Cerebral Cortex, 2018; 1–15

doi: 10.1093/cercor/bhy243
Original Article

ORIGINAL ARTICLE

Homology and Specificity of Natural Sound-Encoding in Human and Monkey Auditory Cortex

Julia Erb [1,2,8], Marcelo Armendariz[3], Federico De Martino[1,2], Rainer Goebel[1,2], Wim Vanduffel[3,4,5,6] and Elia Formisano[1,2,7]

[1]Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD Maastricht, The Netherlands, [2]Maastricht Brain Imaging Center (MBIC), 6200 MD Maastricht, The Netherlands, [3]Laboratorium voor Neuro-en Psychofysiologie, KU Leuven, 3000 Leuven, Belgium, [4]MGH Martinos Center, Charlestown, MA 02129, USA, [5]Harvard Medical School, Boston, MA 02115, USA, [6]Leuven Brain Institute, 3000 Leuven, Belgium, [7]Maastricht Center for Systems Biology (MaCSBio), 6200 MD Maastricht, The Netherlands and [8]Department of Psychology, University of Lübeck, 23562 Lübeck, Germany

Address correspondence to Julia Erb, Maria-Goeppert-Str. 9a, 23562 Lübeck, Germany. Email: julia.erb@maastrichtuniversity.nl orcid.org/0000-0002-3440-7269

Julia Erb, Marcelo Armendariz and Federico De Martino are Co-first authors
Wim Vanduffel and Elia Formisano are Co-senior authors

Abstract

Understanding homologies and differences in auditory cortical processing in human and nonhuman primates is an essential step in elucidating the neurobiology of speech and language. Using fMRI responses to natural sounds, we investigated the representation of multiple acoustic features in auditory cortex of awake macaques and humans. Comparative analyses revealed homologous large-scale topographies not only for frequency but also for temporal and spectral modulations. In both species, posterior regions preferably encoded relatively fast temporal and coarse spectral information, whereas anterior regions encoded slow temporal and fine spectral modulations. Conversely, we observed a striking interspecies difference in cortical sensitivity to temporal modulations: While decoding from macaque auditory cortex was most accurate at fast rates (> 30 Hz), humans had highest sensitivity to ~3 Hz, a relevant rate for speech analysis. These findings suggest that characteristic tuning of human auditory cortex to slow temporal modulations is unique and may have emerged as a critical step in the evolution of speech and language.

Key words: functional MRI, primate auditory cortex, rhesus macaque, spectrotemporal modulations, tonotopy

Introduction

How does the cortical processing of sounds compare between human and nonhuman primates? Previous studies have assessed the functional organization of the primate auditory cortex using animal electrophysiology (Merzenich and Brugge 1973; Morel et al. 1993; Kosaki et al. 1997; Rauschecker 1997; Bendor and Wang 2008; Kusmierek and Rauschecker 2009) and more recently fMRI in humans (Formisano et al. 2003; Talavage et al. 2004; Humphries et al. 2010; Woods et al. 2010; Da Costa et al. 2011; Striem-Amit et al. 2011; Langers and van Dijk 2012; Moerel et al. 2012) and monkeys (Petkov et al. 2006; Joly et al. 2012, 2014). Using synthetic and natural stimuli, the large-scale tonotopic organization (i.e., the spatially ordered representation of frequency) as well as the functional properties of distinct auditory areas have been established in human (Formisano et al. 2003; Striem-Amit et al. 2011; Langers and van Dijk 2012; Moerel et al.

2012) and nonhuman primates (Bendor and Wang 2008; Joly et al. 2014; Baumann et al. 2015).

At higher processing levels, regions in auditory cortex that are more responsive to conspecific vocalizations than other sound categories have been identified in humans (Belin et al. 2000) and monkeys (Petkov et al. 2008; Ortiz-Rios et al. 2015). Yet, little is known about the representation of "intermediate" acoustic features (e.g., spectrotemporal modulations) in the primate brain, which are relevant for the analysis of complex sounds.

So far, acoustic feature mapping has largely been conducted with simple synthetic sounds such as pure tones and ripples (Petkov et al. 2006; Schonwiesner and Zatorre 2009; Baumann et al. 2015).

Compared to synthetic sounds, natural sounds are beneficial for two reasons. First, they are ecologically valid and engage the auditory cortex in meaningful processing. Second, natural sounds elicit higher responses than artificial stimuli, particularly in nonprimary areas (Theunissen and Elie 2014). Hence, the use of natural sounds has enabled the characterization of stimulus–response functions for neurons that respond poorly to artificial sounds. Spectrotemporal receptive fields (STRFs) estimated using synthetic sounds poorly predict STRFs to natural sounds in some neurons especially of nonprimary areas (Theunissen et al. 2000; Bitterman et al. 2008). Accumulating evidence suggests that auditory cortical processing is adapted to the statistics of natural sounds, for example, vocalizations (Woolley et al. 2005; Rodriguez et al. 2010; Theunissen and Elie 2014).

The acoustic properties of behaviorally relevant sounds likely differ for different species. Thus, cross-species comparisons of acoustic feature processing provide a unique opportunity to differentiate between exclusively human and evolutionarily conserved processes that we share with nonhuman primates. Such information is crucial to ultimately understand the neurobiological origins of speech and language (Wilson et al. 2015). Further, comparative studies are essential to relate single cell properties of auditory cortex obtained in invasive monkey studies with noninvasive human fMRI data (Vanduffel et al. 2014).

The present study combines high-resolution fMRI and computational modeling to examine auditory cortical responses to natural sounds in awake macaque monkeys. We focus on the level of representation of fundamental acoustic features (spectrotemporal modulations) and compare the results to those previously obtained in humans with 7 T fMRI, using an identical paradigm and the same computational modeling approach (Santoro et al. 2014, 2017).

We address 2 specific comparative issues. First, in a voxel-by-voxel encoding analysis we derive topographical cortical maps of acoustic feature preference in the macaque. As mentioned above, tonotopy is a well-established organizational principle of the auditory system in both human and nonhuman primates (Morel et al. 1993; Formisano et al. 2003; Bendor and Wang 2008; Moerel et al. 2013; Joly et al. 2014). However, the topographic organization of other acoustic features, for example, temporal and spectral modulations, remains elusive (Joris et al. 2004), although those are ubiquitous features of natural sounds and crucial for processing behaviorally relevant stimuli such as speech (Drullman et al. 1994; Shannon et al. 1995; Elliott and Theunissen 2009). Converging results support the presence of a topographic organization for temporal modulations in both macaque (Baumann et al. 2015) and human auditory cortex (Langner et al. 1997; Barton et al. 2012; Santoro et al. 2014; Brewer and Barton 2016; Hullett et al. 2016). Conversely, evidence for a spatial representation of spectral modulations in the auditory cortex is more limited (Santoro et al. 2014, 2017).

Second, in a multivariate model-based decoding analysis, we quantify the sensitivity of auditory cortex to distinct acoustic features. Both neuroimaging and behavioral data suggest that humans have highest sensitivity to temporal modulations of approximately 3–4 Hz. Likewise, neuronal populations in human auditory cortex have been shown to preferentially encode temporal modulations in this range (Santoro et al. 2017) relevant for the analysis of speech (Giraud and Poeppel 2012; Luo and Poeppel 2012). Psychoacoustic evidence confirms that humans best detect modulation rates at approximately 4 Hz (Viemeister 1979; Bacon and Viemeister 1985) whereas macaques exhibit highest sensitivity to faster temporal modulations of approximately 30–60 Hz (O'Connor et al. 2011; Massoudi et al. 2014). Hence, an open question is how such differences across species in psychoacoustic sensitivity are reflected in neural response properties.

Materials and Methods

Subjects

Three adult rhesus monkeys participated in the experiment (*macaca mulatta*; referred to as M1–M3; 2 females; aged 6–8 years; 4.1–8.4 kg). Animal care and methods were in accordance with national and European guidelines and were approved by the ethical committee of the "Katholieke Universiteit Leuven." The animals were implanted with 8-channel phased-array coils (Janssens et al. 2012). For the details of the surgical procedures, training of monkeys, and eye-monitoring please refer to (Vanduffel et al. 2002; Fize et al. 2003; Ekstrom et al. 2008). Monkeys had experience performing behavioral tasks and were prepared for awake fMRI sessions. Before scanning sessions, monkeys were trained daily (3–5 weeks) to perform a passive fixation task with the head rigidly fixed in sphinx position to a plastic primate chair.

Stimuli and Experimental Procedure

Stimuli and experimental design were identical to the ones presented to human subjects in previous experiments (Moerel et al. 2012, 2013; Santoro et al. 2014). In brief, the stimuli consisted of 168 natural sounds comprising human speech and vocal sounds, animal cries, tool and environmental sounds. The category of animal sounds contained 5 monkey calls; stimuli are available at dx.doi.org/10.5061/dryad.np4hs. Sounds were sampled at 16 kHz and their duration was cut at 1000 ms. Sound onset and offset were ramped with a 10 ms linear slope, and their energy (root mean square) levels were equalized.

Sounds were presented in the silent gap between acquisitions with a randomly assigned interstimulus interval of 2, 3, or 4 TRs-plus an additional random jitter. Zero trials (trials where no sound was presented) constituted 6% (5%) of the trials in train (test) runs. Sounds were presented through MR-compatible insert earphones (MR Confon, Magdeburg, Germany) at approximately 80 dB SPL.

Each monkey completed as many runs as possible during 4 consecutive days (M1: 60 runs; M2: 80 runs; M3: 57 runs); each run lasted approximately 10 min. For runs with too much movement, we discarded the data of the whole run. The data were subdivided into train and test runs. Of the 168 sounds comprised in the stimulus set, 144 belonged to the train set and 24 to the test set. We had a set of 6 train and 2 test runs that were repeated as many times as possible. In each train run, approximately half of the 144 training stimuli (69–74 sounds) were presented once, such that after a set of 6 train runs, each

training sound had been presented 3 times. In the test runs, the 24 testing stimuli were presented and repeated 3 times per run, such that after a set of 2 test runs, each testing sound had been presented 6 times. Thus, the presentation ratio of train to test sounds was 1:2.

Behavioral Task

During fMRI data acquisition, monkeys performed a visual fixation task on which they were highly trained. Eye position was monitored at 120 Hz, using pupil position and an infrared corneal reflection system (Iscan). To encourage monkeys to maintain fixation of a red dot in the center of a black screen, a juice reward was delivered during continuous visual fixation (>800 ms) of the target (red dot in the center of a black screen) through a plastic tube in intervals of 1–2 s. Humans performed a one-back task on the sounds, indicating when the same sound was repeated (Santoro et al. 2014, 2017). Although the tasks differed between species, both tasks were intended to keep subjects alert during passive listening.

MRI Data Acquisition

Data were acquired on a 3 T Siemens Trio scanner with an AC88-insert gradient. Functional (T2*-weighted) contrast-agent enhanced images were collected using implanted 8-channel phased-array coils (Janssens et al. 2012). Implantation of coils adjacent to the macaque's skull places the coil closer to the measured signal and has been shown to increase the signal-to-noise ratio (SNR) of functional MR images by up to a factor of 5 compared with external coils (Janssens et al. 2012). Before each scanning session, the contrast agent monocrystalline iron oxide nanoparticle (MION) was injected intravenously (6–11 mg/kg). MION has the potential to increase the contrast-to-noise ratio by a factor of ~3 at 3 T (Vanduffel et al. 2001). Note that MION-weighted signal changes were opposite to the sign of the blood-oxygenation-level-dependent (BOLD) response which we accounted for in the modeling procedure (see below).

The experiment had a fast event-related design. T2*-weighted functional data were acquired using an echo planar imaging sequence. The acquisition parameters were as follows: repetition time (TR) = 2600 ms; acquisition time (TA) = 1200 ms; echo time (TE) = 30 ms; number of slices = 33; voxel size = 0.75 mm isotropic. Between subsequent acquisitions, there was a silent gap of 1400 ms during which the sounds were presented. The slices covered the brain transversally from the inferior portion of the anterior temporal pole to the superior portion of the superior temporal gyrus bilaterally.

High-resolution T1-weighted images were acquired for each monkey during a separate session under ketamine–xylazine anesthesia. The anatomical images were acquired prior to coil implantation (see above) using a single radial transmit-receive surface coil and an MPRAGE sequence (TR = 2200 ms, TE = 4.05 ms, flip angle = 13°, number of slices = 208, voxel size = 0.4 mm isotropic). We collected 12–15 whole-brain volumes per anatomical session that were averaged to improve SNR.

Preprocessing

Functional and anatomical data were preprocessed in BrainVoyager QX. We applied slice scan-time correction (using sinc interpolation), 3D motion correction and temporal high-pass filtering of 5 (for monkey M1), 10 (M2), or 15 (M3) cycles per time course to remove low frequency drifts. Inspection of raw data revealed more pronounced low frequency drifts in M2 and M3; therefore, high-pass filtering was adjusted individually to maximize the responses to sounds as observed in a GLM analysis (see below). Functional data were coregistered manually to the anatomical data. To improve the SNR in M2, we excluded the 20 runs which had the lowest t-values in response to sounds (in a GLM which treated all sounds as a single condition) within a bilateral auditory cortex mask. Thus, the number of runs that were analyzed ($n = 60$) was comparable to the other 2 monkeys.

Anatomical scans were segmented into gray matter and white matter. We used the border between gray and white matter to obtain inflated hemispheres of the individual monkeys. Next, cortex-based alignment (Goebel et al. 2006) was performed to align the major sulci and gyri between the 3 monkeys using BrainVoyager 20. This alignment information was used for calculating and displaying group maps.

Computational Modeling

We applied an identical computational modeling approach to the macaque fMRI data as described in Santoro et al. (2014, 2017). Two modeling procedures were applied: In a first univariate encoding analysis, we calculated an MTF for each individual voxel. Thus, we obtained maps of the voxel's preferred features across the auditory cortex. In a second multivariate decoding analysis, data from voxels were jointly modeled within a model-based decoding framework. The combined analysis of signals from multiple voxels increases the sensitivity for stimulus information that is represented in patterns of activity, rather than in individual voxels. It further provides an explicit measure of the amount of information about sound features available in the cortex in the accuracy with which those features can be reconstructed.

Extraction of the Sounds' Frequency-Specific Modulation Content

The modulation content of the stimuli was obtained by filtering the sounds with a biologically plausible model of auditory processing (Chi et al. 2005). This auditory model consists of an early stage that models the transformations that acoustic signals undergo from the cochlea to the midbrain; and a cortical stage that accounts for the processing of the sounds at the level of the auditory cortex. We derived the spectrogram and its modulation content using the "NSL Tools" package (available at http://www.isr.umd.edu/Labs/NSL/Software.htm) and customized Matlab code (The MathWorks Inc.). The sounds' spectrograms were obtained using a bank of 128 overlapping bandpass filters with equal width ($Q_{10\,dB} = 3$), spaced along a logarithmic frequency axis over a range of $f = 180$–7040 Hz. The output of the filter bank was band-pass filtered (hair cell stage). A midbrain stage modeled the enhancement of frequency selectivity as a first-order derivative with respect to the frequency axis, followed by a half-wave rectification and a short-term temporal integration (time constant $\tau = 4$ ms). The auditory spectrogram was further analyzed by the cortical stage, where the modulation content of the auditory spectrogram was computed through a bank of 2D filters selective for a combination of spectral and temporal modulations. The filter bank performs a complex wavelet decomposition of the auditory spectrogram. The magnitude of such decomposition yields a phase-invariant measure of modulation content. The modulation selective filters have joint selectivity for spectral and temporal modulations, and

are directional, that is, they respond either to upward or downward frequency sweeps.

In the univariate encoding analysis, filters were tuned to spectral modulation frequencies of Ω = [0.5, 1, 2, 4] cyc/oct, temporal modulation frequencies of ω = [1, 3, 9, 27] Hz, and frequencies of f = [232, 367, 580, 918, 1452, 2297, 3633, 5746] Hz. Our rationale for this choice of values was first, to use a decomposition roughly covering the temporal and spectral modulations present in the acoustic energy of natural sounds (see Supplemental Fig. S1) and second, to use an identical decomposition as previously used humans (Santoro et al., 2014) in order to obtain comparable best feature maps in both species.

The filter bank output was computed at each frequency along the tonotopic axis and then averaged over time. To avoid overfitting, a reduced modulation representation was obtained. This resulted in a representation with 4 spectral modulation frequencies × 4 temporal modulation frequencies × 8 tonotopic frequencies = 128 parameters to learn. Note that the number of parameters to estimate is thus smaller than the number of observations in the train set (n = 144 sounds). The time-averaged output of the filter bank was averaged across the upward and downward filter directions. Then, we divided the tonotopic axis in ranges with constant bandwidth in octaves and averaged the modulation energy within each of these regions (Santoro et al. 2014).

In the multivariate decoding analysis, filters were tuned to 7 spectral modulation frequencies (Ω = [0.3, 0.5, 0.7, 1.1, 1.7, 2.6, 4] cyc/oct), and 12 temporal modulation frequencies (ω = [1, 1.5, 2.4, 3.7, 5.7, 8.8, 13.6, 21, 32.5, 50.3, 77.7, 120] Hz). This resulted in a representation with 7 spectral modulations × 12 temporal modulations (averaged across 12 upwards and 12 downwards) × 60 tonotopic frequencies = 5040 features. Note that in an initial decoding analysis using an identical decomposition as for the encoding analysis (data not shown), we noticed that the decoding accuracy profile for temporal modulations was high-pass in the macaque. Thus, we extended the upper limit of modulation rates from ω = 30 Hz to ω = 120 Hz to explore at which rate the macaque decoding accuracy profile peaked. Dimensionality and overfitting are not affected by the number of features (Equation 5), because in the multivariate case all voxels in a region are used to fit the variation of each feature independently. Therefore, we were able to the use a more fine-grained resolution for the feature decomposition in the multivariate decoding analysis.

The above described processing steps were applied to all stimuli, resulting in an [N × F] feature matrix S of modulation energy, where N is the number of sounds, and F is the number of features in the modulation representation.

Extraction of the Sounds' Frequency Content

For the univariate encoding analysis, we used a tonotopy model as a control analysis in which the stimulus representation in the frequency domain was obtained using only the first stage of the auditory model. The spectrogram was computed at 128 logarithmically spaced frequency values (f = 180–7040 Hz) and averaged over time.

Estimation of fMRI Responses to Sounds

To estimate responses to sounds, first, a matrix Y [(N × V), V = number of voxels] of the fMRI responses to the sounds was calculated using a voxel-by-voxel general linear model (GLM) analysis (Friston et al. 1995). For each voxel i, the response vector Y_i [(N)] was calculated in 2 steps. First, we performed a deconvolution analysis which treated all stimuli as a single condition in order to estimate the hemodynamic response function (HRF) common to all stimuli. Then, using this HRF and one predictor per sound, we computed the response of single voxels to each sound as beta weight (Kay et al. 2008; Moerel et al. 2012; Santoro et al. 2014). This deconvolution analysis was applied to the human fMRI data (Santoro et al. 2014) and the data of macaque M1. The other 2 monkeys' data (M2 and M3) were noisier. Thus, their data were instead modeled using a canonical HRF which is more robust to noise than a deconvolution analysis. Note that MION-weighted signal changes measured here are opposite to the sign of the BOLD response. Therefore, the sign of the applied HRF was inverted.

Further analyses were performed on voxels with a response to the sounds (thresholds were set at $t > 3$ for M1 and $t > 1$ for M2 and M3 in order to be not too strict in this stage of the analysis) within an anatomically defined mask of auditory core and belt regions.

We further improved the SNR by applying a denoising procedure. Noise regressors were entered into the GLM analysis as implemented in the Matlab-based package *GLMdenoise* (Kay et al. 2013). As an improved SNR was observed only in M1, denoising was applied uniquely to this monkey's data.

Univariate Encoding Analysis: Model Estimation

Based on the training data only, the fMRI activity Y_i [N_{train} × 1] at voxel i was modeled as a linear transformation of the feature matrix S_{train} [N_{train} × F] plus a noise term n [N_{train} × 1]:

$$Y_i = S_{train} C_i + n \quad (1)$$

where N_{train} is the number of sounds in the training set, and C_i is an [N × 1] vector of model parameters, whose elements c_{ij} quantify the contribution of feature j to the overall response of voxel i. Columns of matrices S_{train} and Y_i were converted to standardized z-scores. Therefore, Equation (1) does not include a constant term. The solution to Equation (1) was computed using kernel ridge regression (Hoerl and Kennard 1970). The regularization parameter λ was selected independently for each voxel via generalized cross validation (Golub et al. 1979). The search grid included 32 values between $10^{0.5}$ and 10^{11} logarithmically spaced with a grid grain of $10^{0.33}$.

To obtain more stable estimates of the voxels' feature profiles, this computation was performed 5 times using different subsets of the 144 training sounds. For each iteration, 10 of the training sounds were randomly selected and left out, resulting in subset of 134 sounds on which the estimation was performed. In this way, we obtained 5 estimates of each voxel's feature profile which were averaged across iterations.

Model Evaluation

To evaluate the model's prediction accuracy we performed a sound identification analysis (Kay et al. 2008). To this end, we used the fMRI activity patterns predicted by the model to identify which of the test sounds had been heard. Given the trained model $\tilde{C}$ [F × V], and the feature matrix S_{test} [N_{test} × F] for the test set, the predicted fMRI activity $\hat{Y}_{test}$ [N_{test} × V] for the test sounds was obtained as follows:

$$\hat{Y}_{test} = S_{test} \tilde{C} \quad (2)$$

Then, we computed for each stimulus s_k the correlation between its predicted fMRI activity $\hat{Y}_{test}(sk)$ [1 × V] and all

measured fMRI responses $Y_{test}(s_k)$ [1 × V]. The rank of the correlation between predicted and observed activity for stimulus s_k was used as a measure of the model's ability to correctly match $Y_{test}(s_k)$ with its prediction $\hat{Y}_{test}(sk)$. The rank was then normalized between 0 and 1 as follows to obtain the sound identification score m for stimulus s_k:

$$m(s_k) = 1 - \frac{\text{rank}(s_k) - 1}{N_{test} - 1} \quad (3)$$

Note that $m = 1$ indicates correct match; $m = 0$ indicates predicted activity pattern for stimulus s_i was least similar to the measured one among all stimuli. Normalized ranks (sound identification scores) were computed for all stimuli in the test set, and the model's overall accuracy was obtained as the mean of the sound identification scores across stimuli. Statistical significance of the observed accuracy was assessed using permutation testing. The empirical null-distribution of accuracies was obtained by randomly permuting (200 times) the stimulus labels (i.e., N in matrix Y) and repeating the training and testing procedures. In order to preserve the spatial correlations among cortical locations, the same permutations were applied to all voxels. The regularization parameter was constant across permutations and was set to the value derived when the model was estimated on the unpermuted set of responses. Accuracies were converted to z-scores via Fisher's transformation in order to reduce deviations from normality.

Topographic Maps of Feature Preference

The response profiles for temporal modulation, spectral modulation and frequency were computed as marginal sums of the estimated stimulus-activity mapping function C of the frequency-specific modulation model by summing across irrelevant dimensions. For example, to obtain the temporal modulation transfer function (tMTF), we summed across the spectral modulation and frequency dimension:

$$tMTF(\omega) = \sum_{\Omega} \sum_{f} C(\omega, \Omega, f) \quad (4)$$

To calculate profiles for the spectral modulation transfer function (sMTF) and frequency transfer function (fTF), we correspondingly summed across irrelevant dimensions. The voxels' characteristic values (CTM, CSM, CF) were defined as the point of maximum of the tMTF, sMTF, and fTF, respectively. Cortical maps were generated by color-coding the voxels' preferred values and projecting them onto an inflated representation of the monkey's cortex. To obtain group maps, individual maps were aligned using cortex-based alignment (Goebel et al. 2006) and averaged. Surface maps were smoothed using BrainVoyager (1 iteration).

Multivariate Decoding Analysis: Model Estimation

In the multivariate decoding analysis, we evaluated the fidelity with which regions of interest (ROIs) in auditory cortex encode acoustic features by estimating decoders. We selected 3 ROIs, CM/CL, A1, and R/RT based on individual tonotopic maps (see Results). For each monkey in each ROI, a linear decoder was trained for every feature of the modulation space based on the training data only (Santoro et al. 2017). Each stimulus feature S_j [N_{train} × 1] was modeled as a linear transformation of the multivoxel response pattern Y_{train} [N_{train} × V] plus a bias term b_j and a noise term n [N_{train} × 1] as follows:

$$S_j = Y_{train} C_j + b_j 1 + n \quad (5)$$

where N_{train} is the number of sounds in the training set, V is the number of voxels, 1 is a [N_{train} × 1] vector of ones, and C_j is a [V × 1] vector of model parameters, whose elements c_{ji} quantify the contribution of voxel i to the encoding of feature j. The solution to Equation (5) was computed using kernel ridge regression (Hoerl and Kennard 1990). The regularization parameter λ was determined independently for each feature by generalized cross validation (Golub et al. 1979). The search grid included 42 values between $10^{4.4}$ and $10^{8.5}$ logarithmically spaced with a grid grain of $10^{0.1}$.

Estimation of Multivoxel MTFs

Decoders were estimated on the train runs (see above) and tested on the test runs. Given the trained model $\tilde{C}$ [F × V], and the patterns of fMRI activity for the test sounds Y_{test} [N_{test} × V], the predicted feature matrix activity $\hat{S}_{test}$ [N_{test} × F] for the test sounds was calculated as follows:

$$\hat{S}_{test} = Y_{test} \tilde{C} \quad (6)$$

The features' predictions from the test sets were concatenated and decoders were assessed individually by computing the Pearson's correlation coefficient (r) between the predicted and the actual stimulus features. This resulted in 5040 correlation coefficients, which represented the MTF. To obtain marginal profiles of the MTFs, we averaged across irrelevant dimensions.

For statistical testing, all correlation values were transformed into z-scores by applying Fisher's z-transformation and pooled together. For each stimulus feature we computed the null distribution of correlation coefficients at the single-subject level. Null distributions were obtained by randomly permuting (500 times) the stimulus labels of the reconstructed features and computing the correlation coefficient for each permutation. The empirical chance level of correlation r_{chance} was defined as the mean of the null distribution. The P-value was computed as the proportion of permutations that yielded a correlation equal to or more extreme than the observed one. This procedure was repeated in each monkey. For each feature bin, we counted the number of monkeys in which the correlation between reconstructed and original feature was significantly higher than chance ($P < 0.05$); this is indicated in Supplemental Figure S6A by gray color shading of the background.

Post Hoc Statistical Analysis of Marginal MTFs

Group marginal profiles of MTFs were obtained as the mean of all individual marginal MTFs. To assess the statistical significance of the observations on the MTF's marginal profiles, we performed the following post hoc analyses. For the temporal modulation profile we compared accuracies across all hemispheres pooled across 3 low rates (2.4–5.7 Hz for low frequencies < 2 kHz; 1–2.4 Hz for high frequencies > 2 kHz) 3 high rates (32.5–77.5 Hz for low frequencies; 50–120 Hz for high frequencies). Similarly, for the spectral modulation profile, we compared decoding accuracies across all hemispheres at low (0.3–0.5 cyc/oct) and high scales (2.7–4 cyc/oct) using a Wilcoxon signed-rank test. For the frequency profile, we pooled accuracies across 8 bins as the sound decomposition had higher resolution for frequency than for rate and scale (60 frequencies, 12 rates, 7 scales; see above). We compared decoding accuracies at low (0.3–0.5 kHz) and high frequencies (2.1–3.2 kHz) to accuracies at mid frequencies (0.7–1 kHz), using a Wilcoxon signed-rank test.

To test for an interaction of species and modulation rate on decoding accuracy of temporal modulations in the low frequency range we conducted pairwise comparisons between human and macaque ROIs using 2-way ANOVAs with the factor species (human, macaque) and modulation rate (low [2.4–5.7 Hz] or high [32.5–77.5 Hz] rates).

Results

We acquired high-resolution (0.75 mm isotropic voxels) contrast-agent enhanced fMRI at 3 T (Vanduffel et al. 2001) using implanted phased-array coils (Janssens et al. 2012) in awake macaque monkeys (n = 3). Monkeys listened to a large set of real-life sounds (n = 168), including speech and vocal samples, music pieces, animal cries, tool sounds, and scenes from nature (for modulation content of different stimulus categories see Supplemental Fig. S1A). Macaque fMRI responses to natural sounds were modeled using both univariate encoding (Santoro et al. 2014) and multivariate model-based decoding (Santoro et al. 2017), following methods previously employed for homologous human fMRI data (for details see Materials and Methods). For each monkey, we first calculated a map of the responses to all sounds in auditory cortex (see Materials and Methods) and restricted further analyses to activated voxels ($t > 3$ for M1 and $t > 1$ for M2 and M3 in order to be not too strict at this stage of the analysis). The activation spanned a large area of the superior temporal plane (Supplemental Fig. S2A).

Single-Voxel MTFs

In a first univariate encoding analysis we estimated an encoding model for each of these sound-activated voxels. We compared 2 computational models of auditory processing. The first model describes the responses at each auditory cortex voxel as resulting from the combination of modulation-selective filters, each tuned to a specific spectral modulation, temporal modulation, and frequency (referred to as modulation model). The second model describes the responses at auditory cortex voxels as resulting from a bank of frequency selective filters (referred to as tonotopy model). The tonotopy model reflects the hypothesis that voxels simply contain information about the frequency content of the stimuli.

Based on a subset of fMRI data (training), we estimated an MTF (using the modulation model) or a frequency tuning curve (using the tonotopy model) for each voxel. We then assessed the ability of these models to accurately predict the fMRI responses to sounds of a new, independent dataset (testing data; see Materials and Methods). For both the modulation and tonotopy model, we quantified prediction accuracy by means of an identification analysis. For each sound in the testing set, we predicted the fMRI activity patterns using estimated voxels' MTFs (or tuning curves). We then correlated the predicted response pattern with the activity patterns measured for all other sounds in the test set, resulting in voxel-wise prediction accuracy maps (for average accuracy maps of the modulation model see Supplemental Fig. S2B). The rank of the correlation between a sound's predicted and measured activity was used as identification score. This score ranged between 0 and 1; 0 denotes that the predicted activity pattern for a given stimulus was least similar to the measured one among all test stimuli; 1 denotes correct identification; chance level is 0.5. The average score across all test sounds was used as the model's overall prediction accuracy (see Materials and Methods).

The sound identification accuracy of the modulation model was significantly higher than chance for each individual monkey (Fig. 1A, $P < 0.005$ for monkey M1, $P = 0.04$ for M2 and M3; permutation test), indicating that the model was able to generalize to stimuli not used for parameter estimation. The tonotopy model performed significantly above chance for monkeys M1 ($P = 0.02$) and M3 ($P < 0.005$) and showed a trend for performance above chance for monkey M2 ($P = 0.07$; permutation test). We tested for the difference between the tonotopy and modulation model considering the identification accuracies of all available hemispheres (n = 6; 2-tailed Wilcoxon signed-rank test). The modulation model (median [range] = 0.56 [0.52 0.69]) had significantly higher sound identification accuracies than the tonotopy model (median [range] = 0.54 [0.47 0.68]; $z = 2.21$; $P = 0.028$). This is similar to results in humans, where a significant improvement of the modulation model over the tonotopy model was observed (see Santoro et al. (2014), Fig. 4). Thus, a model accounting for the frequency-specific modulation content of the spectrogram is predicting fMRI responses to natural sounds significantly better than a model solely based on the frequency content of the spectrogram.

Preferred Feature Maps

The cortical topography of voxels' preferred features was investigated by calculating maps of voxels' characteristic frequency (CF), temporal modulation (CTM), and spectral modulation (CSM). We applied previously established analyses of best feature mapping (Moerel et al. 2012, 2013; Santoro et al. 2014). For each feature, the estimated single-voxel MTF was averaged across irrelevant dimensions (e.g., spectral and temporal modulation for the frequency transfer function) and the point of maximum was assigned as the voxel's preferred feature value. Tonotopic cortical maps were obtained by logarithmic mapping of best-frequency values to a red–yellow–green–blue color scale. Spectral and temporal modulation maps were obtained by linear mapping of best-feature values to a yellow–green–blue–purple color scale. Maps were then projected onto an inflated representation of the macaques' cortex (Figs 1B and 2A). To obtain group maps we performed cortex-based alignment and averaged individual surface maps (see Materials and Methods).

Maps of best frequency confirmed the presence of the typical mirror-symmetric tonotopic pattern with multiple reversals of the frequency gradient along the anterior–posterior axis in all 3 monkeys (see Fig. 1B for individual map of M1 and group map and Supplemental Fig. S3 for individual maps of M2 and M3). Tonotopic maps provide a basis for the delineation of auditory fields in the primate auditory cortex (Merzenich and Brugge 1973; Morel et al. 1993; Kosaki et al. 1997; Petkov et al. 2006; Joly et al. 2014; Baumann et al. 2015). Tonotopic gradients were most obvious in the putative core fields A1 and R, particularly in monkey M1 (Fig. 1B, middle panel) and the macaque group map (Fig. 1B, lower panel). Consistent with previous macaque fMRI data (Petkov et al. 2006; Joly et al. 2014; Baumann et al. 2015; for review see Baumann et al. 2013) the high-to-low frequency gradient of A1 started in the midline of the posterior superior temporal plane. The gradient ran anterolaterally parallel to the circular sulcus, where the low frequency region (Fig. 1B, red–yellow color) forming the presumed border between core fields A1 and R was located (Fig. 1B, dashed black line). Anterior to this low frequency area, a high-frequency region was located on the medial side of the superior temporal plane, in the depth of the circular sulcus, presumably delineating the border between rostral (R) and rostrotemporal (RT) core

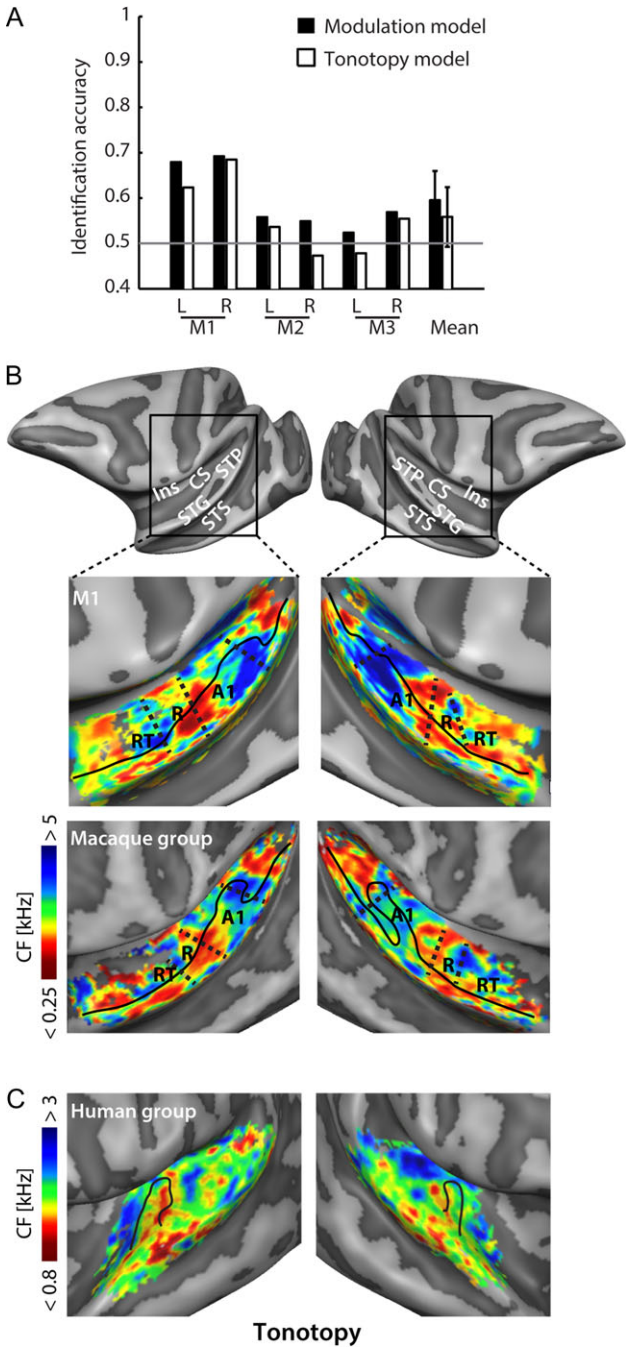

Figure 1. Sound identification accuracies for the modulation and tonotopy model and tonotopic maps. (*A*) Bars indicate the sound identification accuracies for each monkey (M1–M3) and hemisphere (L: left, R: right) separately for the modulation model (black bars) and the tonotopy model (white bars). The last column represents the mean (±SEM) of all 6 hemispheres. Accuracies are normalized between 0 and 1, chance level is 0.5. (*B*) Upper panel: Anatomy is shown as inflated representation of monkey M1's cortex; tonotopic maps are depicted in the cortical region highlighted by the black square (temporal lobe). Middle panel: Individual tonotopic map for monkey M1. Black line illustrates the border of the circular sulcus; dashed black lines delineate the putative borders of auditory fields A1, R, and RT located at the reversals of frequency preference. Lower panel: Macaque group tonotopic map computed as the mean across all monkeys. See also Supplemental Figure S3 for individual tonotopic maps for monkeys M2 and M3. (*C*) Human group tonotopic map modified from Santoro et al. (2014). Black line indicates Heschl's gyrus. Red denotes tuning for low frequencies; blue denotes tuning for high frequencies. CS, circular sulcus; Ins, Insula; STG, superior temporal gyrus; STP, superior temporal plane; STS, superior temporal sulcus; CF, characteristic frequency.

fields. Posterior to presumed area A1, another high frequency region was observed in all monkeys, corresponding to the putative border between A1 and caudomedial (CM) / caudolateral (CL) belt areas (Fig. 1*B*, blue–green color). We observed an additional low frequency area at the posteromedial end of the lateral fissure in 2 monkeys, likely comprising the posterior boundaries of CM and CL (Fig. 1*B*, Supplemental Fig. S3).

Note that putative area borders (dashed black lines) are indicated along the posterior–anterior axis at the frequency preference reversals. In contrast, the distinction between core, belt and parabelt regions along the medial–lateral axis is more problematic, as adjacent core and belt areas share the same tonotopic preference (Joly et al. 2014). Those can be defined on the basis of microanatomical properties (Hackett et al. 2001; Morosan et al. 2001). Auditory core regions, in particular A1, are known to be more densely myelinated than belt areas (Hackett et al. 2001; Hackett 2011; Joly et al. 2014). We estimated myelin maps based on the ratio of T1- over T2-weighted (T1w/T2w) MR images (Glasser and Van Essen 2011; De Martino et al. 2015). The highest myelin content was indeed localized in the posterior region of the lateral sulcus at the high-to-low frequency gradient presumably delineating A1 (Supplemental Fig. S4).

Human tonotopic maps obtained using identical stimuli (Santoro et al. 2014) showed a consistent high–low–high frequency gradient (Fig. 1*C*). A low frequency region was observed in the central region of Heschl's gyrus (HG) presumably marking the boundary between the human homolog of primary fields A1 and R (Fig. 1*C*, red–yellow). This low frequency region was surrounded anteromedially and posteriorly by high frequency regions (Fig. 1*C*, green–blue). The anteromedial high frequency areas clustered on the planum polare (PP). The posterior regions preferring high frequencies covered Heschl's sulcus and planum temporale (PT).

The topography of characteristic spectral and temporal modulations was more variable and complex across monkeys (see Fig. 2*A–C*, and Supplemental Fig. S5 for individual maps of temporal and spectral modulations). However, the group data (Fig. 2*B*,*C*, lower panels) confirmed distinct regional preferences for modulation frequencies. In line with a recent macaque fMRI study (Baumann et al. 2015), preferences for high temporal modulations (high CTM, purple) consistently clustered in the posterior auditory cortex with maxima at the high frequency part of the putative primary field A1 in both hemispheres. Preference for low temporal modulations (low CTM, yellow) was located in anterior–lateral auditory regions (Fig. 2*B*). In contrast, best-spectral-modulation maps indicated that coarse spectral information (low CSM, yellow) was preferably encoded in posterior–medial auditory regions, as opposed to fine spectral information (high CSM, purple) in anterior–lateral auditory regions (Fig. 2*C*). On a qualitative level, best-modulation maps confirmed the systematic organization of fMRI responses to temporal and spectral modulations consistent with the topographic representation observed previously in human data (Fig. 2*D*,*E*).

Supporting the hypothesis of a trade-off between spectral and temporal resolution at the map level (Santoro et al. 2014), we found a negative correlation between voxels' characteristic spectral and temporal modulation for all 3 monkeys (Fisher-transformed Spearman's rank correlation coefficient: median [range] = −0.12 [−0.11 −0.15], $P < 0.001$). Further, we observed a positive correlation between tonotopy and temporal modulation maps, on the one hand, and negative correlation between tonotopy and spectral modulation maps, on the other hand. In other words, sensitivity for low temporal modulations tended

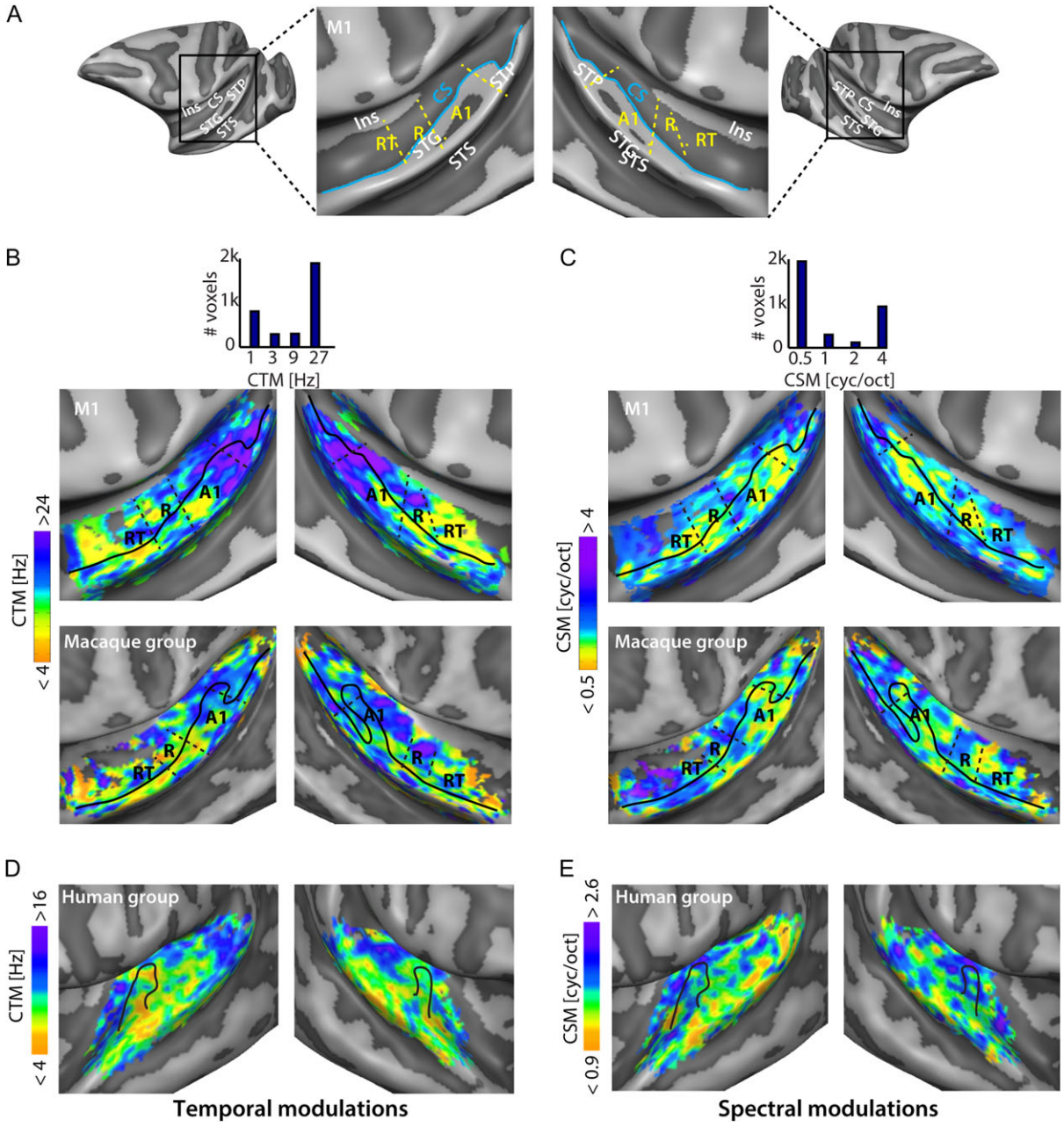

Figure 2. Individual and group maps for temporal and spectral modulations. (*A*) Inflated representation of monkey M1's auditory cortex; dashed yellow lines delineate the putative borders of auditory fields A1, R, and RT based on the reversals in the tonotopic maps from Figure 1*B*. (*B*) Distribution of best temporal modulation responses in auditory cortex of monkey M1. Middle panel: Individual map of preferred temporal modulations for monkey M1. See also Supplemental Figure S5 for individual maps of temporal modulations for monkeys M2 and M3. Lower panel: Macaque group map for preferred temporal modulations obtained as the mean across 3 monkeys. (*C*) Distribution of best spectral modulation responses in auditory cortex of monkey M1. Middle panel: Individual map of preferred spectral modulations for monkey M1. See also Supplemental Figure S5 for individual maps for monkeys M2 and M3. Lower panel: Macaque group map for preferred spectral modulations. (*D*) Human group best feature map for temporal modulations modified from Santoro et al. (2014). (*E*) Human group best feature map for spectral modulations modified from Santoro et al. (2014). Black line indicates Heschl's gyrus. Purple denotes tuning for fast (fine) temporal (spectral) modulations; yellow denotes tuning for slow (coarse) temporal (spectral) features. CS, circular sulcus; Ins, insula; STG: superior temporal gyrus; STP, superior temporal plane; STS, superior temporal sulcus; CTM, characteristic temporal modulation; CSM, characteristic spectral modulation.

to be observed in low frequency regions of auditory cortex (Fisher-transformed Spearman's rank correlation coefficient: median [range] = 0.19 [0.18 0.26], $P < 0.001$; see also Fig. 2) and sensitivity to low spectral modulations in high frequency regions (Fisher-transformed Spearman's rank correlation coefficient: median [range] = −0.38 [−0.29 −0.44], $P < 0.001$).

Multivoxel MTFs

In a second analysis, we ran a multivariate model-based decoding analysis to investigate whether we can decode the spectro-temporal modulations of sounds from fMRI response patterns in the macaque auditory cortex. First, we calculated MTFs for

the whole auditory cortex. Second, we computed MTFs in ROIs along the superior temporal plane corresponding to putative core and belt areas. Three ROIs on the temporal plane were defined using the frequency reversals of individual tonotopic maps (Fig. 1B and Supplemental Fig. S3): putative primary core area A1, rostral and rostrotemporal core areas R/RT and caudomedial and caudolateral belt areas CM/CL (see Supplemental Fig. S7A for correspondence between tonotopy and ROIs). In humans, the following anatomically defined ROIs (Kim et al. 2000) were included: HG, the putative location of primary AC, and adjacent areas PP and PT (Santoro et al. 2017). We calculated the modulation representation for all stimuli (see Materials and Methods). Then, in each subject and each ROI separately, we estimated a linear decoder for each feature of this modulation representation using a subset of fMRI data (training), which resulted in a map of voxels' contributions. This allowed us to investigate how faithfully distinct features in the sounds' spectrotemporal modulation content can be decoded from activation patterns in ROIs. In each subject, we quantified decoding accuracy as Pearson's r between decoded and actual stimulus features. This resulted in a multivoxel MTF per ROI per hemisphere with corresponding marginal profiles of frequency, spectral, and temporal modulation. The group marginal multivoxel MTFs were obtained as the mean of individual MTFs (Fig. 3).

Multivoxel Marginal Profiles for Temporal Modulation

Results in the whole macaque auditory cortex indicated that decoding accuracies for temporal modulations were highest at rates >30 Hz, peaking at 77.5 Hz in the macaque (Supplemental Fig. S6A, left). The profile showed a trough at low modulation rates with a local minimum at 3.7 Hz. Post hoc analyses of the marginal MTF for temporal modulations confirmed that decoding accuracies at 32.5–77.5 Hz were significantly higher than accuracies at 2.4–5.7 Hz ($z = 3.07$, $P = 0.002$, Wilcoxon signed-rank test). Conversely, in the human auditory cortex, the decoding profile for temporal modulations peaked at 3.7 Hz. Accuracies at low modulation rates (2.4–5.7 Hz) were significantly higher than at high rates (32.5–77.5 Hz; $z = 4.7$, $P < 0.001$, Wilcoxon signed-rank test; Supplemental Fig. S6B, left). A 2-way ANOVA confirmed a significant interaction of species (human, macaque) and modulation rate (pooled at rates of 2.4–5.7 Hz and 32.5–77.5 Hz) on accuracy ($F[1, 95] = 10.53$, $P = 0.001$).

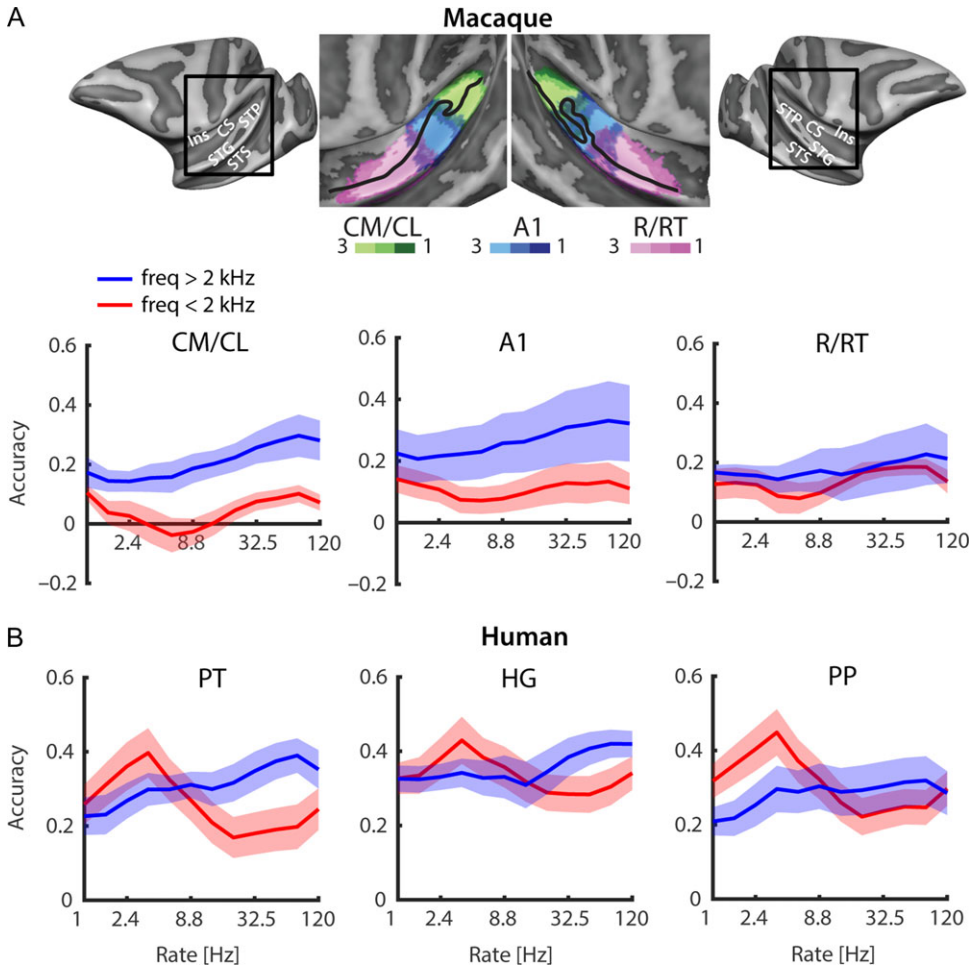

Figure 3. ROI-based multivoxel marginal profiles for temporal modulation. The profiles represent decoding accuracies defined as Pearson's correlation coefficient between decoded and original stimulus features. Marginal transfer functions were obtained by averaging the MTFs across irrelevant dimensions (spectral modulation and frequency). (A) Macaque decoding accuracy profiles were obtained as the mean across the 6 hemispheres of 3 monkeys. (B) Human marginal MTFs in distinct ROIs; 10 hemispheres of 5 human subjects from Santoro et al. (2017) were reanalyzed using an identical sound decomposition as for the macaque data. Shades denote SE across all hemispheres. CM/CL, caudomedial/caudolateral belt areas; A1, primary auditory area; R/RT, rostral/rostrotemporal area; PT, planum temporale; HG, Heschl's gyrus; PP, planum polare.

Next, we computed MTFs for individual ROIs. Because visual inspection of the 3D MTFs in the human (Santoro et al. 2017, Fig. 4B) and monkey (Supplemental Fig. S7B) indicated that decoding accuracy for temporal modulation rates depended on frequency, we computed distinct marginal tuning functions for frequencies < and >2 kHz. In all 3 macaque ROIs, decoding accuracies for temporal modulations for low frequencies (<2 kHz) were highest at 32.5–77.5 Hz and lowest at 2.4–5.7 Hz (Fig. 3A, red profile). Post hoc analyses confirmed that decoding accuracies at 32.5–77.5 Hz were significantly higher than accuracies at 2.4–5.7 Hz in all 3 macaque ROIs (CM/CL: $z = 3.72$, $P < 0.001$; A1: $z = 2.68$, $P = 0.007$; R/RT: $z = 3.07$, $P = 0.002$; Wilcoxon signed-rank test). Conversely, in humans, the profile for temporal modulations at frequencies <2 kHz peaked at 3.7 Hz and was lowest at high rates >21 Hz in PT, HG, and PP (Fig. 3B, red profile, see also Santoro et al. 2017; Wilcoxon signed-rank test comparing 2.4–5.7 Hz and 32.5–77.5 Hz: $z = [3–4.6]$, $P < 0.005$ in all ROIs). The temporal rate profile in macaque ROIs correlated negatively with the profile in all corresponding ROIs on the human temporal plane (Spearman's rho over the 12 modulation rates: CM/CL–PT: –0.59, $P = 0.05$; A1–HG: –0.73, $P = 0.009$; R/RT–PP: –0.85, $P < 0.001$). Comparisons between corresponding human and macaque ROIs using 2-way ANOVAs confirmed a significant interaction of species (human, macaque) and modulation rate (low [2.4–5.7 Hz] or high [32.5–77.5 Hz] rates) on decoding accuracy in all corresponding ROIs (CM/CL–PT: $F[1,95] = 12.23$; $P = 0.001$, A1–HG: $F[1,95] = 4,41$, $P = 0.038$; R/RT–PP: $F(1,95) = 20.46$, $P < 0.001$). These results suggest that cortical tuning to a timescale relevant for the analysis of speech (3–5 Hz) is characteristic of the human but not monkey auditory cortex (see Discussion).

Importantly, the enhanced accuracy of low temporal rates was not an indirect effect of decoding a global preference for speech. When removing speech sounds from the analysis, decoding results remained largely unchanged and the peak of the MTF profile for temporal modulations at ~3 Hz was still observed in all human ROIs (Pearson's r with the original marginal MTF: HG = 0.99, PT = 0.99, PP = 0.98, Santoro et al. 2017). Removal of speech and vocal sounds altered the relative contribution of low and high frequencies to the acoustic energy of the stimulus set, but for temporal and spectral modulations, the consequences on the relative energy distribution were smaller (Supplemental Fig. S1B; for details also refer to Santoro et al. 2017, Supplemental Fig. S4).

Moreover, ANOVAs also showed a significant main effect of species in all ROIs ($F[1,95] = [51.58–65.03]$, $P < 0.001$), indicating that humans had overall higher decoding accuracies than monkeys.

In contrast, the decoding accuracy profile in the temporal modulation dimension for high frequencies (>2 kHz) was high-pass in both species (Fig. 3, blue profiles). Decoding accuracies at low temporal modulations of 1–2.4 Hz were significantly lower than accuracies at high temporal modulations of 50–120 Hz in monkey CM/CL ($z = –3.382$, $P < 0.001$) and A1($z = –2.80$, $P = 0.005$) and all human ROIs (PT: $z = –4.6$, $P < 0.001$, HG: –4.58, $P < 0.001$, PP: $z = –3.14$, $P = 0.002$, Wilcoxon signed-rank test). The profiles in the macaque ROIs correlated positively with the profiles in the corresponding human ROIs for the pairs CM/CL–PT (Spearman's rho = 0.93, $P < 0.001$) and A1–HG (Spearman's rho = 0.8, $P = 0.003$).

Multivoxel Marginal Profiles for Spectral Modulation and Frequency

The macaque decoding accuracy profiles for spectral modulation and frequency were reminiscent of the marginal profiles observed in humans (Santoro et al. 2017, see Supplemental Fig. S6). For the whole macaque auditory cortex, the profile for spectral modulations was low-pass (Supplemental Fig. S6A, middle panel). Decoding accuracies at low spectral scales (0.3–0.5 cyc/oct) were significantly higher than accuracies at high spectral scales (2.6–4 cyc/oct; Supplemental Fig. S6A, middle panel) according to a Wilcoxon signed-rank test ($z = 1.96$, $P = 0.05$). This pattern of results is consistent with results in humans, where we also observed a low-pass decoding accuracy profile for spectral modulations (Supplemental Fig. S6B, middle panel).

In macaques, this result was driven by voxels in A1: Only in A1, but not CM/CL and R/RT, decoding accuracies at low spectral modulations (0.3–0.5 cyc/oct) were significantly higher than accuracies at high spectral modulations (2.6–4 cyc/oct; Supplemental Fig. S8A; $z = 2.2$, $P = 0.03$) similar to human PT, HG, and PP ($z = [2.87–3.85]$; $P < 0.005$, Wilcoxon signed-rank test; Supplemental Fig. S8A, for human results see also Santoro et al. 2017). We did not find a significant interaction of species and spectral modulation (high, low).

For spectral modulations, the median accuracy profile in the whole macaque auditory cortex correlated positively with the acoustic energy in the stimuli (Spearman's rho over the 7 spectral scales = 0.89, $P = 0.01$) indicating that brain responses followed the spectral modulation content of the stimuli. Conversely, for temporal modulations and frequency we found a negative correlation between median accuracy profile in macaque auditory cortex and the acoustic energy in the stimuli (temporal modulations: Spearman's rho over the 12 modulation rates = –0.65, $P = 0.02$; frequency: Spearman's rho over the 60 frequencies = –0.48, $P < 0.001$).

The decoding accuracy profile for frequency for the whole auditory cortex appeared more complex and variable (Supplemental Fig. S6A, right panel). In particular, we observed the highest decoding accuracies at high frequencies (>2 kHz). Decoding accuracies were lowest in the mid frequency range (0.7–1 kHz), similar to the spectral profile observed in humans (Supplemental Fig. S6B, right panel). Post hoc analyses of the frequency profile in the whole macaque auditory cortex confirmed that decoding accuracies at mid frequencies (0.7–1 kHz) were significantly lower than accuracies at high frequencies (2.1–3.2 kHz; $z = –4.2$, $P < 0.001$, Wilcoxon signed-rank test) and accuracies at low frequencies (0.3–0.5 kHz; $z = –3.91$, $P < 0.001$).

In individual ROIs, a similar trough in decoding accuracies for mid frequencies was observed (Supplemental Fig. S8B). In all macaque ROIs, decoding accuracies at high frequencies (2.1–3.2 kHz) were significantly higher than in mid the frequency range (0.7–1 kHz, $z = [4.62–4.77]$, $P < 0.001$, Wilcoxon signed-rank test). Decoding accuracies at low frequencies (0.3–0.5 kHz) were also higher than in the mid frequency range in CM/CL ($z = 2.78$, $P = 0.005$) and R/RT ($z = 4.03$, $P = 0.001$). Post hoc analyses of the human spectral profile in PT, PP and HG confirmed that decoding accuracies at high frequencies were significantly higher than accuracies at mid frequencies ($z = [3.05–6.59]$, $P < 0.005$, Wilcoxon signed-rank test). Accuracies at low frequencies were also significantly higher than at mid frequencies ($z = [4.24–5.42]$, $P < 0.001$, see also Santoro et al. 2017). This frequency profile was not present in the stimuli and might be related to effects of the scanner noise (see Discussion). We observed a significant interaction of species (macaque, human) and frequency (low [0.3–0.5 kHz] or high [2.1–3.2 kHz]) on decoding accuracy in all corresponding ROIs (CM/CL–PT: $F[1, 255] = 11.81$, $P < 0.001$; A1–HG: $F[1, 255] = 11.61$, $P < 0.001$; R/RT–PP: $F[1255] = 4.95$, $P = 0.03$), where macaques had better decoding

accuracies for high frequencies and poorer accuracies for low frequencies compared with humans. This finding may be related to the difference in hearing range in both species (see Discussion).

Discussion

Here, we show evidence for topographic cortical maps of temporal and spectral modulation preference in awake macaques by combining the presentation of natural sounds and computational modeling of high-resolution fMRI data. The large-scale organization of these maps is remarkably similar to the one previously observed in humans using the same natural stimuli (Santoro et al. 2014). We further demonstrate that we can decode the spectrotemporal modulations of natural sounds from macaque fMRI response patterns. Model-based decoding in the temporal modulation domain was most accurate for high rates (>30 Hz) in macaques. This property of the macaque auditory cortex contrasts with the human auditory cortex that exhibits characteristic tuning to slower timescales (~3 Hz) relevant for the analysis of speech sounds.

Homologies in Natural Sound-Encoding: Large-Scale Topographic Maps

We derived topographic maps of acoustic feature preference for frequency, spectral, and temporal modulations. Tonotopic maps exhibited the well-established mirror-symmetric high-low-high frequency gradients across the primary core and surrounding belt areas (Fig. 1B,C) in line with numerous studies in human (Formisano et al. 2003; Talavage et al. 2004; Humphries et al. 2010; Woods et al. 2010; Da Costa et al. 2011; Striem-Amit et al. 2011; Langers and van Dijk 2012; Moerel et al. 2012, 2014) and nonhuman primates (Merzenich and Brugge 1973; Morel et al. 1993; Kosaki et al. 1997; Rauschecker et al. 1997; Bendor and Wang 2008; Joly et al. 2014; Baumann et al. 2013, 2015). Another low frequency region in the posterior end of the lateral fissure was consistently observed which may have been obscured in previous monkey fMRI studies due to an incomplete coverage of the lateral fissure (1–3 oblique slices in Petkov et al. 2006), lower resolution of the data (1.2 mm isotropic in Joly et al. 2014; 1 mm isotropic in Baumann et al. 2015; as compared with 0.75 mm isotropic in the present dataset), or smoothing of the data. When looking at tonotopic maps with higher spatial resolution, and exploring single subject rather than group maps, more frequency reversals than commonly reported become evident (Moerel et al. 2014). The observed redundancy in neuronal populations responding to the same frequency range may enable auditory cortex to generate simultaneous "views" of the spectrogram at distinct spectral resolutions (see below).

In contrast, previous studies have not come to an agreement on the existence of a topographic organization for temporal and spectral modulations in auditory cortex. Here, we observed a posterior-to-anterior high-to-low gradient for temporal modulations across macaque auditory cortex (Fig. 2B). This observation from single-voxel encoding was confirmed by the multivoxel model-based decoding from distinct ROIs: although the general pattern of profiles was similar in caudal and rostral areas, the caudal fields CM/CL had higher decoding accuracies for fast temporal modulations (Fig. 3A). A topographic representation for modulation rate has previously been reported in the inferior colliculus (IC) of cats (Schreiner and Langner 1988). At the level of the cortex, primate fMRI (Baumann et al. 2015) and electrophysiological studies showed that fast temporal acoustic information is preferably encoded in caudal auditory regions (Camalier et al. 2012; Kusmierek and Rauschecker 2014) and slow temporal information in rostral areas R and RT (Liang et al. 2002; Bendor and Wang 2008). Although earlier human fMRI studies failed to observe a clear topography for modulation rate (Giraud et al. 2000; Schonwiesner and Zatorre 2009; Overath et al. 2012; Leaver and Rauschecker 2016), fMRI data from our lab (Santoro et al. 2014) as well as a recent electrocorticography (ECoG) study (Hullett et al. 2016) confirmed the presence of a posterior-to-anterior high-to-low rate gradient in human auditory cortex. Such a cortical modulation filter bank could parse the acoustic information in the stimulus according to its dominant temporal components and would facilitate the encoding of multiple views of the incoming spectrogram at different temporal resolutions (Dau et al. 1997). Multiple cortical representations of sounds are likely critical for simultaneously executing distinct behavioral tasks. For example, encoding of slow temporal modulations is critical for speech comprehension (Elliott and Theunissen 2009) while encoding of fast temporal modulations is hypothesized to facilitate sound localization. Notably, the presence of high modulation rates in human screams (>30 Hz) improves the ability to localize sounds (Arnal et al. 2015). In the framework of an auditory "where"-pathway the observed topography is consistent with an account where fast temporal modulations are represented more prominently in caudal regions of auditory cortex. Those posterior regions have been shown consistently to participate in sound localization (Rauschecker and Tian 2000; Tian et al. 2001; Woods et al. 2006; Rauschecker and Scott 2009; Ortiz-Rios et al. 2017).

In electrophysiology, responses of single neurons to temporal modulations follow 2 different coding principles, namely rate coding and temporal coding (Joris et al. 2004). The rate code represents the average firing rate (i.e., the average number of spikes over a period of time). The temporal code constitutes a measure of phase-locking to the stimulus envelope. Neurons tuned to fast temporal modulations above approximately 50 Hz (Joris et al. 2004) typically code through average spike rate (rate code), whereas neurons tuned to slow temporal modulations below approximately 50 Hz mostly demonstrate synchronization to the sounds' modulations (temporal code; Joris et al. 2004). How the hemodynamic response reflects these 2 types of neuronal coding and how they are related in turn to the observed topographies remains unclear (Joris et al. 2004; Baumann et al. 2015). Advances in computational models integrating results from single-cell recordings (Petkov and Bendor 2016) and fMRI studies in macaques are expected to shed light on this crucial question in the future.

Previous fMRI studies investigated cortical processing of spectral and temporal modulations measuring responses to artificial sounds, for example, amplitude-modulated noise (Baumann et al. 2015) or dynamic ripples (Schonwiesner and Zatorre 2009), whereas we chose to present natural sounds, for the following reasons. Natural sounds are characterized by distinct statistical regularities. As synthetic sounds lack both the behavioral relevance and the statistical structure of natural sounds, they activate auditory cortex differently than under natural listening conditions (Theunissen and Elie 2014): STRFs of auditory neurons differ significantly under natural and synthetic stimulus conditions, in human (Bitterman et al. 2008) and animal electrophysiological experiments (Theunissen et al. 2000). In particular, tonotopic maps obtained with tones or natural sounds are similar in the IC (De Martino et al. 2013) and the primary auditory cortex, but differ more in nonprimary

areas (Moerel et al. 2013). Further, whereas natural stimuli cover a wide range of combination of features, achieving the same resolution with ripples would involve the presentation of 5040 conditions, exceeding experimental time constraints.

Beyond the presentation of natural sounds, our approach approximates natural listening conditions with another respect: Whereas earlier primate experiments collected data in anaesthetized animals (Rauschecker et al. 1995), our data were acquired in awake macaques. Thus, we provide evidence that topographic maps of feature preference can be observed under approximately natural hearing situations. However, natural hearing also involves attention and goal-directed behavior. Whether and how best feature maps are shaped by task demands necessitates further investigation.

Specificity in Natural Sound-Encoding: Multivoxel MTFs

On a methodological level, our first univariate encoding analysis derived cortical maps of relative feature preference by selecting the feature eliciting the highest response at each voxel. These maps are based on the assumption that stimulus features are more accurately encoded when they maximally activate single voxels. However, higher responses may not necessarily mean better encoding. Our second, complementary multivariate decoding analysis relies on measures of information rather than activation levels. It assesses how faithfully stimulus features can be retrieved from fMRI response patterns. Thus, the amount of information that is available in the cortex about a set of stimulus features is explicitly quantified in the accuracy with which those features can be decoded. Accurate decoding of spectrotemporal modulations in test sounds indicates that these modulations are reproducibly mapped into distinct spatial patterns. As data from individual voxels are jointly modeled, the cortex is characterized in a multivariate manner, without the need of integrating results derived from the modeling of individual voxels. This combined analysis of signals from multiple voxels increases the sensitivity for stimulus information that may be represented in patterns of responses, rather than in individual voxels.

In the human primary area HG and adjacent regions PP and PT, decoding accuracy for temporal modulation rates depended on frequency. For low frequencies, we observed a striking species difference: While humans' highest decoding accuracies were observed at the speech-relevant rate of ~3 Hz, the macaque's temporal modulation profile peaked instead at faster rates of > 30 Hz (Fig. 3). Importantly, the selectivity to slow temporal modulations observed in humans was present already in primary auditory (HG) and adjacent areas (PP, PT).

One might argue that the observed specialization of human auditory cortex could be the spurious side effect of selectivity for higher level properties of sound, such as the semantic category. A large fraction of human auditory cortex, especially along the STG, is highly selective for speech, producing larger fMRI responses to speech and voice than to other sounds (Belin et al. 2000; Overath et al. 2015). Therefore, to rule out the possibility that the enhanced decoding accuracy for slow temporal rates could reflect speech selectivity rather than privileged encoding of those rates, a control analysis was performed in humans: all speech and vocal sounds were removed from the stimulus set (Supplemental Fig. S1B). Critically, modulation tuning curves in early auditory areas (HG, PT, PP) were largely unaffected by removing speech sounds from the decoding analysis (Santoro et al. 2017, Supplemental Fig. S6). Three conclusions arise from this result. First, this finding excludes the possibility that tuning profiles could simply mirror the stimulus statistics rather than representing characteristic acoustic tuning properties of neuronal populations in auditory cortex. Second, the result indicates that successful decoding of slow temporal rates was independent of semantic category. It rules out the possibility that low frequency sensitivity in human auditory cortex could reflect "speech selectivity" rather than privileged encoding of slow modulation rates. Third, the finding supports the hypothesis that, in the human brain, the tuning properties of neuronal populations for the analysis of any sound have been shaped by the characteristic acoustic properties of speech.

While lower levels of the auditory system, that is, the IC and medial geniculate body, have been found to express faster modulation rates in a number of species (for review see Joris et al. 2004), the finding of strong responsiveness of macaque auditory cortex to such fast modulation rates is surprising. This result suggests that the cutoff of the cortical MTF is not a purely physiological limit but is rather plastic and reflects the animal's particular acoustic environment. Importantly, supporting evidence for this finding comes from psychoacoustic observations of human listeners performing best in the discrimination of slow temporal modulations in the 2–4 Hz range (O'Connor et al. 2011; Massoudi et al. 2014), whereas the macaques' sensitivity to temporal modulations is highest at fast rates of approximately 30–60 Hz (O'Connor et al. 2011). A plausible explanation for the observed divergence of cortical function across species is that the auditory cortex has evolved to optimize the representational mechanisms for acoustic features of the behaviorally most relevant sounds, species-specific vocalizations. In anesthetized marmosets, neuronal A1 responses showed highest synchronization to the repetition rate of natural calls (approximately 8 Hz), irrespective of whether species-specific vocalizations (Wang et al. 1995) or amplitude-modulated tones (Nagarajan et al. 2002) were presented, indicating that most neurons were tuned to the natural repetition rate of calls. For humans, speech is arguably the most relevant sound category. Consistently, the modulation spectrum of speech peaks at low temporal modulation rates of approximately 3–4 Hz, corresponding to the rate of syllables (Giraud and Poeppel 2012). Similarly, for macaques, vocalizations are considered essential for intraspecies communication (Hauser et al. 2002). The modulation spectrum of macaque vocalizations is broader, and more importantly, the highest variance is contained in the energy of higher modulation rates (Cohen et al. 2007; Joly et al. 2012). Such high temporal modulations in the frequency range of 30–120 Hz elicit the percept of roughness (Joris et al. 2004). They are abundant in, for example, screams and alarm signals (Arnal et al. 2015) and the environmental and nature sounds of the current stimulus set (Supplemental Fig. S1A).

A caveat of our approach was the task difference between species. Both species performed unspecific tasks to keep subjects awake but for reasons of feasibility, macaques performed a visual fixation and humans a one-back task. One might argue that the human sensitivity to slow modulations in the theta range (4–8 Hz) could be interpreted as working memory related rather than a specialization for species-specific communication signals given the task differences and given earlier ECoG investigations that identified semantic memory related theta oscillatory signals in human auditory cortex (Canolty et al. 2006). However, it seems unlikely that an unspecific vigilance task would exclusively affect the cortical response to temporal modulations (but not the other acoustic dimensions). To date, our understanding of how hemodynamic effects relate to

electrophysiological oscillatory findings is still limited. Studies directly looking at the relation between the BOLD signal and oscillatory activity in different frequency bands find the strongest correlation of the hemodynamic response with power of oscillations in the high gamma (60–80 Hz) rather than the theta range (Niessing et al. 2005; Scheeringa et al. 2011).

The macaques' decoding accuracy profiles for spectral modulation and frequency were reminiscent of the marginal profiles observed in humans (Santoro et al. 2017). In the spectral modulation domain, the accuracy profile correlated with the spectral modulation energies in the stimuli, indicating that neural responses followed the spectral modulation content of the sounds. The marginal MTF for frequency revealed a decrease in decoding accuracies for frequencies at 1 kHz that may be related to the scanner noise (Santoro et al. 2017). In the clustered fMRI acquisition, sounds were presented during silent gaps between scans. Thus, the acoustic scanner noise between stimulus presentations may have interacted with the response to the auditory stimulation, through, for example, adaptation of the neuronal population preferring sound frequencies in the range of the scanner noise or saturation of the hemodynamic response. In contrast to humans who had higher sensitivity to low frequencies, macaque's decoding accuracies were enhanced at high frequencies (>2 kHz) possibly due to macaques' larger hearing range up to 32 kHz (Pfingst et al. 1978).

Although stimuli were identical, decoding accuracy profiles differed between human and nonhuman primates, excluding the possibility that observed tuning profiles could simply mirror the acoustic properties of the presented natural stimuli. Rather, our results provide evidence that decoding accuracies represent characteristic tuning of neuronal populations in auditory cortex (see above). Importantly, by comparing human data to the present results in macaques, we have shown that tuning to speech-relevant modulation frequencies is specific to the human brain and is already present in early auditory areas. The finding supports a tight link between modulation tuning and speech: we show that tuning of the auditory cortex to slow modulation rates is uniquely human and propose that this characteristic may have evolved as a function of the acoustic properties of human speech. Thus, it is plausible that even basic processing mechanisms in auditory cortex have evolved to selectively amplify acoustic features and optimize the representation of species-specific vocalizations.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Authors' Contributions

Conceptualization: E.F., F.D.M., and W.V.; Methodology: J.E., A.M., E.F., F.D.M., and W.V.; Software: F.D.M., R.G.; Formal analysis: J.E., F.D.M.; Investigation: A.M.; Resources: E.F., F.D.M., A.M., J.E., and W.V; Writing—original draft: J.E.; Writing—review and editing: J.E., E.F., F.D.M., W.V., A.M.; Visualization: J.E., F.D.M., E.F.; Supervision: E.F.; Funding acquisition: E.F., W.V.

Funding

European Union's Horizon 2020 Framework Programme for Research and Innovation under Grant Agreement No. 785907 (Human Brain Project SGA2); Maastricht University; the Dutch Province of Limburg; the Netherlands Organization for Scientific Research (NWO Grant 453-12-002 to E.F.); the Research Foundation Flanders (FWO-Flanders) G0007.12; KU Leuven Programme Financing PFV/10/008 and C14/17/109; and the Hercules foundation.

Notes

We thank Vittoria de Angelis, Giancarlo Valente and Michelle Moerel for useful discussions on data analysis. We thank three anonymous reviewers for their valuable comments. *Conflict of Interest*: The authors declare no competing financial interests.

References

Arnal LH, Flinker A, Kleinschmidt A, Giraud AL, Poeppel D. 2015. Human screams occupy a privileged niche in the communication soundscape. Curr Biol. 25:2051–2056.

Bacon SP, Viemeister NF. 1985. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. Audiology. 24:117–134.

Barton B, Venezia JH, Saberi K, Hickok G, Brewer AA. 2012. Orthogonal acoustic dimensions define auditory field maps in human cortex. Proc Natl Acad Sci USA. 109:20738–20743.

Baumann S, Joly O, Rees A, Petkov CI, Sun L, Thiele A, Griffiths TD. 2015. The topography of frequency and time representation in primate auditory cortices. eLife. 4:e03256.

Baumann S, Petkov CI, Griffiths TD. 2013. A unified framework for the organization of the primate auditory cortex. Front Syst Neurosci. 7:11.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. Nature. 403:309–312.

Bendor D, Wang X. 2008. Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. J Neurophysiol. 100:888–906.

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.

Brewer AA, Barton B. 2016. Maps of the auditory cortex. Annu Rev Neurosci. 39:385–407.

Camalier CR, D'Angelo WR, Sterbing-D'Angelo SJ, de la Mothe LA, Hackett TA. 2012. Neural latencies across auditory cortex of macaque support a dorsal stream supramodal timing advantage in primates. Proc Natl Acad Sci USA. 109:18168–18173.

Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, Berger MS, Barbaro NM, Knight RT. 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science. 313:1626–1628.

Chi T, Ru P, Shamma SA. 2005. Multiresolution spectrotemporal analysis of complex sounds. J Acoust Soc Am. 118:887–906.

Cohen YE, Theunissen F, Russ BE, Gill P. 2007. Acoustic features of rhesus vocalizations and their representation in the ventrolateral prefrontal cortex. J Neurophysiol. 97:1470–1484.

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.

Dau T, Kollmeier B, Kohlrausch A. 1997. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. J Acoust Soc Am. 102:2906–2919.

De Martino F, Moerel M, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E. 2013. Spatial organization of

frequency preference and selectivity in the human inferior colliculus. Nat Commun. 4:1386.
De Martino F, Moerel M, Xu J, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E. 2015. High-resolution mapping of myeloarchitecture in vivo: localization of auditory areas in the human brain. Cereb Cortex. 25:3394–3405.
Drullman R, Festen JM, Plomp R. 1994. Effect of temporal envelope smearing on speech reception. J Acoust Soc Am. 95: 1053–1064.
Ekstrom LB, Roelfsema PR, Arsenault JT, Bonmassar G, Vanduffel W. 2008. Bottom-up dependent gating of frontal signals in early visual cortex. Science. 321:414–417.
Elliott TM, Theunissen FE. 2009. The modulation transfer function for speech intelligibility. PLoS Comput Biol. 5:e1000302.
Fize D, Vanduffel W, Nelissen K, Denys K, Chef d'Hotel C, Faugeras O, Orban GA. 2003. The retinotopic organization of primate dorsal V4 and surrounding areas: a functional magnetic resonance imaging study in awake monkeys. J Neurosci. 23:7395–7406.
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.
Friston KJ, Frith CD, Turner R, Frackowiak RS. 1995. Characterizing evoked hemodynamics with fMRI. Neuroimage. 2:157–165.
Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. 2000. Representation of the temporal envelope of sounds in the human brain. J Neurophysiol. 84:1588–1598.
Giraud AL, Poeppel D. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 15:511–517.
Glasser MF, Van Essen DC. 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J Neurosci. 31:11597–11616.
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.
Golub G, Heath M, Wahba G. 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics. 21:215–223.
Hackett TA. 2011. Information flow in the auditory cortical network. Hear Res. 271:133–146.
Hackett TA, Preuss TM, Kaas JH. 2001. Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. J Comp Neurol. 441:197–222.
Hauser MD, Chomsky N, Fitch WT. 2002. The faculty of language: what is it, who has it, and how did it evolve? Science. 298:1569–1579.
Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 12:55–67.
Hoerl AE, Kennard RW. 1990. Ridge-regression—degrees of freedom in the analysis of variance. Commun Stat Simulat. 19: 1485–1495.
Hullett PW, Hamilton LS, Mesgarani N, Schreiner CE, Chang EF. 2016. Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. J Neurosci. 36:2014–2026.
Humphries C, Liebenthal E, Binder JR. 2010. Tonotopic organization of human auditory cortex. Neuroimage. 50:1202–1211.
Janssens T, Keil B, Farivar R, McNab JA, Polimeni JR, Gerits A, Arsenault JT, Wald LL, Vanduffel W. 2012. An implanted 8-channel array coil for high-resolution macaque MRI at 3T. Neuroimage. 62:1529–1536.
Joly O, Baumann S, Balezeau F, Thiele A, Griffiths TD. 2014. Merging functional and structural properties of the monkey auditory cortex. Front Neurosci. 8:198.
Joly O, Ramus F, Pressnitzer D, Vanduffel W, Orban GA. 2012. Interhemispheric differences in auditory processing revealed by fMRI in awake rhesus monkeys. Cereb Cortex. 22:838–853.
Joris PX, Schreiner CE, Rees A. 2004. Neural processing of amplitude-modulated sounds. Physiol Rev. 84:541–577.
Kay KN, Naselaris T, Prenger RJ, Gallant JL. 2008. Identifying natural images from human brain activity. Nature. 452: 352–355.
Kay KN, Rokem A, Winawer J, Dougherty RF, Wandell BA. 2013. GLMdenoise: a fast, automated technique for denoising task-based fMRI data. Front Neurosci. 7:247.
Kim JJ, Crespo-Facorro B, Andreasen NC, O'Leary DS, Zhang B, Harris G, Magnotta VA. 2000. An MRI-based parcellation method for the temporal lobe. Neuroimage. 11:271–288.
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.
Kusmierek P, Rauschecker JP. 2009. Functional specialization of medial auditory belt cortex in the alert rhesus monkey. J Neurophysiol. 102:1606–1622.
Kusmierek P, Rauschecker JP. 2014. Selectivity for space and time in early areas of the auditory dorsal stream in the rhesus monkey. J Neurophysiol. 111:1671–1685.
Langers DR, van Dijk P. 2012. Mapping the tonotopic organization in human auditory cortex with minimally salient acoustic stimulation. Cereb Cortex. 22:2024–2038.
Langner G, Sams M, Heil P, Schulze H. 1997. Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. J Comp Physiol [A]. 181:665–676.
Leaver AM, Rauschecker JP. 2016. Functional topography of human auditory cortex. J Neurosci. 36:1416–1428.
Liang L, Lu T, Wang X. 2002. Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J Neurophysiol. 87:2237–2261.
Luo H, Poeppel D. 2012. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. Front Psychol. 3:170.
Massoudi R, Van Wanrooij MM, Van Wetter SM, Versnel H, Van Opstal AJ. 2014. Task-related preparatory modulations multiply with acoustic processing in monkey auditory cortex. Eur J Neurosci. 39:1538–1550.
Merzenich MM, Brugge JF. 1973. Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res. 50:275–296.
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.
Moerel M, De Martino F, Formisano E. 2014. An anatomical and functional topography of human auditory cortical areas. Front Neurosci. 8:225.
Moerel M, De Martino F, Santoro R, Ugurbil K, Goebel R, Yacoub E, Formisano E. 2013. Processing of natural sounds: characterization of multipeak spectral tuning in human auditory cortex. J Neurosci. 33:11888–11898.
Morel A, Garraghty PE, Kaas JH. 1993. Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. J Comp Neurol. 335:437–459.
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.
Nagarajan SS, Cheung SW, Bedenbaugh P, Beitel RE, Schreiner CE, Merzenich MM. 2002. Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. J Neurophysiol. 87:1723–1737.
Niessing J, Ebisch B, Schmidt KE, Niessing M, Singer W, Galuske RA. 2005. Hemodynamic signals correlate tightly with synchronized gamma oscillations. Science. 309:948–951.
Ortiz-Rios M, Azevedo FA, Kusmierek P, Balla DZ, Munk MH, Keliris GA, Logothetis NK, Rauschecker JP. 2017. Widespread and opponent fMRI signals represent sound location in macaque auditory cortex. Neuron. 93:971–983.e974.
Ortiz-Rios M, Kusmierek P, DeWitt I, Archakov D, Azevedo FA, Sams M, Jaaskelainen IP, Keliris GA, Rauschecker JP. 2015. Functional MRI of the vocalization-processing network in the macaque brain. Front Neurosci. 9:113.
Overath T, McDermott JH, Zarate JM, Poeppel D. 2015. The cortical analysis of speech-specific temporal structure revealed by responses to sound quilts. Nat Neurosci. 18:903–911.
Overath T, Zhang Y, Sanes DH, Poeppel D. 2012. Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: fMRI evidence. J Neurophysiol. 107:2042–2056.
O'Connor KN, Johnson JS, Niwa M, Noriega NC, Marshall EA, Sutter ML. 2011. Amplitude modulation detection as a function of modulation frequency and stimulus duration: comparisons between macaques and humans. Hear Res. 277:37–43.
Petkov CI, Bendor D. 2016. Neuronal mechanisms and transformations encoding time-varying signals. Neuron. 91:718–721.
Petkov CI, Kayser C, Augath M, Logothetis NK. 2006. Functional imaging reveals numerous fields in the monkey auditory cortex. PLoS Biol. 4:e215.
Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. 2008. A voice region in the monkey brain. Nat Neurosci. 11:367–374.
Pfingst BE, Laycock J, Flammino F, Lonsbury-Martin B, Martin G. 1978. Pure tone thresholds for the rhesus monkey. Hear Res. 1:43–47.
Rauschecker JP. 1997. Processing of complex sounds in the auditory cortex of cat, monkey, and man. Acta Otolaryngol Suppl. 532:34–38.
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.
Rauschecker JP, Tian B, Pons T, Mishkin M. 1997. Serial and parallel processing in rhesus monkey auditory cortex. J Comp Neurol. 382:89–103.
Rodriguez FA, Chen C, Read HL, Escabi MA. 2010. Neural modulation tuning characteristics scale to efficiently encode natural sound statistics. J Neurosci. 30:15969–15980.
Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E. 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. PLoS Comput Biol. 10:e1003412.
Santoro R, Moerel M, De Martino F, Valente G, Ugurbil K, Yacoub E, Formisano E. 2017. Reconstructing the spectrotemporal modulations of real-life sounds from fMRI response patterns. Proc Natl Acad Sci USA. 114:4799–4804.
Scheeringa R, Fries P, Petersson KM, Oostenveld R, Grothe I, Norris DG, Hagoort P, Bastiaansen MC. 2011. Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. Neuron. 69:572–583.
Schonwiesner 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.
Schreiner CE, Langner G. 1988. Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. J Neurophysiol. 60:1823–1840.
Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. Science. 270:303–304.
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.
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.
Theunissen FE, Elie JE. 2014. Neural processing of natural sounds. Nat Rev Neurosci. 15:355–366.
Theunissen FE, Sen K, Doupe AJ. 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci. 20:2315–2331.
Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. 2001. Functional specialization in rhesus monkey auditory cortex. Science. 292:290–293.
Vanduffel W, Fize D, Mandeville JB, Nelissen K, Van Hecke P, Rosen BR, Tootell RB, Orban GA. 2001. Visual motion processing investigated using contrast agent-enhanced fMRI in awake behaving monkeys. Neuron. 32:565–577.
Vanduffel W, Fize D, Peuskens H, Denys K, Sunaert S, Todd JT, Orban GA. 2002. Extracting 3D from motion: differences in human and monkey intraparietal cortex. Science. 298:413–415.
Vanduffel W, Zhu Q, Orban GA. 2014. Monkey cortex through fMRI glasses. Neuron. 83:533–550.
Viemeister NF. 1979. Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am. 66:1364–1380.
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.
Wilson B, Kikuchi Y, Sun L, Hunter D, Dick F, Smith K, Thiele A, Griffiths TD, Marslen-Wilson WD, Petkov CI. 2015. Auditory sequence processing reveals evolutionarily conserved regions of frontal cortex in macaques and humans. Nat Commun. 6:8901.
Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X. 2010. Functional properties of human auditory cortical fields. Front Syst Neurosci. 4:155.
Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH. 2006. Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. J Neurophysiol. 96:3323–3337.
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.