli et al., 2011) using an in-house Java version of GingerALE (Eickhoff et al., 2009; Turkeltaub et al., 2012; www.brainmap.org) — a publicly available meta-analysis tool based on the activation likelihood estimation (ALE) method developed by Turkeltaub et al. (2002). Second, we modified the traditional approach by using GingerALE in conjunction with a matching procedure aimed at equating study numbers for the factors and factor levels specified above. This was done to prevent our results from being confounded by the contribution of different methodological approaches as captured by the factors *Structural Complexity* and *Contrast Category*. For example, research on human vocalization has explored the phoneme/syllable, word, sentence and discourse levels relatively evenly such that the number of available studies for these levels of *Structural Complexity* is fairly balanced (Table 1). In comparison, research on music has explored the tone level more frequently than other levels of *Structural Complexity*. Thus, a traditional, unmatched comparison of human vocalizations and music might reveal processing differences that arise from differences in *Structural Complexity* rather than *Sound Class*. Our second analysis addressed this concern. The results reported in this paper were significant with both the traditional and the modified approach. Both approaches are described below.

Table 1
Number of studies for variables of interest.

	Human vocalizations	Music	Environmental sounds
Contrast category 1: stimulus > rest			
Structural Complexity 1	17	25	
Structural Complexity 2	23	3	21
Structural Complexity 3	19	3	
Structural Complexity 4	10	2	
Contrast Category 2: stimulus > different stimulus from the same sound class			
Structural Complexity 1	16	10	
Structural Complexity 2	27	6	14
Structural Complexity 3	15	2	
Structural Complexity 4	2	4	
Contrast Category 3: stimulus > different stimulus from a different sound class			
Structural Complexity 1	15	30	
Structural Complexity 2	32	7	18
Structural Complexity 3	43	2	
Structural Complexity 4	13	3	
Contrast Category 4: stimulus > same stimulus in another task			
Structural Complexity 1	5	8	
Structural Complexity 2	11	2	2
Structural Complexity 3	11	0	
Structural Complexity 4	2	1	
Total[a]	196	89	43

[a] The total number of studies listed here is NOT the sum of studies listed above. This is because some studies performed multiple contrasts and included stimuli of different sound classes or complexity levels. Hence, the sum of the total number of studies using vocalizations, music and environmental sounds, respectively, is greater than the total number of studies (N = 297) included in this meta-analysis.

For the traditional approach, activation foci for a particular *Sound Class* were organized by study and Gaussian widths were calculated as described by Eickhoff et al. (2009). Subsequently, a modeled activation (MA) map of activation likelihood was computed for each study by taking the maximum probability associated with any one focus of the study. The voxelwise union of the probabilities in all the MA maps was then used to calculate an activation likelihood estimation (ALE) map indicating the likelihood of a given voxel to be activated (Turkeltaub et al., 2012). The ALE map was tested against a null-distribution reflecting a random spatial association of MA maps across studies (Eickhoff et al., 2009). Unless stated otherwise, we used the false discovery rate (FDR) method for multiple comparison correction with an FDR value of 0.05 (Genovese et al., 2002). For the purpose of describing the activations for the different *Sound Class*es, we subjected the thresholded ALE maps to the cluster analysis implemented in GingerALE with a minimum cluster volume size of 15 mm^3. We choose a minimum that only removed single-voxel clusters in order to avoid overly stringent thresholding and to enable a thorough exploration of *Sound Class* related effects. For the purpose of describing *Sound Class* differences and differences between human vocalizations of different *Structural Complexity*, we subjected the thresholded ALE maps to the subtraction algorithm implemented in GingerALE with an FDR corrected p-value of 0.05. The results of this subtraction then underwent the same cluster analysis as was described for the thresholded ALE maps above.

The modified analysis approach used a similar basic procedure as the traditional approach with the exception that study numbers were matched across conditions of interest before being subjected to GingerALE. Specifically, we tried to match the number of studies for levels of *Structural Complexity* and *Contrast Category* in the comparisons of interest (i.e., human vocalizations vs music; human vocalizations vs environmental sounds; human vocalizations of different structural complexity levels) as to eliminate the influence of these factors on the results. To this end, we checked which of the two conditions in a comparison had the smaller number of studies for a given level of *Structural Complexity* and *Contrast Category* (Table 1). For example, the comparison of human vocalizations and music was collapsed over all levels of *Structural Complexity* and *Contrast Category*, while keeping the absolute number of studies for the levels of these factors comparable between voice and music. We used the smaller number of studies (i.e., either the number of studies available for human vocalizations or music) to randomly select the same number of studies from the other condition thus generating stimulus sets that were partially fixed and partially random. As can be inferred from Table 1, for human vocalizations we would select *all* available studies for *Structural Complexity* (SC) level 1/ *Contrast Category* (CC) level 1, SC level 1/CC level 3, SC level 1/CC level 4, and SC level 4/CC level 2. Thus, the studies selected for these levels were fixed. For the remaining combinations of *Structural Complexity* and *Contrast Category*, we would randomly pick studies from the number of available human vocalization studies to match the smaller number of available music studies. To make full use of the available data and to reduce the impact of outliers, we randomly created 100 such data sets for each

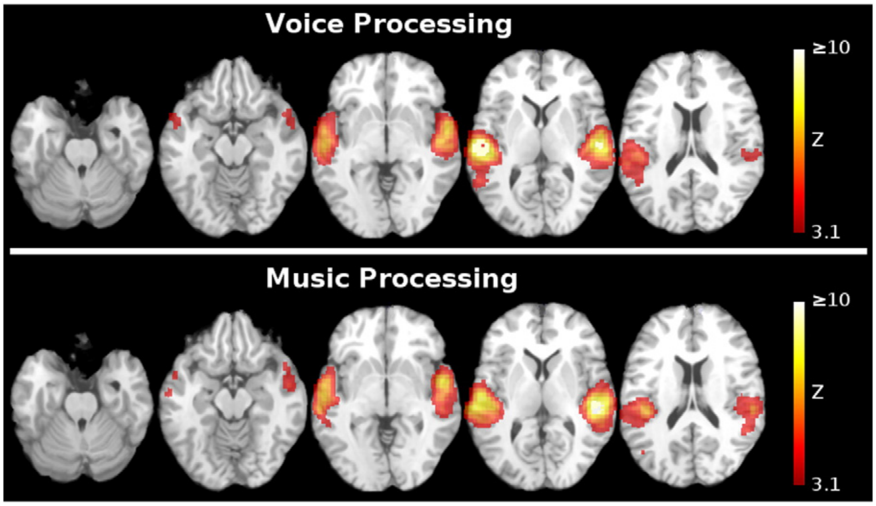

Fig. 1. Z-maps with significant activation clusters for human vocalizations and music in neurological space (left = left). Z-coordinates for the axial slices are −24, −14, −4, 6, and 10 from left to right, respectively.

condition. Returning to our example of the human vocalization/music comparison, we would repeat 99 times the partially fixed/partially random selection of studies for human vocalizations and music as described above.

To examine the activation patterns in individual conditions, the data sets derived in the previous step were subjected to GingerALE, which produced thresholded ALE-maps as well as corresponding z-maps for each data set. If appropriate (i.e., this did not apply for environmental sounds as explained further below), we then performed a conjunction analysis on the 100 z-maps considering a voxel significant if it was activated above threshold in 95 of the 100 maps. The resulting conjunction map was subjected to the cluster analysis implemented in GingerALE with a minimum cluster volume size of 15 mm^3. Between condition comparisons were conducted in two ways. First, we performed 100 subtraction analyses by subtracting the 100 maps created for one condition (e.g., human vocalizations) from the map(s) created for the other condition (e.g., music) in a given comparison. In the comparison of human vocalizations and music and the comparison of the different voice complexity levels, we picked a map in one condition and randomly selected without replacement a map from the other condition for subtraction. In the comparison of human vocalizations and environmental sounds, we performed 100 subractions using the 100 voice maps and the one map computed for environmental sounds. For environmental sounds, study numbers were consistently smaller than for human vocalizations such that they always entered the final contrast set and no random selection was possible. Subtractions were done using the algorithm implemented in GingerALE with an uncorrected p-value of 0.05 and the results again subjected to a conjunction analysis, which identified voxels that were significant in 95 out of 100 maps. Finally, the cluster analysis implemented in GingerALE was applied to the conjunction map using a minimum cluster volume size of 15 mm^3.

By controlling for major confounds such as *Structural Complexity* and *Contrast Category* and conducting 100 randomly sampled analyses of which 95 had to be significant, this second approach was very rigorous. Moreover, it was potentially too conservative as the GingerALE subtraction algorithm already conducts 5000 permutations in which activations from the two conditions are randomly drawn and compared thereby controlling for differences in study numbers (note that this control does not address methodological confounds between the conditions of interest). Thus, we chose a relatively liberal uncorrected p-value of 0.05 here. To ensure that this liberal p-value did not lead to false-positive activations, we only report activations that were significant with both the traditional approach using an FDR corrected p-value and the modified approach.

Results

Voice and music processing

As described in the Methods section, voice and music processing were assessed by collapsing over the different levels of *Structural Complexity* and *Contrast Category* while keeping the number of studies for each level of these factors comparable between the two *Sound Classes*. Moreover, we first examined each *Sound Class* separately and subsequently performed a subtraction.

Human vocalizations produced two clusters (Fig. 1, Table 2, Supplementary Material Map1). The first and larger cluster was located in the left hemisphere with the largest extrema centering in the MTG and four other extrema centering along the STG. The cluster extended from primary auditory cortex to STG/STS/MTG reaching into the posterior insula, anteriorly into the temporal pole and posteriorly into the supramarginal gyrus. A second cluster was located in the right hemisphere with three extrema along the STG. Like the left hemisphere cluster, it extended from auditory cortex to STG/STS/MTG and reached into the posterior insula, the anterior pole and the supramarginal gyrus. For vocalizations and the other sound types examined here, we conducted a lateralization analysis. To this end, we flipped the coordinates for voice activations on the X-axis for all 100 stimulus sets and subtracted each flipped set from its original. The results were then subject to a conjunction analysis as described in the Methods section. This approach revealed no significant differences for vocalizations between the left and right hemispheres. However, with a lower significance criterion,[1] we observed a small cluster of greater activation likelihood in the left hemisphere. The cluster was located in the medial aspect of the primary auditory cortex.

Music produced four clusters (Fig. 1, Table 2, Supplementary Material Map2). The largest one was located in the right hemisphere and comprised extrema in STG/STS/MTG. It extended from the primary auditory cortex to the posterior insula, the anterior pole and the supramarginal gyrus. The first music cluster was thus comparable to the voice clusters reported above. The second music cluster was located in the left hemisphere. Its extrema were widely distributed across primary auditory cortex, STG/STS, MTG and even included the ITG. Apart from these extrema,

---

[1] All analyses were conducted 100 times drawing randomly from a pool of available studies. Activations were considered significant, if they emerged in more than 95 cases (see the Methods section for more details). For the analysis reported here, we reduced this threshold to 80 cases.

Table 2
Comparison of vocalizations and music.

Cluster	Activation extrema (BA)	Talairach (x,y,z)	Extent (mm^3)	Z	Included subregions (BA)
Human vocalizations					
1	L MTG (21)	−53,−22,5	27,856	Inf	L AC (41,42)
	L STG (22)	−54,−52,6		4.1	L temporal pole (38)
	L STG (39)	−46,−52,16		3.4	L MTG (37)
	L posterior insula (13)	−52,−44,18		3.2	L supramarginal gyrus (40)
	L STG (22)	−58,−48,14		3.2	L postcentral gyrus (43)
2	R STG (22)	53,−16, 2	22,488	Inf	R MTG (21)
	R STG (38)	48,8,−10		4.1	R AC (41,42)
	R STG (38)	54,12,−16		3.1	R posterior insula (13)
					R temporal pole (38)
					R supramarginal gyrus (40)
Music					
1	R MTG (21)	51,−27,6	27,008	Inf	R AC (41,42)
	R STG (22)	54,−6,−6		7.6	R posterior insula (13)
	R MTG (21)	54,−44,10		5.3	R supramarginal gyrus (40)
	R STG (38)	50,12,−8		3.8	R MTG (39)
	R STG (38)	48,8,−16		3.3	
2	L STG (22)	−56,−17,2	26,120	Inf	L AC (42)
	L STG (22)	−60,−28,6		7.8	L posterior insula (13)
	L AC (41)	−44,−28,10		7.6	L temporal pole (38)
	L AC (41)	−54,−26,8		7.5	L supramarginal gyrus (40)
	L AC (41)	−48,−26,8		7.2	
	L MTG (37)	−52,−52,2		3.4	
	L MTG (21)	−58,−40,−4		3.4	
	L ITG (37)	−56,−56,0		3.4	
	L MTG (21)	−50,0,−16		3.3	
3	L fusiform gyrus (37)	−44,−54,−10	96	3.1	NA
4	L MTG (39)	−42,−68,18	88	3.4	NA
Human vocalizations > music					
1	L AC (42)	−60,−20,10	128	1.9	NA
	L AC (41)	−56,−18,12		1.8	
2	L MTG	−52,−32, 0	16	1.7	NA
Human vocalizations > music (syllables and tones only)					
1	R MTG	62,0,−8	88	1.86	NA

the second music cluster also included secondary auditory cortex, posterior insula, anterior pole, and supramarginal gyrus. The third and the forth clusters were located in the left fusiform gyrus and the left MTG, respectively. A statistical comparison between hemispheres revealed two regions that were more likely activated in the left than in the right hemisphere and two regions that were more likely activated in the right than in the left hemisphere (Table 3). Compared to activations in the right hemisphere, activations in the left hemisphere extended more posteriorly, dorsally and medially in the auditory cortex as well as more ventrally and laterally in the MTG. Please note that the extension in the auditory cortex compares to the subthreshold lateralization effect observed for vocalizations. Compared to activations in the left hemisphere, activations

Table 3
Lateralization patterns — activation likelihood clusters that are greater in one hemisphere as compared to the other.

Cluster	Activation extrema (BA)	Talairach (x,y,z)	Extent (mm^3)	Z	Included subregions (BA)
Human vocalizations (matched for comparison with music)					
1	L AC (41)[a]	−42,−32,10	112	1.8	NA
Music					
1	L Insula (13)	−40,−28,20	456	2.1	L AC (41)
2	L MTG (21)	−66,−32,−2	288	2	L MTG (22)
3	R MTG (21)	49,−32,−1	184	2	R MTG (22)
4	R AC (42)	64,−18,12	72	1.7	NA
Human vocalizations (matched for comparison with environmental sounds)					
1	L AC (41)	−34,−30,10	40	1.7	NA
Environmental sounds					
1	L STG (22)	−54,−47,17	3224	3.7	L STG (13, 39)
	L MTG (21)	−49,−41,13		3.5	L supramarginal gyrus (40)
					L AC (41)
2	L postcentral gyrus (40)	−52,−24,16	208	2.2	L AC (41)
3	L MTG (21)	−58,−20,2	144	1.8	L STG (22)
					L AC (41)
4	R fusiform gyrus (37)	50,−44,−8	112	2.3	R MTG (20)
5	L AC (41)	−36,−30,16	64	1.9	L insula (13)

[a] This activation showed at a lower significance criterion. It was significant in at least 80 of the 100 analyses conducted for vocalizations. For all other tests activations had to be significant in at least 95 of the 100 analyses conducted.

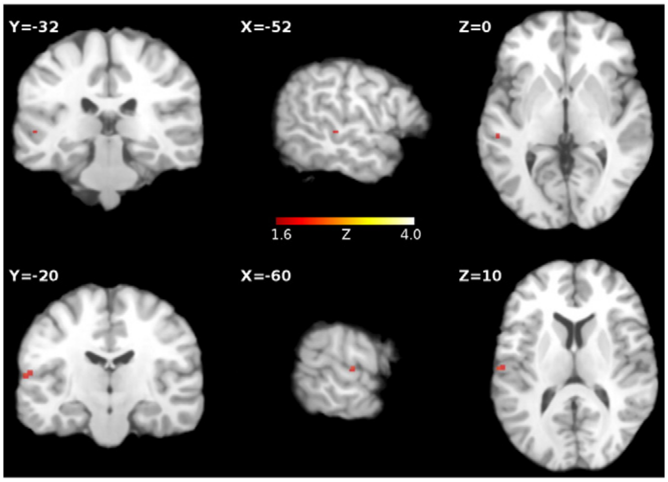

Fig. 2. Z-maps with significant activation clusters for the subtraction analysis of human vocalizations and music in neurological space (left = left). Areas marked in red reflect greater activation likelihood for human vocalizations as compared to music. Coordinates for coronal, sagittal and axial slices are provided in white.

in the right hemisphere extended more laterally in the auditory cortex as well as more ventrally and medially in the MTG.

A subtraction analysis of human vocalizations and music revealed only two small clusters (Fig. 2, Table 2, Supplementary Material Map3). Vocalizations were more likely than music to activate the left primary/secondary auditory cortex and the left STS/MTG. A comparison with previously identified voice sensitive regions was achieved through visual inspection of a voice localizer developed and used by Belin and others (Belin et al., 2000; Frühholz et al., 2012). As this localizer was in MNI space, Talairach coordinates from cluster peaks of the present study were converted to MNI coordinates in Matlab using non-linear transformation (http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach). Visual identification of these coordinates on the MNI voice localizer indicated that neither of the two clusters fell neatly within putative voice regions. Instead, the first clusters bordered dorsally and the second cluster ventrally. There were no brain regions more likely activated for music than for vocalizations.

An additional subtraction analysis restricted to spoken syllables and tones could not replicate the two voice clusters suggesting that they were driven by higher order language processing. Compared to tones, spoken syllables or phonemes were more likely to activate an area in the right MTG. Again, this region failed to clearly fall within previously identified voice regions (Supplementary Material Map4). No regions were more strongly activated for tones than syllables.

Voice and environmental sound processing

As for the comparison between voice and music, we matched the number of studies contributing to the comparison between voice and environmental sounds. Additionally, we matched *Structural Complexity* and *Contrast Category* by restricting the selection of voice studies to those of spoken words and by equating study numbers at the four *Contrast Category* levels as specified in the Methods section. Therefore, the voice studies included here differ from those used for the music comparison.

Nevertheless, we obtained voice activations comparable to those reported above. Specifically, there were two clusters (Fig. 3, Table 4, Supplementary Material Map5). The larger, left hemisphere cluster had one extrema in the mid STG, one in the posterior insula and one in the posterior STG. The cluster included primary and secondary auditory cortices, postcentral gyrus, STS, and MTG reaching anteriorly into the temporal pole and posteriorly into the supramarginal gyrus. The second cluster was located in the right hemisphere with four

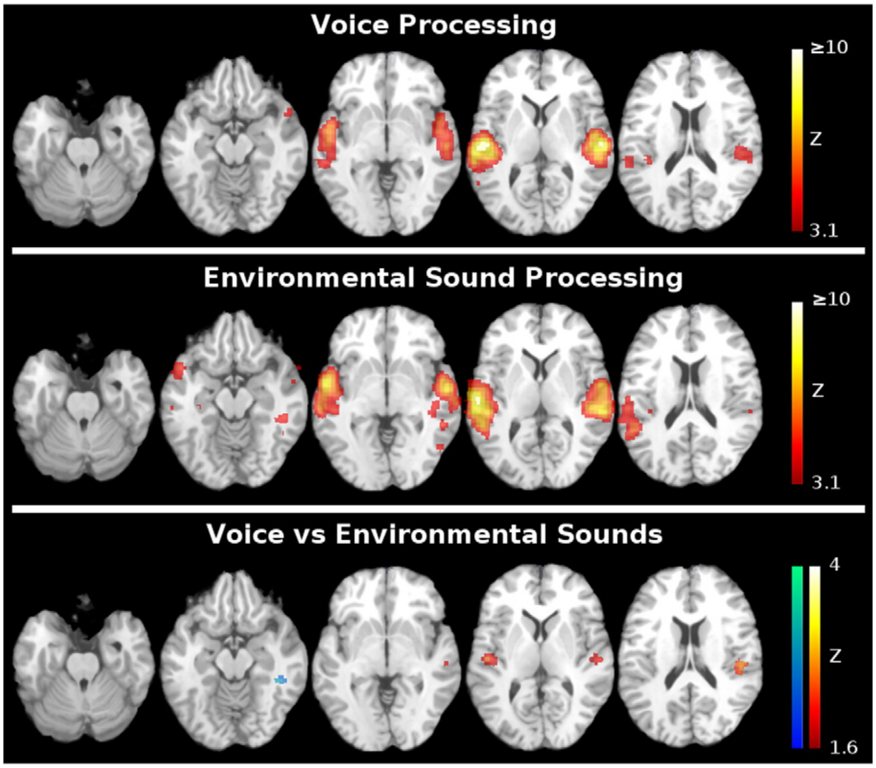

Fig. 3. Z-maps with significant activation clusters for human vocalizations (i.e., words), environmental sounds, and human vocalizations versus environmental sounds in neurological space (left = left). For the latter maps, greater activation likelihood for vocalizations is illustrated in red/yellow, while greater activation likelihood for environmental sounds is illustrated in blue/green. Z-coordinates for the axial slices are −24, −14, −4, 6, and 10 from left to right, respectively.

Table 4
Comparison of vocalizations and environmental sounds.

Cluster	Activation extrema (BA)	Talairach (x,y,z)	Extent (mm^3)	Z	Included subregions (BA)
Human vocalizations					
1	L STG (22)	−52,−21,5	19,088	Inf	L AC (41,42)
	L posterior insula (13)	−36,−38,18		3.4	L MTG (20, 21, 37)
	L STG (22)	−52,−54,6		3.2	L temporal pole (38)
					L supramarginal gyrus (40)
					L postcentral gyrus (43)
2	R STG (22)	56,−21,4	16,080	Inf	R AC cortex (41,42)
	R STG (22)	50,−12,4		7.1	R MTG (21)
	R STG (22)	50,2,−4		4.9	R temporal pole (38)
	R STG (38)	52,10,16		3.7	R posterior insula (13)
					R precentral gyrus (40,43)
Environmental sounds					
1	L MTG (21)	−58,−24,4	26,576	Inf	L AC (41,42)
	L STG (22)	−54,−6,−4		8.0	L temporal pole (38)
	L STG (22)	−54,−12,0		7.6	L posterior insula (13)
	L STG (22)	−48,−32,4		7.2	L MTG (37,39)
	L MTG (21)	−52,−42,6		6.7	L postcentral gyrus (43)
	L STG (22)	−62,−40,16		3.6	L ITG (19)
2	R STG (22)	56,−10,−2	18,360	7.6	R AC (42)
	R AC (41)	58,−30,6		7.5	R posterior insula (13)
	R MTG (21)	54,−44,−2		4.2	R precentral gyrus (43)
	R fusiform gyrus (37)	46,−38,−12		4.1	R fusiform gyrus (20)
	R temporal pole (38)	50,8,−10		3.8	R parahippocampal gyrus (36)
3	R temporal pole (38)	30,10,−30	344	4.1	NA
4	R MTG (21)	62,8,−16	216	3.5	NA
	R temporal pole (38)	56,12,−20		3.5	
5	R ITG (19)	50,−64,−4	208	3.9	R middle occipital gyrus (37)
6	L temporal pole (38)	−40,6,−16	56	3.2	NA
7	L MTG (21)	−40,−4,−32	40	3.2	NA
8	L ITG (20)	−56,−28,−16	40	3.2	NA
9	L parahippocampal gyrus (36)	−30,−26,−14	32	3.2	NA
10	L fusiform gyrus (37)	46,−50,−14	16	3.1	NA
Human vocalizations > environmental sounds					
1	R AC (41)	48,−28,18	1784	2.8	R AC (42)
					R STG (22)
					R posterior insula (13)
					R MTG (21)
					R postcentral gyrus (40)
2	L STG (22)	−48,−20,4	1296	2.9	L MTG (21)
	L posterior insula (13)	−40,−20,6		2.0	
	L AC (41)	−42,−30,8		1.9	
3	L AC (41)	−38,−38,10	32	2	NA
Environmental sounds > human vocalizations					
1	R fusiform gyrus (20)	40,−40,−14	472	2.8	R fusiform gyrus (37)
					R parahippocampal gyrus (36)

extrema distributed along the STG. It included primary and secondary auditory cortices, posterior insula, precentral gyrus, STS, and MTG reaching anteriorly into the temporal pole. A lateralization analysis revealed again a small significant cluster in the medial aspect of the auditory cortex that was more likely to be activated in the left as compared to the right hemisphere (Table 3). This cluster was comparable to the sub-threshold cluster observed in the lateralization analysis for vocalizations and the significant cluster for music reported above.

Ten clusters emerged from the analysis of environmental sounds (Fig. 3, Table 4, Supplementary Material Map6). The first cluster was located in the left hemisphere and was about 40% larger than the corresponding vocal cluster. Its extrema lay in mid and posterior STG and posterior insula. It included primary and secondary auditory cortices, MTG reaching anteriorly into the temporal pole and posteriorly into the supramarginal gyrus. The second cluster was located in the left hemisphere with extrema in the primary auditory cortex, STG/STS, MTG, temporal pole and fusiform gyrus. It also included secondary auditory cortex, posterior insula, precentral gyrus, and parahippocampal gyrus. Other, smaller clusters were located in the temporal pole, MTG and ITG of the right and left hemispheres as well as the left parahippocampal gyrus. A lateralization analysis revealed significantly greater activation likelihood in the left as compared to the right hemisphere. Specifically, there were four clusters more strongly activated in the left as compared to the right hemisphere. They were located in STG, MTG, postcentral gyrus, and auditory cortex, respectively (Table 3). There was only one small cluster with greater activation likelihood in the right as compared to the left hemisphere. It was located in the fusiform gyrus.

The subtraction of voice and environmental sound activations revealed the following results. Voice was more likely to activate auditory regions in the right and left hemispheres (Supplementary Material Map7). The right hemisphere cluster peaked in the primary auditory cortex and included secondary auditory cortex, STG/STS, posterior insula, MTG and the postcentral gyrus. The smaller left hemisphere cluster had three extrema located in the STG, primary auditory cortex and posterior insula. It reached into the STS and MTG. In a comparison with previously reported voice-sensitive regions, we observed the clusters identified here to be located more dorsally. Nevertheless, both, especially the right-hemisphere cluster bled into putative voice regions (Fig. 3). A third, small cluster was located in the left primary auditory cortex outside putative voice regions. Environmental sounds were more likely than vocalizations to activate the right fusiform gyrus (Fig. 3, Table 4, Supplementary Material Map8).

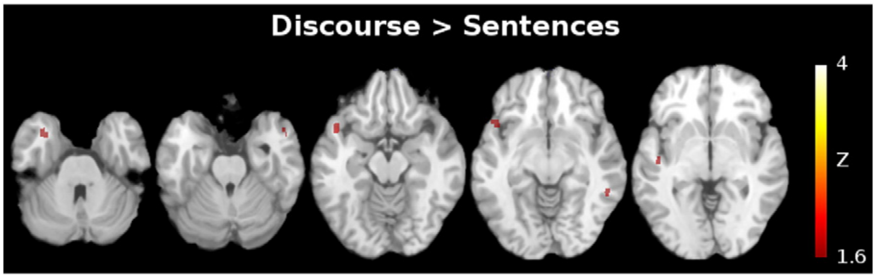

Fig. 4. Z-maps with significantly greater activation likelihood for human vocal discourse as compared to spoken sentences in neurological space (left = left). Z-coordinates for the axial slices are −29, −24, −14, −9, and −4 from left to right, respectively.

Structural complexity effects during vocal processing

An analysis of structural complexity effects would be of interest for sounds other than vocalizations. However, due to limitations of the available data set, such an analysis was feasible for vocalizations only. For the latter, we hypothesized an increase in activation likelihood with increasing structural complexity. To test this, we compared increasing levels of complexity matched for contrast type and study number. This comparison was non-significant for phonemes/syllables versus words and for words versus sentences. However, we observed five significant clusters for the comparison of sentences versus discourse (Fig. 4, Table 5, Supplementary Material Map9). The first two clusters were located in the left temporal pole reaching into the inferior frontal gyrus and the uncus. The other three clusters were located in the right ITG, the right temporal pole/MTG and the left STG, respectively. All these activations were greater for discourse than sentences. No significant effects were observed for the opposite contrast.

Discussion

The present study sought to test current notions about temporal lobe functional specialization. Of interest was whether and how activations elicited to human vocalizations differ from those elicited to other sounds and how a sound's structural complexity shapes the spatial pattern of activation. In the following we will discuss our findings with respect to prominent sound class specific processing proposals in the literature as well as proposals regarding neuroanatomical pathways specialized for certain sound attributes or informational content.

Sound class specific processing

For social animals like humans, sounds produced by conspecifics are probably more relevant than other sounds. Therefore, some have postulated the evolution of brain mechanisms that are particularly adapted to within-species auditory communication (Belin, 2006; Petkov et al., 2009). We tested this assertion by comparing vocalizations with other sounds. In a comparison with music, vocalizations were more likely to activate the left primary and secondary auditory cortices reaching into the STG and a small region within the left STS/MTG. In a comparison with environmental sounds, vocalizations were more likely to activate auditory cortex bilaterally with activations extending into the STS and MTG. Relative to the comparison with music, the latter effects differed somewhat in that they were larger and more likely to overlap with putative voice regions. Together, these results imply differential processing of the voice. Contrary to previous work (Belin, 2006; Petkov et al., 2009), however, they do not suggest voice regions that extend anteriorly along the STG/STS. The STG/STS involvement observed here differed morphologically from voice activations in previous studies. It was primarily an extension of auditory cortex activations and showed more consistently (i.e., in comparison with both music and environmental sounds) in the left than in the right hemisphere. Based on this and the observation that non-meaningful syllables did not activate the regions mentioned above when compared with single tones, we conclude that they primarily serve linguistic or language specific acoustic processing as opposed to vocal processing (see discussion below).

What could be reasons for the small and relatively inconsistent voice regions found in the present meta-analysis? First, it is possible that robust indicators for voice-sensitive processing emerge only when vocalizations are contrasted with environmental sounds. Moreover, the music typically used in neuroimaging research (e.g., instrumental melodies) may be too similar to the voice to elicit substantial auditory processing differences (Escoffier et al., 2012; Fedorenko et al., 2009; Slevc et al., 2009). Although we cannot exclude this possibility, we would like to note that research, which directly compared the processing of human and non-human animal vocalizations, revealed larger activity in the STG/STS for the former relative to the latter in

Table 5
Comparison of vocalizations of different complexity levels.

Cluster	Activation extrema (BA)	Talairach (x,y,z)	Extent (mm^3)	Z	Included subregions (BA)
(1) Syllables > (2) words *ns*					
(2) Words > (1) syllables *ns*					
(2) Words > (3) sentences *ns*					
(3) Sentences > (2) words *ns*					
(3) Sentences > (4) discourse *ns*					
(4) Discourse > (3) sentences					
1	L temporal pole (38)	−48,6,−14	488	2.1	L inferior frontal gyrus (47)
2	L STG (38)	−30,2,−32	216	1.7	NA
	L uncus (28)	−28,4−32			
	L STG (38)	−32,8,−30			
3	R ITG (20)	56,−46,−8	120	1.8	NA
	R subgyral (37)	52,−48,−8		1.8	
4	R temporal pole (38)	50,10,−22	80	1.7	NA
	R MTG (21)	50,4,−24		1.7	
5	L STG (22)	−46,−20,−4	40	1.8	NA

human listeners (Fecteau et al., 2004). This suggests that voice sensitive regions emerge even in a comparison with acoustically similar stimuli.

Therefore, we would like to provide a second possible explanation for the present results. This explanation derives from the nature of putative voice regions. Specifically, there is evidence from single cell-recordings in non-human primates that these regions are not "pure" in the sense that they support voice processing only (Perrodin et al., 2011). Instead, they seem to maintain a mixture of voice-sensitive cells and cells that show no voice sensitivity. Moreover, while the voice-sensitive cells are particularly excited by vocal stimulation, they also respond to non-vocal stimulation. Thus, previous observations of voice regions in human functional imaging studies may be explained by differences in activation strength rather than the location of activations elicited by vocalizations relative to other sounds.

The present study was designed to identify differences in location of activations rather than activation strength. Using activation likelihood, it compared the spatial distribution of voice activations with the spatial distribution of activations from other sounds obtained in separate contrasts and studies. Thus, it could identify differences in spatial activation patterns but was essentially blind to differences in activation strength. Given that this approach resulted in little evidence for dedicated voice regions, particularly in the comparison between voice and music, the present findings corroborate the notion that the temporal lobe serves as an all-purpose sound processor that is simply more excited by human vocalizations. Due to their particular relevance, vocalizations, like other social stimuli such as faces or interpersonal touch (Escoffier et al., 2010; Schirmer et al., 2011; Vuilleumier et al., 2001), may trigger bottom-up and/or top-down shifts in attention that enhance sensory processing.

Support for this reasoning comes from past research and the present study. Specifically, past research showed that emotional vocalizations elicit stronger activations in putative voice regions than do neutral vocalizations that are arguably less relevant (Grandjean et al., 2005; Schirmer et al., 2008). Furthermore, the present study found smaller differences between vocalizations and music than between vocalizations and environmental sounds. Given that the kind of music typically used in the studies that were examined here (e.g., instrumental melodies) was more similar to vocalizations in terms of acoustic properties and social relevance than were environmental sounds (Escoffier et al., 2012; Remedios et al., 2009), music may also be more likely than environmental sounds to engage both bottom-up and/or top-down attention to enhance sensory processing.

A second hypothesis tested in the present meta-analysis was that speech and music processing are lateralized to the left and right hemispheres, respectively. In line with previous evidence, we observed a left-lateralization for speech. However, this lateralization effect was relatively small and while significant at the word level showed only subthreshold when collapsed across phonemes/syllables, words, sentences and discourse. Its location in the medial aspect of the primary auditory cortex suggests that it arises from hemispheric specialization in acoustic feature processing (Zatorre and Belin, 2001) as opposed to linguistic processing.

Surprisingly, this meta-analysis failed to reveal a clear right hemisphere dominance for music. Although parts of secondary auditory cortex and the MTG had a greater activation likelihood in the right as compared to the left hemisphere, there were also two temporal regions that showed the opposite pattern. One region was located in the primary auditory cortex and overlapped with the lateralization cluster obtained for vocalizations. The other region was located in the MTG. Both regions were numerically greater than the regions that showed greater activation likelihood in the right hemisphere. Thus, left hemisphere mechanisms seem to be equally if not more important than right hemisphere mechanisms in the processing of music. The failure to observe a right lateralization for music may be due to our restricted focus on the temporal lobe. Perhaps such lateralization arises primarily in frontal (Peretz et al., 1997; Yasui et al., 2009) and/or parietal regions (Confavreux et al., 1992).

Notably, activations to both vocalizations and music were less lateralized than activations to environmental sounds. Besides a left lateralization in the medial aspect of the primary auditory cortex that overlapped with the lateralization of vocalizations and music, environmental sounds showed a left lateralization for large parts of the STG and MTG. This conflicts with long standing notions of brain lateralization that saw the left hemisphere specialized for speech (Broca, 1863; Wernicke, 1874). However, these notions were derived primarily from patients with hemispheric lesions rather than from more recent neuroimaging research. Moreover, a re-visitation of deficits incurred by hemispheric lesions has begun to challenge old models of hemispheric specialization. Among others such efforts revealed that certain language functions including the comprehension of prosody, irony and discourse require both left and right hemisphere mechanisms (Pell, 2006; Schirmer et al., 2001; Winner et al., 1998). Based on this and the present meta-analysis it seems that vocalizations and music are more likely to recruit processing in both hemispheres than are environmental sounds. Being less complex and non-communicative, the latter may be sufficiently represented by left hemisphere mechanisms with only a partial contribution from the right hemisphere.

Sound attribute specific processing

The anterior pathway

Previous work suggests that auditory processing involves an anteriorly directed pathway contained within the STG/STS supporting sound recognition (Belin and Zatorre, 2000; Rauschecker and Tian, 2000). The anterior extension of this pathway has been associated with factors such as acoustic complexity (Binder et al., 2000; Schönwiesner et al., 2005), sound intelligibility (Scott et al., 2000), and combinatory demands (Friederici, 2012; Lau et al., 2008). In line with this, the present study observed an anterior extension of activity within the STG/STS for all sounds tested. Its function was assessed in two ways. The first, and perhaps less appropriate, way was to compare speech with other sounds matched for contrast type and combinatory complexity. The rationale for this comparison was that speech may be acoustically more complex, due to its consonant/vowel structure and rich harmonic changes (Price, 2010) and that this acoustic complexity would lead to greater activation. In line with this, we observed greater activation in the auditory cortex extending into the STG when comparing all vocalizations against music and spoken words against environmental sounds. Notably, however, these activations failed to demarcate a clear anterior pathway but merely extended laterally and ventrally from the primary auditory cortex. Moreover, no STG differences were observed in a contrast of phonemes/syllables and tones suggesting that the anterior pathway may be sensitive to sound features other than acoustic complexity.

A second way of elucidating anterior pathway function was to compare vocalizations of different structural complexity levels. Surprisingly, no differences emerged between syllables, words and sentences suggesting that they equally engage superior temporal cortex anterior to primary sensory regions (but see DeWitt and Rauschecker, 2012). As one would expect sentences to engage more combinatorial processes than words and syllables, the functions of the anterior pathway also do not seem combinatorial in nature. Instead this pathway may support sound recognition in general and was hence activated across the different sound classes examined here. This would be in line with evidence that the anterior pathway is sensitive to sound intelligibility (Scott et al., 2000). Additionally, the anterior pathway may play a role in discourse semantic processing. The latter can be inferred from the present finding that the bilateral temporal pole is more likely to be activated in response to spoken discourse relative to other speech sounds. As this activation was accompanied by activity in MTG and ITG one may speculate that the bilateral temporal pole is part of a widespread semantic network that integrates anterior and ventral processing streams. This proposition is in line with the discovery of white matter fiber connections within the temporal lobe that connect MTG and ITG with anterior temporal lobe and eventually inferior frontal regions via the external capsule (Saur et al., 2008; Weiller et al., 2011).

The posterior pathway

All sounds investigated in the present study activated regions extending posteriorly from the auditory cortex along the STG/STS. Such a posterior extension has been reported before and associated with a range of functions. These functions include spatial processing (Rauschecker and Tian, 2000; Zatorre et al., 2002), early mapping of sound to meaning (Lau et al., 2008), perception of spectral change (Belin and Zatorre, 2000) and the activation of somatosensory and motor representations for sound embodiment (Rauschecker and Scott, 2009). Due to methodological constraints, not all proposed functions could be examined here. However, one may consider implications of the present results for the processing of spectral change, sound-to-meaning mapping and sound embodiment.

A potential role of the posterior pathway in spectral change may be inferred from a comparison of sounds that entail much spectral change with sounds that entail little spectral change. Here, we approximated this comparison by contrasting spoken syllables with tones. As mentioned above, this contrast failed to activate the STG and thus did not demarcate a posterior pathway. A potential role of the posterior pathway in sound-to-meaning mapping, may be inferred from a comparison of speech and other sounds. Such a comparison might isolate sound-to-meaning mapping because listeners routinely assign meaning to speech but are less likely to do this for other sounds. Moreover, sound-to-meaning mapping should be least likely for music where notes do not neatly map onto semantic concepts and intermediate for environmental sounds whose source or underlying action is potentially recognizable (e.g., door closing). If true, involvement of the posterior pathway should be greatest for voice followed by environmental sounds and then music processing. The present findings do not fit this pattern. Although we found differences between voice and environmental sound processing, there were no differences between voice and music suggesting that the posterior pathway does not specifically support sound-to-meaning mapping. However, as we cannot exclude the possibility that sound-to-meaning mapping is equally recruited for voice and music (Koelsch, 2005), rejecting sound-to-meaning mapping as a function for the posterior pathway may be premature.

To gain insights into a potential role of the posterior pathway in sound embodiment, we considered comparing sounds that are readily embodied with sounds that are not. We reasoned that most of us are highly proficient in imitating our native language, able to hum along a tune but would have difficulties imitating environmental sounds such as thunder or rain. Moreover, as the social relevance of these sounds varies, so may the likelihood that we attempt imitation or embodiment. If true and the posterior pathway supports sound embodiment, it should be more likely to be activated by vocalizations and music relative to environmental sounds. In line with this, we found no differences between vocalizations and music, but a significant difference between vocalizations and environmental sounds in both hemispheres with the right hemisphere difference extending more posteriorly than the left.

The ventral pathway

The last pathway to be mentioned here extends from primary auditory cortex ventrally into the MTG and the inferior temporal lobe. It has been proposed to support the storage of lexical representations and/or conceptual information (Chandrasekaran et al., 2011; Hickok and Poeppel, 2000; Lau et al., 2008). To test these proposals, we compared speech with music and environmental sounds as this should isolate lexical processing most likely to occur for speech. In accordance with our predictions, we found almost identical areas in the left MTG when speech was compared with music and when it was compared with environmental sounds. Interestingly, while the reversed comparison between music and speech was non-significant, the comparison between environmental sounds and speech returned an area, even more ventral than the above effect, in the right fusiform gyrus. We suspect that this arose from a greater likelihood of environmental sounds to activate visual conceptual information. Environmental sounds always reference concrete concepts, whereas speech references both concrete and, hard to visualize, abstract concepts (e.g., peace). Further indication for the ventral pathway's role in conceptual processing comes from the comparison of speech at different levels of combinatory complexity. Specifically, conceptually rich discourse activated right ITG more than sentences, words and syllables. Moreover, this area was in relative proximity to the area identified for environmental sounds suggesting that both serve similar roles.

Conclusions

Meta-analyses overcome many problems of individual-study neuroimaging research such as small sample sizes and the increased likelihood of false positive results (Yarkoni et al., 2010). However, they also bear inherent dangers associated with data reduction (e.g., consideration of activation peaks rather than all significant voxels) and potential confounds (i.e., systematic differences in studies examining two, to be contrasted, conditions). Therefore, meta-analyses cannot replace carefully designed individual studies but simply provide an additional, birds-eye viewpoint.

From this viewpoint, we failed to observe some of the sound class specific activation patterns reported in the literature and could not confirm many of the functions hitherto linked to anterior, posterior and ventral processing streams. However, we found evidence for these streams as well as some associated neurofunctional proposals. Based on this evidence it seems that all sounds activate an anterior stream that likely supports more general perceptual and/or discourse-level semantic processes. A posterior stream appears to contribute to sound embodiment, whereas a ventral stream appears relevant for the mapping of sounds to meaning and associated visual representations. One characteristic, common across these three streams, is that they exist in both hemispheres and are fairly bilaterally engaged especially for highly relevant sounds such as vocalizations and music. A second characteristic is that they are activated relatively independently of sound class and sound attributes such as acoustic complexity, spectral change or combinatory complexity. Instead, all streams rely on stored memory representations of perceptual, sentient or conceptual nature against which incoming sounds are matched for identification. As these representations emerge through interactions with one's auditory environment, the gross spatial organization of auditory processing within the temporal lobe is unlikely to depend on preexisting sound classes. Rather, it seems to depend on an all-purpose auditory brain system and a lifetime of sonic experiences.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2012.06.025.

Acknowledgments

We would like to thank Nicolas Escoffier for commenting on an earlier version of this manuscript. This work was supported by the NUS Young Investigator Award (WBS R581-000-066-101) and the A*Star SERC grant (0921570130) conferred to AS. This study was also supported by the Swiss National Science Foundation (SNSF 105314_124572/1 by DG) and by the National Center for Competence in Research in Affective Sciences at the University of Geneva.

References

Alain, C., Arnott, S.R., Hevenor, S., Graham, S., Grady, C.L., 2001. "What" and "where" in the human auditory system. Proc. Natl. Acad. Sci. U. S. A. 98 (21), 12301–12306.
Arnott, S.R., Binns, M.A., Grady, C.L., Alain, C., 2004. Assessing the auditory dual-pathway model in humans. Neuroimage 22 (1), 401–408.

Arnott, S.R., Grady, C.L., Hevenor, S.J., Graham, S., Alain, C., 2005. The functional organization of auditory working memory as revealed by fMRI. J. Cogn. Neurosci. 17 (5), 819–831.

Belin, P., 2006. Voice processing in human and non-human primates. Philos. Trans. R. Soc. Lond. B Biol. Sci. 361 (1476), 2091–2107.

Belin, P., Zatorre, R.J., 2000. "What", "where" and "how" in auditory cortex. Nat. Neurosci. 3 (10), 965–966.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403 (6767), 309–312.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., Possing, E.T., 2000. Human temporal lobe activation by speech and nonspeech sounds. Cereb. Cortex 10 (5), 512–528.

Broca, P., 1863. Localization des fonctions cerebrales. Siege du language articule. Bull. Soc. Anthropol. 4, 208.

Chandrasekaran, B., Chan, A.H.D., Wong, P.C.M., 2011. Neural processing of what and who information in speech. J. Cogn. Neurosci. 23 (10) (Oct), 2690–2700.

Confavreux, C., Croisile, B., Garassus, P., Aimard, G., Trillet, M., 1992. Progressive amusia and aprosody. Arch. Neurol. 49 (9), 971–976.

De Santis, L., Clarke, S., Murray, M.M., 2007. Automatic and intrinsic auditory "what" and "where" processing in humans revealed by electrical neuroimaging. Cereb. Cortex 17 (1), 9–17.

DeWitt, I., Rauschecker, J.P., 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. U. S. A. 109 (8), E505–E514.

Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum. Brain Mapp. 30 (9), 2907–2926.

Escoffier, N., Sheng, D.Y.J., Schirmer, A., 2010. Unattended musical beats enhance visual processing. Acta Psychol. (Amst) 135 (1), 12–16.

Escoffier, N., Zhong, J., Schirmer, A., Qiu, A., 2012. Emotional expressions in voice and music: same code, same effect? Hum. Brain Mapp. (n/a–n/a).

Fecteau, S., Armony, J.L., Joanette, Y., Belin, P., 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage 23 (3), 840–848.

Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., Gibson, E., 2009. Structural integration in language and music: evidence for a shared system. Mem. Cognit. 37 (1), 1–9.

Friederici, A.D., 2012. The cortical language circuit: from auditory perception to sentence comprehension. Trends Cogn. Sci. (Regul. Ed.) 16 (5), 262–268.

Frühholz, S., Ceravolo, L., Grandjean, D., 2012. Specific brain networks during explicit and implicit decoding of emotional prosody. Cereb. Cortex 22 (5) (May), 1107–1117.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage 15 (4), 870–878.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M.L., Scherer, K.R., Vuilleumier, P., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 8 (2), 145–146.

Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4 (4), 131–138.

Kim, H., 2011. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. Neuroimage 54 (3), 2446–2461.

Koelsch, S., 2005. Neural substrates of processing syntax and semantics in music. Curr. Opin. Neurobiol. 15 (2), 207–212.

Laird, A.R., Lancaster, J.L., Fox, P.T., 2005. BrainMap: the social evolution of a human brain mapping database. Neuroinformatics 3 (1), 65–78.

Lamm, C., Decety, J., Singer, T., 2011. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 54 (3), 2492–2502.

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. 9 (12), 920–933.

Pell, M.D., 2006. Cerebral mechanisms for understanding emotional prosody in speech. Brain Lang. 96 (2), 221–234.

Peretz, I., Belleville, S., Fontaine, S., 1997. Dissociations between music and language functions after cerebral resection: a new case of amusia without aphasia. Can. J. Exp. Psychol. 51 (4), 354–368.

Perrodin, C., Kayser, C., Logothetis, N.K., Petkov, C.I., 2011. Voice cells in the primate temporal lobe. Curr. Biol. 21 (16), 1408–1415.

Petkov, C.I., Logothetis, N.K., Obleser, J., 2009. Where are the human speech and voice regions, and do other animals have anything like them? Neuroscientist 15 (5), 419–429.

Price, C.J., 2010. The anatomy of language: a review of 100 fMRI studies published in 2009. Ann. N. Y. Acad. Sci. 1191 (1), 62–88.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12 (6), 718–724.

Rauschecker, J.P., Tian, B., 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 97 (22), 11800–11806.

Remedios, R., Logothetis, N.K., Kayser, C., 2009. Monkey drumming reveals common networks for perceiving vocal and nonvocal communication sounds. Proc. Natl. Acad. Sci. U. S. A. 106 (42), 18010–18015.

Romanski, L.M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P.S., Rauschecker, J.P., 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat. Neurosci. 2 (12), 1131–1136.

Sabatinelli, D., Fortune, E.E., Li, Q., Siddiqui, A., Krafft, C., Oliver, W.T., Beck, S., Jeffries, J., 2011. Emotional perception: meta-analyses of face and natural scene processing. Neuroimage 54 (3), 2524–2533.

Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., Weiller, C., 2008. Ventral and dorsal pathways for language. PNAS 105 (46), 18035–18040.

Schirmer, A., Kotz, S.A., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn. Sci. (Regul. Ed.) 10 (1), 24–30.

Schirmer, A., Alter, K., Kotz, S.A., Friederici, A.D., 2001. Lateralization of prosody during language production: a lesion study. Brain Lang. 76 (1), 1–17.

Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., Friederici, A.D., 2008. When vocal processing gets emotional: on the role of social orientation in relevance detection by the human amygdala. Neuroimage 40 (3), 1402–1410.

Schirmer, A., Teh, K.S., Wang, S., Vijayakumar, R., Ching, A., Nithianantham, D., Escoffier, N., Cheok, A.D., 2011. Squeeze me, but don't tease me: human and mechanical touch enhance visual attention and emotion discrimination. Soc. Neurosci. 6 (3), 219–230.

Schönwiesner, M., Rübsamen, R., von Cramon, D.Y., 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur. J. Neurosci. 22 (6), 1521–1528.

Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123 (Pt 12), 2400–2406.

Slevc, L.R., Rosenberg, J.C., Patel, A.D., 2009. Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. Psychon. Bull. Rev. 16 (2), 374–381.

Tian, B., Reser, D., Durham, A., Kustov, A., Rauschecker, J.P., 2001. Functional specialization in rhesus monkey auditory cortex. Science 292 (5515), 290–293.

Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. NeuroImage 16 (3), 765–780.

Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., Fox, P., 2012. Minimizing within-experiment and within-group effects in Activation Likelihood Estimation meta-analyses. Hum. Brain Mapp. 33 (1), 1–13.

Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J., 2001. Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. Neuron 30 (3), 829–841.

Weiller, C., Bormann, T., Saur, D., Musso, M., Rijntjes, M., 2011. How the ventral pathway got lost — and what its recovery might mean. Brain Lang. 118 (1–2), 29–39.

Wernicke, C., 1874. Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis. Max Cohn & Weigert, Breslau.

Winner, E., Brownell, H., Happé, F., Blum, A., Pincus, D., 1998. Distinguishing lies from jokes: theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. Brain Lang. 62 (1), 89–106.

Yarkoni, T., Poldrack, R.A., Van Essen, D.C., Wager, T.D., 2010. Cognitive neuroscience 2.0: building a cumulative science of human brain function. Trends Cogn. Sci. 14 (11), 489–496.

Yasui, T., Kaga, K., Sakai, K.L., 2009. Language and music: differential hemispheric dominance in detecting unexpected errors in the lyrics and melody of memorized songs. Hum. Brain Mapp. 30 (2), 588–601.

Zatorre, R.J., Belin, P., 2001. Spectral and temporal processing in human auditory cortex. Cereb. Cortex 11 (10), 946–953.

Zatorre, R.J., Bouffard, M., Ahad, P., Belin, P., 2002. Where is "where" in the human auditory cortex? Nat. Neurosci. 5 (9), 905–909.