Brain & Language 116 (2011) 83–90

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Distributed processing and cortical specialization for speech and environmental sounds in human temporal cortex

Robert Leech [a,*], Ayse Pinar Saygin [b]

[a] Department of Experimental Medicine, Imperial College London, Hammersmith Hospital, Du Cane Road, London, W12 0NN, UK
[b] Department of Cognitive Science, Neurosciences Program and Center for Research in Language, University of California, San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:
Accepted 15 November 2010
Available online 16 December 2010

Keywords:
fMRI
Multivariate
Temporal cortex
Auditory

ABSTRACT

Using functional MRI, we investigated whether auditory processing of both speech and meaningful non-linguistic environmental sounds in superior and middle temporal cortex relies on a complex and spatially distributed neural system. We found that evidence for spatially distributed processing of speech and environmental sounds in a substantial extent of temporal cortices. Most importantly, regions previously reported as selective for speech over environmental sounds also contained distributed information. The results indicate that temporal cortices supporting complex auditory processing, including regions previously described as speech-selective, are in fact highly heterogeneous.

1. Introduction

For centuries, researchers have been exploring whether there are specific areas of the human brain that give our species the capacity for language. Over the past 20 years, functional neuroimaging has made it possible to study the localization of language functions non-invasively (e.g., Binder, Frost, Hammeke, Cox, et al., 1997; Wise et al., 1991). Previous work has shown that regions in left superior temporal cortex exhibit activation for speech sounds in relation to other auditory stimuli (Binder et al., 2000; Humphries, Kimberley, Buchsbaum, & Hickok, 2001; Scott, Blank, Rosen, & Wise, 2000; Thierry, Giraud, & Price, 2003). Consequently, left superior temporal regions have been implicated in prelexical processing of speech (Scott et al., 2000). In addition, a network of superior temporal regions have also been delineated for processing the human voice (e.g., Belin, Zatorre, & Ahad, 2002; Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Lewis et al., 2009; see also Altmann, Doehrmann, and Kaiser (2007) for superior temporal responses to non-human vocalizations).

Nevertheless, the extent to which left superior temporal cortex is dedicated to the neural processes required for speech or the human voice rather than general acoustic processing is still uncertain. More broadly, identifying category specific responses in the human brain using neuroimaging can be difficult, and there is no consensus on what would constitute as evidence for category specific neural responses (Pernet, Schyns, & Demonet, 2007). A region that only differs in activation from baseline for speech, can be thought of as speech-specific. However, this stringent criterion is often relaxed, and substantial increases in activation for speech or multiple dissociations between speech and other categories may be taken as evidence for selectivity (Pernet et al., 2007). In contrast, regions that show increased neural response for a range of acoustic stimuli, but with only a slight bias for speech may be better thought of as speech-preferential. In neuroimaging studies of auditory processing, superior temporal regions have been traditionally reported as speech or voice-specific (Belin et al., 2000; Scott et al., 2000) with more recent studies suggesting some category-selective responses (e.g., Altmann et al., 2007; Engel et al., 2009; Leaver & Rauschecker, 2010). However, other studies suggest these regions overlap with areas that activate robustly for non-speech tasks such as melody and pitch processing (for a review see Price, Thierry, & Griffiths, 2005). These areas also respond strongly (and bilaterally) to other complex auditory stimuli such as meaningful environmental sounds (e.g., a dog barking or a car starting) (Dick et al., 2007; Lewis et al., 2004, 2009), suggesting only more general speech-preferential processing within superior temporal regions. Indeed, distinct cortical networks including superior temporal regions have been identified for different classes of environmental sounds; e.g., Lewis, Brefczynski, Phinney, Janik, and DeYoe (2005) describe separate pathways for processing animal sounds and tools. Furthermore, lesions in left middle and superior temporal cortex are associated with correlated deficits in both speech and environmental sound processing (Saygin, Dick, Wilson, Dronkers, & Bates, 2003; Saygin, Leech, & Dick, 2010; Schnider, Benson, Alexander, & Schnider-Klaus, 1994).

One possibility is that association regions that have shown increased activation for a particular category of sounds such as speech using PET and fMRI may reflect general high-level auditory

* Corresponding author.
E-mail address: r.leech@imperial.ac.uk (R. Leech).

0093-934X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2010.11.001

processing common to many classes of sounds, but with a slight bias for speech on aggregate (because of differences from other stimuli in terms of acoustical, semantic, or attentional features). If this is the case, these cortical regions may not be spatially homogeneous (i.e., a consistent preference for speech over other classes of sounds across voxels within a region). Instead, the underlying high-level auditory processing would reveal itself as a heterogeneous spatially varying pattern of speech-preferential and other-sound-preferential voxels. This would be reflected in small local biases in the sensitivity to e.g., different auditory features, which when averaged across all voxels, demonstrate a slight bias for speech. Indeed, recent approaches to functional neuroimaging data analysis have suggested that language and other complex sounds processing may rely on complex underlying computations (Formisano, De Martino, Bonte, & Goebel, 2008; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009; but see Op de Beeck, 2010). However, these studies have only indirectly investigated the heterogeneity of the spatial signal measured in fMRI studies of complex auditory processing and have not directly addressed how this relates to the speech-preferential patterns that have been found in previous studies.

In this paper, using functional magnetic resonance imaging (fMRI), we address the extent to which auditory processing of both speech and complex, meaningful, non-linguistic environmental sounds within superior temporal regions relies on heterogeneous processing. Instead of standard univariate measures of effect size, we used multivariate statistics that are sensitive to spatially distributed patterns of activation: We asked not whether there is greater activation for one condition (i.e., speech) over another condition (i.e., environmental sounds), but rather whether there is sufficient information to distinguish between the two conditions across multiple voxels irrespective of which condition is most active in any given voxel (Kriegeskorte & Bandettini, 2007; Kriegeskorte, Goebal, & R. Bandettini, 2006). Henceforth, we will refer to the commonly used univariate methods as the ''activation approach'' and our multivariate methods as the ''information approach''.

The information approach has recently shown distributed patterns in inferior temporal cortex for processing complex visual objects and faces (Kriegeskorte, Formisano, Sorger, & Goebel, 2007). Here we asked whether complex auditory processing similarly relies on distributed processing within superior temporal regions. We compared the information approach to an activation-based analysis to ask: (i) whether there are regions containing distributed information that are not detected using standard approaches; (ii) whether regions identified as language-preferential based on univariate analyses may nevertheless contain additional distributed information, evidence of more heterogeneous processing.

Environmental sounds are meaningful and acoustically complex sounds; as such they constitute a good class of stimuli to compare with speech to investigate complex auditory processing (Saygin, Dick, & Bates, 2005). One limitation with comparing speech and environmental sounds is that non-speech environmental sounds contain greater spectrotemporal variability. Although some previous studies have attempted to address this issue (e.g., Thierry et al., 2003), equating spectrotemporal complexity across speech and other classes of sounds is always imperfect. Therefore, in this study we limit ourselves to investigating the coarse differences between speech and a broad range of acoustically complex environmental sounds. The purpose of this study is not to ask whether speech differs from environmental sounds along specific acoustic dimensions, or controlling for specific types of spectrotemporal complexity. Instead, by comparing speech and environmental sound using multivoxel pattern analysis techniques we gain insight into the general style of processing involved in complex acoustic perception of meaningful sounds. Specifically, we investigate the extent to which processing is distributed or focal and whether regions implicated in this processing are more heterogeneous than activation-based analyses suggest. If heterogeneous activation for complex auditory processing is widespread across auditory association regions, then this challenges how we think about the underlying neural processing; even labeling given regions of superior temporal lobe as preferentially active for a given auditory class may be an oversimplification.

Note that the question asked in this study is not the typical one in most fMRI studies, i.e., is there more activation for one sound class than another in a given voxel. Indeed, the information approach does not distinguish between activation due to environmental sounds or speech. Rather we asked about both the level of activation, and the spatial similarity of neighboring voxels, i.e., do adjacent voxels have similar differences in activation for different sound classes. Differences in detection between the activation and information approaches would be due to differences in the spatial homogeneity of the neighboring voxels, not how active a single voxel or cluster of voxels was. To summarize our reasoning, if the underlying spatial signal is approximately the same size or greater than the Gaussian kernel (in this case 6 mm), then the standard activation approach should have greater sensitivity for detection. However, if the actual underlying signal is more distributed within a region (i.e., if the ''salt and pepper'' pattern of contrasts in activation visible in unsmoothed data is truly representative of spatially varying signal rather than merely noise) then the smoothed absolute-t value (the information-based analysis) should be more sensitive (see also Kriegeskorte et al., 2006, 2007). Both approaches enjoy the benefit of increased signal detection afforded by spatial smoothing, however, the smoothed-absolute-t analysis is tolerant of heterogeneous patterns of activation. Furthermore, the fact that the different classes of stimuli were not matched along acoustic dimensions is likely to make hypothesis testing about the heterogeneity of speech more, not less conservative, since large acoustic differences are likely to make activation differences larger and more robust in each individual voxel.

2. Methods

2.1. Participants

Seven neurologically healthy right-handed participants aged 25–40 took part in the study. All subjects reported normal hearing and gave written informed consent in accordance with local ethics.

2.2. Imaging procedure

Participants were scanned using a 3T GE Excite scanner and a phased-array head coil at the Center for Functional Magnetic Resonance Imaging at University of California San Diego. We used a standard single-shot echo planar T2$*$-weighted gradient echo pulse sequence (TR = 2400 ms, TE = 27.4 ms, flip angle = 90°, linear auto-shim) and acquired 31 interleaved slices covering the whole brain (3.75 × 3.75 × 3.8 mm voxels, 0 mm gap). We also acquired a structural image from each participant (MPRAGE, TR = 10.5 ms, TE = 4.8, 1 × 1 × 1.5 mm voxels).

2.3. Experiment design

The experiment featured a mixed block design with three block types: everyday environmental sounds, speech sounds (moderate frequency bi or tri-syllabic nouns, or verb phrases , all recorded in a sound-proof booth), or silence (the baseline condition). A subset of the environmental sounds have been used previously in an fMRI study (Dick & et al., 2007). Both the environmental sound and speech stimuli were taken from a behavioral norming study (Saygin et al., 2005) with sounds in both classes being approximately equally easy

to recognize behaviorally. A list of the environmental sounds (along with a range of acoustic measures) and speech sounds used in this study can be found in Supplementary materials. Blocks were 24 s in duration and there were eight blocks for each condition.

Sounds were presented dichotically. Participants listened to the stimuli with eyes closed and were instructed to listen carefully and try to comprehend each sound. Sounds were presented rapidly following each other with 100 ms ISI and an additional 250 ms between blocks. Prior to the actual scan, sound volume was adjusted individually for each participant such that the stimuli were loud enough to hear clearly over the scanner noise and with earplugs, but not too loud to cause discomfort. Each participant was scanned in two runs, each lasting approximately 5 min.

2.4. Analyses

2.4.1. Whole-brain group activation analyses

To situate the present study with previous work investigating the neural correlates of speech and environmental sound processing, we ran a standard whole-brain group activation analysis using FSL software (www.fmrib.ox.ac.uk/fsl). For this analysis, functional images were realigned to correct for small head movements (Jenkinson & Smith, 2001) and then smoothed with a 6-mm full-width half-maximum Gaussian filter to increase SNR. The time series data were pre-whitened to remove temporal auto-correlation (Woolrich, Ripley, Brady, & Smith, 2001). Images were then entered into a general linear model by separately convolving speech and environmental sound blocks with a double-gamma canonical hemodynamic response function (Glover, 1999). Rest trials formed the implicit baseline condition. In addition, temporal derivatives and estimated motion parameters were included as covariates of no interest to increase statistical sensitivity. First level results were transformed into standard space using a 12 degree-of-freedom affine registration to the MNI152 template. At the second level, a paired *t*-test was used to calculate group effects using a mixed-effect model (Beckmann, Jenkinson & Smith, 2003). Activations were thresholded at $Z > 2.3$, and were considered significant at $p < 0.05$ using a cluster-wise significance test (Friston et al., 1994). Four contrasts were calculated, (i) speech > rest; (ii) environmental sounds > rest; (iii) speech > environmental sound; (iv) environmental sound > speech.

2.4.2. Individual information and activation analyses

Image preprocessing and statistical analysis were performed using Analysis of Functional Neuroimages (AFNI) software (Cox, 1996) and MATLAB, including the afni_matlab toolbox. For all analyses, the two runs were concatenated (yielding 288 volumes), spatially registered for motion correction using a six-dimensional affine transformation. The analyses were performed in each subjects' native space unless noted otherwise.

Both the information- and activation-based of analyses were restricted to a large region of interest in temporal cortex to reduce the computational load required by the randomization resampling methods used for inference. We defined a broad anatomical mask encompassing all of superior and middle temporal cortex (regions that have been previously associated with speech and environmental sound auditory processing, e.g., Dick et al., 2007; Price et al., 2005; Thierry et al., 2003), defined using the Harvard-Oxford probabilistic atlas in fslview. This mask was then warped into each subject's native space using flirt (Jenkinson, Bannister, Brady, & Smith, 2002).

Two distinct analysis procedures were used:

(i) *Activation analysis*

There were two sets of activation analyses, identical apart from using either smoothed data or unsmoothed data. For the spatially smoothed analysis, we used a Gaussian kernel of 6 mm full-width at half-maximum (FWHM). The AFNI program 3dDeconvolve was used to fit a general linear model at each voxel. The model estimated parameters for each of the two non-baseline conditions (speech, environmental sounds), as well as three parameters for each run to account for mean, linear and quadratic trends. A *t*-statistic contrasting speech with environmental sounds was calculated for each voxel.

(i) Information analysis – smoothed absolute-*t* statistic

The same general linear model was calculated as in (i), except using the unsmoothed data. We then calculated the absolute values of the resulting speech versus sound *t*-statistics at each voxel. By taking the absolute value of the *t*-statistic we lose all information about the direction of activation (i.e., whether speech or environmental sound is preferentially activate in a given voxel). This *t*-statistic map was spatially smoothed with a 6 mm FWHM Gaussian kernel. If we had used a voxel sphere for smoothing, the smoothed absolute-t statistic would be equivalent to the Euclidean distance (Kriegeskorte & Bandettini, 2007). Note that because of the Gaussian spatial smoothing, spatial resolution is sacrificed for statistical power, and any increase in detection power in the information analysis may be the result of heterogeneous voxels further than the 6 mm width of the Gaussian kernel.

To investigate the coarse spatial distribution of the pattern of preferential activation, a spherical searchlight was centered on each voxel. A voxel was given a value of 1 if any voxels in the surrounding neighborhood had above threshold voxels in both classes (i.e., evidence of some degree of heterogeneity) and a 0 otherwise. This measure of variability in preferential activation was investigated at two levels of spatial resolution: either a 2- (4 mm) or 3-voxel (6 mm) sphere around each voxel. Thus, this was a simple quantification of local variability in heterogeneous preferential processing.

2.5. Significance testing

We tested for significance using randomization. A null-distribution of test statistics was calculated for each voxel by shuffling the labels (speech, environmental sounds) of each non-baseline block 1000 times. Each of these 1000 randomized time courses was then analyzed as in (i) and (ii) above. The distribution of randomized test statistics was then sorted and the rank order (divided by 1000) of the real test statistic provided a *p*-value.

The randomization was applied to both the activation and information-based analyses to ensure that the two could be appropriately compared quantitatively (even though the standard *t*-statistic could also be assessed using *t*-distributions, this would be a biased comparison due to the greater sensitivity of parametric methods).

To correct for multiple comparisons, a false discovery rate (Genovese, Lazar & Nichols, 2002) (FDR, $q < 0.05$) threshold was applied to the resulting *p*-value maps. To facilitate unbiased comparison between the activation and the information-based approaches, the same *p*-value was used to threshold both approaches using the lower, and more conservative FDR value of the two. Those voxels surviving the FDR threshold were then used in subsequent analyses (one subject had no voxels surviving the FDR threshold and so a threshold of $p < 0.01$ uncorrected was used instead, the general pattern of results were unchanged with or without the inclusion of this subject). We counted the supra-threshold voxels (given the use of randomization techniques on different underlying statistics). Although a single threshold was used, the results were qualitatively similar at different thresholds (e.g., $p < 0.05$ uncorrected).

2.6. Region of interest analysis

We chose two sets of regions of interest (ROIs) based on previous work: As speech-preferential areas, five theoretically-motivated regions of interest were defined using the peak coordinates previously reported to be more active for speech than environmental sounds in left superior temporal cortex (Price, Thierry & Griffiths, 2005). The ROIs were constructed by creating a 10 mm radius sphere in standard space centered on each of the peak voxels and were transformed back into each subject's native space. To investigate voice-preferential processing, 10 theoretically-motivated ROIs were defined by placing 5 mm radius spheres centered on peak coordinates taken from Belin et al. (2000). We used a smaller radius in order to prevent overlap because the ROIs from Belin et al. (2000) were situated close to each other along the upper bank of the STG. The number of active FDR corrected voxels for the activation-based and information-based analyses for each participant were counted within each of these ROIs.

3. Results

3.1. Group-based whole-brain activation analysis

Fig. 1 presents the group whole-brain activation analyses. Consistent with previous studies we found that when compared to rest, both speech and environmental sounds led to bilateral activation (see also Table 1) in superior and middle temporal cortical regions, encompassing primary and association auditory cortices. Bilateral superior and middle temporal regions showed stronger responses for speech sounds compared with environmental sounds, and medial temporal and thalamic regions showed the opposite preference.

3.2. Individual information and activation-based analyses

Subsequently, we applied both the activation-based and the information-based analyses described above to each participant's data. First, we investigated whether there were regions of temporal cortex that contained a distributed response patterns differentiating environmental sounds and speech sounds that were not detected with the activation approach. Second, we explored whether regions previously identified as preferential for speech sounds showed heterogeneous activation profiles.

3.3. Activation versus information analyses

As shown in Fig. 2, across superior and middle temporal regions, we found considerably more voxels containing information distinguishing speech and environmental sounds using the information-based approach (red voxels), than could be detected using the activation based approach (yellow and green voxels). This was true for every individual participant. Fig. 2 shows only the left hemisphere for each of the seven participants – qualitatively similar results were found for each subject's right hemisphere.

Looking at the proportion of post-threshold voxels averaged across subjects (Fig. 2b) for the two types of analyses, we observed that the majority of voxels (55%) were shared between both analyses, with an additional 40% unique to the information-based approach, set against 5% unique for the activation-based analysis. The pattern was similar in every subject and was highly significant, assuming the null hypothesis that the activation and information-based approaches are equally likely to yield unique voxels.

3.4. Speech and voice sensitive superior temporal cortex

Overall, the information approach revealed a greater extent of auditory and auditory association cortex distinguishing speech and environmental sounds than estimated using activation-based analyses. We next narrowed our focus on portions of temporal cortex most closely identified with speech processing and voice processing in previous studies, to ask whether there was distributed information present in these regions. Fig. 3 illustrates five ROIs defined based on previous studies (Price et al., 2005) and the average number of super-threshold voxels for the information and activation analyses within these regions. Aggregated across all regions, and within regions B, C, D and E there were significantly more super-threshold voxels in the information analysis than the activation analysis (all regions, Wilcoxon sign-rank = 14, $p < 0.01$; region B, Wilcoxon sign-rank = 10.5, $p < 0.05$; region C, Wilcoxon sign-rank = 7.5, $p < 0.05$; region D, Wilcoxon sign-rank = 7.5, $p < 0.05$; region E, Wilcoxon sign-rank = 10.5, $p < 0.05$). In anterior temporal

Table 1
Summary of whole-brain group based analysis. The locations given are the MNI coordinates of the maximum *z*-statistic for each contrast.

Location of max z	Cluster volume (mm^3)	*p*-value	*X*	*Y*	*Z*
Speech > rest					
Left STG	4712	$p < 0.0001$	−62	−6	−2
Right STG	3179	$p < 0.0001$	58	−10	−4
Env sound > rest					
Left STG	6165	$p < 0.0001$	−66	−36	14
Right STG	2730	$p < 0.0001$	54	−16	6
Occipital Fusiform	1886	$p < 0.0001$	−32	−74	−26
Speech > env sound					
Left STG	2399	$p < 0.0001$	−62	−16	−2
Right MTG/STS	982	$p < 0.01$	58	−14	−10
Env sound > speech					
PT	3131	$p < 0.0001$	−32	−36	10

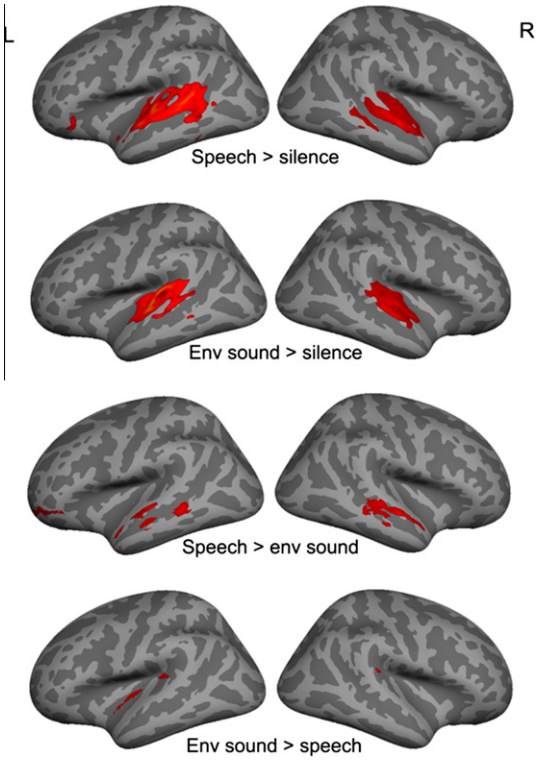

Fig. 1. Whole-brain group analyses, projected onto an average surface using Freesurfer (Dale, Fischl, & Sereno, 1999). Red voxels mark clusters of activation that are significant at $p < 0.01$ (cluster-wise corrected).

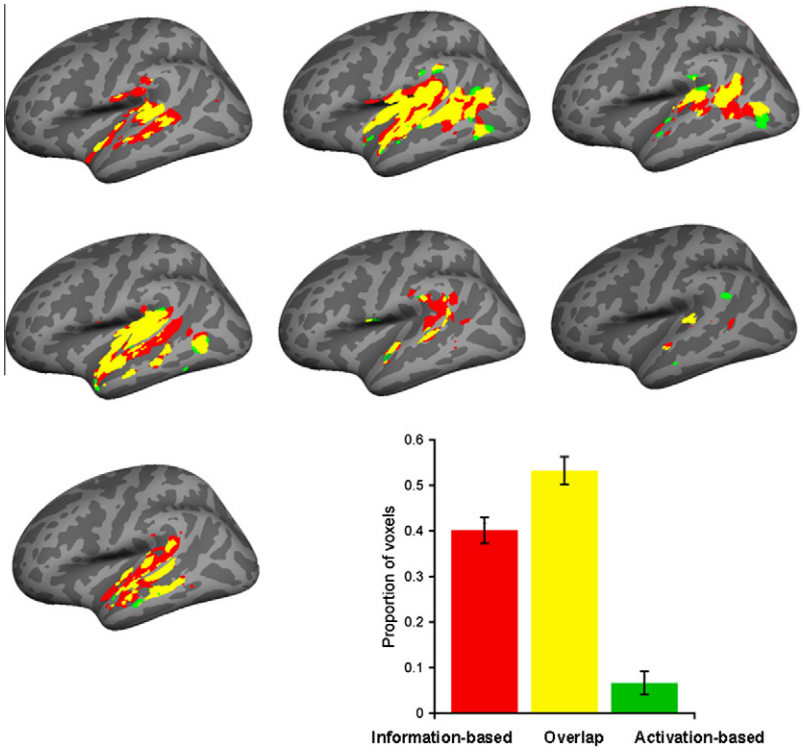

Fig. 2. The left hemispheres of the seven subjects projected onto an average surface using Freesurfer (Dale et al., 1999). Red are FDR corrected super-threshold voxels for the information-based analysis, green are for the activation-based analysis, and yellow are the voxels common to both analyses. Mean proportion of super-threshold voxels for both information and activation-based analyses across the seven subjects. Red is the proportion of active voxels for the analysis, yellow is the proportion of voxels shared by both analyses, green is the proportion of analyses in the activation-based analysis.

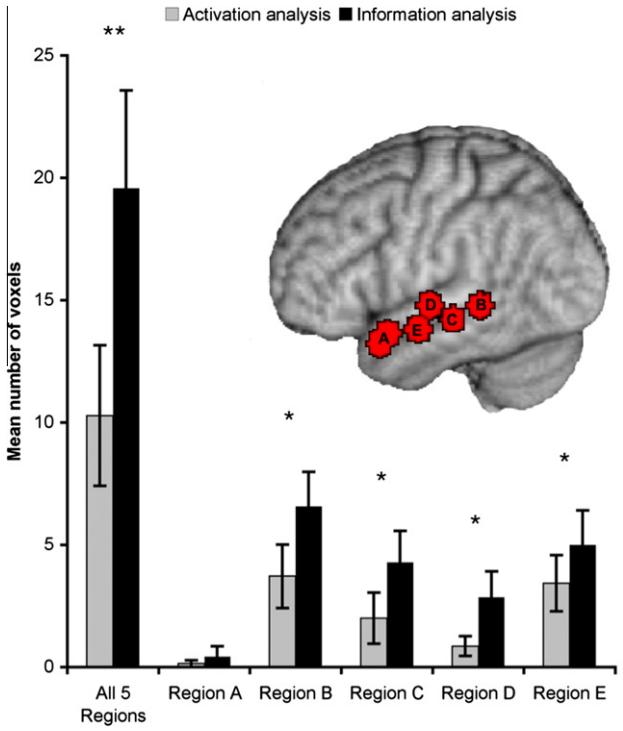

Fig. 3. The number of super-threshold voxels (in native space) in the activation-based analysis (grey) and for the information-based analysis (black) for five regions of interest previously identified (Price et al., 2005) for preferential activation for speech over environmental sound, * is significant at $p < 0.05$; and ** is significant at $p < 0.001$.

region A, the data were less reliable since there were very few post-threshold voxels, with only one subject for region A. (Failing to find active voxels in this area is consistent with previous comparisons of fMRI with PET activation on speech processing tasks (Devlin et al., 2000).) Fig. 4 illustrates the same approach centered on 10 ROIs that have been found to be preferential to the human voice over a range of non-vocal stimuli (Belin et al., 2000). In eight of the regions, we found marginally or significantly more post-threshold voxels using the information approach compared with the activation approach (right anterior STS 2, Wilcoxon sign-rank = 13, $p < 0.05$; right middle STS, Wilcoxon sign-rank = 9, $p < 0.1$; right middle temporal gyrus, Wilcoxon sign-rank = 10, $p < 0.1$; right posterior STS 1, Wilcoxon sign-rank = 10.5, $p < 0.05$; right posterior STS 2, Wilcoxon sign-rank = 7.5, $p < 0.05$; left middle temporal gyrus, Wilcoxon sign-rank = 11, $p < 0.05$; left middle STS, Wilcoxon sign-rank = 10.5, $p < 0.05$; left poterior STS, Wilcoxon sign-rank = 10.5, $p < 0.05$). The most anterior right superior temporal ROI, and the left planum temporale were the only regions not demonstrating a statistical increase in detection for the information analysis.

3.5. Spatial distribution of preferential processing

We present the patterns of preferential activation using unsmoothed data (in contrast to the 6 mm Gaussian kernel used in the other analyse) in Fig. 5a. These data highlight two things: first, that there is preferential processing of environmental sounds at the individual level in superior temporal regions that is averaged out at the group level. This provides evidence, at the level of the individual, of different superior temporal regions that are preferential to both environmental sound and to speech processing. Second, the spatial heterogeneity detected by the information analyses

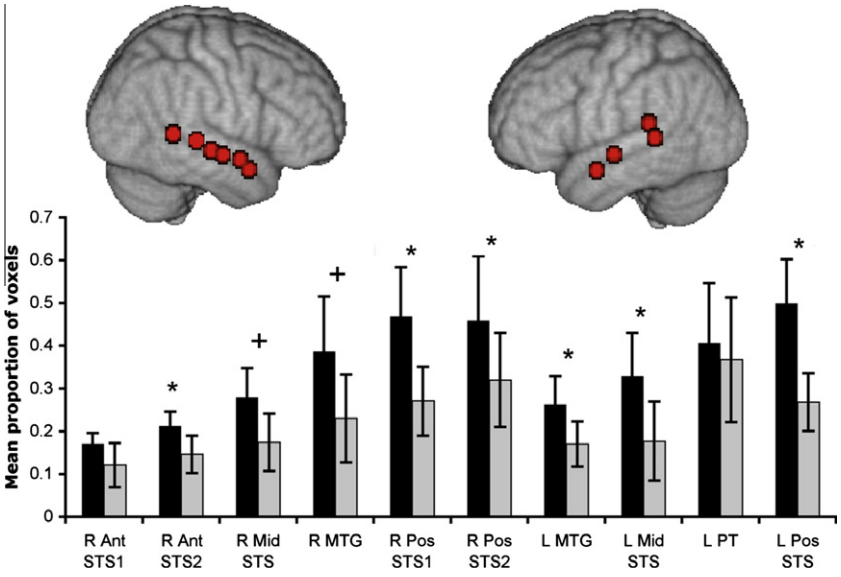

Fig. 4. The number of super-threshold voxels (in native space) in the activation-based analysis (grey) and for the information-based analysis (black) for 10 regions of interest (5 mm spheres) from (Belin et al., 2000) for preferential activation for vocal over non-vocal sounds. * is significant at $p < 0.05$; and + is significant at $p < 0.1$. STS = superior temporal sulcus; MTG = middle temporal gyrus; PT = planum temporale; Pos = posterior, Ant = anterior.

appears to be the result of small clusters of interspersed voxels preferential for either environmental sound or speech processing, and that these patterns vary substantially from individual to individual.

Fig. 5b depicts a simple quantification of the spatial heterogeneity of the activation patterns for environmental sound or speech stimuli. This analysis complements the comparison of information with activation analyses, and emphasizes that heterogeneous processing is widespread across temporal cortex with considerable individual variability. This spatial and within-subject variability is averaged out in traditional activation analyses.

4. Discussion

The most striking finding from the present study is the extent of processing heterogeneity for spectrotemporally complex meaningful sounds across human temporal cortex. Post-threshold voxels unique to the information analysis contained a mixture of voxels in their local neighborhood, some of which had preferential activation for speech and others for environmental sounds. The underlying spatial pattern was therefore less focal, reflecting a more distributed pattern of preferential responses. Evidence for heterogeneous processing was most marked in superior temporal regions (including primary and association auditory cortex) typically associated with both basic and complex sound processing. Whereas previous studies have suggested that areas of temporal cortex are specific or selective for processing the human voice and speech, our results suggest that these regions should, at most, be regarded as preferential for these classes of stimuli. Voxels in regions reported as selective for speech or voice processing are intermixed with voxels that are preferentially activate for other complex auditory processing (e.g., Belin et al., 2000; Dick et al., 2007; Price et al., 2005; Saygin et al., 2003; 2004). This heterogeneity of processing is not observed with univariate fMRI analyses due to substantial spatial smoothing as a preprocessing step. Typically, auditory and auditory association cortices are thought to process multiple types of stimuli, but with regional preferences (manifested as increased activation) for one type of sound over others (e.g., meaningful speech, the human voice, or different classes of environmental sounds). The present study suggests that these aggregated regional preferences mask more subtle variability in patterns of activation, suggesting highly heterogeneous acoustic processing.

It is important to note that this result does not reflect greater sensitivity of the information approach in general; the information approach is actually less sensitive than the activation approach in the case where the underlying signal is focal (see Kriegeskorte et al., 2006). Even restricting the analyses to specific regions of interest that have previously been associated with speech or voice processing, we observed distributed information reflecting both speech and environmental sound processing. More generally, human superior and middle temporal regions appeared to be heterogeneous in how they process complex auditory stimuli.

The work presented here contrasted neural response to speech and a range of environmental sounds that varied in their source (e.g., machine, human, animal, event, music, vehicle) as well as in their acoustical properties (e.g., sound length, pitch, harmonicity, intensity). Previous studies had revealed substantial regional differences in brain responses for processing different kinds of environmental sounds (e.g., animal, human and machine sounds, Engel et al., 2009; Lewis et al., 2009). This variability may also explain why we do not observe more regional double dissociations between environmental sounds and speech in our group analysis (Fig. 1). Furthermore, the variability also implies that we do not necessarily identify regions that discriminate between sound categories *per se*. However, our analyses were intended to explore the spatial heterogeneity of processing complex sounds, rather than to specify neural regions that process a specific class of sounds. Indeed, comparing speech sounds to such a variety of different environmental sounds, a priori, reduces the likelihood of finding voxels that preferentially respond to environmental sounds, and so reduces the likelihood of finding improved detection ability for the information relative to the activation analysis. As such, given that we do find widespread differences between information and activation analyses, it is likely that this same pattern of distributed preferential patterns of activation would also occur for sets of acoustic stimuli designed to investigate more specific questions.

The presence of local voxels with heterogeneous profiles of activation is consistent with several accounts of complex sound processing. One possibility is that listening to spectrotemporally complex auditory stimuli relies on many distinct sub-regions varying in size (e.g., from an individual voxel to large clusters) that are

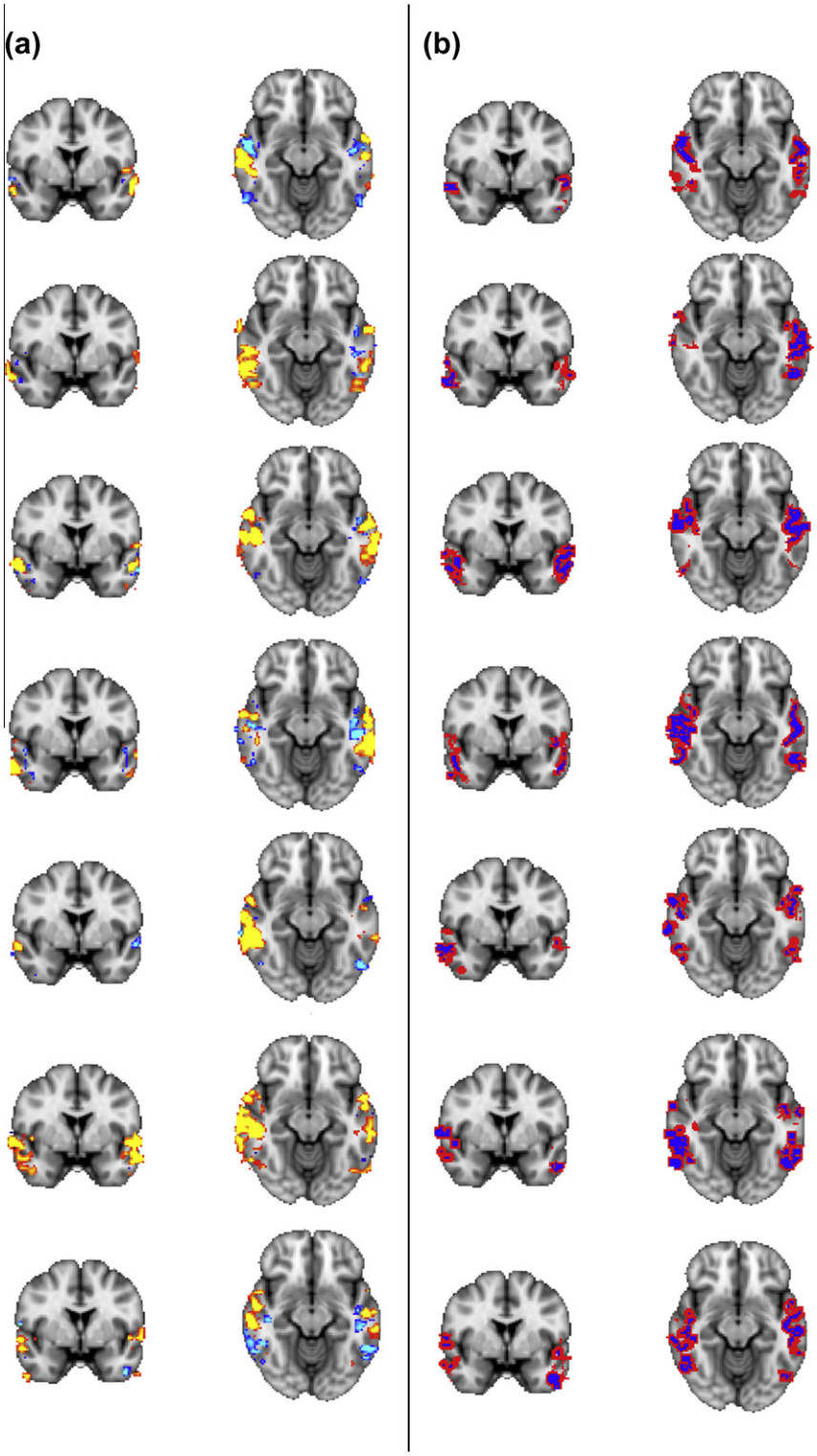

Fig. 5. (a) Individual activation maps for speech > environmental sounds (yellow) and environmental sounds > speech (blue) using *t*-statistics ($p < 0.05$ uncorrected) calculated without smoothing the data. (b) Voxels with heterogeneous speech and environmental sounds processing in their neighborhood (defined using a spherical searchlight). Red is a sphere of three voxels radius and blue is a two voxel radius spherical searchlight.

intermixed with regions better suited for processing other types of complex auditory stimuli. It is possible that these sub-regions perform distinct speech-relevant tasks (e.g., auditory categorization (e.g., Desai, Liebenthal, Waldron, & Binder, 2008) or different aspects of spectrotemporal analysis tuned to speech (e.g., Zatorre & Belin, 2001) that in concert support speech. Under this account, the observed distributed processing is evidence of far more sub-regional variation and distinct processes than reported in previous studies. The unsmoothed pattern of activation presented in Fig. 3a suggests that to some degree this may be the case, with clusters of voxels preferential to speech sounds interspersed with clusters of voxels preferential to environmental sounds.

An alternative possibility is that the pattern of preferential activation for speech or environmental sounds we observe reflects an

underlying auditory processing system that is highly distributed at much higher resolution. Under this account, the recorded greater activation for speech or environmental sounds in a given voxel is actually the tip of the iceberg in terms of underlying computational processing. A single voxel in the current experiment measures approximately 50 mm^3 and cortical tissue of this size can contain many millions of neurons. A voxel showing preferential activation for speech may reflect what is actually a mild bias aggregated across a large distributed system of underlying neurons, a minority of which would better serve non-speech environmental sound processing (cf. Kamitani & Tong, 2005; Haynes & Rees, 2005 for a similar account of orientation coding in primary visual cortex). Recent high-resolution neuroimaging work has provided stronger evidence for this kind of fine-scaled distribution of information within the visual modality (Swisher & et al., 2010), and using higher-resolution fMRI in the auditory domain may also reveal a much finer-grained level of heterogeneous processing.

Finally, the extent of focal versus distributed processing may be relevant to how the auditory comprehension breaks down following neurological insult (Saygin et al., 2003, 2010; Schnider et al., 1994). Focal lesions may have more easily interpretable mappings between site of damage and behavioral change in cortical regions where there is less distributed processing, whereas regions with more distributed patterns of activation potentially have more opaque lesion to symptom mappings, and possibly have greater resilience to damage as predicted from artificial neural network simulations (e.g., Rumelhart & McClelland, 1986). Future work is needed to compare information and activation based approaches across different brain regions on a variety of fMRI tasks, and to bring these together with neuropsychological profiles of stoke and neurodegenerative patients.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bandl.2010.11.001.

References

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex, 17*, 2601–2608.

Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *Neuroimage, 20*, 1052–1063.

Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research, 13*, 17–26.

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

Binder, J. R. et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*, 512–528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., et al. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17*, 353–362.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers in Biomedical Research, 29*, 162–173.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage, 9*, 179–194.

Desai, R., Liebenthal, E., Waldron, E., & Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *Journal of Cognitive Neuroscience, 20*, 1174–1188.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H. E., et al. (2000). Susceptibility-induced loss of signal: Comparing PET and fMRI on a semantic task. *Neuroimage, 11*, 589–600.

Dick, F. et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing. *Journal of Cognitive Neuroscience, 19*, 799–816.

Engel, L. R. et al. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage, 47*, 1778–1791.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "What"? brain-based decoding of human voice and speech. *Science, 322*, 970–973.

Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., Frackowiak, R., et al. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping, 2*(4), 189–210.

Glover, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage, 9*, 416–429.

Genovese, C., Lazar, N., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate* 1. *Neuroimage, 15*(4), 870–878.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature neuroscience, 8*, 686–691.

Humphries, C., Kimberley, T., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *NeuroReport, 12*, 1749–1752.

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis, 5*, 143–156.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion cor-rection of brain images. *Neuroimage, 17*(2), 825–841.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience, 8*(5), 679–685.

Kriegeskorte, N., & Bandettini, P. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage, 38*, 649–662.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). *Proceedings of the National Academy of Sciences of the United States of America, 103*, 3863–3868.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 20600–20605.

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: Effects of acoustic features and auditory object category. *Journal of Neuroscience, 30*, 7604.

Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., & DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *Journal of Neuroscience, 25*, 5148–5158.

Lewis, J. W., Talkington, W. J., Walker, N. A., Spirou, G. A., Jajosky, A., Frum, C., et al. (2009). Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. *Journal of Neuroscience, 29*, 2283–2296.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex, 14*, 1008–1021.

Op de Beeck, H. P. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage, 49*, 1943–1948.

Pernet, C., Schyns, P. G., & Demonet, J. F. (2007). Specific, selective or preferential: Comments on category specificity in neuroimaging. *NeuroImage, 35*, 991–997.

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences, 9*, 271–276.

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Foundations, vol. 1*. MIT Press.

Saygin, A. P., Dick, F., & Bates, E. (2005). An online task for contrasting auditory processing in the verbal and nonverbal domains and norms for college-age and elderly subjects. *Behavior Research Methods, 37*, 99–110.

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Shared neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain, 126*, 928–945.

Saygin, A. P., Leech, R., & Dick, F. (2010). Nonverbal auditory agnosia with lesion to Wernicke's area. *Neuropsychologia, 41*(1), 107–113.

Schnider, A., Benson, F., Alexander, D. N., & Schnider-Klaus, A. (1994). Nonverbal environmental sound recognition after unilateral hemispheric stroke. *Brain, 117*, 281–287.

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*, 2400–2406.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology, 19*, 498–502.

Swisher, J. D. et al. (2010). Multiscale pattern analysis of orientation-selective activity in the primary visual cortex. *Journal of Neuroscience, 30*, 325–333.

Thierry, G., Giraud, A. L., & Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron, 38*, 499–506.

Wise, R., Chollet, F., Hadar, U., Friston, K., Hoffner, E., & Frackowiak, R. (1991). Distribution of cortical neural networks involved in word comprehension and word retrieval. *Brain, 114*, 1803–1817.

Woolrich, M., Ripley, B., Brady, M., & Smith, S. (2001). Temporal autocorrelation in univariate linear modeling of FMRI data. *Neuroimage, 14*(6), 1370–1386.

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