Cerebral Cortex Advance Access published July 16, 2012

Cerebral Cortex
doi:10.1093/cercor/bhs162

Abstract Encoding of Auditory Objects in Cortical Activity Patterns

Bruno L. Giordano[1,2], Stephen McAdams[2], Robert J. Zatorre[3], Nikolaus Kriegeskorte[4] and Pascal Belin[1,5]

[1]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK [2]Department of Music Research, Centre for Interdisciplinary Research in Music, Media and Technology, McGill University, Montreal, QC, Canada [3]Montréal Neurological Institute, McGill University, Montreal, QC, Canada, [4]Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK and [5]International laboratories for Brain, Music and Sound (BRAMS), Universite de Montreal & McGill University, Montreal, QC, Canada

Address correspondence to Bruno L. Giordano, Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK. Email: brunog@psy.gla.ac.uk

The human brain is thought to process auditory objects along a hierarchical temporal "what" stream that progressively abstracts object information from the low-level structure (e.g., loudness) as processing proceeds along the middle-to-anterior direction. Empirical demonstrations of abstract object encoding, independent of low-level structure, have relied on speech stimuli, and non-speech studies of object-category encoding (e.g., human vocalizations) often lack a systematic assessment of low-level information (e.g., vocalizations are highly harmonic). It is currently unknown whether abstract encoding constitutes a general functional principle that operates for auditory objects other than speech. We combined multivariate analyses of functional imaging data with an accurate analysis of the low-level acoustical information to examine the abstract encoding of non-speech categories. We observed abstract encoding of the living and human-action sound categories in the fine-grained spatial distribution of activity in the middle-to-posterior temporal cortex (e.g., planum temporale). Abstract encoding of auditory objects appears to extend to non-speech biological sounds and to operate in regions other than the anterior temporal lobe. Neural processes for the abstract encoding of auditory objects might have facilitated the emergence of speech categories in our ancestors.

Keywords: categorization, condition-rich design, fMRI, multivariate information-based mapping, temporal cortex

Introduction

Two questions are at the heart of theories concerning the cortical processing of naturally occurring auditory objects: 1) Which low-level features drive neural processing and 2) how do computations lead to abstract semantic categories robust to large variations in the low-level features? These questions continue to stir a lively debate within the domain of auditory neuroscience. For example, there is a lack of consensus concerning which low-level features are represented in the cortex (Schönwiesner and Zatorre 2009; Recanzone and Cohen 2010), and processing models based on the concept of spectrotemporal receptive fields have had only partial success in accounting for the cortical responses to naturalistic sounds (Machens et al. 2004; Bar-Yosef and Nelken 2007). Further, the empirical evidence for abstract cortical encoding independent of low-level structure is limited to a very specific class of stimuli, harmonic sounds such as human vocalizations or musical tones (Hasson et al. 2007; Leaver and Rauschecker 2010; Kilian-Hütten et al. 2011), whereas studies of highly diverse naturalistic sounds (e.g., animal vocalizations vs. human-action sounds such as sawing wood; Lewis et al. 2005) did not test for abstract encoding. As a consequence, it is still unknown whether the abstraction of cortical representations is a general functional principle that operates for all classes of naturalistic auditory objects. We addressed these questions by analyzing the extent to which the fine-grained spatial functional magnetic resonance imaging (fMRI) patterns measured for highly heterogeneous environmental sounds selectively encoded information about the low-level and category-membership features (Fig. 1).

There is a wide consensus on the hierarchical nature of the auditory cortex, which relies on modules that progressively abstract from the low-level structure to optimize the analysis of auditory objects (e.g., "what/where model," Romanski et al. 1999). However, 1) There is currently a large disagreement on the location of such abstract sound-processing modules and 2) the majority of the empirical evidence on the existence of such modules has been collected by focusing on a restricted number of the auditory objects that our brain analyzes during our daily life, that is, human vocalizations and, more specifically, speech. Focusing on the location of abstract-processing modules, it has, for instance, been observed that the primary auditory cortex (A1) retains less information about the spectrotemporal structure but more information about abstract properties such as stimulus identity, than does inferior colliculus (Chechik et al. 2006). Consistently, a recent study of perceptually ambiguous speech argues for abstract object-like representations in the early auditory cortex (Kilian-Hütten et al. 2011). Abstract-processing aspects have also been attributed to the planum temporale (PT), an area hypothesized to be involved in matching stored spectrotemporal templates to the incoming sound information (Griffiths and Warren 2002) and to implement a process that abstracts from the fine-grained spectral shape of the incoming signal (Warren et al. 2005; Kumar et al. 2007). Other influential studies locate abstract modules in the anterior temporal cortex, part of a ventral pathway involved in the recognition of auditory objects (Romanski et al. 1999; Rauschecker and Tian 2000; Scott et al. 2000; Davis and Johnsrude 2003; Hasson et al. 2007; Rauschecker and Scott 2009; Goll et al. 2010; Leaver and Rauschecker 2010, for studies involving speech stimuli). Other empirical investigations observe instead abstract categorical processes in the posterior superior temporal sulcus (pSTS; e.g., Desai et al. 2008; Okada et al. 2010, for speech stimuli and Warren et al. 2006, for voices), or even argue for abstract representation in the entire sound-sensitive cortex, including areas traditionally

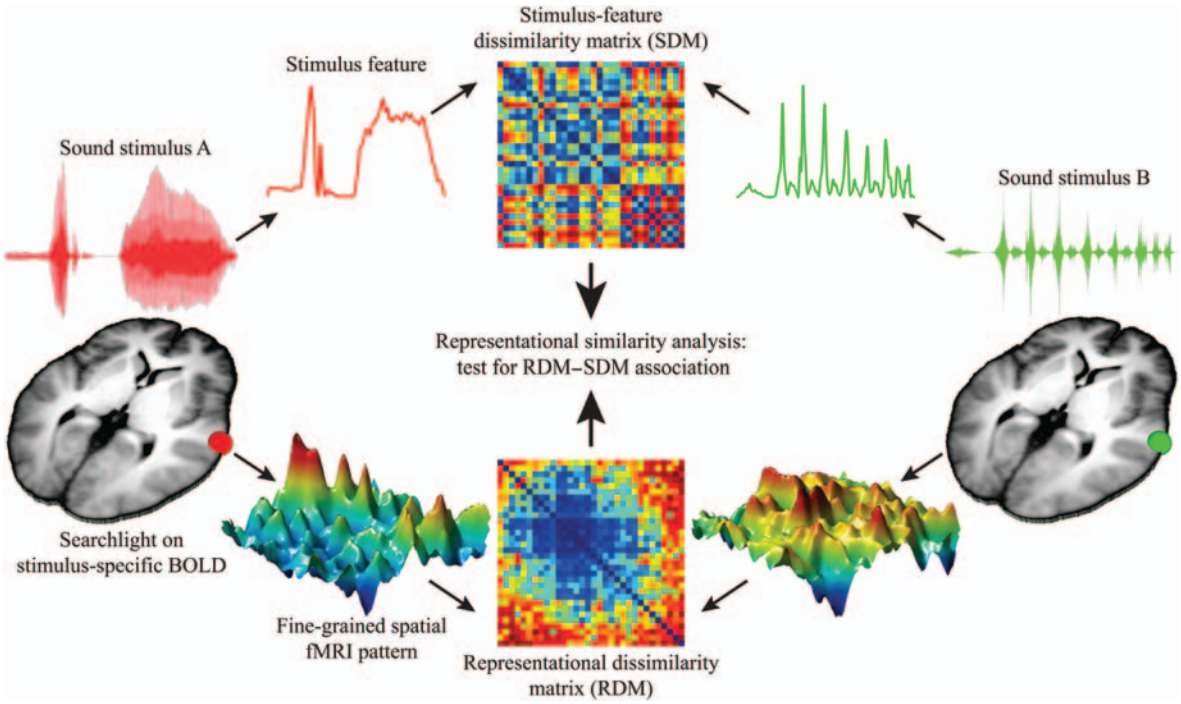

Figure 1. Pipeline for the analysis of representational similarity (RSA). For each stimulus, the fine-grained spatial distribution of the blood oxygenation level dependent (BOLD) effect is extracted within a spherical searchlight (radius = 6.25 mm). For each pair of stimuli, the dissimilarity between the fine-grained spatial fMRI patterns is defined as 1 minus the correlation between their voxel-specific estimates of the BOLD effect inside the searchlight. The complete square dissimilarity matrix computed for each pair of stimuli is termed RDM. The RDM is finally correlated with a SDM. The analysis is repeated for each possible gray matter location of the spherical searchlight. In this study, the RSA method was adopted to assess what stimulus features drive the dissimilarity of BOLD responses within the searchlight. We considered 24 potential proxy measures that capture the low-level sensory dissimilarity (12 SDMs), and the higher level category-membership dissimilarity (12 SDMs). Significant encoding of a stimulus feature in the spatial fMRI pattern was inferred if the correlation and partial correlation between the RDM and SDM had the same sign and were both significant. The partial correlation between the target SDM and the RDM was computed after removing from both the variance they shared with all of the non-target SDMs (see text).

assumed to be involved in low-level processing (Formisano et al. 2008; Staeren et al. 2009). An important aspect of previous studies further obscures our understanding of abstract sound encoding in the cortex: They have almost exclusively focused on a relatively homogeneous class of stimuli, that is, highly harmonic vocalizations (e.g., speech or animal vocalizations) and musical instrument tones (Leaver and Rauschecker 2010). This aspect of past studies not only reduces the general validity of current stances on abstract sound processing in the cortex (i.e., it is unknown whether the cortex represents abstractly many non-speech naturalistic auditory objects), but it is also associated with potential methodological problems. Indeed, given the high plasticity and context-dependence of auditory cortical computations (Ulanovsky et al. 2003; Jääskeläinen et al. 2007; Asari and Zador 2009), it is likely that the presence of a large number of speech-like stimuli within the experimental context triggers the activation of language-related abstract-processing modules. More importantly, reverse hierarchy theories of perceptual processing (Ahissar et al. 2009) predict a stronger emphasis on abstract representation for sets of sounds that are relatively similar in the low-level structure (e.g., speech and crying baby, both of which are highly harmonic) than for sets of sounds that are highly heterogeneous in the low-level structure (e.g., inharmonic crackling fire and crying baby). For these reasons, it is currently unknown whether abstract processing constitutes a general functional principle implicated in the cortical processing of all classes of naturalistic sounds. To address these issues, we used a condition-rich design with a stimulus set that is heterogeneous in both low-level and category structures. The stimulus set included distinctions between living and non-living, human and non-human and vocal and non-vocal sounds, very few human vocalizations, and no speech stimuli (Supplementary Table S1).

The cortical encoding of highly diverse categories of naturalistic non-speech auditory objects has been investigated by several fMRI and electroencephalography studies (Belin et al. 2000, 2002; Fecteau et al. 2004, 2005; Lewis et al. 2005, 2006, 2010; Pizzamiglio et al. 2005; Kraut et al. 2006; Murray et al. 2006; Altmann et al. 2007; Kaplan and Iacoboni 2007; Doehrmann et al. 2008; Galati et al. 2008; Engel et al. 2009; De Lucia et al. 2010; Leaver and Rauschecker 2010). Despite showing a cortical sensitivity for object categories, this literature is unable to prove the existence of abstract cortical encoding modules because it has not assessed the extent to which category sensitivities can be accounted for by systematic between-category differences in the low-level structure (e.g., a neural network that is selectively activated by vocalizations and not by tool-action sounds might simply be processing a reliable low-level difference between these categories, namely, harmonicity [HNR]). For this reason, it is largely unknown whether the previous category-sensitivity results were the product of abstract cortical encoding based on high-level features that optimize the categorization of sound stimuli even in the absence of systematic between-category differences in the low-level structure. A notable exception to this trend is the recent study by Leaver and Rauschecker (2010), which considered 6 different low-level

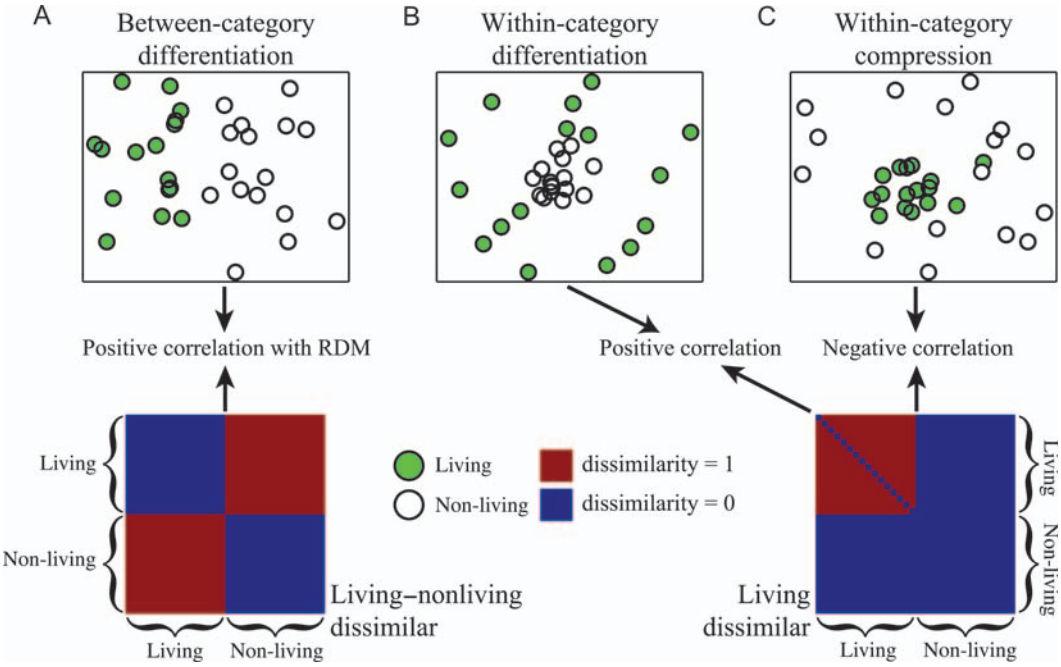

Figure 2. Cortical encoding of category-membership information. We tested for 3 different effects of category-membership information on the dissimilarity of the spatial fMRI patterns. These effects are exemplified in the top 3 panels for the living versus non-living category-membership distinction. In each of these panels, the dissimilarity between the spatial fMRI patterns for different stimuli (circles) is represented as the distance between stimuli within a 2-dimensional Euclidean space where stimuli that evoke largely different fMRI patterns lie farther apart. Similar Euclidean representations can be computed by analyzing RDMs with multidimensional scaling (MDS) models. (*A*) The cortical patch represents the distinction between the living and non-living categories. This outcome of between-category differentiation leads to a positive correlation between the RDM and a living–nonliving dissimilar matrix equal to 0 if 2 sounds are both living or non-living and equal to 1 if 1 sound is living and the other is non-living. (*B*) The cortical patch enhances the distinctions between members of the living category. This outcome of within-category differentiation leads to a positive correlation between the RDM and a living-dissimilar matrix equal to 1 if 2 sounds are both living and equal to 0 otherwise. (*C*) The cortical patch suppresses distinctions between members of the living category. This outcome of within-category compression leads to a negative correlation between the RDM and the living-dissimilar matrix.

alternative hypotheses for the cortical encoding of auditory-object categories. Importantly, however, this study investigated only highly harmonic sounds (human and animal vocalizations and musical instrument tones), and is, for this reason, characterized by the same general validity and methodological caveats that we noted for speech-based investigations of abstract cortical encoding. In the present study, we extended the approach of Leaver and Rauschecker by considering a larger number of low-level features describing both the long-term and the time-varying sound structure (e.g., Giordano and McAdams 2006; Giordano, Rocchesso, et al. 2010 for the psychophysical relevance of the temporal structure of naturalistic sounds and, e.g., Zatorre and Belin 2001; Poeppel 2003; Boemio et al. 2005; Schönwiesner et al. 2005; Zatorre and Gandour 2008, for their cortical processing). The characterization of the low-level features adopted in this study is, to date, the most extensive among those carried out in previous brain imaging studies of naturalistic auditory objects. Our category-encoding tests thus take into account a comparatively large number of low-level alternative hypotheses about the nature of the cortical sensitivity to categories (Fig. 4 for the cortical encoding of the low-level features considered in this study).

We measured the encoding of object-category and low-level features in the stimulus-specific fine-grained spatial fMRI patterns. For this purpose, we adopted the multivariate method of representational similarity analysis (RSA; Kriegeskorte, Mur, Bandettini 2008), previously applied only in the study of visual object processing (Kriegeskorte, Mur, Ruff, et al. 2008; Fig. 1 for our analysis pipeline). The RSA method assesses the encoding of stimulus-feature information in representational dissimilarity matrices (RDMs), measuring the dissimilarity of the spatial fMRI patterns for different stimuli. The RSA method combines the high statistical sensitivity of multivariate classification methods (Kriegeskorte et al. 2006; Kriegeskorte and Bandettini 2007; Staeren et al. 2009), with a greater flexibility in testing the encoding of feature structures whose representation cannot be easily verified with classical massively univariate analyses. For example, within a univariate framework, category encoding is assumed when the blood oxygenation level dependent (BOLD) response for the exemplars of 1 category differs significantly from that for a different reference category. Importantly, this "activation"-based method can measure between-category differences, but cannot detect within-category effects (e.g., the neural responses for different face pictures are more similar to each other than to the neural responses for pictures of cats; Haxby et al. 2001). The "information-based" RSA approach made it possible to assess a larger number of abstract category-membership effects: 1) Cortical representation of the distinction between categories (between-category differentiation); 2) cortical enhancement of the diversity of the stimuli within the same category (within-category differentiation); and 3) cortical suppression of the differences of stimuli within the same category (within-category compression; Figs 2 and 5).

During scanning, participants were presented with highly identifiable environmental sounds (Supplementary Table S1) and carried out a 1-back repetition-detection task. Analyses relied on the measurement of the association between RDMs and stimulus-feature dissimilarity matrices (SDMs; Figs 1 and 2,

and see Material and Methods). Twelve low-level SDMs were derived from the time-varying: 1) Loudness; 2) spectral centroid, a measure of the perceived brightness of a sound; 3) pitch; and 4) harmonic-to-noise ratio (HNR) or, in short, harmonicity, a measure of the amount of periodicity. For each of the 4 time-varying features, we computed 3 SDMs by considering: 1) The median value across time; 2) the amount of temporal change (interquartile range dissimilarity, IQR); and 3) the overall pattern of temporal variation (cross-correlation dissimilarity). Twelve object-category SDMs were derived from the following distinctions: 1) Living versus non-living; 2) human versus non-human; 3) vocal versus non-vocal. One group of object-category SDMs assessed between-category differences in spatial fMRI patterns (Fig. 2*A*); another group assessed within-category effects such as a large similarity of the fMRI patterns for same-category stimuli (e.g., highly similar fMRI patterns for living sounds; Fig. 2*B*,*C*). Variance-decomposition methods made it possible to measure the selective encoding of each of the stimulus features independently of their covariation with non-target features (both low-level and object category). Thus tested, cortical selectivities for object-category features were taken as evidence of abstract cerebral encoding of auditory objects. Based on this methodology, we were able to measure the cortical encoding of several low-level features (Fig. 4) and, most importantly, we observed the abstract cortical encoding of sound-object categories (Fig. 5).

Materials and Methods

Stimuli

Sound stimuli were selected from those investigated by Giordano, McDonnell, et al. (2010). Following standard practices, sounds were equalized in root mean square (RMS) level. Note, however, that cortical activation does not appear to follow the physical intensity of a sound but rather its loudness (Langers et al. 2007), and that RMS equalization does not guarantee constant loudness because it does not take into account the changes in sensitivity across spectral frequencies (Moore 2003). Sounds were 3 s in duration. Sounds from Giordano, McDonnell, et al. (2010) shorter than 3 s were replaced by an alternate excerpt generated by a similar event selected from a variety of commercial databases of sound effects (e.g., Sound Ideas 2004). We selected 32 stimuli: 16 living sounds, generated by the vibration of an object that is part of the body of a living being, such as hands in "clapping hands" and 16 non-living sounds; 16 human-action sounds, generated as a consequence of the motor activity of a human being (such as in "hammering nail") and 16 non-human sounds; 8 vocal sounds, generated as the consequence of the vibrations in the larynx or syrinx (e.g., "croaking frogs") or which included such vibrations (e.g., "panting man") and 24 non-vocal sounds. The human–non-human categorical distinction was perfectly orthogonal to the living–non-living distinction. The vocal–non-vocal distinction was orthogonal to the human–non-human distinction within the category of living sounds (by definition, all vocal sounds are living sounds). Given these stimulus selection constraints, we randomly extracted 20 million stimulus sets from the available samples. Among these random selections, we chose for the final set sounds that: Maximized the average identification performance and had a minimum identification performance score of 50% correct, as measured by Giordano, McDonnell, et al. (2010); minimized the across-sound standard deviation [SD] of the peak of the time-varying level in dB SPL; did not include significant between-category differences (e.g., living versus non-living) in peak dB SPL and in identifiability ($P \geq 0.1$). The measures of identification performance considered during the sound selection process were collected by Giordano, McDonnell, et al. (2010). Identification performance was also measured with the participants in this experiment, subsequent to the scanning session. Based on these measures, all sounds were very accurately identified (average correct = 94%; SD = 6%; minimum correct = 79%), and no differences in identification performance emerged between living and non-living, human and non-human, and vocal and non-vocal sounds ($t \leq 1.74$; $P \geq 0.09$). Supplementary Table S1 reports the properties of the selected stimuli.

Stimulus-Feature Dissimilarity Matrices

We computed 12 matrices quantifying the pairwise dissimilarity of the sound stimuli relative to various category attributes. The strategy followed to compute category-feature dissimilarity matrices is exemplified in Fig. 2. The first 6 category dissimilarities focused on each of the following categorical dimensions in turn: Living versus non-living; human versus non-human; vocal versus non-vocal. The living–non-living dissimilar matrix equaled 1 if 2 sounds did not belong to the same category (i.e., one sound living, the other non-living) and 0 if the 2 sounds belonged to the same category (i.e., both sounds living or both sounds non-living). The living-dissimilar matrix equaled 1 if the 2 sounds were living and 0 otherwise. A third non-living-dissimilar matrix, equal to 1 if the 2 sounds are non-living and 0 otherwise was not considered to avoid problems with the partial-correlation analyses (see below). Indeed, the sum of this third matrix with the living-nonliving dissimilar and living-dissimilar matrices is a constant. As such, the correlation of any of these 3 matrices with a fourth dependent variable would equal 0 after the other 2 matrices are partialed out of the correlation. We adopted the same approach to compute the following matrices: Human–non-human dissimilar; human dissimilar; vocal–non-vocal dissimilar; vocal dissimilar. The final 6 dissimilarities considered the intersection of the 3 main categorical distinctions. Because all vocal sounds are, by definition, living sounds, the intersection of the 3 main categorical distinctions defined 6 independent classes of sound stimuli (e.g., living–human–vocal sounds). The dissimilarity corresponding to each of these 6 intersection classes equaled 1 if both sounds were members of the same intersection class (e.g., both were living–human–vocal sounds) and 0 otherwise. It should be noted that among these matrices, only the living–non-living, human–non-human, and vocal–non-vocal matrices were capable of assessing the differentiation between 2 sound categories. All the other matrices could instead model effects specific to a single category, specifically either a comparatively higher differentiation or a comparatively higher similarity of the sounds within the category of interest.

We computed 12 matrices quantifying the dissimilarity of different low-level properties of the sound stimuli. We initially quantified the time-varying profile of 4 different low-level features: Loudness in sones, defined for each frame of analysis as the sum of the specific loudness for the different cochlear filters; spectral centroid in Equivalent Rectangular Bandwidth-rate units (ERB; Moore and Glasberg 1983), defined as the specific-loudness–weighted average of the spectral frequency; HNR, defined as the ratio of the periodic-to-non-periodic energy in the sound signal (HNR) in dB; pitch in ERB-rate units. Time-varying loudness and spectral centroid were derived from the time-varying specific loudness of the sound signals, as computed according to the model of Glasberg and Moore (2002). Time-varying HNR and pitch were computed using the Praat software (Boersma and Weenink 2009). The temporal resolution of each of the time-varying features was 1 ms. We derived 3 dissimilarity matrices for each of the 4 time-varying sound features by using 1 of 3 different mathematical operators. 1) The first 4 dissimilarity matrices measured the absolute value of the difference in the median of the time-varying feature between each pair of sounds. Median dissimilarities focus on the time-independent scale of the sound features. 2) The next 4 matrices measured the absolute value of the difference in the interquartile range of the time-varying feature between each pair of sounds. The interquartile range dissimilarities focus on the amount of temporal change of the sound features. 3) The last 4 dissimilarity matrices measure the between-sounds diversity in the entire pattern of temporal variation, independently of scale (e.g., high- vs. low median pitch). To this purpose, dissimilarity was defined as 1 minus the maximum cross-correlation between the time-varying feature

measured on sounds A and B (e.g., time-varying loudness for both sounds). The cross-correlation was normalized so as to yield a value of −1 for the cross-correlation between 1 signal and its negative at lag 0, and a value of 1 for the cross-correlation between 1 signal and its replica (i.e., autocorrelation) at a temporal lag of 0. In order to yield a scale-independent measure of the dissimilarity between the time-varying profiles, time-varying features were range normalized between 0 and 1 before being analyzed with the cross-correlation algorithm. Importantly, the cross-correlation measures of dissimilarity are independent of onset-time differences between 2 sounds. Finally note that the number of possible low-level measures of sound dissimilarity is in principle infinite because multiple basic representations of the acoustical structure and multiple mathematical or statistical operators can be adopted to quantify the differences between 2 sounds with 1 single number (Peeters et al. 2011, for an extensive list of acoustical features). In this study: 1) We equated the number of low-level and category-membership dissimilarities to avoid skewing the likelihood of observing significant encoding of features belonging to 1 of these 2 groups (e.g., consider a study with 100 low-level and 1 category-membership dissimilarity); 2) we considered plausible models of how the auditory system computes the temporal variation of 4 basic sensory attributes; 3) we applied the same set of (simple) statistical operators to each of the 4 features.

Participants

Twenty subjects took part in this study (10 females, 10 males; age = 23.8 years, SD = 4.8 years; average years of experience with English language = 20.6 years, SD = 5.6 years; number of native English speakers = 11). All participants had limited musical training (years of music performance experience = 2.6, SD = 4.8 years), had normal hearing as assessed with a standard audiometric procedure (Martin and Champlin 2000; ISO 2004), and were right handed (average laterality quotient = 74.3, SD = 17.6) as assessed with an Edinburgh handedness inventory (Oldfield 1971). Informed consent was obtained from all individuals, and the protocol was approved by the Ethics Committee of the University of Glasgow.

fMRI Task

Participants performed a 1-back repetition-detection task, that is, were requested to press a key when they heard 2 subsequent presentations of the same stimulus. On each block of trials, participants were presented with the 32 stimuli in random order and with 1 repetition of 2 of the 32 stimuli, for a total of 34 stimulus presentations per block. At the end of each block, participants were presented with 6 subsequent silent stimuli of 3 s duration each. Throughout the experiment, each participant carried out 16 blocks of trials. Throughout the experiment, we had a total of 32 subsequent repetitions of 1 sound, 1 for each of the stimuli. The entire scanning session lasted approximately 60 min.

fMRI Data Acquisition

Participants were scanned with a Siemens 3 Tesla Tim Trio scanner (Siemens, Erlangen, Germany), using a 12-channel head coil. Sound stimuli were presented through electrostatic headphones (Nordic Neuro Lab, Bergen, Norway) at a level of 68 dB SPL. The time to repetition (TR) was 5 s, composed of a 2-s acquisition time and a 3-s silent period during which sound stimuli were played on a silent background. No stimulus-onset jittering was used, and the silent period of 3 s between acquisitions was occupied in its entirety by the auditory stimulation (i.e., inter-stimulus interval = TR and stimulus duration = TR − acquisition time). Each brain volume contained 31 slices of 2.2-mm thickness with an inter-slice distance of 2.75 mm in an axial orientation along the direction of the temporal lobe, providing near full-brain coverage (part of the superior prefrontal cortex and the posterior part of the occipital cortex were not acquired in several subjects and were thus excluded from analysis). The in-plane voxel size was $2 \times 2\ mm^2$ (64×64 matrix). A whole-brain, high-resolution, structural T_1-weighted MP-RAGE image (192 sagittal slices, 256×256 matrix size, $1 \times 1 \times 1\ mm^3$ voxel size) was also acquired to characterize the subjects' anatomy.

fMRI Data Analysis

All analyses were carried out using SPM8 and custom Matlab code. Functional images were slice-time corrected to the onset of the first slice and spatially realigned using a 6-parameter affine transformation. High-resolution T_1 images for each of the participants were coregistered to the average functional image and segmented into gray matter, white matter, and cerebrospinal fluid.

The first step of the analysis pipeline involved fitting a first-level native-space generalized linear model (GLM) to the unsmoothed functional images for each participant (Kriegeskorte, Mur, Bandettini 2008; Kriegeskorte, Mur, Ruff, et al. 2008). The GLM focused on gray matter voxels, as identified on the basis of the segmented T_1 scan, and included 33 conditions, 1 for each of the 32 sound stimuli and 1 for sound repetitions and key presses. The GLM also included head-motion parameters estimated during the spatial realignment step, and an intercept term modeling activation during the implicit silent baseline condition. Stimulus-specific BOLD effects were estimated by convolving the sound-stimulus onsets with the canonical hemodynamic response function.

The second step aimed at extracting RDMs (Kriegeskorte, Mur, Bandettini 2008; Kriegeskorte, Mur, Ruff, et al. 2008), measuring the (scale-independent) dissimilarity between the fine-grained spatial distribution of the BOLD effect for each pair of stimuli. Given a target center voxel, we extracted the stimulus-specific BOLD estimates from the contrast images of each of the 32 sounds (sound–silence) inside a spherical volume or searchlight (Kriegeskorte et al. 2006). We chose a searchlight radius of 6.25 mm (89 voxels), because previous simulation studies showed that searchlight radii containing a similar number of voxels optimize the discrimination of experimental conditions (Kriegeskorte et al. 2006). The dissimilarity between the spatial pattern of activation for 2 sounds was thus computed as 1 minus the Pearson correlation between the voxel-specific BOLD estimates for the 2 sounds within the searchlight (Kriegeskorte, Mur, Bandettini 2008; Kriegeskorte, Mur, Ruff, et al. 2008). RDMs were extracted for each gray matter center voxel, provided that at least 50% of the spherical volume included gray matter voxels. Participant-specific RDMs were normalized to MNI space using the normalization parameters obtained from the segmentation procedure and were smoothed using a Gaussian kernel (6-mm full-width at half-maximum, FWHM). In practice, for each of the participants, the normalization and smoothing algorithms were applied independently to each of the 496 maps measuring the dissimilarity between each of the 496 pairs of the 32 sound stimuli. The normalization and smoothing steps were necessary for carrying out random-effects analyses. After normalization voxels were 2 mm^3 in size.

The third and final step of the analysis pipeline adopted a random-effects approach to test the association between RDMs and SDMs (Fig. 1), independently of variance common to the different SDMs (both low-level and category-feature dissimilarities). At the first level, we thus computed the Spearman rank correlation between the RDMs on the one hand and each of the SDMs on the other. For each SDM, this procedure yielded 1 rank- correlation map for each participant. Correlation maps were transformed into Z-maps by applying the variance-stabilizing Fisher Z-transform. For each SDM, we then computed 1 random-effects t-test to assess whether the correlation between the RDMs and the SDMs was significantly different than 0 at the group level [degrees of freedom (df) = 19; $P < 0.0001$, extent threshold = 20 voxels]. This initial correlational test ignored the variance common to the different SDMs. We thus repeated the same random-effects analysis approach by considering the partial Spearman correlation between RDMs and each of the SDMs (df = 19 for t-test; $P < 0.0001$, extent threshold = 20 voxels). The partial correlation analysis discarded all sources of variance shared between the particular SDM and all the other SDMs. Note that the partial-correlation analysis can potentially reveal significant effects even in the absence of a significant correlation, and that the sign of correlations and partial correlations can potentially differ. Both of these potential

results have an unclear functional interpretation. We thus finally considered for each of the SDMs the conjunction of the correlation and partial-correlation group-level *t*-tests ($P < 0.0001$, and extent threshold = 20 voxels for both correlation and partial correlation, Nichols et al. 2005), where the correlation and partial correlation had the same sign. The conjunction analysis thus revealed that those areas where both the correlation and partial correlation between the RDMs and the SDM were significantly different than zero and had the same sign (Figs 4 and 5). Note that an alternative analysis approach based on the smoothed and normalized correlation maps computed in native space yielded highly similar results to those presented in this study, based on smoothed and normalized RDMs.

Results

An initial analysis measured the anatomical overlap in the sensitivity to category and low-level features. This analysis did not discard the variance that was common among all the stimulus features and was meant to illustrate some problematic aspects of studies of the encoding of object categories that do not consider low-level explanations (Fig. 3). The second and third analyses assessed the cortical selectivity for low-level and object-category features, respectively (Figs 4 and 5). For each of the features, significant selectivities were measured by a simultaneous correlation and partial correlation between the RDMs and the SDM (both correlation and partial correlation with same sign). The partial correlation discarded the variance common between the SDM for the target feature and non-target category and low-level SDMs. The third analysis thus assessed the presence of cortical modules that encode categories of auditory objects abstractly. Table 1 summarizes the results of analyses 2–3.

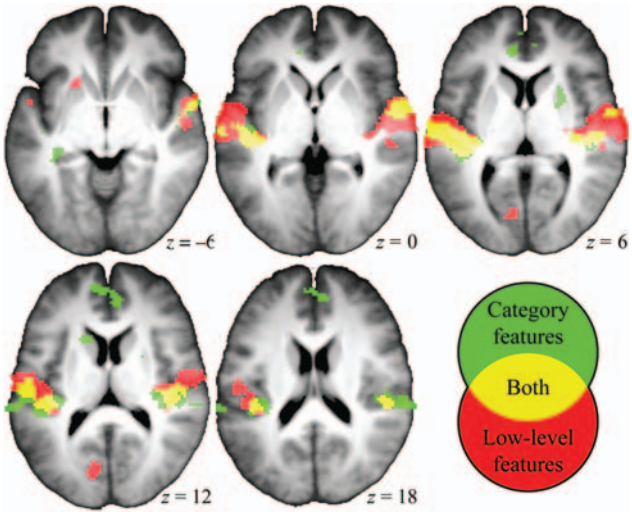

Figure 3. The anatomical overlap of regions sensitive to both category and low-level features (yellow) exemplifies the potential interpretational problems that arise from a study of category encoding independent of the assessment of low-level alternative hypotheses. The measurement of cortical sensitivity is based on the (unpartialed) correlation between RDMs and category or low-level SDMs, and thus considers both the variance specific to each feature, and that shared between the features. With feature-rich sets of naturalistic stimuli, like the 1 in this study, the cortical sensitivity to sound categories can simply be mediated by systematic category differences in neurally relevant low-level structures. Color codes—significant correlation of RDM with: red = only low-level features (one or several); green = only category features (one or several); yellow = at least 1 low-level feature and at least 1 category feature ($P < 0.0001$, uncorrected, extent threshold = 20).

Large Extents of the Temporal Cortex are Sensitive to Both Object-Category and Low-Level Features

Figure 3 shows part of the cortical regions in which we observed a significant correlation between the RDMs and at least 1 of the object-category or low-level features ($P < 0.0001$, uncorrected). In particular, areas marked in yellow are sensitive to both low-level and object-category structure (significant RDM correlation with at least 1 category SDM and at least 1 low-level SDM). Large portions of the bilateral temporal cortex are characterized by dual category/low-level sensitivity, including Heschl's gyrus (HG), whose medial two-thirds are classically assumed to correspond to the core primary auditory fields (Rademacher et al. 1993; Morosan et al. 2001), and the superior temporal plane both anterior and posterior to HG, that is, the PT and the aSTG. The functional meaning of these results is uncertain, however: It could, for example, be the simple product of a statistical association between low-level and object-category features.

Fine-Grained Spatial fMRI Patterns in Both the Temporal and Extratemporal Cortex Selectively Encode Several Low-Level Features

Figure 4 shows those regions where we observed a simultaneous correlation and partial correlation between the RDMs and the low-level feature SDMs ($P < 0.0001$, uncorrected for both; partial correlation discards the variance common between target low-level and non-target low-level features and object-category features). These regions meet the stringent statistical criteria of feature selectivity (e.g., Hall and Plack 2009). We observed encoding of: 1) The median value of the time-varying pitch in a large temporal cluster in both hemispheres, comprising the lateral aspects of HG and including the most anterior portions of the PT; 2) the median value of the time-varying loudness in a large patch of the left auditory cortex extending laterally from the middle portion of HG to the anterior aspects of the left PT; 3) the amount of temporal change of the spectral centroid (spectral centroid IQR SDM) in the most medial aspect of the right HG and the right PT; 4) the median value of the time-varying HNR in the right-lateralized temporal cortex (posterior superior temporal gyrus [pSTG]/STS) and in a bilateral frontal cluster comprising the medial frontal gyrus (medFG) and the anterior cingulate cortex (ACC), with a peak effect in the right hemisphere; 5) the overall pattern of temporal variation of loudness (loudness cross-correlation SDM) in the anterior aspect of the superior parietal lobule (SPL).

The Right Planum Temporale and Posterior Superior Temporal Gyrus Represent Abstract Categories of Auditory Objects

Figure 5 shows those regions where both the correlation and partial correlation between RDMs and category-feature SDMs were significant ($P < 0.0001$, uncorrected for both). These regions are potentially involved in the abstract representation of categories of auditory objects because none of the low-level features we considered explains their encoding in the cortex. We observed encoding of: 1) The category of living sounds, comprising both vocal and non-vocal sounds, as well as human and non-human sounds, in the medial right PT, bordering the medial HG; 2) the category of human sounds, comprising both vocal and non-vocal sounds, as well as both

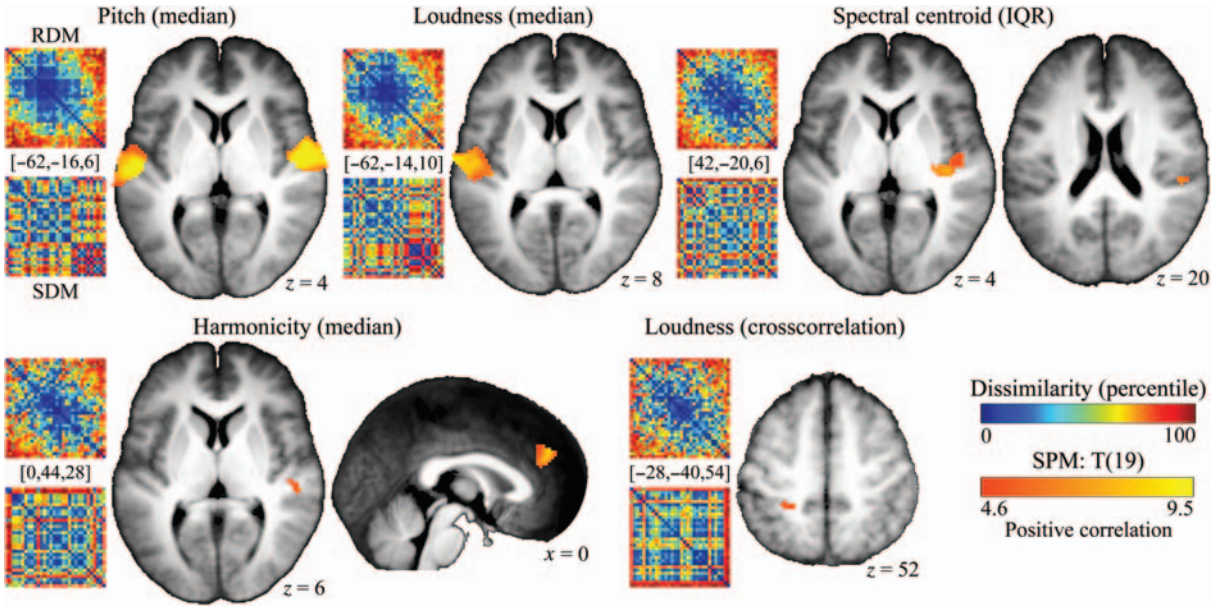

Figure 4. Selective encoding of low-level features is revealed after partialing out of the variance shared with other low-level features and with category features. For each of the low-level features revealed by this analysis, larger low-level differences result in a larger dissimilarity of the fine-grained spatial fMRI patterns (positive RDM-SDM correlation). Clusters show the statistical parametric map (*t*-test) for testing a significant correlation between RDMs and low-level feature SDMs only within those cortical regions characterized by a significant partial correlation between the same low-level SDM and the RDM with the same sign as the correlation ($P < 0.0001$, uncorrected; extent threshold = 20 voxels for both the correlation and partial correlation tests). The statistical parametric maps are overlaid onto the group-average T_1. IQR = interquartile range of time-varying low-level feature. For each of the features, we display the group-average RDM for the peak effect and the relative SDM (numbers between square parentheses = MNI coordinates). The RDMs are displayed in anti-Robinson form (dissimilarity increases moving away from the diagonal); the relative SDM is resorted accordingly. Note the analyses reported in this manuscript have been carried out on the participant-specific RDMs, and not on the group-average RDMs. Group-average RDMs are reported in this figure in order to give a clearer picture of the across-participants structure in the RDM that most likely reflects the stimulus feature at hand.

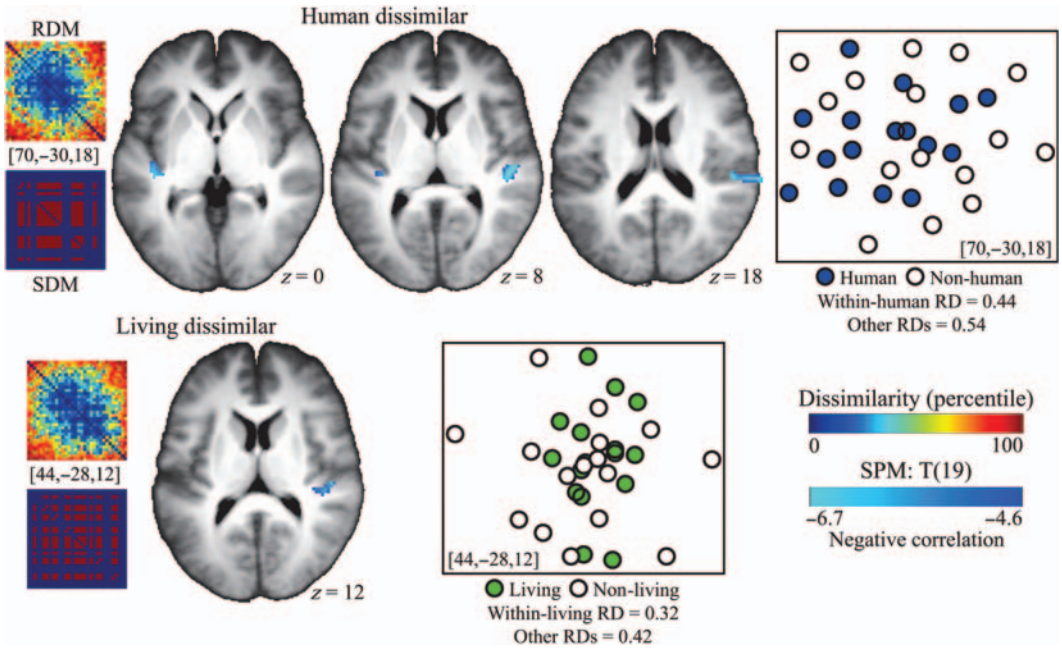

Figure 5. Selective abstract encoding of sound categories is revealed after partialing out of the variance shared with other category features and with low-level features. See Fig. 4 for the meaning of statistical parametric maps and of dissimilarity matrices, and Fig. 2 for details on category SDMs. This analysis reveals 2 temporal regions where all human-action sounds give rise to similar fine-grained spatial fMRI patterns, the right PT/pSTG and left medHG (negative correlation of RDM with human-dissimilar SDM; top slices). A similar effect is observed for the category of living sounds in the right PT/medHG (negative correlation of RDM with living-dissimilar SDM; bottom slice). For both human and living sounds, we thus observe a significant within-category effect (significantly lower diversity of spatial fMRI patterns for sounds within the same category; Fig. 2*B*, *C*) that differs qualitatively from the between-category differences detected with "activation-based" univariate analyses (Fig. 2*A*). To further illustrate the observed within-category effects, we display for each of the category features a 2-dimensional multidimensional scaling (MDS) model of the group-average RDM for the spherical searchlight centered at the peak-correlation voxel (numbers between square parentheses = MNI coordinates). Within the MDS models, points separated by a large distance represent stimuli associated with highly dissimilar fMRI patterns. Note the higher proximity of human compared with non-human sounds (top) and of living compared with non-living sounds (bottom), representing a higher dissimilarity of fMRI patterns within each of these categories. MDS space captions = average representational dissimilarities (RD) for various categories of interest.

Table 1

Summary of random-effects representational similarity analyses: Correlations between stimulus dissimilarities and RDMs, masked for the significant partial correlation with the same sign ($P < 0.0001$, uncorrected, df = 19, extent threshold = 20 voxels for both correlation and partial correlation)

		Left hemisphere						Right hemisphere					
Location	BA	*Z*	Vox	*x*	*y*	*z*	ρ	*Z*	Vox	*x*	*y*	*z*	ρ
Pitch (median)													
lHG/PT	41	5.69	718	−62	−16	6	0.27	–	–	–	–	–	–
lHG/PT	41	–	–	–	–	–	–	5.67	669	68	−10	2	0.18
Loudness (median)													
HG/mSTG/aPT	41/42	5.41	680	−62	−14	10	0.19	–	–	–	–	–	–
Spectral centroid (interquartile range)													
medHG	41	–	–	–	–	–	–	4.92	178	42	−20	6	0.27
PT	41	–	–	–	–	–	–	4.38	32	54	−30	20	0.25
Harmonicity (median)													
medFG/ACC	9/32	–	–	–	–	–	–	4.83	236	0	44	28	0.28
pSTS/pSTG	22/42	–	–	–	–	–	–	4.55	31	48	−22	4	0.12
Loudness (cross-correlation)													
Anterior SPL	7	4.18	24	−28	−40	54	0.32	–	–	–	–	–	–
Human dissimilar													
PT/pSTG	22/42	–	–	–	–	–	–	4.73	228	70	−30	18	−0.24
medHG	41	4.38	99	−44	−20	0	−0.34	–	–	–	–	–	–
Living dissimilar													
PT/medHG	41	–	–	–	–	–	–	4.74	80	44	−28	12	−0.24

Note: The ρ columns report the Spearman correlation between group-average RDMs on the 1 hand and 1 specific SDM on the other.
BA = Brodmann area; *Z* = *Z*-score; Vox = number of voxels in cluster; pSTG/pSTS = posterior superior temporal gyrus/sulcus; mSTG = middle STG; HG = Heschl's gyrus; lHG = lateral HG; HS = Heschl's sulcus; medHG = medial HG; ACC = anterior cingulate cortex; medFG = medial frontal gyrus; SPL = superior parietal lobule; PT = planum temporale; aPT = anterior planum temporale.

living and non-living sounds, in the right pPT/pSTG, and in the most medial aspect of the left HG. In both cases, we detected a significant within-category effect measuring a greater similarity of the spatial fMRI patterns within the living and human categories (Fig. 2, for more details on the strategy adopted to assess the encoding of object-category features).

Discussion

Our study aimed to assess the abstract encoding of categories of non-speech naturalistic auditory objects independently of systematic fingerprints of their low-level structure. The main goal of this study also led us to assess the cortical encoding of a large number of low-level features of naturalistic sounds. We adopted a condition-rich design coupled with information-based analyses of the spatial fMRI patterns. The selectivity for both object-category and low-level features was assessed after partialing out their shared variance. Selective encoding was observed for several low-level features: The brain imaging of naturalistic sounds can represent a powerful instrument for characterizing the signal-processing architecture of the cortex. In both hemispheres, posterior temporal regions, among which the PT, appeared to encode abstractly the categories of living and human sounds. These results 1) reveal domain-general processes for the abstract encoding of auditory objects; 2) motivate a revised hierarchical processing model of increasing information abstraction with the distance from A1 both in the anterior and posterior directions (Rauschecker and Tian 2000; Peelle et al. 2010); 3) suggest that part of the by-product of the template-matching process that takes place in the PT (Griffiths and Warren 2002) is abstract in nature.

Accurate Models of the Cortical Encoding of Sound Categories Should Consider Their Low-Level Structure

Our current understanding of the processing of naturalistic auditory objects in the human cortex focuses on the encoding of categorical structure and largely disregards low-level features (see, e.g., Rauschecker et al. 1995 and Leaver and Rauschecker 2010, for exceptions in the human and animal literature, respectively). The presence of large cortical patches sensitive to both low-level and object-category features (Fig. 3) exemplifies the ambiguity of this approach. The observation of a dual low-level/object-category sensitivity can indeed have 2 interpretations. First, it might be the simple product of a statistical association between categories of objects and low-level features. Secondly, it might be the product of multifunctionality, that is, of the simultaneous encoding of low-level and object-category features in the same neural population (see Bizley et al. 2009, for encoding of multiple low-level features in the same neural populations in A1 in the ferret). Variance decomposition methods, like those adopted in the current study, are necessary to decide between these alternative hypotheses and to assess abstract object encoding independently of the rich low-level structure of naturalistic sounds.

Abstract Representation of Biological Sounds

We observed cortical encoding of 2 categories of auditory objects: Living (right medHG/PT) and human-action auditory objects (left HG and right pPT/pSTG; Fig. 5). In both cases, objects belonging to the same category emerged as evoking similar spatial fMRI patterns. This particular within-category effect could not be detected with conventional univariate analysis approaches, and reveals that the cortical encoding of object categories does not necessarily rely on the ability to tell apart different category exemplars. The comparatively large set of low-level features considered in this study cannot explain these results because the measurement of category encoding ignored the variance common between object-category and low-level features. It is possible that additional low-level features, not considered in this study, account for these results. Another interpretation, however, is that these areas encode high-level abstract features optimized for the

processing of object-category information. Note that various factors might determine whether abstract cortical encoding of sound categories occurs (e.g., salience-related attentional processes, Kayser et al. 2005; identification-related processes, Kilian-Hütten et al. 2011; or in general, the perceptual set that governs how a listener approaches the heard sounds, Liebenthal et al. 2003). It is thus significant that in this study abstract category encoding emerged in the absence of experimentally induced biases toward focusing on a particular source of information: Participants were free to carry out the 1-back repetition-detection task by focusing on, for example, low-level or category-related information. As such, the observed abstract category-encoding effects might potentially be indicative of cortical processing strategies active outside the laboratory. Consistently with this interpretation, previous psychophysical investigations demonstrated a cognitive bias toward processing living sounds by focusing on high-level semantic information (Giordano, McDonnell, et al. 2010). Interestingly, both the human-action and living categories comprise events of a biological origin. We thus argue that abstract processing is a general functional principle of the auditory brain that operates for both speech and non-speech ecologically relevant biological auditory objects. Notably, the location of category-selective modules in our study is consistent with previous observations of abstract processing of speech and human vocalizations in the posterior temporal cortex (e.g., Warren et al. 2006; Desai et al. 2008; Okada et al. 2010). Our results thus complement the notion of abstract object encoding in the anterior temporal lobe (e.g., Belin et al. 2000; Davis and Johnsrude 2003; Hasson et al. 2007; Goll et al. 2010; Leaver and Rauschecker 2010) and suggest a very simple hierarchical model according to which information abstraction in the temporal lobe grows with the distance from A1 both in the posterior and anterior directions (see Fig. 1*B* in Peelle et al. 2010, for an earlier proponent of this hypothesis).

Two aspects of our study represent a departure from previous empirical investigations on the cortical encoding of naturalistic auditory objects. First, previous studies rarely considered low-level alternative hypotheses. Secondly, where previous studies relied on univariate analyses of the voxel-specific BOLD response, our study focused on the multivariate analysis of fine-grained spatial fMRI patterns. In the face of these differences, it is thus significant that the location of abstract modules revealed in this study is consistent with the results from previous category-encoding studies. For example, the representation of the living category in the right medHG/PT is evocative of previous observations of sensitivity to vocalizations, particularly animal vocalizations, in the middle temporal gyrus (e.g., Lewis et al. 2006; Altmann et al. 2007; Doehrmann et al. 2008). More consistently, the representation of the human-action category in the left HG agrees with the results of Kaplan and Iacoboni (2007) and Doehrmann et al. (2008) and with the observations of Leaver and Rauschecker (2010) of the abstract encoding of musical instrument sounds (generated as a consequence of human actions) in the same left-HG area. Finally, the representation of human-action events in the posterior temporal cortex is consistent with the previous reports by Lewis et al. (2006), Murray et al. (2006), Kaplan and Iacoboni (2007), and Doehrmann et al. (2008). Surprisingly, the results of our study did not confirm the hypothesis of a middle anterior superior temporal sulcus center that selectively processes human vocal sounds (Belin et al. 2000, 2002; Gervais et al. 2004; Grandjean et al. 2005; Ethofer et al. 2009; Leaver and Rauschecker 2010). One of the potential sources for our null result is the difference in analysis strategies. Future studies will thus be necessary to disentangle this issue. Alternatively, our null result could arise from the low number of human vocalizations in our stimulus set (12.5% of the total), which was not large enough to promote the abstract encoding of this category (see, e.g., Ulanovsky et al. 2003; Asari and Zador 2009; King and Nelken 2009, for short-term plasticity and context-dependence in auditory system), or, more simply, to make a reliable measurement of abstract human-vocal representations possible. Similar explanations might account for the absence in this study of category-encoding effects related to the vocal versus non-vocal distinction at large.

Cortical Labeling of Sound Categories

We assessed 3 different effects of category-membership information on the dissimilarity of the spatial fMRI patterns, independent of low-level information: 1) Between-category differentiation; 2) within-category differentiation; and 3) within-category compression. Each of these effects can have a different functional interpretation: 1) The cortical patch represents abstractly the distinction between different categories; 2) the cortical patch is involved in the fine processing of stimuli within the category of interest leading to an enhancement of their diversity that go beyond that afforded by low-level information; and 3) the cortical patch codes whether stimuli are exemplars of the category of interest, leading to a compression of the diversity of same category exemplars beyond that afforded by the low-level structure. Our analyses did not provide evidence for the former 2 effects. It is interesting to note that previous studies of naturalistic visual objects did reveal encoding of between-category distinctions in spatial patterns of activation (Kriegeskorte, Mur, Ruff, et al. 2008). As such, it remains to be seen whether our failure to observe between-category effects is indicative of a specific property of pattern-based encoding in the auditory brain or, instead, stems from the specifics of our experimental methodology (e.g., choice of stimuli). More importantly, we measured within-category compression for both the living and human-action categories, and in several mid-to-posterior temporal areas, among which the PT (Fig. 5). Interestingly, the PT is considered to implement acoustics-dependent processes that match stored spectrotemporal templates to the incoming sensory information (Griffiths and Warren 2002; Warren et al. 2005; Kumar et al. 2007). Our results thus suggest that although the matching process implemented by the PT relies on the analysis of sensory information, part of the end product of this process is abstract in nature. We thus argue that by matching incoming low-level patterns to stored templates, the PT facilitates a process of labeling auditory objects as members of specific categories. In post-PT stations of information relay, this labeling information could, for example, facilitate the discrimination of different categories, and promote similar processing pipelines for same category exemplars.

Brain Imaging of Naturalistic Sounds Makes it Possible to Assess the Encoding of Multiple Low-Level Features

The majority of studies on the cortical encoding of low-level features investigate synthetic stimuli that differ along a restricted number of acoustical dimensions, often one. From the methodological point of view, single-feature studies are not capable of measuring cortical selectivity because they do not consider the effects of variations in extraneous features (Hall and Plack 2009; Bizley and Walker 2010). In general, brain mapping of naturalistic sounds is a powerful instrument for the study of low-level feature encoding because: 1) It exploits the rich low-level structure of naturally occurring stimuli that likely shapes the neural processing starting from the auditory nerve (Lewicki 2002); 2) it makes it possible to measure encoding of multiple sound features within a single experiment; 3) it makes it possible to test for cortical selectivity. In this study, we considered 12 different low-level features and observed selectivity for 5 of them. Overall, these analyses confirm previous hypotheses of a pitch-encoding center in the lateral HG (e.g., Zatorre 1988) and of a right-lateral bias for the processing of spectral features (spectral centroid, but also HNR; e.g., Zatorre and Belin 2001). We also observed left-lateralized encoding of loudness in the temporal cortex, a rarely observed functional hemispheric asymmetry that could originate from the exclusive allocation of right-hemisphere resources to the processing of the rich spectral structure of the stimuli in the current study. Finally, the left SPL appeared to encode selectively the pattern of temporal variation of loudness. This result might be indicative of a role of this low-level feature in the cortical analysis of auditory and multimodal tool-action events (e.g., a series of loudness impulses for hammering nail versus less abrupt temporal variations for sawing wood; see Lewis 2006, for a review). In the following, we discuss the results for each of the low-level features in more detail.

Pitch

The median of the time-varying pitch was encoded bilaterally in an area of the temporal cortex that includes the lateral HG. A significant body of evidence supports the hypothesis of a general pitch-encoding center in the lateral HG (Zatorre 1988; Johnsrude et al. 2000; Gutschalk et al. 2002; Patterson et al. 2002; Bendor and Wang 2006; Hyde et al. 2008; Foster and Zatorre 2010). The general validity of this position has recently been criticized on the grounds that some of the studies consistent with this hypothesis are carried with synthetic stimuli from the same class (iterated ripple noises, Hall and Plack 2009). Our results thus provide strong support for the hypothesis of a general pitch-encoding center in the lateral HG, because the stimuli investigated in this study are highly diverse in their low-level structure.

Spectral Centroid

The right medHG and PT encoded the amount of temporal variation of the spectral centroid. In general, these results agree with the hypothesis of finer spectral processing abilities in the right temporal cortex (Zatorre and Belin 2001; Schönwiesner et al. 2005; Warren et al. 2005; Jamison et al. 2006; Kumar et al. 2007; Obleser et al. 2008; see Zatorre and Gandour 2008, for a review; Altmann et al. 2010) and with that of a specialization of the PT in the analysis of time-varying spectral patterns (Griffiths and Warren 2002; Zatorre and Belin 2005). The right-lateralized encoding of this feature is, at least apparently, at odds with the frequent observation of a left-hemisphere specialization for the analysis of the temporal variation of spectral information (Zatorre and Belin 2001; Poeppel 2003; Boemio et al. 2005; Schönwiesner et al. 2005; Zatorre and Gandour 2008). However, we note that several studies did reveal encoding of the spectrotemporal variation in both the right and left temporal cortices. They also show that hemispheric asymmetries generally emerge as a function of the rate of spectrotemporal variation rather than of overall temporal variation (left-hemispheric specialization for faster rates, e.g., Belin et al. 1998; Boemio et al. 2005; see also Obleser et al. 2008), whereas our spectral-centroid IQR measure captures the amount of temporal variation across slower and faster rates.

Harmonicity

The median HNR was encoded in the right pSTG/STS and in the bilateral ACC/medFG (peak effect in the right hemisphere). The right lateralization of the pSTG/STS encoding of this feature is perhaps suggestive of cortical computations based on spectrum-matching processes rather than on an analysis of the temporal structure of the incoming waveform (cf., right-hemispheric advantage for spectral processing; see above). From the psychophysical point of view, HNR accounted for the dissimilarity ratings of environmental sounds in Gygi et al. (2007), and for the tool versus animal categorization in Lewis et al. (2005). Within the brain-imaging literature, ACC has been reported to differentiate between highly harmonic voiced speech and less harmonic whispered speech (Schulz et al. 2005). Notably, only 2 studies investigated systematically the cortical encoding of HNR (Lewis et al. 2009; Leaver and Rauschecker 2010). Consistently with our results, both of these studies observed right-temporal sensitivity to HNR, although in more anterior regions. It should be emphasized, however, that the cortical representation of HNR appears to be largely dependent on the investigated sound set (cf. variability of HNR-sensitive centers for animal vocalizations and iterated ripples noises in Lewis et al. 2009). Given the paucity of studies on the cortical processing of HNR, statements about a general processing center are premature. Given the high relevance of HNR for the behavioral evaluation of heterogeneous sets of environmental sounds (e.g., Gygi et al. 2007), it is plausible that the participants in this experiment focused on this same low-level feature when carrying out the 1-back repetition-detection task inside the scanner (e.g., answer "repetition" if 2 subsequent stimuli have highly similar HNR values; note that the task did not explicitly impose constraints on the response strategies). As such, the encoding of the median HNR in the ACC might be the product of task-related processes: This cortical area is indeed part of a "salience network" involved in decisional processes based, for instance, on sensory information (Seeley et al. 2007) and in the processing of errors and conflicts (Menon et al. 2001; Ridderinkhof et al. 2004). Furthermore, it is hypothesized to be part of a network that supports focal auditory activity (Hunter et al. 2006).

Loudness

Two loudness features were encoded in the left hemisphere: The median of the time-varying loudness in the primary auditory cortex, extending also to anterior planum temporale, and

the overall pattern of time-varying loudness in aSPL. Among various studies carried out with synthetic sounds, only Brechmann et al. (2002) observed a clear left-lateralized bias for the processing of the overall loudness of a sound signal, whereas partial agreement emerges concerning the role of the PT in the processing of this property (see Ernst et al. 2008, for a review). Among the various factors that might explain these divergences, it might be speculated that the higher complexity of the spectral structure of the sounds in the current experiment strengthened the right-lateral bias for the processing of spectral properties at the expense of the processing of energetic features such as loudness in the same hemisphere. Focusing on the role of the left SPL in differentiating between temporal patterns of loudness variation, it is interesting to note that this area appears to be involved in the processing of tool-action events in the motor, visual, and auditory domains (Lewis et al. 2005, see Lewis 2006, for a review; and Giusti et al. 2010, for cortical processing of action sounds in left SPL; see Griffiths 2008; Rauschecker and Scott 2009; Recanzone and Cohen 2010, for role of SPL in dorsal pathway), and in the online updating of actions (Tunik et al. 2008). Notably, psychophysical investigations of naturalistic sounds suggest that the identification of the actions carried out on an object relies primary on the temporal patterning of the sound signals (e.g., bouncing versus breaking of a glass bottle, Warren and Verbrugge 1984). As such, the role of SPL in differentiating between temporal loudness patterns might potentially subserve processes of sound-based motor control and of sensory-motor transformation.

Making Sense of a Variable Environment

In this study, we revealed that the spatial patterns of activation in various regions of the temporal cortex label auditory objects as exemplars of 2 ecologically relevant categories: Sounds generated by vibrating living objects, and action sounds involving a human agent. The exact nature of the neural processes at the basis of this result remains to be detailed. For example, category encoding might be product a non-linear feature-combination analysis that merges information from multiple low-level features (Sadagopan and Wang 2009). Independently of the exact nature of the neural code, it is important to emphasize that it appears to be independent of between-sound differences along various fundamental dimensions of auditory sensation such as loudness, pitch, and timbre-related dimensions such as spectral centroid and HNR (note the investigation in this study of different measures of dissimilarity along each of these dimensions). As such, the categorization code appears to be optimized for carrying out a job that is very important for an adaptive organism: Recognizing basic properties of the objects that populate the environment in the face of variations along several attributes of the input sensory information (King and Nelken 2009). In our ancestors, general-purpose abstract encoding mechanisms that serve to extract biologically relevant auditory-object information might thus have spurred the development of increasingly sophisticated strategies for the robust categorical processing of calls, ultimately resulting in the emergence of phonetic analysis processes at the basis of the speech ability.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work was supported by the Marie Curie Intra-European Fellowships program (FP7 PEOPLE-2011-IEF, project BrainInNaturalSound to B.L.G. and P.B.), by the Biotechnology and Biological Sciences Research Council (grant BB/E003958/1 to P.B.), by the Economic and Social Research Council-MC (grant RES-060-25-0010 to P.B.), by the Canada Research Chair in Music Perception and Cognition (S.Mc.A.), and by the Natural Sciences and Engineering Research Council of Canada (RGPIN 312774-2010 to S.Mc.A.).

Notes

Conflict of Interest: None declared

References

Ahissar M, Nahum M, Nelken I, Hochstein S. 2009. Reverse hierarchies and sensory learning. Philos Trans R Soc B. 364:285–299.
Altmann CF, Doehrmann O, Kaiser J. 2007. Selectivity for animal vocalizations in the human auditory cortex. Cereb Cortex. 17:2601–2608.
Altmann CF, Júnior CGO, Heinemann L, Kaiser J. 2010. Processing of spectral and amplitude envelope of animal vocalizations in the human auditory cortex. Neuropsychologia. 48:2824–2832.
Asari H, Zador A. 2009. Long-lasting context dependence constrains neural encoding models in rodent auditory cortex. J Neurophysiol. 102:2638–2656.
Bar-Yosef O, Nelken I. 2007. The effects of background noise on the neural responses to natural sounds in cat primary auditory cortex. Front Comput Neurosci. 1:3.
Belin P, Zatorre RJ, Ahad P. 2002. Human temporal-lobe response to vocal sounds. Cogn Brain Res. 13:17–26.
Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. Nature. 403:309–312.
Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y. 1998. Lateralization of speech and auditory temporal processing. J Cogn Neurosci. 10:536–540.
Bendor D, Wang X. 2006. Cortical representations of pitch in monkeys and humans. Curr Opin Neurobiol. 16:391–399.
Bizley J, Walker K. 2010. Sensitivity and selectivity of neurons in auditory cortex to the pitch, timbre, and location of sounds. Neuroscientist. 16:453–469.
Bizley J, Walker K, Silverman B, King A, Schnupp J. 2009. Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. J Neurosci. 29:2064–2075.
Boemio A, Fromm S, Braun A, Poeppel D. 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci. 8:389–395.
Boersma P, Weenink D. 2009. Praat: doing phonetics by computer version 5.1.05 [computer program]. http://www.fon.hum.uva.nl/praat/. Last date retrieved 1 May 2009.
Brechmann A, Baumgart F, Scheich H. 2002. Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. J Neurophysiol. 87:423–433.
Chechik G, Anderson M, Bar-Yosef O, Young E, Tishby N, Nelken I. 2006. Reduction of information redundancy in the ascending auditory pathway. Neuron. 51:359–368.
Davis M, Johnsrude I. 2003. Hierarchical processing in spoken language comprehension. J Neurosci. 23:3423–3431.
De Lucia M, Clarke S, Murray MM. 2010. A temporal hierarchy for conspecific vocalization discrimination in humans. J Neurosci. 30:11210–11221.

Desai R, Liebenthal E, Waldron E, Binder J. 2008. Left posterior temporal regions are sensitive to auditory categorization. J Cogn Neurosci. 20:1174–1188.
Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF. 2008. Probing category selectivity for environmental sounds in the human auditory brain. Neuropsychologia. 46:2776–2786.
Engel L, Frum C, Puce A, Walker N, Lewis J. 2009. Different categories of living and non-living sound-sources activate distinct cortical networks. Neuroimage. 47:1778–1791.
Ernst S, Verhey J, Uppenkamp S. 2008. Spatial dissociation of changes of level and signal-to-noise ratio in auditory cortex for tones in noise. Neuroimage. 43:321–328.
Ethofer T, Van De Ville D, Scherer K, Vuilleumier P. 2009. Decoding of emotional information in voice-sensitive cortices. Curr Biol. 19:1028–1033.
Fecteau S, Armony JL, Joanette Y, Belin P. 2004. Is voice processing species-specific in human auditory cortex? An fMRI study. Neuroimage. 23:840–848.
Fecteau S, Armony JL, Joanette Y, Belin P. 2005. Sensitivity to voice in human prefrontal cortex. J Neurophysiol. 94:2251–2254.
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.
Foster NEV, Zatorre RJ. 2010. A role for the intraparietal sulcus in transforming musical pitch information. Cereb Cortex. 20:1350–1359.
Galati G, Committeri G, Spitoni G, Aprile T, Di Russo F, Pitzalis S, Pizzamiglio L. 2008. A selective representation of the meaning of actions in the auditory mirror system. Neuroimage. 40:1274–1286.
Gervais H, Belin P, Boddaert N, Leboyer M, Coez A, Sfaello I, Barthélémy C, Brunelle F, Samson Y, Zilbovicius M. 2004. Abnormal cortical voice processing in autism. Nat Neurosci. 7:801–802.
Giordano BL, McAdams S. 2006. Material identification of real impact sounds: effects of size variation in steel, glass, wood and plexiglass plates. J Acoust Soc Am. 119:1171–1181.
Giordano BL, McDonnell J, McAdams S. 2010. Hearing living symbols and nonliving icons: category-specificities in the cognitive processing of environmental sounds. Brain Cogn. 73:7–19.
Giordano BL, Rocchesso D, McAdams S. 2010. Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. J Exp Psychol Hum Percept Perform. 36:462–479.
Giusti MA, Bozzacchi C, Pizzamiglio L, Di Russo F. 2010. Sight and sound of actions share a common neural network. Eur J Neurosci. 32:1754–1764.
Glasberg BR, Moore BCJ. 2002. A model of loudness applicable to time-varying sounds. J Audio Eng Soc. 50:331–342.
Goll JC, Crutch SJ, Warren JD. 2010. Central auditory disorders: toward a neuropsychology of auditory objects. Curr Opin Neurol. 23:617–627.
Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P. 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat Neurosci. 8:145–146.
Griffiths TD. 2008. Sensory systems: auditory action streams? Curr Biol. 18:R387–R388.
Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. Trends Neurosci. 25:348–353.
Gutschalk A, Patterson R, Rupp A, Uppenkamp S, Scherg M. 2002. Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage. 15:207–216.
Gygi B, Kidd GR, Watson CS. 2007. Similarity and categorization of environmental sounds. Percept Psychophys. 69:839–855.
Hall D, Plack C. 2009. Pitch processing sites in the human auditory brain. Cereb Cortex. 19:576–585.
Hasson U, Skipper J, Nusbaum H, Small S. 2007. Abstract coding of audiovisual speech: beyond sensory representation. Neuron. 56:1116–1126.
Haxby J, Gobbini M, Furey M, Ishai A, Schouten J, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science. 293:2425–2430.
Hunter M, Eickhoff S, Miller T, Farrow T, Wilkinson I, Woodruff P. 2006. Neural activity in speech-sensitive auditory cortex during silence. Proc Natl Acad Sci USA. 103:189–194.
Hyde KL, Peretz I, Zatorre RJ. 2008. Evidence for the role of the right auditory cortex in fine pitch resolution. Neuropsychologia. 46:632–639.
ISO. 2004. Acoustics – reference zero for the calibration of audiometric equipment – part 8: reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (ISO 389–8) (Tech. Rep.). Geneva: International Organization for Standardization.
Jääskeläinen I, Ahveninen J, Belliveau J, Raij T, Sams M. 2007. Short-term plasticity in auditory cognition. Trends Neurosci. 30:653–661.
Jamison H, Watkins K, Bishop D, Matthews P. 2006. Hemispheric specialization for processing auditory nonspeech stimuli. Cereb Cortex. 16:1266–1275.
Johnsrude I, Penhune V, Zatorre RJ. 2000. Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain. 123:155–163.
Kaplan JT, Iacoboni M. 2007. Multimodal action representation in human left ventral premotor cortex. Cogn Process. 8:103–113.
Kayser C, Petkov CI, Lippert M, Logothetis NK. 2005. Mechanisms for allocating auditory attention: an auditory saliency map. Curr Biol. 15:1943–1947.
Kilian-Hütten N, Valente G, Vroomen J, Formisano E. 2011. Auditory cortex encodes the perceptual interpretation of ambiguous sound. J Neurosci. 31:1715–1720.
King AJ, Nelken I. 2009. Unraveling the principles of auditory cortical processing: can we learn from the visual system? Nat Neurosci. 12:698–701.
Kraut MA, Pitcock JA, Calhoun V, Li J, Freeman T, Hart J, Jr. 2006. Neuroanatomic organization of sound memory in humans. J Cogn Neurosci. 18:1877–1888.
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. Information-based functional brain mapping. Proc Natl Acad Sci USA. 103:3863–3868.
Kriegeskorte N, Mur M, Bandettini P. 2008. Representational similarity analysis – connecting the branches of systems neuroscience. Front Syst Neurosci. 2:4.
Kriegeskorte N, Mur M, Ruff D, Kiani R, Bodurka J, Esteky H, Tanaka K, Bandettini PA. 2008. Matching categorical object representations in inferior temporal cortex of man and monkey. Neuron. 60:1126–1141.
Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD. 2007. Hierarchical processing of auditory objects in humans. PLoS Comput Biol. 3:e100.
Langers D, van Dijk P, Schoenmaker E, Backes W. 2007. fMRI activation in relation to sound intensity and loudness. Neuroimage. 35:709–718.
Leaver AM, Rauschecker JP. 2010. Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. J Neurosci. 30:7604–7612.
Lewicki MS. 2002. Efficient coding of natural sounds. Nat Neurosci. 5:356–363.
Lewis JW. 2006. Cortical networks related to human use of tools. Neuroscientist. 12:211–231.
Lewis JW, Brefczynski JA, Phinney RE, Jannik JJ, DeYoe ED. 2005. Distinct cortical pathways for processing tool versus animal sounds. J Neurosci. 25:5148–5158.
Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe EA. 2006. Lefties get it "right" when hearing tool sounds. J Cogn Neurosci. 18:1314–1330.
Lewis JW, Talkington WJ, Puce A, Engel LR, Frum C. 2010. Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. J Cogn Neurosci. 23:2079–2101.
Lewis JW, Talkington WJ, Walker NA, Spirou GA, Jajosky A, Frum C, Brefczynski-Lewis JA. 2009. Human cortical organization for processing vocalizations indicates representation of harmonic structure as a signal attribute. J Neurosci. 29:2283–2296.

Liebenthal E, Binder JR, Piorkowski RL, Remez RE. 2003. Short-term reorganization of auditory analysis induced by phonetic experience. J Cogn Neurosci. 15:549–558.
Machens C, Wehr M, Zador A. 2004. Linearity of cortical receptive fields measured with natural sounds. J Neurosci. 24:1089–1100.
Martin FN, Champlin CA. 2000. Reconsidering the limits of normal hearing. J Am Acad Audiol. 11:64–66.
Menon V, Adleman NE, White CD, Glover GH, Reiss AL. 2001. Error-related brain activation during a Go/NoGo response inhibition task. Hum Brain Mapp. 12:131–143.
Moore BCJ. 2003. An introduction to the psychology of hearing. 5th ed. San Diego (CA): Academic Press.
Moore BCJ, Glasberg BR. 1983. Suggested formulae for calculating auditory–filter bandwidths and excitation patterns. J Acoust Soc Am. 74:750–753.
Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. Neuroimage. 13:684–701.
Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S. 2006. Rapid brain discrimination of sounds of objects. J Neurosci. 26:1293–1302.
Nichols T, Brett M, Andersson J, Wager T, Poline JB. 2005. Valid conjunction inference with the minimum statistic. Neuroimage. 25:653–660.
Obleser J, Eisner F, Kotz SA. 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J Neurosci. 28:8116–8123.
Okada K, Rong F, Venezia J, Matchin W, Hsieh I, Saberi K, Serences JT, Hickok G. 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb Cortex. 13:1–7.
Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 9:97–113.
Patterson R, Uppenkamp S, Johnsrude I, Griffiths TD. 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron. 36:767–776.
Peelle J, Johnsrude I, Davis M. 2010. Hierarchical processing for speech in human auditory cortex and beyond. Front Hum Neurosci. 4:51.
Peeters G, Giordano BL, Susini P, Misdariis N, McAdams S. 2011. The timbre toolbox: extracting acoustic descriptors from musical signals. J Acoust Soc Am. 130:2902–2916.
Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D'Amico S, Di Russo F. 2005. Separate neural systems for processing action- or non-action-related sounds. Neuroimage. 24:852–861.
Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as asymmetric sampling in time. Speech Commun. 41:245–255.
Rademacher J, Caviness VS, Steinmetz H, Galaburda AM. 1993. Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. Cereb Cortex. 3:313–329.
Rauschecker JP, Scott SK. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci. 12:718–724.
Rauschecker JP, Tian B. 2000. Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci USA. 97:11800–11806.
Rauschecker JP, Tian B, Hauser M. 1995. Processing of complex sounds in the macaque nonprimary auditory cortex. Science. 268:111–114.
Recanzone G, Cohen Y. 2010. Serial and parallel processing in the primate auditory cortex revisited. Behav Brain Res. 206:1–7.
Ridderinkhof K, Ullsperger M, Crone E, Nieuwenhuis S. 2004. The role of the medial frontal cortex in cognitive control. Science. 306:443–447.
Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci. 1999:1131–1136.
Sadagopan S, Wang X. 2009. Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. J Neurosci. 29:11192–11202.
Schönwiesner M, Rübsamen R, von Cramon DY. 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. Eur J Neurosci. 22:1521–1528.
Schönwiesner M, Zatorre RJ. 2009. Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. Proc Natl Acad Sci USA. 106:14611–14616.
Schulz GM, Varga M, Jeffires K, Ludlow CL, Braun AR. 2005. Functional neuroanatomy of human vocalization: an H215O PET study. Cereb Cortex. 15:1835–1847.
Scott SK, Blank CC, Rosen S, Wise RJS. 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain. 123:2400–2406.
Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, Reiss AL, Greicius MD. 2007. Dissociable intrinsic connectivity networks for salience processing and executive control. J Neurosci. 27:2349–2356.
Sound Ideas. 2004. Series 6000 DVD Combo Sound Effects Library. Ontario (Canada): Richmond Hill.
Staeren N, Renvall H, De Martino F, Goebel R, Formisano E. 2009. Sound categories are represented as distributed patterns in the human auditory cortex. Curr Biol. 19:498–502.
Tunik E, Ortigue S, Adamovich SV, Grafton ST. 2008. Differential recruitment of anterior intraparietal sulcus and superior parietal lobule during visually guided grasping revealed by electrical neuroimaging. J Neurosci. 28:13615–13620.
Ulanovsky N, Las L, Nelken I. 2003. Processing of low-probability sounds by cortical neurons. Nat Neurosci. 6:391–398.
Warren JD, Jennings AR, Griffiths TD. 2005. Analysis of the spectral envelope of sounds by the human brain. Neuroimage. 24:1052–1057.
Warren JD, Scott SK, Price CJ, Griffiths TD. 2006. Human brain mechanisms for the early analysis of voices. Neuroimage. 31:1389–1397.
Warren WH, Verbrugge RR. 1984. Auditory perception of breaking and bouncing events: a case study in ecological acoustics. J Exp Psychol Hum Percept Perform. 10:704–712.
Zatorre RJ. 1988. Pitch perception of complex tones and human temporal-lobe function. J Acoust Soc Am. 84:566–572.
Zatorre RJ, Belin P. 2005. Auditory cortex processing streams: where are they and what do they do? In: Josef S, Merzenich MM, editors. Plasticity and signal representation in the auditory system. New York (NY): Springer. p. 277–290.
Zatorre RJ, Belin P. 2001. Spectral and temporal processing in human auditory cortex. Cereb Cortex. 11:946–953.
Zatorre RJ, Gandour JT. 2008. Neural specializations for speech and pitch: moving beyond the dichotomies. Philos Trans R Soc B. 363:1087–1104.