Processing of Natural Sounds: Characterization of Multipeak Spectral Tuning in Human Auditory Cortex

Michelle Moerel,^{1,2} Federico De Martino,^{1,2,3} Roberta Santoro,^{1,2} Kamil Ugurbil,³ Rainer Goebel,^{1,2} Essa Yacoub,³ and Elia Formisano^{1,2}

¹Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, The Netherlands,

²Maastricht Brain Imaging Center (MBIC), 6229 EV Maastricht, The Netherlands, and ³Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, Minneapolis, Minnesota 55455

We examine the mechanisms by which the human auditory cortex processes the frequency content of natural sounds. Through mathematical modeling of ultra-high field (7 T) functional magnetic resonance imaging responses to natural sounds, we derive frequency-tuning curves of cortical neuronal populations. With a data-driven analysis, we divide the auditory cortex into five spatially distributed clusters, each characterized by a spectral tuning profile. Beyond neuronal populations with simple single-peaked spectral tuning (grouped into two clusters), we observe that ~60% of auditory populations are sensitive to multiple frequency bands. Specifically, we observe sensitivity to multiple frequency bands (1) at exactly one octave distance from each other, (2) at multiple harmonically related frequency intervals, and (3) with no apparent relationship to each other. We propose that beyond the well known cortical tonotopic organization, multi-peaked spectral tuning amplifies selected combinations of frequency bands. Such selective amplification might serve to detect behaviorally relevant and complex sound features, aid in segregating auditory scenes, and explain prominent perceptual phenomena such as octave invariance.

Introduction

The sounds and auditory scenes we encounter every day consist of a rich and complex combination of frequencies. At the sensory periphery (cochlea) and in the subcortical auditory relays, sound frequency bands are selectively processed in spatially segregated channels. At the early stages of the auditory processing hierarchy, neurons with similar frequency preference cluster together and form a cochleotopic or tonotopic map (King and Nelken, 2009). This topographic organization of sound frequency is maintained at the level of the auditory cortex, where multiple tonotopic maps can be discriminated (Merzenich and Brugge, 1973; Merzenich et al., 1973; Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012).

How frequency information is represented and processed at the level of the human auditory cortex beyond this tonotopic representation remains largely unknown. Results from invasive studies of animal audition suggest that next to their preferred frequency, cortical neurons exhibit sensitivity to additional fre-

quency bands (bat, Fitzpatrick et al., 1993; cat, Sutter and Schreiner, 1991; Noreña et al., 2008; and marmoset, Kadia and Wang, 2003; Sadagopan and Wang, 2009). This sensitivity to multiple frequency bands is shaped by the acoustic environment. For example, in the bat neurons were found that showed facilitative responses to the frequencies in the bat's pulse and echo, which this animal uses during the search and pursuit of insects (Fitzpatrick et al., 1993). In marmosets and songbirds, neurons finely tuned to informative features of conspecific vocalizations were reported (Wang et al., 1995; Woolley et al., 2005). It has been hypothesized that these neurons with sensitivity to multiple frequency bands could serve as complex feature detectors, signaling the presence of a specific and informative combination of behaviorally relevant frequencies. Such neuronal tuning may play a crucial role in the creation of an abstract, higher level sound representation (deCharms et al., 1998; Wang, 2007; Sadagopan and Wang, 2009). Alternatively, from a predictive coding perspective, the representation of auditory objects with complicated (polymodal) spectral representations would be required to provide predictions of frequency specific auditory input, at low levels of the auditory hierarchy (Friston, 2005; Winkler et al., 2009).

Is a similar mechanism of complex spectral tuning in place in the human auditory cortex? In this study, we use ultra-high field functional magnetic resonance imaging (fMRI; 7 T) to measure brain responses to a large set of natural sounds and extract spectral profiles throughout the human auditory cortex. Next, we use a data-driven algorithm to cluster these spectral profiles and extract five spatially distributed functional subdivisions, each characterized by a specific spectral tuning profile. Simple single-peaked spectral profiles are complemented by profiles with sen-

Received Nov. 15, 2012; revised May 24, 2013; accepted June 8, 2013.

Author contributions: M.M., F.D.M., and E.F. designed research; F.D.M. and E.Y. performed research; R.S. and R.G. contributed unpublished reagents/analytic tools; M.M., F.D.M., and E.F. analyzed data; M.M., F.D.M., R.S., K.U., R.G., E.Y., and E.F. wrote the paper.

This work was supported by Maastricht University, the Netherlands Organization for Scientific Research (Topical Grant 021-002-102, M.M.), and Innovative Research Incentives Scheme Vidi Grant 452-04-330, E.F.), the National Institutes of Health (grants P41 EB015894, P30 NS076408, and S10 RR26783), and the WM KECK Foundation. We thank A. Goulas and M. Frost for comments and discussions.

Correspondence should be addressed to Michelle Moerel, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: michelle.moerel@maastrichtuniversity.nl.

DOI:10.1523/JNEUROSCI.5306-12.2013

Copyright © 2013 the authors 0270-6474/13/3311888-11\$15.00/0

sitivity to multiple frequency bands. Specifically, we observe selectivity to frequency bands (1) at exactly one octave lag, (2) at harmonically related frequency intervals, and (3) with no apparent relationship to each other. We propose that this multipeaked spectral tuning may amplify selected combinations of frequency bands, possibly serving to detect behaviorally relevant sound features, aiding in segregating auditory scenes and underlying prominent perceptual phenomena such as octave perception.

Materials and Methods

Subjects. Five subjects (median age = 32, three males) participated in the main study, and additionally took part in a separate localizer experiment (see below). The subjects had no history of hearing disorder or neurological disease, and gave informed consent before commencement of the measurements. The Institutional Review Board for human subject research at the University of Minnesota granted approval for the study.

Stimuli. In the main study, the stimuli consisted of recordings of 168 natural sounds, and included human vocal sounds (both speech and nonspeech, e.g., baby cry, laughter, coughing), animal cries (e.g., monkey, lion, horse), tool sounds and musical instruments (e.g., keys, scissors, piano, flute), and scenes from nature (e.g., rain, wind, thunder). Sounds were sampled at 16 kHz and their duration was cut at 1000 ms. In addition to the main study, we collected localizer data in the same subjects. In the localizer, the stimuli consisted of sounds grouped into eight conditions (three tones and five semantic category conditions). We analyzed only the responses to the tones. Amplitude-modulated tones were created in MATLAB (8 Hz, modulation depth of 1) with a carrier frequency of 0.45, 0.5, and 0.55 kHz for the low-frequency condition; 1.35, 1.5, and 1.65 kHz for the middle frequency condition; and 2.25, 2.5, and 2.75 kHz for the high-frequency conditions. Sounds were sampled at 16 kHz and their duration was cut at 800 ms.

In both the main study and the localizer, sound onset and offset were ramped with a 10 ms linear slope and their energy (root mean square) levels were equalized. Inside the scanner, sounds were presented with an MR compatible audio system based on air tubes (Avotec; linear frequency transfer function up to ~4 kHz) at ~60 dB. Before starting the measurement, sounds were played to the subject and individual sound intensity was further adjusted to equalize their perceived loudness.

MRI. Data were acquired on a 7 T whole-body system driven by a Siemens console using a head gradient insert operating at up to 80 mT/m with a slew rate of 333 T/m/s. A head RF coil (single transmit, 16 receive channels) was used to acquire anatomical (T1 weighted) and functional (T2* weighted blood oxygenation level-dependent) images. T1 weighted (1 mm³) images were acquired using a modified magnetization-prepared rapid acquisition gradient echo sequence (TR = 2500 ms; TI = 1500 ms; TE = 3.67 ms). Proton density (PD) images were acquired together with the T1 weighted images using an interleaved acquisition. The PD images were used to minimize inhomogeneities in T1 weighted images (Van de Moortele et al., 2009). Acquisition time for anatomy was ~7 min. T2* weighted functional data were acquired using an echo planar imaging sequence in which time gaps were placed after the acquisition of each volume.

The main experiment was designed according to a fast event-related scheme. The acquisition parameters were as follows: TR = 2600 ms; TA = 1200 ms; TE = 30 ms; number of slices = 31; GRAPPA acceleration $\times 3$; partial Fourier 6/8; voxel size = $1.5 \times 1.5 \times 1.5$ mm³. Note that between subsequent acquisitions, there was a silent gap of 1.4 s during which the sounds were presented. Slices covered the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally. The experiment consisted of eight runs. Six of these runs were “training” runs, in which 144 sounds were presented three times in total (i.e., each sound was presented in half of the training runs). The two remaining runs were “testing” runs, in which 24 different sounds were presented six times in total (three times per run). Within each run, sounds were randomly spaced at a jittered interstimulus interval of 2, 3, or 4 TRs and presented, with additional random jitter, in the silent gap between acquisitions. Zero trials (trials where no sound was presented, 6% or 5% of the trials in training and

testing runs, respectively) and catch trials (trials in which the preceding sound was repeated, 6% and 3% of the trials in training and testing, runs respectively) were included. Subjects were instructed to perform an incidental task that was used to maintain attentional set, and were required to respond with a button press when a sound was repeated. Catch trials were excluded from the analysis. Each run lasted ~10 min.

The localizer was designed according to a blocked scheme. The acquisition parameters were as follows: TR = 3000 ms; TA = 1500 ms; TE = 30 ms; number of slices = 44; GRAPPA acceleration $\times 3$; partial Fourier 6/8; voxel size = $1.5 \times 1.5 \times 1.5$ mm³, silent gap = 1500 ms. Slices covered the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally. The localizer consisted of six runs, and in each run two blocks of each condition were presented. In each block, lasting 18 s, six sounds of the same condition were presented (one sound per TR, presented in the silent gap). Blocks of acoustic stimulation were separated from each other by 12 s of silence. Each run lasted ~9 min.

Functional and anatomical data were analyzed with BrainVoyager QX. Preprocessing consisted of slice scan-time correction (with sinc interpolation), temporal high-pass filtering (removing drifts of seven cycles or less per run for the main study and three cycles or less per run for the localizer), and 3D motion correction (trilinear/sinc interpolation; each volume was aligned to the first volume of run 1). Additional spatial smoothing (Gaussian kernel with full-width at half-maximum = 2 mm) and temporal smoothing (two consecutive data points) were applied to the localizer data only. Functional data were coregistered to the anatomical data and normalized in Talairach space (Talairach and Tournoux, 1988). This spatial normalization is a piecewise linear operation. Specifically, a linear transformation is applied to each of the 12 parts of the brain defined by manually selected anatomical landmarks (the anterior and posterior commissure and the anterior, posterior, superior, inferior, left, and right extreme points; standard BrainVoyager procedure). Functional data were resampled (with sinc interpolation) in the normalized space at a resolution of 1 mm isotropic. Anatomical volumes were also used to derive gray matter segmentations indicating the border between white and gray matter. Using this border, inflated hemispheres of the individual subjects were obtained.

Characterization of frequency tuning in human auditory cortex. The primary aim of this study was to characterize, quantitatively, the different forms of frequency tuning in auditory cortex. This analysis can be summarized in three stages, which are described below in detail (Fig. 1). We used a heuristic clustering analysis based upon the similarity of frequency tuning among different voxels (Stage 3, see below). This analysis used voxel-by-voxel estimates of spectral response profiles (see Stage 2) calculated from the responses to natural stimuli (see Stage 1).

Stage 1: extraction of sounds' frequency content. In Stage 1, we characterized the sounds used as stimuli in our experiment by their frequency content (Fig. 1, Stage 1). The sounds were filtered through a biologically plausible computational model of auditory processing from the cochlea to the midbrain (NSL Tools package, available at <http://www.isr.umd.edu/Labs/NSL/Software.htm>; Chi et al., 2005; Fig. 1, Stage 1). Within this model, sound waveforms are passed through a bank of 128 overlapping bandpass filters with equal width ($Q_{10dB} = 3$). This results in the mathematical representation of training sounds S in terms of an $N \times F$ matrix $\mathbf{W}$ of coefficients, where N = number of sounds, and F = the number of resulting frequency bins ($N = 144$ and $F = 128$; center frequencies were logarithmically spaced and ranged from 0.2 to 7 kHz).

Stage 2: computation of a voxel's spectral response profile. In Stage 2, we calculated the voxels' spectral profile based on their responses to natural sounds, using customized MATLAB code (www.mathworks.com; Fig. 1, Stage 2). We followed methodological procedures similar to the ones previously described for the analyses of visual responses to natural scenes (Kay et al., 2008b; Naselaris et al., 2009, 2011), which we have recently adapted and validated for the analysis of natural sounds (Moerel et al., 2012).

Based on the training data only, we calculated the voxel's spectral preference (matrix $\mathbf{R}$ [$F \times V$], where V = number of voxels) using the fMRI response matrix $\mathbf{Y}$ [$N \times V$] (Kay et al., 2008a) and the frequency representation of the sounds $\mathbf{W}$ [$N \times F$] obtained in Stage 1. For each

voxel j , its frequency profile R_j $[(F \times 1)]$ was obtained as the relevance vector machine solution to the linear problem:

$$Y_j = W \cdot R_j \quad (1)$$

where each element i of the vector R_j describes the contribution of the frequency bin i to the overall response of voxel j (Moerel et al., 2012). We performed this computation in fivefold cross-validation, by using only a subset of the training sounds in each fold to compute matrix R (10 sounds of the 144 training sounds were left out at random; per fold, R was computed on 134 of the training sounds). In that way, we obtained five estimates of each voxel's spectral profile (one for each of the five cross-validations). The spectral profile averaged across cross-validations was used as input to the clustering analysis. This cross-validation procedure served to obtain more stable estimates of the voxels spectral profiles.

Next, we assessed the general validity of these frequency tuning estimates by assessing their capability to predict fMRI responses to novel (testing) stimuli. These novel stimuli were the 24 sounds, which were presented in the two testing runs of the experiment (completely separate from the 144 training sounds, which were presented in the six training runs). Although this part of analysis is not necessary for our (clustering) analysis of functionally segregated auditory responses, it ensures that the spectral response profiles do not only account for the responses to the specific sounds used for their estimation but can also be used to describe fMRI responses to a different set of natural sounds. Specifically, we predicted the response $\hat{Y}_j$ to the 24 test sounds as follows:

$$\hat{Y}_j = W \cdot R_j \quad (2)$$

where W $[N \times F]$ is the frequency representation of the test sounds ($N = 24$). We evaluated prediction accuracy by concatenating $\hat{Y}_j$ across voxels resulting in predicted responses $\hat{Y}$ $[N \times V]$, and computing correlation matrix C $[N \times N]$ between predicted and measured responses to the test sounds. Next, for each sound i we obtained the rank r_i of $C_{i,:}$. Note that a rank of 1 indicates perfect prediction, while a rank of N would be the worst outcome. Prediction accuracy P_i of each sound i was defined as follows:

$$P_i = 1 - \left(\frac{r_i - 1}{N - 1} \right). \quad (3)$$

Values of P_i range between 0 and 1, with 1 being perfect prediction. Prediction accuracy was calculated as the average across N test sounds. To test whether this accuracy was significantly greater than zero, we compared it to the null distribution obtained using permutation testing. Specifically, we permuted V in R (number of permutations = 1000) and used this permuted matrix to predict the response to test sounds. The actual prediction accuracy values were considered to be significantly higher than chance if their values were within the upper 5% interval of the empirical null distribution. Note that in this analysis prediction accuracy is assigned as the accuracy of predicting the response to a novel sound based on the entire auditory cortex. Consequently, we only obtain one value of prediction accuracy per subject and cannot estimate how prediction accuracy varies across the auditory cortex (Moerel et al., 2012).

Figure 1. Encoding approach to estimate spectral response profiles. In the first stage, sound features were extracted as the output of a computational model mimicking early auditory processing (Chi et al., 2005). In the second stage, the measured brain responses Y to the sounds' spectral components W allowed estimating, through regularized regression, the spectral profiles R of all cortical locations (N = number of sounds; F = number of frequencies; V = number of locations/voxels). In the third stage, we divided the voxels into distinct clusters. We computed the autocorrelation function of each voxel's profile, and correlated voxels to each other to represent their similarity. The data-driven Louvain clustering algorithm provided a separation of this network into clusters, each consisting of a centroid and a map (see Materials and Methods).

Stage 3: extraction of spectral clusters. To characterize the complex, multipeaked spectral tuning of voxels throughout the auditory cortex (i.e., voxels' spectral tuning beyond their main frequency peak), we performed the following clustering analysis (Fig. 1, Stage 3). First, we computed the normalized autocorrelation function (autocorrelation at zero lag equal to 1) of each profile R_j (averaged across the five cross-validations). Note that by computing the profiles' autocorrelation function, we obtained a representation of spectral profiles that reflects spectral modulations and is independent of the frequency of the main peak. Second, we used a data-driven clustering algorithm to obtain the different "types" of spectral profiles present in the auditory cortex. There are several ways in which these separate clusters could have been identified. Commonly used approaches include the k -means clustering algorithm (Beckmann et al., 2009; Kim et al., 2010) and spectral clustering (Kelly et al., 2010). Here we used the so-called Louvain module detection algorithm, as implemented in the Brain Connectivity Toolbox (<http://www.brain-connectivity-toolbox.net>, an open-access MATLAB network analysis toolbox; Fig. 1, Stage 3; Blondel et al., 2008; Rubinov and Sporns, 2011). This algorithm stems from graph theory and has been applied in various contexts, such as in the characterization of biological

Figure 2. Group tonotopic maps. **A**, Anatomy of the left and right hemisphere superior temporal plane. Light/dark colors indicate the location of gyri and sulci, respectively. The abbreviations indicate the location of the first transverse sulcus (FTS), intermediate sulcus (SI), second Heschl's gyrus (HG2), Heschl's sulcus (HS), second Heschl's sulcus (HS), planum temporale (PT), superior temporal gyrus (STG), and superior temporal sulcus (STS). White dotted lines indicate the location of HG. **B**, Group maps of tonotopy, representing the mean across individual tonotopic maps aligned in fCBA space. The group map is shown for voxels that are included in ≥ 3 individual maps.

or social networks (Fortunato, 2010). Recently, it has been successfully applied to neuroimaging data (Barnes et al., 2010; Rubinov and Sporns, 2011; Goulas et al., 2012).

The Louvain algorithm has several advantages compared with more common clustering algorithms (such as *k*-means clustering). Three advantages were crucial for this study specifically. First, the algorithm does not require the a priori specification of the number of resulting clusters (as is required with for example the *k*-means algorithm or fuzzy clustering). This was important for the current study, as there was no clear evidence with regard to the number of clusters that should underlie the data. Second, in a recent study in which different clustering algorithms were compared, this algorithm was identified among the best performing ones (Lancichinetti and Fortunato, 2009). Finally, the algorithm is designed to analyze large networks in a fast and efficient manner. This quality is crucial to us, as we project many signals ($>20,000$ voxels per subject) on a small number of dimensions. As more commonly used methods (such as *k*-means clustering) lack these advantages, it is expected to be challenging for these algorithms to obtain similar results.

Our application of the Louvain algorithm starts with the calculation of a $V \times V$ pairwise correlation matrix of the autocorrelations of the voxels response profiles, where V is the number of auditory responsive voxels. Across subjects, V varied between $\sim 20,000$ and 25,000 voxels. The sparseness of this correlation matrix was set at 0.3 by sorting the correlation values and setting the lowest 70% to zero, and the resulting matrix contained only positive values. The Louvain algorithm considers the correlation matrix as a graph, where the V voxels are the nodes. The correlation between the autocorrelation profiles of the i th and j th voxels is an undirected edge (i.e., a weighted connection) between the i th and j th node. Modules (clusters) within the correlation matrix are assumed to have more edges within the module (i.e., higher correlation) than expected if the nodes were randomly connected. This is quantified by modularity value Q (Newman, 2006):

$$Q = \sum_{c=1}^{n_c} \left[\frac{e_c}{m} - \left(\frac{d_c}{2m} \right)^2 \right], \quad (4)$$

where n_c is the number of clusters, m is the total degree (number of edges) of the graph, e_c is the number of edges joining the nodes within cluster c , and d_c is the total number of edges of the nodes belonging to cluster c . A good cluster should have a higher value for the first fraction than for the second fraction. The algorithm seeks to maximize the modularity value Q .

Several assumptions accompany our implementation of the Louvain clustering analysis. First of all, we apply the clustering algorithm to a $V \times V$ correlation matrix, computed by correlating the autocorrelation profile of each voxel with that of every other voxel. As a measure of correlation, Pearson's correlation coefficient was used. As such, we can only estimate linear relations between nodes of the graph. Second, all algorithms that maximize modularity suffer from the resolution limit (Fortunato and Barthélemy, 2007), which is the failure to identify clusters smaller than a minimum scale. Although the Louvain algorithm has been shown to be less sensitive to this confound than most modularity maximization algorithms, we may be biased to the detection of large clusters.

As the output of this algorithm is stochastic, it finds a slightly different solution from run to run. To evaluate the stability of the obtained solutions, we ran the algorithm multiple times ($N = 100$ repetitions). Additionally, we evaluated the output of the Louvain algorithm after randomizing the correlation matrix (by randomizing the phase of the Fourier transformed matrix, and computing the inverse Fourier transform; $N = 100$ repetitions; Zalesky et al., 2012). The current choice of N repetitions was a trade-off between manageable computation times and obtaining a reasonable amount of information regarding stability of obtained results. We evaluated the output of the Louvain algorithm by the modularity of the obtained subdivision, its stability in number of resulting clusters across runs, and its stability across runs (measured as estimated mutual information across repetitions), both when computed on the data and on the randomized network (Rubinov and Sporns, 2011).

After this quality check on the output of the Louvain algorithm, we computed the characteristic profile (i.e., the centroid) of each resulting cluster as follows. We normalized each voxel's spectral profile by expressing all amplitude and frequency values as ratios with respect to its amplitude maximum and the frequency at that maximum. Thus, the abscissa and ordinate became, respectively, $F_n = F/F_{\max}$ and $A_n = A/A_{\max}$, where A_n and F_n were the normalized values (Schwartz et al., 2003). F_n ranged between $0.1 \cdot F_{\max}$ and $10 \cdot F_{\max}$, divided in 134 logarithmically spaced bins. By normalizing the spectral profiles in this manner, both their maxima and additional frequency peaks below and above the maxima were aligned across voxels. Each cluster's centroid was computed by averaging the normalized spectral profiles of the voxels in that cluster.

To compute group clustering results, we first ascertained which clusters corresponded to each other across subjects. To that end, we correlated the centroids of each subject to the centroids of subject 1. Corresponding clusters across subjects were identified as those clusters whose centroids correlated highest (mean [SD] correlation between matching and non-matching centroids = $0.99 [9.7 \cdot 10^{-3}]$ and $0.78 [0.05]$, respectively). Group centroids were obtained by averaging matching centroids across subjects. At each frequency bin, values of the group centroids were tested for significant deviation from zero using a one-sample *t* test (five observations per frequency, where each observation is data from a subject). Resulting significance values were corrected for multiple comparisons (i.e., the number of frequency bins and number of group centroids) using FDR correction.

Cluster spatial maps were obtained as follows. We computed the correlation of each voxel's autocorrelation profile to the corresponding clusters' centroid (hard clustering; each voxel belongs to one cluster only), reflecting the amount to which that voxel represented the cluster. Individual subject maps were created by smoothing maps on the surface and thresholding the maps ($r > 0.55$; cluster threshold > 10). We investigated the lateralization of each cluster by comparing its size in the left hemisphere to the size in the right hemisphere. Differences in cluster size across hemispheres were tested for significance by performing a paired

Figure 3. Individual tonotopic maps. Tonotopic maps for each of the five subjects, extracted as the maximum of the voxels' spectral profiles. Red and blue colors indicate regions preferring low and high frequencies, respectively. The maps in the lower right corner show the consistency in tonotopic pattern across subjects. White dotted lines indicate the location of HG.

samples t test across the five subjects. Next, the spatial layout of individual subject maps was quantified by dividing each hemisphere in five anatomically defined subregions [Heschl's gyrus (HG), planum polare (PP), planum temporale (PT), rostral and caudal superior temporal gyrus (rSTG and cSTG), respectively] based on Kim et al., 2000; see Fig. 7A). We computed the proportion of each cluster located in each of the anatomical regions. Finally, group maps were created by transforming individual maps into functional informed cortex-based alignment (fCBA) space (see below), averaging matching maps across subjects (threshold on single subject maps $r > 0.60$), and smoothing resulting maps on the group surface (repeat value = 1).

Computation of tonotopic maps. Per cross-validation, we obtained a tonotopic map by considering the characteristic frequency of a voxel as the frequency corresponding to the maximum of the coefficients in R_j . These maps were averaged across cross-validations to create one map of tonotopy per subject. Tonotopic cortical maps were obtained by logarithmic mapping of best-frequency values to a red-yellow-green-blue color scale (Figs. 2, 3).

We explored the variability in the tonotopic pattern across subjects by bringing each individual tonotopy map in a normalized space (fCBA space, see below) and normalizing values in each map between 0 and 1. Next, for each combination of subjects, for each voxel j , we computed the normalized difference $diff_j$ as follows:

$$diff_j = \frac{|subj_{s,j} - subj_{s+1,j}|}{subj_{s,j} + subj_{s+1,j}}. \quad (5)$$

A maximum difference in frequency between the two maps would result in $diff_j = 1$. The group tonotopic variability was obtained as the median across all subject pairs.

From the localizer, tonotopic maps were extracted in the following manner. Using a single subject general linear model analysis with a standard hemodynamic response model (Friston et al., 1995), we computed the responses to the three center frequencies (.5, 0.1.5, and 2.5 kHz) in all six runs separately. Voxels that showed a significant response to the sounds were selected ($Q[FDR] < 0.05$), and the response to the three tones was z -normalized across these voxels. For each voxel, its best frequency was determined in sixfold cross-validation (one run was left out in each fold). If the estimated best frequency had a majority across folds (three or more occurrences), the voxel was color coded accordingly.

Functional cortex-based alignment. Alignment across subjects for the purpose of computing group maps is particularly challenging in the auditory cortex, due to considerable interindividual differences in macro-anatomical landmarks and related functional responses (Da Costa et al., 2011; Penhune et al., 1996). Here, we optimized the alignment across subjects by performing fCBA. fCBA complements standard CBA (Goebel et al., 2006; Frost and Goebel, 2012) by allowing individually defined functional regions, in addition to the major sulci and gyri, to drive the across-subject alignment.

We based fCBA on tonotopic maps obtained from the localizer. In each subject and hemisphere, we delineated the low-frequency region consistently present in the vicinity of Heschl's gyrus as region of interest. fCBA was partially driven by this functional region (weighting decreased over itera-

Figure 4. Representative spectral profiles. Response profiles, i.e., the voxels' sensitivity to frequencies spanning 5.3 octaves (~ 0.2 to 7 kHz), of four representative voxels per cluster are shown (data of subject 5). Examples of the (A) broadly tuned, single peaked cluster; (B) cluster with simple tuning and pronounced inhibitory sidebands surrounding the main frequency peak; (C) multipeaked cluster with variable distance between frequency peaks; (D) multipeaked octave cluster; and (E) multipeaked harmonic cluster.

tions), and partially by anatomical information (weighting increased over iterations). The resulting alignment information was used for calculating and displaying group tonotopic maps and cluster group maps (see below) as obtained from the main study. Note that fCBA was based on the localizer data only, and therefore does not bias the interpretation of the group results of the main experiment.

Results

Estimation of voxels' spectral profiles and extraction of tonotopic maps

As expected, sounds evoked significant activation in a large expanse of the superior temporal cortex. The activated region included early auditory areas along the HG and surrounding regions on the PP, PT, STG, and parts of the superior temporal

sulcus (STS). Based on the estimated voxels' spectral profiles (see Materials and Methods), responses to new sounds could be predicted significantly above chance in each subject ($p < 0.01$ in each subject; prediction accuracy for subject 1: subject 5 = 0.68, 0.67, 0.76, 0.65, and 0.76; mean prediction accuracy [SD] = 0.70 [0.05]), illustrating that these profiles accurately characterize the voxels' frequency tuning.

Based on the maximum of each voxel's spectral tuning curve, tonotopic maps were extracted across subjects and hemispheres (group and single subject maps are displayed in Figs. 2, 3, respectively). Consistent with results of previous studies, we observed a large low-frequency region near HG, bordered anteriorly (on HG and first transverse sulcus, FTS) and posteriorly (on Heschl's sulcus, HS, and anterior PT) by regions preferring higher frequencies (Formisano et al., 2003; Humphries et al., 2010; Da Costa et al., 2011; Moerel et al., 2012). Beyond Heschl's region, on PP, PT, and STG/STS additional frequency clusters were present (Figs. 2, 3). These regions may reflect additional frequency gradients, possibly indicating the location of belt and parabelt auditory fields [see Moerel et al. (2012) for a thorough discussion on this hypothesis]. Tonotopic maps were most stable on HG and regions on PT, and varied most on the lateral extremities of the map (i.e., STG; see lower right maps in Fig. 3). Furthermore, maps were more stable in the left than the right hemisphere (median normalized difference of left and right hemisphere = 0.26 and 0.34, respectively).

Tonotopic maps reflect only the main peak in the voxels' spectral profiles, and thereby represent well those voxels that displayed a simple, single-peaked frequency profile (for example, Fig. 4A). However, most voxels had a more complex multipeaked frequency preference (Fig. 4C–E). Multipeaked voxels were present both on HG (Fig. 4C, second column) and beyond HG and on PT and STG/STS.

Cluster extraction with Louvain algorithm

To examine the spectral profiles beyond their main tonotopic peak, we used a data-driven algorithm to divide the voxels into separate clusters according to their type of spectral tuning (Fig. 1, Stage 3; Blondel et al., 2008; Rubinov and Sporns, 2011). The algorithm's solution had a high modularity in each subject (mean[SE] = 0.25[$2.0 \cdot 10^{-3}$]), a stable number of extracted clusters across repeated runs of the algorithm (median[SE] = 5[0.02]), and high estimated mutual information across resulting partitions (mean[SE] = 0.87[0.03]). When running the same algorithm on a randomized network, modularity decreased (mean[SE] = 0.19[0.02]), variability in number of extracted clusters increased (median[SE] = 3[1.09]), and estimated mutual information

across resulting partitions decreased ($\text{mean}[\text{SE}] = 0.17[3.0 \cdot 10^{-3}]$), ensuring that the clustering algorithm provided a satisfactory partitioning.

Cluster centroids: characteristic tuning profiles

The algorithm divided auditory cortical voxels into five clusters, based on their spectral tuning. Each of the five clusters contained, respectively, 19.9, 17.8, 19.0, 16.7, and 26.6% of total number of voxels (average across subjects). Resulting characteristic profiles (i.e., centroids) of each resulting cluster were highly consistent across subjects ($\text{mean}[\text{SD}]$ correlation between matching and non-matching centroids = $0.99[9.7 \cdot 10^{-3}]$ and $0.78[0.05]$, respectively). The first centroid (Fig. 5A, left column) included voxels with broadly tuned frequency profiles (e.g., voxels in Fig. 4A). The second centroid (Fig. 5B, left column) described a cluster with significant positive peaks around $0.25 \cdot F_{\text{max}}$ and $4 \cdot F_{\text{max}}$, corresponding to a distance of two octaves (i.e., three harmonics) between the maximum frequency and additional peaks. Voxels included in this cluster often showed one distinct frequency peak flanked by pronounced attenuation sidebands (e.g., voxels in Fig. 4B). The third centroid (Fig. 5C, left column) displayed significant positive frequency peaks around $0.35 \cdot F_{\text{max}}$ and $3 \cdot F_{\text{max}}$, and additionally around $0.1 \cdot F_{\text{max}}$ and $9 \cdot F_{\text{max}}$. Thus, this centroid displayed sensitivity to additional frequency bands at a distance of two harmonics from the maximum frequency. Voxels in this cluster often had multipeaked spectral profiles, with no clear relationship between the peaks (e.g., voxels in Fig. 4C). The fourth centroid (Fig. 5D, left column) had significant positive peaks located at 0.25 , 0.5 , and additionally at $2 \cdot F_{\text{max}}$. Voxels clustered into this cluster were selective to frequency bands at octave interval (e.g., voxels in Fig. 4D). Thus, the voxels belonging to this fourth cluster were not sensitive to all harmonic distances, but instead displayed a specific selectivity to octave frequency lags. We refer to this cluster as the “octave cluster.” Finally, the fifth centroid (Fig. 5E, left column) displayed a large number of localized significant positive peaks, at 0.35 , 0.5 , 0.65 , 1.5 , 2 , 3 , 4 , 5 , and $7 \cdot F_{\text{max}}$. Interestingly, voxels belonging to this cluster were sensitive to additional frequency bands at multiple harmonics of the lowest frequency peak. For example, the voxels in the first and second column of Figure 4E display sensitivity to $0.2/0.4/0.6$ kHz and to $0.5/1.0/1.5$ kHz, respectively. We refer to this cluster as the “harmonic cluster.”

Cluster spatial maps

After investigating the five centroids, we explored the cortical location of each cluster at group level and in individual subjects

Figure 5. Group cluster maps. Centroids (left) and group maps (right) characterizing the five extracted clusters (rows A–E consecutively display the broad cluster, cluster with attenuation bands, multipeaked cluster with no clear relationship between peaks, octave cluster, and harmonic cluster). The centroids’ main peak represents the main frequency peak across voxels, and additional peaks show the presence of sensitivity to additional frequency bands at consistent spectral intervals. Positive and negative significant deviations from zero (FDR corrected for multiple comparisons) are color coded in green and blue, respectively. The maps show the correlation of each voxel’s profile to the clusters’ centroid, averaged across subjects. Maps are displayed on an fCBA-based reconstruction of the temporal cortex, and white dotted lines indicate the location of HG.

(Figs. 5, right column, Fig. 6, Fig. 7). As we did not observe any significant lateralization effects at group level, the spatial patterns in the left and right hemisphere are described together. At group level, we observed that the first cluster (Figs. 5A, 6, green) occupied posterior primary auditory cortex (PAC; Fig. 6, black dotted circles) and cSTG. The second cluster (Figs. 5B, 6, yellow) occupied parts of medial HG, medial HS, and PT. The third cluster (Figs. 5C, 6, red) was located on HG (occupying the lateral part of PAC). The octave cluster (Figs. 5D, 6, blue) was located lateral to PAC, occupying the STG. Finally, the harmonic cluster was located anterolateral to PAC (Figs. 5E, 6, purple), and occupied a small region on PP.

The spatial distribution of the clusters in individual subjects was quite variable (compare maps in Fig. 6), reflected in the

Figure 6. Group and individual cluster maps. Resulting cluster maps as extracted from each subject are shown. Additionally, group cluster maps are displayed. The colors show maps belonging to clusters 1–5 in green, yellow, red, blue, and purple, respectively (i.e., octave and harmonic cluster shown in blue and purple, respectively). Maps are smoothed on the surface (repeat value = 1). Consequently maps are thresholded based on the voxels' correlation to the centroid (>0.82 / >0.55 for group and individual subject maps, respectively) and cluster thresholded (minimum threshold = 10 voxels). White dotted lines indicate the location of HG. Black dotted lines indicate the location of the PAC, identified as the main high–low–high tonotopic gradient in the proximity of HG in individual tonotopic maps.

histograms of Figure 7 by the spread of voxels in each cluster across anatomical regions. However, the main peaks in these histograms followed the pattern observed in the group maps. That is, the first cluster occupied the posterior locations PT and cSTG. The second cluster was localized in medial locations, occupying regions in PP, HG, and PT. The only exception to the group pattern was seen for the third cluster (Fig. 6, red), which showed a peak in the histograms within rSTG additional to the peak in HG. The octave and harmonic cluster occupied the STG (both rostral and caudal) and PP, respectively, in accordance with the patterns observed in the group maps (compare group map in Fig. 6 with Fig. 7).

Discussion

In this study, we used ultra-high field fMRI to extract neuronal populations' spectral profiles based on their responses to natural sounds. With a data-driven analysis, we identified five clusters of spectral tuning types. Two of these clusters displayed simple single-peaked spectral profiles, containing a broadly or narrowly tuned frequency peak. However, ~60% of neuronal populations throughout auditory cortex (divided between the three remaining clusters) displayed complex frequency selectivity beyond their main tonotopic frequency peak. We identified neuronal populations with sensitivity to multiple frequency bands (1) at exactly one octave distance from each other, (2) at harmonically related frequency intervals, and (3) with no apparent relation-

ship. We propose that beyond the well known tonotopic organization of the auditory cortex, this multipeaked spectral tuning concurs to define the representation space of natural sounds in the human auditory cortex.

Characterization of cortical spectral tuning based on responses to natural sounds

By modeling the fMRI responses in terms of the spectral content of many natural sounds, we calculated the voxels' frequency profiles. These frequency profiles were subsequently used to extract clusters and centroids. Using natural sounds for feature mapping (i.e., estimating the voxels' response profile) has several advantages over the use of artificial sounds. First, as natural sounds inherently engage auditory cortical neurons in meaningful and behaviorally relevant processing, they may be optimal for studying the functional architecture of higher order auditory areas. Second, previous studies showed that response profiles estimated with artificial sounds do not predict responses to natural sounds well (Machens et al., 2004; Bitterman et al., 2008). Consequently, the use of natural sounds was advised in mapping response features (Theunissen et al., 2000). Third, the method implemented here allows the simultaneous estimation of the voxels' feature preference (e.g., tuning to multiple frequency peaks). As neuronal populations show nonlinearities in their feature tuning, such simultaneous estimation is of paramount importance.

Figure 7. Spatial distribution of clusters. **A**, Anatomy of the left and right hemisphere superior temporal plane (left) and subdivision into five regions (based on Kim et al., 2000; right). Overlap in defined regions across three to five subjects is shown in a range of dark to light hues. **B**, HG, PP, PT, rSTG, and cSTG are indicated in blue, red, green, yellow, and orange, respectively. For each cluster, the proportion of voxels in each of the anatomical regions is shown. Clusters 1–5 consecutively display the broad cluster, cluster with attenuation bands, multipeaked cluster with no clear relationship between peaks, octave cluster, and harmonic cluster. Error bars indicate SE across subjects.

By mapping the main peak of the voxels' spectral profiles, we showed frequency selective regions throughout the superior temporal cortex. In agreement with previous fMRI studies, we observed a large low-frequency area in the vicinity of the HG, surrounded anteriorly and posteriorly by regions preferring higher frequencies (Formisano et al., 2003; Talavage et al., 2004; Da Costa et al., 2011). Although the exact relationship between these main frequency gradients and auditory fields remains debated (Humphries et al., 2010; Langers and van Dijk, 2012), these regions most likely reflect the location of the human primary auditory cortex (i.e., the homologs of monkey primary fields AI and R; Kosaki et al., 1997; Hackett et al., 1998). Beyond these main frequency gradients, we observed frequency selective clusters in regions that, based on their underlying anatomy, probably reflect belt and parabelt cortex (PP, PT, and STG/STS). Tonotopic maps in these regions are less frequently reported, yet these results replicate findings of at least two previous studies (Striemi-Amit et al., 2011; Moerel et al., 2012).

Types of multipeaked spectral tuning

In a data-driven manner, we identified five clusters of spectral tuning types. Voxels within two of these clusters displayed an overall simple tuning profile, including voxels with a broad frequency peak, and more narrowly tuned voxels with attenuating sidebands. This type of spectral tuning may reflect the traditionally reported neuronal populations with single peaks. Voxels belonging to these clusters were found throughout the superior temporal plane. Functionally, this spectral tuning might be rele-

vant for capturing the overall frequency content of incoming sounds.

Beyond these simple spectral tuning types, we observed three clusters (~60% of neuronal populations throughout auditory cortex) that displayed complex frequency selectivity beyond their main tonotopic frequency peak. Specifically, we observed a cluster in which voxels displayed selectivity to multiple octave frequency lags. Based on their spectral profiles, we predict that these octave-tuned voxels, clustered most densely along the STG, respond in similar manner to tones with frequencies at a 2:1 ratio. Consequently, they could elicit the percept of octave generalization. A hard-wired octave representation in the brain is in accordance with the widespread occurrence, early onset, and generalization beyond the human species of octave perception (Demany and Armand, 1984; Wright et al., 2000; Randel, 2003). Moreover, our findings of octave-based tuning are in accordance with invasive recordings in various mammals, which observed octave tuning at neuronal level (Brosch et al., 1999; Kadia and Wang, 2003; Noreña et al., 2008; Brosch and Schreiner, 2000).

Functional relevance of neuronal sensitivity to harmonic structure

Next, we observed a complex cluster with spectral tuning that finely followed sensitivity to multiple harmonic lags. As the natural sounds most important to humans (i.e., vocalizations, music) are harmonically structured (Ross et al., 2007; Noreña et al., 2008), these multipeaked spectral profiles may be functionally relevant for amplifying the spectral content of those sounds spe-

cifically. In everyday life, we are constantly exposed to sounds from different sources, whose spectral components overlap with each other. To perceive sounds we must segregate them into separate auditory objects, a task that is referred to as “auditory scene analysis” (Bregman, 1990). We accomplish this task by grouping components of the same source together based on various fundamental features of complex sounds, such as harmonicity (Darwin, 1997). For example, a stack of harmonically related frequencies is perceptually fused into a single tone (Michey and Oxenham, 2010; Borchert et al., 2011). The multipeaked cortical neuronal populations could underlie the perceptual mechanism of fusing harmonically related components. That is, the harmonic tuning within this spectral cluster could group sound components for the purpose of object segregation, or significantly enhance harmonic components relative to background noise without harmonic structure (Kadia and Wang, 2003).

Finally, voxels in the third multipeaked cluster, most densely clustered on the middle region on the PAC, displayed additional sensitivity to multiple frequency bands with no clear relationship to each other. The clusters’ centroid displayed sensitivity to frequency bands at 1.6 octaves from the main peak, consistent with the main peaks’ second harmonic. At present, we cannot identify an acoustic or behavioral correlate of the multiple frequency bands to which the voxels in this cluster are sensitive.

Neuronal substrate underlying complex population tuning

Using fMRI, we observed spectral tuning at the level of neuronal populations. Various types of spectral tuning at the level of individual neurons could underlie our observations. At present, we cannot draw conclusions regarding the nature of the sensitivity to multiple frequency bands that we observe in the human auditory cortex. First, the observed sensitivity to multiple frequency bands may reflect true multipeaked neuronal tuning, as has been observed in a number of studies and species (e.g., bat, Fitzpatrick et al., 1993; cat, Noreña et al., 2008; and monkey, Kadia and Wang, 2003). Second, the sensitivity to multiple frequency bands could result from combination-sensitive neuronal mechanisms (Wang et al., 2005; Sadagopan and Wang, 2009). As opposed to truly multipeaked neurons, combination-sensitive neurons only display sensitivity to multiple frequencies when probed with that exact frequency combination (nonlinear response mechanism). Third, as each measured voxel in the current study included thousands of neurons, we may observe the complex average spectral profile of many simple, single-peaked neurons. In that case, the observed multipeaked tuning would only emerge at a population level of neuronal responses.

To advance our understanding regarding the functional relevance of the cortical spectral sound representation, a well controlled exploration of nonlinear and combination-sensitive tuning to the multiple frequency bands within these profiles is needed. Furthermore, as previous studies show that feature tuning in auditory cortex is highly affected by changes in task, context, and attention (Fritz et al., 2003; Atiani et al., 2009), the challenge for future work is to investigate changes in spectral profiles during these manipulations. Our results can guide future explorations of complex spectral tuning in human auditory cortex, and the methods described here provide a means for this endeavor.

References

- Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA (2009) Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron* 61:467–480. [CrossRef Medline](#)

- Barnes KA, Cohen AL, Power JD, Nelson SM, Dosenbach YB, Miezin FM, Petersen SE, Schlaggar BL (2010) Identifying basal ganglia divisions in individuals using resting-state functional connectivity in MRI. *Front Syst Neurosci* 4:18. [CrossRef Medline](#)
- Beckmann M, Johansen-Berg H, Rushworth MF (2009) Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J Neurosci* 29:1175–1190. [CrossRef Medline](#)
- Bitterman Y, Mukamel R, Malach R, Fried I, Nelken I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451:197–201. [CrossRef Medline](#)
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* P10008.
- Borchert EM, Michéyl C, Oxenham AJ (2011) Perceptual grouping affects pitch judgements across time and frequency. *J Exp Psychol Hum Percept Perform* 37:257–269. [CrossRef Medline](#)
- Bregman AS (1990) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: MIT.
- Brosch M, Schreiner CE (2000) Sequence sensitivity of neurons in cat primary auditory cortex. *Cereb Cortex* 10:1155–1167. [CrossRef Medline](#)
- Brosch M, Schulz A, Scheich H (1999) Processing of sound sequences in macaque auditory cortex: response enhancement. *J Neurophysiol* 82:1542–1559. [Medline](#)
- Chi T, Ru P, Shamma SA (2005) Multiresolution spectrotemporal analysis of complex sounds. *J Acoust Soc Am* 118:887–906. [CrossRef Medline](#)
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RS, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl’s gyrus. *J Neurosci* 31:14067–14075. [CrossRef Medline](#)
- Darwin CJ (1997) Auditory grouping. *Trends Cogn Sci* 1:327–333. [CrossRef Medline](#)
- deCharms RC, Blake DT, Merzenich MM (1998) Optimizing sound features for cortical neurons. *Science* 280:1439–1443. [Medline](#)
- Demany L, Armand F (1984) The perceptual reality of tone chroma in early infancy. *J Acoust Soc Am* 76:57–66. [CrossRef Medline](#)
- Fitzpatrick DC, Kanwal JS, Butman JA, Suga N (1993) Combination-sensitive neurons in the primary auditory cortex of the mustached bat. *J Neurosci* 13:931–940. [Medline](#)
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859–869. [CrossRef Medline](#)
- Fortunato S (2010) Community detection in graphs. *Phys Rep* 486:75–174. [CrossRef](#)
- Fortunato S, Barthélemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci U S A* 104:36–41. [CrossRef Medline](#)
- Friston K (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360:815–836. [Medline](#)
- Friston KJ, Frith CD, Turner R, Frackowiak RS (1995) Characterizing evoked hemodynamics with fMRI. *Neuroimage* 2:157–165. [CrossRef Medline](#)
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223. [CrossRef Medline](#)
- Frost MA, Goebel R (2012) Measuring structural-functional correspondence: spatial variability of specialized brain regions after macro-anatomical alignment. *Neuroimage* 59:1369–1381. [Medline](#)
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27:392–401. [CrossRef Medline](#)
- Goulas A, Uylings HB, Stiers P (2012) Unravelling the intrinsic functional organization of the human lateral frontal cortex: a parcellation scheme based on resting state fMRI. *J Neurosci* 32:10238–10252. [CrossRef Medline](#)
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivision of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475–495. [CrossRef Medline](#)
- Humphries C, Liebenthal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50:1202–1211. [CrossRef Medline](#)
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multipeaked tuning characteristics. *J Neurophysiol* 89:1603–1622. [Medline](#)
- Kay KN, David SV, Prenger RJ, Hansen KA, Gallant JL (2008a) Modeling

- low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. *Hum Brain Mapp* 29:142–156. [CrossRef Medline](#)
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008b) Identifying natural images from human brain activity. *Nature* 452:352–355. [CrossRef Medline](#)
- Kelly C, Uddin LQ, Shehzad Z, Margulies DS, Castellanos FX, Milham MP, Petrides M (2010) Broca's region: linking human brain functional connectivity data and non-human primate tracing anatomy studies. *Eur J Neurosci* 32:383–398. [CrossRef Medline](#)
- Kim JH, Lee JM, Jo HJ, Kim SH, Lee JH, Kim ST, Seo SW, Cox RW, Na DL, Kim SI, Saad ZS (2010) Defining functional SMA and pre-SMA subregions in human MFC using resting state fMRI: functional connectivity-based parcellation method. *Neuroimage* 49:2375–2386. [CrossRef Medline](#)
- Kim JJ, Crespo-Facorro B, Andreasen NC, O'Leary DS, Zhang B, Harris G, Magnotta VA (2000) An MRI-based parcellation methods for the temporal lobe. *Neuroimage* 11:271–288. [CrossRef Medline](#)
- King AJ, Nelken I (2009) Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nat Neurosci* 12:698–701. [CrossRef Medline](#)
- Kosaki H, Hashikawa T, He J, Jones EG (1997) Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in macaque monkeys. *J Comp Neurol* 386:304–316. [CrossRef Medline](#)
- Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. *Phys Rev E Stat Nonlin Soft Matter Phys* 80:056117. [CrossRef Medline](#)
- Langers DR, van Dijk P (2012) Mapping the tonotopic organization of the human auditory cortex with minimally salient acoustic stimulation. *Cereb Cortex* 22:2024–2038. [CrossRef Medline](#)
- Machens CK, Wehr MS, Zador AM (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24:1089–1100. [CrossRef Medline](#)
- Merzenich MM, Brugge JF (1973) Representation of the cochlear partition on the superior temporal plane of the macaque monkey. *Brain Res* 50:275–296. [CrossRef Medline](#)
- Merzenich MM, Knight PL, Roth GL (1973) Representation of cochlea within primary auditory cortex in the cat. *Brain Res* 63:343–346. [CrossRef Medline](#)
- Micheyl C, Oxenham AJ (2010) Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. *Hear Res* 266:36–51. [CrossRef Medline](#)
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32:14205–14216. [CrossRef Medline](#)
- Naselaris T, Prenger RJ, Kay KN, Oliver M, Gallant JL (2009) Bayesian reconstruction of natural images from human brain activity. *Neuron* 63:902–915. [CrossRef Medline](#)
- Naselaris T, Kay KN, Nishimoto S, Gallant JL (2011) Encoding and decoding in fMRI. *Neuroimage* 56:400–410. [CrossRef Medline](#)
- Newman ME (2006) Modularity and community structure in networks. *Proc Natl Acad Sci U S A* 103:8577–8582. [CrossRef Medline](#)
- Noreña A, Górvitch B, Pienkowski M, Shaw G, Eggermont JJ (2008) Increasing spectrotemporal sound density reveals an octave-based organization in cat primary auditory cortex. *J Neurosci* 28:8885–8896. [CrossRef Medline](#)
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex* 6:661–672. [CrossRef Medline](#)
- Randel DM ed (2003) *The Harvard dictionary of music*, Ed 4. Cambridge, MA: Harvard UP.
- Ross D, Choi J, Purves D (2007) Musical intervals in speech. *Proc Natl Acad Sci U S A* 104:9852–9857. [CrossRef Medline](#)
- Rubinov M, Sporns O (2011) Weight-conserving characterization of complex functional brain networks. *Neuroimage* 56:2068–2079. [CrossRef Medline](#)
- Sadagopan S, Wang X (2009) Nonlinear spectrotemporal interactions underlying selectivity for complex sounds in auditory cortex. *J Neurosci* 29:11192–11202. [CrossRef Medline](#)
- Schwartz DA, Howe CQ, Purves D (2003) The statistical structure of human speech sounds predicts musical universals. *J Neurosci* 23:7160–7168. [CrossRef Medline](#)
- Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One* 6:e17832. [CrossRef Medline](#)
- Sutter ML, Schreiner CE (1991) Physiology and topography of neurons with multi-peaked tuning curves in cat primary auditory cortex. *J Neurophysiol* 65:1207–1226. [CrossRef Medline](#)
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical.
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91:1282–1296. [CrossRef Medline](#)
- Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20:2315–2331. [CrossRef Medline](#)
- Van De Moortele PF, Auerbach EJ, Olman C, Yacoub E, Uğurbil K, Moeller S (2009) T1 weighted brain images at 7 tesla unbiased for proton density, T2 contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. *Neuroimage* 46:432–446. [CrossRef Medline](#)
- Wang X (2007) A sharper view from the top. *Nat Neurosci* 10:1509–1511. [CrossRef Medline](#)
- Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol* 74:2685–2706. [CrossRef Medline](#)
- Wang X, Lu T, Snider RK, Liang L (2005) Sustained firing in auditory cortex evoked by preferred stimuli. *Nature* 435:341–346. [CrossRef Medline](#)
- Winkler I, Denham SL, Nelken I (2009) Modeling the auditory scene: predictive regularity representations and perceptual objects. *Trends Cogn Sci* 13:532–540. [CrossRef Medline](#)
- Woolley SM, Fremouw TE, Hsu A, Theunissen FE (2005) Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8:1371–1379. [CrossRef Medline](#)
- Wright AA, Rivera JJ, Hulse SH, Shyan M, Neiwirth JJ (2000) Music perception and octave generalization in rhesus monkeys. *J Exp Psychol Gen* 129:291–307. [CrossRef Medline](#)
- Zalesky A, Fornito A, Bullmore E (2012) On the use of correlation as a measure of network connectivity. *Neuroimage* 60:2096–2106. [CrossRef Medline](#)