Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI

Marc Schönwiesner^{a,b,1} and Robert J. Zatorre^{b,c}

^aDepartment of Psychology, University of Montreal, Montreal, Canada H2V 4P2; ^bLaboratory for Brain, Music, and Sound Research, Montreal, Quebec, Canada H2V 4P2; and ^cMontreal Neurological Institute, McGill University, Montreal, Quebec, Canada H3A 2B4

Communicated by Brenda Milner, McGill University, Montreal, Canada, July 12, 2009 (received for review January 13, 2008)

Are visual and auditory stimuli processed by similar mechanisms in the human cerebral cortex? Images can be thought of as light energy modulations over two spatial dimensions, and low-level visual areas analyze images by decomposition into spatial frequencies. Similarly, sounds are energy modulations over time and frequency, and they can be identified and discriminated by the content of such modulations. An obvious question is therefore whether human auditory areas, in direct analogy to visual areas, represent the spectro-temporal modulation content of acoustic stimuli. To answer this question, we measured spectro-temporal modulation transfer functions of single voxels in the human auditory cortex with functional magnetic resonance imaging. We presented dynamic ripples, complex broadband stimuli with a drifting sinusoidal spectral envelope. Dynamic ripples are the auditory equivalent of the gratings often used in studies of the visual system. We demonstrate selective tuning to combined spectro-temporal modulations in the primary and secondary auditory cortex. We describe several types of modulation transfer functions, extracting different spectro-temporal features, with a high degree of interaction between spectral and temporal parameters. The overall low-pass modulation rate preference of the cortex matches the modulation content of natural sounds. These results demonstrate that combined spectro-temporal modulations are represented in the human auditory cortex, and suggest that complex signals are decomposed and processed according to their modulation content, the same transformation used by the visual system.

auditory system | cortical representation | dynamic ripples

A major goal of neuroscience is to discover mechanisms of encoding and recoding of sensory information in the brain. Many of these mechanisms have first been described in the visual system. It is unclear in how far auditory and visual processing is based on common principles, but a variety of intriguing similarities suggest that such principles exist: conservation of topography of the sensory epithelium (retinotopy, tonotopy) in low level structures, layout of higher structures in broad cortical processing streams (1, 2), common principles of perceptual organization such as grouping and illusory continuity (3), similar motion after effects (4, 5), and similar organization of memory (6). An established model of visual object recognition is a processing hierarchy in which successive stages are selective for increasingly complex features by combining the output of simpler feature detectors, starting with patchy spatial modulation frequency filters (7). In auditory neuroscience there is no consensus yet about the suitable set of low-level features. If the analogy to the visual system holds, then spectro-temporal modulation rate detectors are likely to be included in this set. Selectivity to other types of modulations has been studied in some detail in animal models and humans, especially for frequency and amplitude modulations (8–15). Selectivity for combined spectro-temporal modulations is inherently better suited to capture features of natural time-varying sounds than frequency and amplitude

modulation separately. In visual neuroscience sinusoidal luminance gratings have been used extensively to characterize spatial frequency tuning (7, 16, 17). An auditory equivalent of a visual grating is the dynamic ripple, a complex broadband stimulus with a sinusoidal spectral envelope that drifts along the logarithmic frequency axis over time (18–20). Dynamic ripples represent some of the spectro-temporal complexity of ecologically relevant sounds, whereas at the same time satisfying the formal requirements for deriving receptive fields (21). If auditory cortical cells act as spectro-temporal modulation rate filters, then their response to dynamic ripples should provide an excellent functional characterization. Dynamic ripples are useful in measuring spectro-temporal modulation transfer functions (MTFs) and spectro-temporal receptive fields (STRF, the two-dimensional Fourier transform of the MTF; 22) of neurons in the inferior colliculus, thalamus, and auditory cortex of various species. Dynamic ripples have been used in neurophysiological studies to characterize responses across cortical auditory fields in various species (23–26), to demonstrate short-term task-dependent plasticity (27–29), to study spectro-temporal modulation tuning as a mechanism for auditory object discrimination (26), and to predict responses to novel sounds (30, 31). Langers and colleagues (32) found dynamic ripples to be effective stimuli in a human neuroimaging experiment, but could not construct MTFs because of the low number of stimuli.

Here we apply the measurement of MTFs to the human auditory cortex using fMRI to characterize and map its spectro-temporal response properties, and provide a first step to relate neuroimaging results in humans to the available cellular data. The motivation for this study is to adapt a quantitative method from single unit neurophysiology to the human auditory system that can then be applied to questions of, for instance, learning and plasticity that are difficult or impossible to address in animals. Predictions can be made about MTF features in the human auditory cortex on the basis of its responses to other complex sounds and results from electrophysiological recordings in animals: Cortical neurons prefer relatively low temporal modulation frequency (13, 33). Modulation detection thresholds measured with dynamic ripples in humans also show a low-pass function (34). Average cortical and behavioral modulation rate preference has been linked to the low-pass amplitude modulation content of natural sounds (26, 34–36). We thus expect that the average human cortical MTF will show a low-pass filter characteristic. The majority of measured STRFs in the mammalian auditory cortex have relatively simple shapes, consisting

Author contributions: M.S. and R.J.Z. designed research; M.S. performed research; M.S. analyzed data; and M.S. and R.J.Z. wrote the paper.

The authors declare no conflict of interest.

¹To whom correspondence should be addressed at: Laboratory for Brain, Music and Sound Research, Pavilion 1420 Mont-Royal, Montreal, Quebec, Canada H2V 4P3. E-mail: marc.schoenwiesner@umontreal.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907682106/DCSupplemental.

of a low number of excitatory and inhibitory domains (33). In a neuroimaging experiment, one voxel necessarily includes a great number of neurons, and we thus expect to measure complex MTFs resulting from superposition of many MTFs. If MTFs are correlated over small areas of cortex, we would nevertheless expect to measure consistent MTFs at the voxel level.

Results

We presented dynamic ripples of varying spectral and temporal modulation rates to healthy human listeners and measured cortical responses with high-resolution functional magnetic resonance imaging. From those responses we reconstructed two-dimensional spectro-temporal modulation transfer functions of individual voxels in the auditory cortex.

Modulation Transfer Functions. The auditory cortex on the superior temporal plane responded highly selectively to different combinations of spectral and temporal modulation rates. The responses to all ripples measured in a given voxel were assembled into a square matrix with responses to increasing spectral modulation rates (Ω) ordered along the vertical axis, and responses to increasing temporal modulation rates (ω) along the horizontal axis (Fig. 1C). This representation of the response magnitude as a function $R(\Omega, \omega)$ is the modulation transfer function, or ripple transfer function (MTF), of a voxel. A total of 18,759 MTFs were obtained in the experiment, one from each voxel exceeding a statistical threshold of $P < 0.0001$ in the sound vs. silence contrast.

MTFs of individual voxels exhibited a variety of spectro-temporal response patterns. Fig. 1D–H shows exemplars of MTFs that represent frequent patterns (chosen by an unsupervised classification algorithm): focal responses with a single peak of varying width (Fig. 1D), broad tuning along one axis but sharp tuning along the other axis (Fig. 1E and F), response structures that traverse diagonally across spectral and temporal modulation rates (Fig. 1G), and disjoint MTFs with multiple peaks (Fig. 1H).

We verified this analysis using conventional k-means clustering. Mean cluster MTFs fell into the same categories (Fig. S1). The percentages of voxels in the categories were: focal response 29%, broad spectral and narrow temporal pass-band 18.7%, broad temporal and narrow spectral pass-band 19.8%, diagonal 14.7%, and disjoint 17.8%.

We quantified the reliability of voxel MTFs against measurement noise as well as the stability of MTFs over time by computing the correlation between MTFs in the even and odd trials as well as between the first and second scanning session for each voxel in each individual. All individuals showed a high degree of reproducibility in both cases (Fig. S2A and B).

To further test our main hypothesis of selective tuning to spectro-temporal modulations in the human auditory cortex we counted the number of voxels responding best for each of the ripple conditions. For each ripple, between 35 and 2,300 voxels were found that responded selectively to that stimulus (Fig. S3A). To quantify the robustness of these results we generated, for all 49 ripple conditions, cross-validated average MTFs across all voxels with a given preferred ripple. The preferred ripple was found in one half of the data (even trials) and MTFs were extracted and averaged in the other half of the data. A clear response peak was found in the mean MTF for nearly every ripple condition, despite the low number of stimulus repetitions in each data half (Fig. S4).

Response Properties. We derived three descriptors of the responses from the individual MTFs: selectivity, inseparability, and direction preference. Response selectivity was quantified as the difference in response magnitude between the preferred ripple and the mean of the remaining responses (excluding neighbors). The measure was cross-validated by finding the

Fig. 1. Construction of MTFs. (A) Dynamic ripple stimuli consist of sinusoidal modulations along time and logarithmic frequency. (B) Each square is a spectrogram of a dynamic ripple (same format as in A), ordered along the spectral density (Ω) and temporal (ω) modulation rate axes. Experimental conditions were constructed by a 2-by-2 stimuli running average (black square with arrow), resulting in 49 conditions on a 7-by-7 grid. (C) Color-coded response magnitudes in a single voxel to all 49 ripple conditions, ordered according to the stimulus grid. This representation is the 2D modulation transfer function (MTF) of the voxel. (D–H) Exemplars of MTFs of single voxels in individuals, selected unsupervised by affinity propagation, and loosely ordered by response shape. (D) focal responses, (E) responses with weak selectivity along the temporal axis, and high selectivity along the spectral axis, (F) vice versa, (G) diagonal response pattern, (H) disjoint responses.

preferred ripple in one half of the data and computing selectivity in the other half. Response selectivity ranged from 0.25 to 0.52 (95% confidence interval) with a median of 0.37 (Fig. 2A; individual data in Fig. S2C).

A characteristic of an MTF that can be readily compared with data from single neurons is the degree of separability. A separable MTF can be fully described by its spectral and temporal components. A low degree of separability implies a significant interaction between spectral density and temporal modulation rate. We quantified the degree to which an MTF can be separated using an index of inseparability developed by (37) for ripple transfer functions in single auditory neurons. The index (α_{SVD}) increases from 0 to 1 with MTF inseparability. Our α_{SVD} ranged from 0.17 to 0.6 (95% confidence interval) with a median of 0.42 (Fig. 2B; individual data in Fig. S2D). The distribution of highly inseparable regions across the cortical surface was patchy, appeared inconsistent across participants and hemispheres, and did not correlate with the distribution of other parameters, except for a trend for larger α_{SVD} in voxels with high modulation rate preference (Fig. S3B). To test for

Fig. 2. Ripple response properties: histograms of (A) response selectivity, (B) inseparability index, and (C) ripple direction preference. The overwhelming preference for rising ripples is likely the result of a long-term adaptation to the large number of falling ripples in the current experiment. (D) The statistical t map of activation produced by the preferred ripple vs. silence superimposed on a rendering the average left and right temporal lobe surface. In (A–C), gray lines indicate the resampling distribution of the parameters in 10,000 random permutations of responses within and between participants (mean and standard error). (E) The average modulation transfer function of the human auditory cortex exhibits a spectral and temporal low-pass characteristic. The color code shows the mean percentage BOLD signal change across all active voxels for each ripple condition.

selectivity to the ripple direction, we compared the responses to the 8 rising ripples that were included in the experiment to the eight matched falling ripples (Fig. 2C). Practically all voxels responded stronger to the rising than to the falling ripples. Several participants reported a perceptual “pop out” of the few rising ripples. This strongly indicates a confound because of long-term adaptation to the more numerous falling ripples, and we did not analyze this parameter further. Nevertheless, direction-specific adaptation is qualitative evidence for ripple direction tuning.

Fig. 2D shows the average MTF across all voxels and participants. The dominant feature is a low pass modulation rate filter in spectral and temporal directions with a drop of response energy above ≈ 10 Hz and 1.33 c/o. This low-pass characteristic approximately matches the modulation rate content of speech (36) and detection thresholds for spectro-temporal modulations (34).

Topography of the Responses. Responses to the ripple stimuli originated from primary and secondary auditory cortex on and around Heschl’s gyrus (HG, Fig. 2D; individual t maps in Fig. S5A). There was practically no significant response from higher auditory cortices, such as posterior planum temporale (PT) or superior temporal sulcus.

One of our initial hypotheses was about differences in tuning to spectral and temporal modulation rates between primary and higher auditory fields and between the hemispheres. To visualize the regional distribution of tuning characteristics, we mapped preferred ripples onto temporal lobe surfaces (Fig. 3). The group map and about half of the individual hemispheres show a tendency for a gradient from high preferred temporal rates on medial HG to high spectral rates on lateral HG (maps split for preferred spectral density and temporal rate available as Fig. S5B and C).

To quantify these differences, we defined regions of interest based on individual macroanatomy covering the left and right medial two-thirds of HG, PT, and planum polare including lateral HG. We calculated the mean preferred modulation rates across active voxels in these regions and tested for significant differences between regions by randomly exchanging ROI labels of voxels 10,000 times to sample the distribution of rate differences (permutation test, all significant differences reported at $P < 0.05$). The preferred temporal rate was highest in medial HG (3.7 Hz) and dropped significantly toward the PT (2.8 Hz) and planum polare (3.1 Hz). The preferred spectral density, on the other hand, showed no difference between medial HG and PT (0.8 c/o), and increased significantly toward the planum polare/lateral HG region (1.1 c/o). Splitting these responses according to hemisphere showed a significant difference in the preference for higher spectral rates in the right lateral HG and planum polare (1.2 c/o) than in the left one (0.99 c/o, $P < 0.0001$). No significant regional differences were found in ripple selectivity or inseparability index. These parameters nevertheless showed variations within the macroanatomical regions that were highly consistent across even and odd trials within each individual.

Discussion

We demonstrated selective tuning to combined spectro-temporal modulations in the human auditory cortex. The high degree of selectivity suggests that conjunctions of spectral and temporal modulations play an important role in auditory processing.

Some important caveats should be mentioned: when comparing neuronal MTFs recorded in animal auditory cortex to MTFs measured here in the human auditory cortex, it should be kept in mind that the two methods measure different parts of the response. The hemodynamic signal is nonlinearly coupled to the local field potential (38) and spiking activity in the auditory

Fig. 3. Maps of the preferred ripple, color-coded with a logarithmic 2D color map superimposed on a rendering of the left and right temporal lobe surfaces. Areas with reddish colors respond best to high spectral densities (Ω); areas with bluish colors respond best to high temporal rates (ω). Green, corresponding to a conjunction of high spectral temporal rates, is relatively rare. In the group map and 8 of the 14 hemispheres (1R, 2R, 3R+L, 5R+L, and 7R+L) reddish colors (higher spectral densities) are more likely found around lateral HG, whereas the purple that marks low spectro-temporal rates appears predominantly on the medial end of HG.

cortex (39), but we would argue that much of our analysis only requires a monotonic, not a linear, coupling between the neural response and the BOLD signal to allow inference of spectro-temporal tuning properties from measured voxel MTFs. The classification of the MTF shape by k-means clustering and affinity propagation as well as the preferred ripple, and preferred spectral density and temporal modulation rates are independent of the shape of a monotonic coupling function. The calculation of selectivity and inseparability is not as robust, but we drew inferences only from the distribution of these parameters across all voxels.

fMRI has a low spatial resolution and we consequently record population activity within voxels. If MTFs of auditory neurons within a voxel were entirely uncorrelated the resulting voxel MTF would be flat. The fact that we were able to observe relatively focal voxel MTFs indicates that neuronal response properties are coherent over small patches of auditory cortex. This conclusion is supported by recordings in animals (40, 41) and cytoarchitectonical studies in humans (42, 43), suggesting that the auditory cortex is organized in architectonic and functional modules that comprise large numbers of neurons with similar response properties.

The BOLD signal also has low temporal resolution. With direct electrode recordings it is possible to record the magnitude and phase of the spike rate fluctuation phase-locked to the temporal modulation rate of the ripples, whereas we measured a correlate of the long-term average magnitude of the response. However, the temporal resolution of BOLD fMRI does not limit our ability to measure responses to high modulation rates; the drop-off in the average MTF at higher modulation rates reflects cortical sensitivity, not methodological constraints. An advantage of our approach is the ability to collect data from all auditory regions simultaneously, something impossible to achieve in single-unit neurophysiology. Given the methodological differences is not surprising that neuroimaging has had limited success in reproducing findings from cellular electrophysiology. Our results illustrate a successful adaptation of an experimental paradigm from electrophysiology to neuroimaging.

Cortical Low Modulation Rate Preference Matches That of Human Habitat Sounds. The majority of measured MTFs show selectivity to low spectral or temporal modulation rates, and only a few voxels responded selectively to combinations of high spectral and temporal modulation rates. The average cortical modulation rate preference shows a low-pass characteristic along the spectral and temporal dimensions that is comparable with the distribution of spectro-temporal modulation rates in speech and environmental sounds (36) and to detection thresholds for spectro-temporal modulations (34). The result also fits with the preferences for low rates in amplitude modulated sounds (13, 14, 44). Low-pass modulation selectivity was also observed in animal recordings and likened to the prevalence of low modulation rates in natural and communication sounds (26, 35, 36). A match between neural response properties and the statistics of relevant stimuli indicates efficient coding (45), and it has been demonstrated for various statistical parameters in the visual and auditory modalities of different species (46–50). Such a match in the human data suggests that the human auditory system has increased sensitivity to commonly encountered modulations that allows their efficient encoding. Our data do not specify whether this match occurs because of evolutionary, developmental, or experience-dependent mechanisms. These possibilities can be explored by mapping responses and manipulating participants' acoustic input.

MTF Shapes. We observed a wide variety of spectro-temporal response patterns in individual voxels. The measured MTFs, as well as all response parameters derived from them, were highly reliable and repeatable in cross-validation and resampling tests.

We can therefore be reasonably confident that the MTFs reported here do not represent noise, but rather the underlying spectro-temporal tuning of patches of human auditory cortex.

On average our MTFs show a higher degree of inseparability (0.42) than the ones measured in single units (ferret: ≈ 0.25 ; Fig. 13 in 37 but note that this is a plot of separability within and across quadrants whereas we plot quadrant separability only). Simon and colleagues (51, their Table 1) report 85% of separable STRFs in the primary auditory cortex of awake ferrets. Only $\approx 20\%$ of voxel MTFs in awake humans are clearly separable ($\alpha_{\text{SVD}} < 0.3$). The higher degree of inseparability in the human auditory cortex may partly be the result of the vastly higher spectro-temporal complexity of speech compared with animal vocalizations. It may also partly result from the superposition of MTFs, which inherently increases inseparability. A methodological consequence of high inseparability is that measuring spectral and temporal transfer functions separately (as done in previous neuroimaging studies) will fail to capture the high degree of spectro-temporal interaction on the voxel level in the human auditory cortex.

Topography of Spectro-Temporal Tuning. The four main results on the distribution of MTF features across the cortical surface are: a) responses are mainly from primary and secondary, but not higher auditory cortex, b) there is no topographic map of spectral or temporal tuning, c) the topography of the responses is surprisingly different between individual brains, and d) there are differences in the average modulation frequency tuning between cortical areas and hemispheres that agree with previous results.

Dynamic ripples elicited strong responses from the primary and secondary auditory cortex on and around HG, but no or very little response from the posterior PT and superior temporal gyrus or sulcus. This is notable, because spectro-temporally structured broadband sounds typically activate those regions strongly. The lack of activity in higher areas may be because of at least two differences between dynamic ripples and natural sounds: the lower acoustical complexity, and the absence of behavioral significance. Concerning the first point, higher areas might integrate information across the modulation spectrum by summing responses units with simple MTFs, and thus would not respond strongly to single ripples. This integration does not appear to happen at the transition between primary and secondary auditory cortex, because it would be accompanied by an increase in inseparability, which we did not observe. Concerning the second point, higher auditory areas do not faithfully represent the physical properties of sounds, but rather the relation between a sound and its behavioral implications. Those areas respond much stronger to sounds that have some behavioral significance than to meaningless sounds that are acoustically matched or identical (52, 53).

We did not detect simple large-scale gradients of ripple preference along macroanatomical features. The same is true for the other MTF parameters, response selectivity and inseparability. However, ripple preference and the other parameters did fluctuate across the cortical surface on a finer scale. The spatial distribution of this fluctuation was stable against cross-validation, and may thus reflect genuine small-scale changes in these parameters across the auditory cortex, perhaps relating to clusters within isofrequency contours in tonotopically organized auditory cortex (reviewed in ref. 54). A similar organizational principle has been observed in the visual system, where a large-scale retinotopic map is superimposed by smaller-scale feature maps—ocular dominance stripes, pinwheels of orientation selective cells, and cytochrome oxidase blobs.

Even though similar regions of auditory cortex were active in all individual brains, the spatial distributions of the preferred modulation frequency are very different. It is possible that the exact layout of the map is determined during development or by environmental interactions.

We found small variations in the average modulation tuning in different cortical regions, consistent with previous findings. The mean preferred temporal modulation rate dropped from the primary to secondary auditory cortex, whereas the spectral density increased toward secondary auditory cortex on lateral HG, especially on the right side. The lateral HG has been implicated with complex pitch processing (55–57), a function that might benefit from sensitivity to finer spectral detail. The direction of this asymmetry (favoring the right) agrees with previous results (58–61), but it is important to note that the differences observed here cancel out almost exactly in the average over all of auditory cortex. The regional differences in modulation tuning may therefore be more accurately described as specializations of certain auditory fields rather than global hemispheric specializations. In the only previous neuroimaging study using dynamic ripples (32), all auditory cortex responded maximally to the lowest presented combination of spectral and temporal modulation rates, and regional variation was only observed after normalizing the responses in each of the ripple conditions to their maximum across auditory cortex. The gradients found after normalization are similar to the large-scale pattern reported here—higher temporal modulation rates around medial HG, and high spectral density around anterolateral HG.

Modulation Tuning as Mechanism for Sensory Encoding. The tuning to spectro-temporal modulation frequencies in the primary and secondary auditory cortex is comparable with the tuning for spatial frequencies in low-level visual cortex, suggesting a common organizational principle. This similarity goes deeper than the conceptual level; as observed by Shamma (62), many of the filter characteristics and detection thresholds are remarkably well matched between the two modalities. In the visual system spatial frequency filters appear to be an efficient way to capture the salient (statistically independent) features of natural images. Independent component analysis applied to natural images results in such filters (63). A similar analysis of speech sounds results in comparable filters in the time-frequency domain, localized ripple-like patches (64). From a theoretical point of view, the independent components of a stimulus set provide a sparse and thus efficient encoding mechanism, independent of the modality of the stimulus set. The use of a common neural encoding strategy in the visual and auditory system via the activity of modulation band filters would also facilitate the combination of information from both modalities. In fact, exactly this strategy of independent-component-based encoding has been proposed by engineers as a solution for so-called sensory fusion, when different types of data, for instance sounds and images, have to be combined to enable transfer over a common channel (65). Furthermore, spectro-temporal modulation filters enable a multiscale encoding of local dynamic sound features (filters are sensitive to different timescales and spectral density scales). This mechanism provides a natural explanation for perceptual spectral and temporal scale invariances in audition (for review and a mathematical formulation of multiscale auditory coding see 62).

In conclusion, the high degree of modulation rate selectivity in primary and secondary auditory cortex, the different types of MTFs extracting different spectro-temporal features, and the match between the overall selectivity and the modulation content of human habitat sounds suggest that spectro-temporal modulation rate analysis plays a key role in cortical sound processing. These results extend the concept of modulation frequency selectivity developed in theoretical and electrophysiological studies in animal models to humans. The developed method is well suited to study mechanisms of human cortical reorganization related to, for instance, attention or musical training.

Methods

Participants and Stimuli. Seven people (5 male) between 22 and 31 years of age, with no history of hearing disorder or neurological disease, participated in the experiment after having given informed consent. The experimental procedures were approved by the MNI ethics committee.

Dynamic ripple stimuli with a bandwidth of 5 octaves (150 Hz to 4.8 kHz) and a modulation depth of 0.9 were generated according to (37, Eq. 1; see also *SI Text*). Falling ripple stimuli (spectral profile drifts downward over time) for all combinations of the following values were generated: ripple velocities (ω) = 1.6, 2.4, 3.6, 5.4, 8.0, 12.2, 18.2, and 27.4 Hz, and ripple densities (Ω) = 0.16, 0.4, 0.8, 1.2, 1.5, 1.8, 2.2, and 2.5 cycles/octave (Fig. 1). To detect ripple direction tuning 8 additional rising ripples were generated. During analysis we found that the unequal number of falling and rising ripples introduces a strong confound because of long-term adaptation. The responses to the rising ripples were therefore not suitable to calculate ripple direction selectivity. The total of 72 dynamic ripple stimuli, each 4 s long and gated with 10 ms raised-cosine ramps, were constructed off-line using the software package Matlab (The MathWorks Inc.) at 22.05-kHz sampling rate. Stimuli were played back binaurally via MR-compatible high-fidelity headphones (MR Confon) at 70 to 80 dB SPL, adjusted for individual comfort level.

Procedure. Each acquisition of a functional volume was followed by a silent period of 3 s, then a stimulus was played for 4 s, and then the following functional volume was acquired in 1 s, completing an 8 s repetition interval ('sparse imaging', see *SI Text*). The stimuli were organized in blocks of five repetitions (40 s) of either the same ripple stimulus or silence. Each stimulus block was presented twice during the experiment (10 presentations of each stimulus); the silent baseline block was repeated eight times (40 presentations of silence). These blocks were played in pseudorandom order with balanced transition probabilities. We presented 760 repetition intervals in total ($8 \times 8 + 8 = 72$ stimuli $\times$ 10 repeats + 40 silent trials). These were split evenly into two sessions with two runs each. Including the structural and preparatory scans, each session lasted ≈ 75 min. Participants were taken out of the scanner for ≈ 90 min between sessions. To control participants' attention inside the scanner they were instructed to listen passively to the sounds and focus on a visual task unrelated to sound presentation. Participants fixated gaze on a cross in the center of the visual field. The cross was infrequently (on average once every 20 s) replaced for 80 ms by either a digit or, with 20% probability, the letter Z, in which case participants were instructed to press a button.

Imaging Protocol. Functional imaging was performed on a 3 Tesla scanner (Trio, Siemens) using an echo-planar imaging sequence with sparse sampling to avoid acoustical noise artifacts (gradient echo; repetition time: 8 s, echo time: 36 ms; flip angle: 90°). Thirteen slices (resolution 1.5×1.5 mm, 2.5-mm thick, 192-mm field of view) were oriented parallel to the lateral sulcus to cover the superior temporal plane. This orientation includes HG, PT, planum polare, and the superior temporal gyrus and sulcus. A standard whole-brain structural volume with 1 mm³ voxels was also obtained.

Data Analysis. Functional data were corrected for motion and blurred with a 2-mm Gaussian kernel. Statistical analysis was based on general linear models as implemented in the FMRistat toolbox (66). Regions of significant activation were determined by comparing the response in the sound conditions to the silent baseline condition. We extracted relative response magnitudes to the different stimulus conditions in each voxel by comparing each condition with the average of the other conditions. To gain sufficient signal-to-noise ratio, responses to 2×2 neighboring stimuli were averaged (Fig. 1B). This resulted in 49 sample points in a 7-by-7 grid on the $[\Omega, \omega]$ plane, with 40 repetitions per condition. Responses were ordered in a 2D matrix, $R(\Omega, \omega)$, according to the position of the stimuli in the ripple parameter plane—this representation is the magnitude of the 2D spectro-temporal modulation transfer function (MTF).

MTFs were classified according to shape unsupervised by k-means clustering and affinity propagation (see *SI Text*). The spatial distribution of MTF parameters was visualized on cortical surface renderings of individual brains and a group-average surface. We quantified regional differences of MTF parameters by manually segmenting the superior temporal plane into three regions of interests (ROIs) in both hemisphere and comparing the mean parameter value across these regions. We segmented the medial two-thirds of HG (according to 67, 68) as approximation of the primary auditory field (69); the PT (according to 70), and the remaining anterior superior temporal plane (planum polare and lateral HG).

ACKNOWLEDGMENTS. We like to thank the members of the McConnel Brain Imaging Centre where the scanning was done. This work was supported by the

Canadian Institutes of Health Research and the German National Academy of Sciences.

- Mishkin M, Ungerleider LG, Macko KA (1983) Object vision and spatial vision: Two cortical pathways. *Trends Neurosci* 6:414–417.
- 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.
- Bregman AS (1990) in *Auditory Scene Analysis: The Perceptual Organization of sound* (The MIT Press, Cambridge, MA).
- Shu ZJ, Swindale NV, Cynader MS (1993) Spectral motion produces an auditory after-effect. *Nature* 364:721–723.
- Barlow HB, Hill RM (1963) Evidence for a physiological explanation of the waterfall phenomenon and figural after effects. *Nature* 200:1345–1347.
- Visscher KM, Kaplan E, Kahana MJ, Sekuler R (2007) Auditory short-term memory behaves like visual short-term memory. *PLoS Biol* 5:e56.
- De Valois KK, De Valois RL, Yund EW (1979) Responses of striate cortex cells to grating and checkerboard patterns. *J Physiol* 291:483–505.
- Bieser A, Muller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: Neural responses to amplitude-modulated sounds. *Exp Brain Res* 108:273–284.
- Kay RH, Matthews DR (1972) On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. *J Physiol* 225:657–677.
- Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. *Physiol Rev* 84:541–577.
- Eggermont JJ (1998) Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. *J Neurophysiol* 80:2743–2764.
- Eggermont JJ (2002) Temporal modulation transfer functions in cat primary auditory cortex: Separating stimulus effects from neural mechanisms. *J Neurophysiol* 87:305–321.
- Giraud AL, et al. (2000) Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 84:1588–1598.
- Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14:731–740.
- Schreiner CE, Urbas JV (1988) Representation of amplitude modulation in the auditory cortex of the cat. II. Comparison between cortical fields. *Hear Res* 32:49–63.
- Frazor RA, Albrecht DG, Geisler WS, Crane AM (2004) Visual cortex neurons of monkeys and cats: Temporal dynamics of the spatial frequency response function. *J Neurophysiol* 91:2607–2627.
- Kagan I, Gur M, Snodderly DM (2002) Spatial organization of receptive fields of V1 neurons of alert monkeys: Comparison with responses to gratings. *J Neurophysiol* 88:2557–2574.
- Kowalski N, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76:3503–3523.
- Kowalski N, Depireux DA, Shamma SA (1996) Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *J Neurophysiol* 76:3524–3534.
- Shamma SA, Versnel H, Kowalski N (1995) Ripple analysis in ferret primary auditory cortex. I. Response characteristics of single units to sinusoidally rippled spectra. *Aud Neurosci* 1:233–254.
- Klein DJ, Simon JZ, Depireux DA, Shamma SA (2006) Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *J Comput Neurosci* 20:111–136.
- Aertsen AM, Johannesma PI (1981) The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol Cybern* 42:133–143.
- Kowalski N, Versnel H, Shamma SA (1995) Comparison of responses in the anterior and primary auditory fields of the ferret cortex. *J Neurophysiol* 73:1513–1523.
- Linden JF, Liu RC, Sahani M, Schreiner CE, Merzenich MM (2003) Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol* 90:2660–2675.
- Sen K, Theunissen FE, Doupe AJ (2001) Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 86:1445–1458.
- 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.
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223.
- Fritz JB, Elhilali M, Shamma SA (2005) Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci* 25:7623–7635.
- Fritz J, Elhilali M, Shamma S (2005) Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res* 206:159–176.
- Versnel H, Shamma SA (1998) Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J Acoust Soc Am* 103:2502–2514.
- Andoni S, Li N, Pollak GD (2007) Spectrotemporal receptive fields in the inferior colliculus revealing selectivity for spectral motion in conspecific vocalizations. *J Neurosci* 27:4882–4893.
- Langers DR, Backes WH, van Dijk P (2003) Spectrotemporal features of the auditory cortex: The activation in response to dynamic ripples. *Neuroimage* 20:265–275.
- Miller LM, Escabi MA, Read HL, Schreiner CE (2002) Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol* 87:516–527.
- Chi T, Gao Y, Guyton MC, Ru P, Shamma S (1999) Spectro-temporal modulation transfer functions and speech intelligibility. *J Acoust Soc Am* 106:2719–2732.
- Escabi MA, Miller LM, Read HL, Schreiner CE (2003) Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus. *J Neurosci* 23:11489–11504.
- Singh NC, Theunissen FE (2003) Modulation spectra of natural sounds and ethological theories of auditory processing. *J Acoust Soc Am* 114:3394–3411.
- Depireux DA, Simon JZ, Klein DJ, Shamma SA (2001) Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol* 85:1220–1234.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
- Mukamel R, et al. (2005) Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309:951–954.
- Read HL, Winer JA, Schreiner CE (2001) Modular organization of intrinsic connections associated with spectral tuning in cat auditory cortex. *Proc Natl Acad Sci USA* 98:8042–8047.
- Schreiner CE, Read HL, Sutter ML (2000) Modular organization of frequency integration in primary auditory cortex. *Annu Rev Neurosci* 23:501–529.
- Galuske RA, Schlote W, Bratzke H, Singer W (2000) Interhemispheric asymmetries of the modular structure in human temporal cortex. *Science* 289:1946–1949.
- Tardif E, Clarke S (2001) Intrinsic connectivity of human auditory areas: A tracing study with Dil. *Eur J Neurosci* 13:1045–1050.
- Rees A, Green GG, Kay RH (1986) Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man. *Hear Res* 23:123–133.
- Barlow HB (1961) Possible principles underlying the transformation of sensory messages. In *Sensory Communication*, ed Rosenblith WA (MIT Press, Cambridge, MA), pp 217–234.
- Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. *J Opt Soc Am A* 4:2379–2394.
- Lewicki MS (2002) Efficient coding of natural sounds. *Nat Neurosci* 5:356–363.
- Machens CK, Gollisch T, Kolesnikova O, Herz AV (2005) Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron* 47:447–456.
- Rieke F, Bodnar DA, Bialek W (1995) Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory afferents. *Proc Biol Sci* 262:259–265.
- Schwartz O, Simoncelli EP (2001) Natural signal statistics and sensory gain control. *Nat Neurosci* 4:819–825.
- Simon JZ, Depireux DA, Klein DJ, Fritz JB, Shamma SA (2007) Temporal symmetry in primary auditory cortex: Implications for cortical connectivity. *Neural Comput* 19:583–638.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312.
- Lewis JW, et al. (2004) Human brain regions involved in recognizing environmental sounds. *Cereb Cortex* 14:1008–1021.
- Read HL, Winer JA, Schreiner CE (2002) Functional architecture of auditory cortex. *Curr Opin Neurobiol* 12:433–440.
- Zatorre RJ (1988) Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am* 84:566–572.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–776.
- Griffiths TD, Büchel C, Frackowiak RS, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. *Nat Neurosci* 1:422–427.
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946–953.
- Schönwiesner M, Rübsem R, von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci* 22:1521–1528.
- Hall DA, et al. (2002) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12:140–149.
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389–395.
- Shamma S (1996) Auditory cortical representation of complex acoustic spectra as inferred from the ripple analysis method. *Network: Computation in Neural Systems* 7:439–476.
- Bell AJ, Sejnowski TJ (1997) The “independent components” of natural scenes are edge filters. *Vision Res* 37:3327–3338.
- Abdallah SA, Plumbley MD (2001) If the independent components of natural images are edges, what are the independent components of natural sounds? In *International Conference on Independent Component Analysis and Blind Signal Separation* (San Diego, CA), pp 534–539.
- Salam FM, Erten G (1999) Sensory fusion by principal and independent component decomposition using neuronal networks. In *International Conference on Multisensor Fusion and Integration for Intelligent Systems* (IEEE, Taipei, Taiwan).
- Worsley KJ, et al. (2002) A general statistical analysis for fMRI data. *Neuroimage* 15:1–15.
- Penhune VB, Zatorre RJ, MacDonald JD, Evans AC (1996) Interhemispheric anatomical differences in human primary auditory cortex: Probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb Cortex* 6:661–672.
- Leonard CM, Puranik C, Kuldau JM, Lombardino LJ (1998) Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: Where is it? *Cereb Cortex* 8:397–406.
- Morosan P, et al. (2001) Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13:684–701.
- Westbury CF, Zatorre RJ, Evans AC (1999) Quantifying variability in the planum temporale: A probability map. *Cereb Cortex* 9:392–405.