Behavioral/Systems/Cognitive

Human Cortical Organization for Processing Vocalizations Indicates Representation of Harmonic Structure as a Signal Attribute

James W. Lewis,[1,2,3] William J. Talkington,[1,2,3] Nathan A. Walker,[1,2,3] George A. Spirou,[2,3,4] Audrey Jajosky,[1,2,3] Chris Frum,[1,2,3] and Julie A. Brefczynski-Lewis[1,5]

[1]Center for Advanced Imaging, [2]Sensory Neuroscience Research Center, and Departments of [3]Physiology and Pharmacology, [4]Otolaryngology, and [5]Radiology, West Virginia University, Morgantown, West Virginia 26506

The ability to detect and rapidly process harmonic sounds, which in nature are typical of animal vocalizations and speech, can be critical for communication among conspecifics and for survival. Single-unit studies have reported neurons in auditory cortex sensitive to specific combinations of frequencies (e.g., harmonics), theorized to rapidly abstract or filter for specific structures of incoming sounds, where large ensembles of such neurons may constitute spectral templates. We studied the contribution of harmonic structure to activation of putative spectral templates in human auditory cortex by using a wide variety of animal vocalizations, as well as artificially constructed iterated rippled noises (IRNs). Both the IRNs and vocalization sounds were quantitatively characterized by calculating a global harmonics-to-noise ratio (HNR). Using functional MRI, we identified HNR-sensitive regions when presenting either artificial IRNs and/or recordings of natural animal vocalizations. This activation included regions situated between functionally defined primary auditory cortices and regions preferential for processing human nonverbal vocalizations or speech sounds. These results demonstrate that the HNR of sound reflects an important second-order acoustic signal attribute that parametrically activates distinct pathways of human auditory cortex. Thus, these results provide novel support for the presence of spectral templates, which may subserve a major role in the hierarchical processing of vocalizations as a distinct category of behaviorally relevant sound.

Key words: **auditory cortex; fMRI; tonotopy; harmonics-to-noise ratio; speech; auditory object**

Introduction

In the mammalian auditory system, recognizing and ascribing meaning to real-world sounds relies on a complex combination of both "bottom-up" and "top-down" grouping cues that segregate sounds into auditory streams, and ultimately lead to the perception of distinct auditory events or objects (Wang, 2000; Cooke and Ellis, 2001; Hall, 2005). To increase signal processing efficiency, different classes of sound may be directed along specific cortical pathways based on relatively low-level signal attributes. In humans, animal vocalizations, as a category of sound distinct from hand-tool sounds, are reported to more strongly activate the left and right middle superior temporal gyri (mSTG), independent of whether or not the sound is correctly perceived, and independent of handedness (Lewis et al., 2005, 2006; Lewis, 2006; Altmann et al., 2007). Consequently, at least portions of the mSTG appear to process bottom-up acoustic signal features, or primitives, characteristic of vocalizations as a distinct category of sound. However, what organizational principles, beyond tonotopic organizations derived from cochlear processing, might generally facilitate segmentation and recognition of vocalizations?

One such second-order acoustic signal attribute is the sound's harmonic structure, which can be quantified by the harmonics-to-noise ratio (HNR) (Boersma, 1993; Riede et al., 2001). Sounds with greater HNR value generally correlate with the perception of greater pitch salience. For instance, a snake produces a hiss with a very low HNR value, near that of white noise (Fig. 1*a*,*b*). In contrast, sounds such as a wolf howl, and some artificially created iterated rippled noise sounds (IRNs) (see Materials and Methods), tend to have a more tonal quality and greater pitch salience, being comprised of more prominent harmonically related frequency bands ("frequency stacks") that persist over time (hear supplemental Audios 1–10, available at www.jneurosci.org as supplemental material). In mammals, the harmonic structure of vocalizations stem from air flow causing vibrations of the vocal folds in the larynx, resulting in periodic sounds (Langner, 1992; Wilden et al., 1998). In other species, this process similarly involves soft vibrating tissues such as the labia in the syrinx of birds, or phonic lips in the nose of dolphins, which underscores the ethological importance of this basic mechanism of "vocal" harmonic sound production for purposes of communication. HNR

Received Aug. 30, 2008; revised Jan. 14, 2009; accepted Jan. 15, 2009.

This work was supported by the National Center for Research Resources–National Institutes of Health–Centers of Biomedical Research Excellence Grant P20 RR15574 (to the Sensory Neuroscience Research Center of West Virginia University and subproject to J.W.L.). We thank Dr. Bill Shofner for Matlab code to construct IRN stimuli, Dr. Tobias Riede for suggestions on HNR calculations, and Doug Ward and Gerry Hobbs for assistance with statistical analyses. We thank Dr. Robert Cox for continual development of AFNI, and Dr. David Van Essen, Donna Hanlon, and John Harwell for continual development of cortical data analysis and presentation with Caret.

Correspondence should be addressed to Dr. James W. Lewis, Department of Physiology and Pharmacology, P.O. Box 9229, West Virginia University, Morgantown, WV 26506. E-mail: jlewis@hsc.wvu.edu.

DOI:10.1523/JNEUROSCI.4145-08.2009

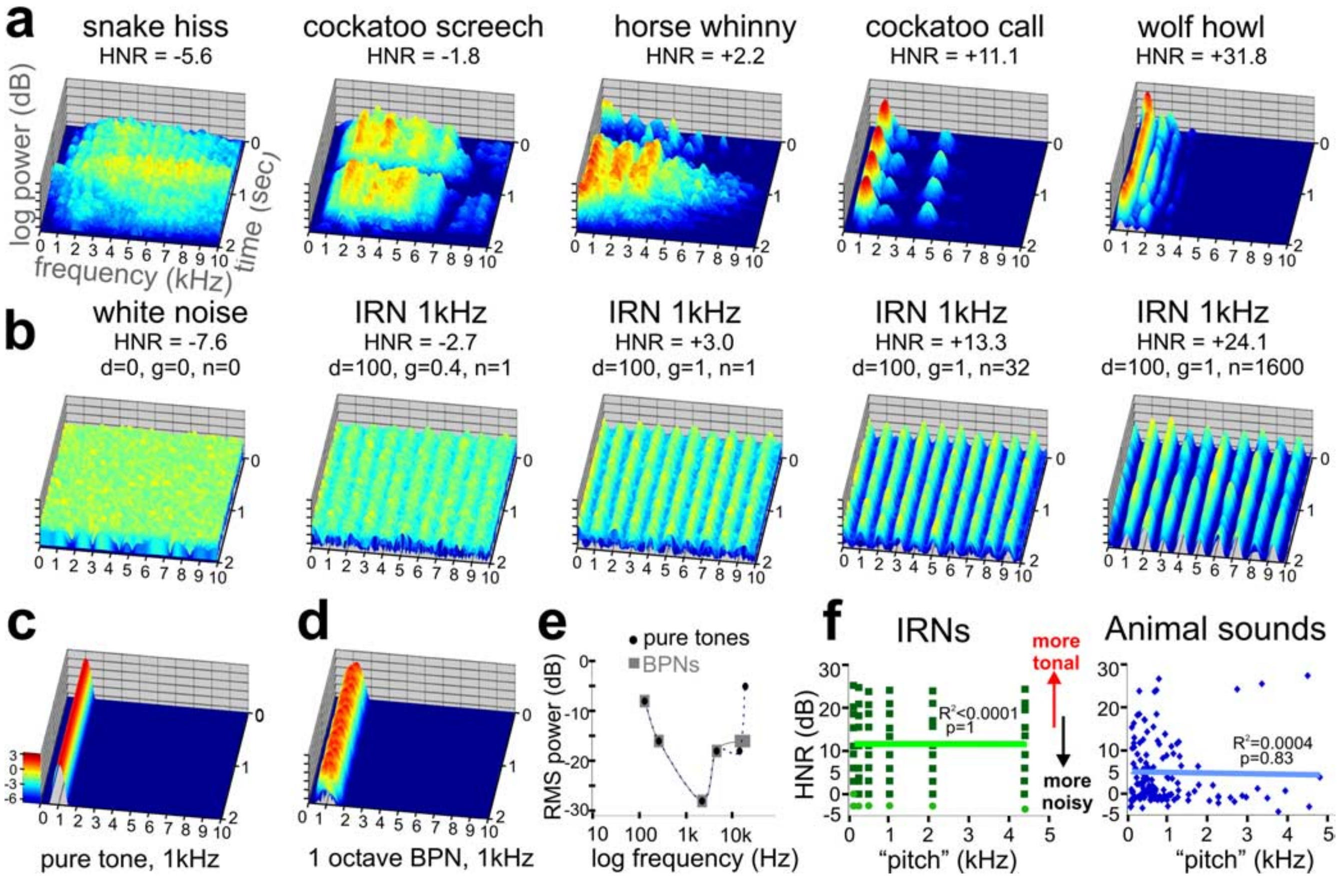

Figure 1. Sound stimulus attributes. ***a***, Three-dimensional spectrograms of five vocalizations (2 s duration), including one from a snake, two from birds, and two from mammals. In all plots, the frequency was limited to 10,000 Hz for illustration purposes, and the *z*-axis represents log-power (relative intensity, scale in ***c*** in log exponentials). The HNR value for each sound is indicated. ***b***, Spectrograms of IRNs derived from one white noise sample (leftmost panel). The IRNs with greater HNR value correlate with more prominent frequency bands (peaks) at all harmonics (1 kHz in these examples) and had a more tonal quality. ***c***, ***d***, Spectrograms of an example PT (***c***) and an example BPN (***d***) used for the FDRR/tonotopy localizer scans. Note the similarity of these peaks to those of the IRNs. ***e***, Audiometric profile used to match perceived loudness of PT and BPN stimuli for the FDRR localizer scans. ***f***, Charts comparing "estimated pitch" versus HNR value of IRNs and animal vocalizations. Light green dots depict IRNs for which a pitch could not be accurately estimated computationally, although was determined by the IRN delay. There was no significant linear correlation between the pitch and HNR value for either stimulus set.

measures have proven useful for analyzing features of animal vocal production (Riede et al., 2001, 2005). In humans, HNR measures have also been used clinically to monitor recovery from voice pathologies (Shama et al., 2007), and used to assess signal characteristics of different forms of speech, such as sarcasm (Cheang and Pell, 2008). We previously reported that the "global" HNR values for human and animal vocalizations were substantially greater than for other categories of natural sound, suggesting that this could be a critical signal attribute that is explicitly processed in cortex to facilitate sound segmentation and categorization of vocalizations (Lewis et al., 2005).

Moreover, HNR is an attractive signal attribute to study from the perspective of neural mechanisms for auditory object or sound-source segmentation. Because harmonically structured sounds are comprised of specific combinations of acoustic peaks of energy at different frequencies (cf. Fig. 1*b–d*), HNR sensitivity could potentially build off of tonotopically organized representations, thereby increasing receptive field complexity, similar to intermediate processing stages in the cortex for other sensory modalities. In several animal species (e.g., frogs, birds, bats, and primates), neurons in auditory cortex, or analogous structures, show facilitative responses to specific combinations of frequencies, notably including the harmonic structures typically found in conspecific vocalizations (Lewicki and Konishi, 1995; Rauschecker et al., 1995; Medvedev et al., 2002; Medvedev and Kanwal, 2004; Petkov et al., 2008). Ensembles of "combination-sensitive" neurons could filter for or extract harmonic features (or primitives). Such representations may reflect elements of theorized spectro-temporal templates that serve to group spectral and temporal components of a sound-source, resulting in coherent percepts (Terhardt, 1974; Medvedev et al., 2002; Kumar et al., 2007). In humans, a substantial portion of auditory cortex presumably is, or becomes, optimized for processing human vocalizations and speech (Belin et al., 2000; Scott, 2005). Thus, the presentation of sounds with parametrically increasing harmonic structure (HNR value), approaching those typical of speech sounds, should grossly lead to the recruitment of greater numbers of, or greater activity from, combination-sensitive neurons. If observed, this would provide evidence for HNR sensitivity, and thus support for spectral templates in representing a neural mechanism for extracting and streaming vocalizations.

The above working model indicates that HNR-sensitive regions, based on combination-sensitive neural mechanisms, would require input from multiple frequency bands. Thus, HNR-sensitive regions, to minimize cortical wiring, should be located along or just outside of tonotopically organized areas, and so we mapped tonotopic functional landmarks in some individuals. Additionally, this hierarchical model indicates that HNR-sensitive regions should largely be located along the cortical surface between tonotopically organized regions and regions preferential for human-produced vocalizations. Thus, as additional functional landmarks, we also mapped cortices sensitive to human nonverbal vocalizations and to speech.

Materials and Methods

Participants. We studied 16 right-handed adult English speaking participants (age 18–39 years; 10 women), who underwent one to five of our scanning paradigms. All participants were free of neurological, audiological, or medical illness, had normal structural MRI and audiometric examinations, and were paid for their participation. Informed consent was obtained following guidelines approved by the West Virginia University Institutional Review Board.

IRN stimuli. As one measure for studying harmonic structure as an isolated signal attribute, we used iterated rippled noises, or IRNs (Yost, 1996; Shofner, 1999), which have previously been used to study pitch and pitch salience processing in other human neuroimaging studies (Griffiths et al., 1998; Patterson et al., 2002; Penagos et al., 2004; Hall et al., 2005). By delaying and adding segments of white noise back to itself, IRN sounds with periodic harmonic structure can be constructed, producing sounds perceived to have a tonal quality embedded in white noise (Fig. 1*b*; hear supplemental Audios 6–10, available at www.jneurosci.org as supplemental material). Wideband noise was systematically altered by temporal rippling, using custom Matlab code (V7.4, Mathworks) [Dr. William Shofner (Indiana University, Bloomington, IN), personal communication]. IRN stimuli were generated (44.1 kHz, 16-bit, monaural, ~6 s duration) by a cascade of operations delaying and adding back to the original noise ("IRNO" in the terminology of Yost, 1996), with a given gain (*g*, ranging 0 to +1, in steps of 0.1) a delay (*d*, 0.25, 0.5, 1, 2, 4, and 8 ms), and a wide range of number of ripple iterations (*n*; including 1, 2, 4, 8, 16, 32, 64, 100, 200, 300, … to 2000). The perceived pitch of the IRN changes inversely with delay, and we included pitches of 125, 250, 500, 1000, 2000, and 4000 Hz, which were chosen to complement tonotopic mapping of cortex (see paradigm 1). Increasing the number of iterations and/or gain qualitatively increases the clarity or strength of the perceived pitch (Penagos et al., 2004), which appears to be highly correlated with the harmonic content of the sound. In contrast to earlier studies using IRNs (ibid), we examined HNR measures of IRNs (see below), effectively manipulating pitch depth along the dimension of harmonic content. We created a much larger set of IRNs (~1700) so as to span a wide range of HNR values (supplemental Fig. 1, available at www.jneurosci.org as supplemental material). We then selected 63 IRNs to evenly sample across the dimension of HNR in steps of 3 dB HNR [trimmed to 2.00 s duration, and matched for overall root mean square (RMS) power: −12.0 ± 0.2 dB]. More importantly, the quantitative HNR measure could be applied to behaviorally relevant real-world sounds (see below), and thus we sought to test a much wider range of IRN stimuli than have previously been studied, being comparable in HNR ranges observed for animal and human vocalizations.

Animal and human vocalization stimuli. We collected 160 professionally recorded animal vocalizations (Sound Ideas), which were typically recorded using stereo microphones containing two directional monaural microphones (44.1 or 48 kHz, 16-bit). Only one channel (left) was retained (down-sampled to 44.1 kHz) to remove binaural spatial cues (Cool Edit Pro v1.2, Syntrillium Software, now owned by Adobe), and the monaural recording was presented to both ears. Sounds included a wide variety of animals producing sound through a vocal tract or analogous structure. Care was taken to select sounds derived from only one animal with relatively little background or ambient noise, and to avoid aliasing, clipping and reverberation that could introduce spectrogram artifacts (Wilden et al., 1998). Most sounds were trimmed to 2.0 ± 0.2 s duration, although a few sounds were of shorter duration (minimum 1.6 s) to allow for more natural sounding acoustic epochs. Sound stimuli were ramped in intensity 20 ms to avoid spectral transients at onset and offset. Most of the animal sounds were matched in total RMS power to the IRN stimuli (at −12 dB). However, since some of the vocalization recordings included quiet or silent gaps, the overall intensity was necessarily lower for some stimuli to avoid clipping (mean= −12.6 dB, range −8.2 to −20 dB total RMS power). Human spoken phrases and nonverbal vocalizations used in functional MRI (fMRI) paradigm 5 were collected using the same techniques described above.

As part of an analysis of the potential behavioral relevance of the global HNR value of human vocalizations, we also recorded adult-to-adult and adult-to-infant speech from 10 participants, using professional recording equipment (44.1 kHz, 16-bit, monaural) in a sound isolation booth. Each participant was provided with a brief script of topics for conversation, including describing weekend plans to another adult, and speaking to a baby (a baby doll was present) in an effort to make him smile. The script also included speaking onomatopoetic words describing different subcategories of animal vocalizations, including phrases such as "a hissing snake" and "a growling lion." The stress phonemes, such as the "ss" in hiss, were selected and subjected to the same HNR analysis as the other sound stimuli, as described below.

HNR calculation. We analyzed and calculated HNR values of all sound stimuli using freely available phonetic software (Praat, http://www.fon.hum.uva.nl/praat/). The HNR algorithm (below) determined the degree of periodicity within a sound signal, $x(t)$, based on finding a maximum autocorrelation, $r'_x(\tau_{\max})$, of the signal at a time lag (τ) greater than zero (Boersma, 1993):

$$\text{HNR (in dB)} = 10^{*} \log_{10}\frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})}.$$

This measure quantified the acoustic energy of the harmonics that were present within a sound over time, $r'_x(\tau_{\max})$, relative to that of the remaining "noise", $1 - r'_x(\tau_{\max})$, which represents nonharmonic, irregular, or chaotic acoustic energy. Three parameters influence the estimate of the harmonic structure of a sound, including a time step (10 ms), minimum pitch cutoff for its fundamental (75 Hz minimum pitch, 20 kHz ceiling), and periods per window (1 per window). As extreme examples, white noise yielded an HNR value of −7.6 with the above parameters, while a sample consisting of two pure tones (PTs) (2 kHz and 4 kHz sine waves) produced an HNR value of +65.4.

Although no single set of HNR parameters is ideal for assessing all real-world sound stimuli (Riede et al., 2001) [Dr. Tobias Riede (University of Utah, Salt Lake City, UT), personal communication], the periodic nature of the IRN stimuli lent themselves to a robust HNR value estimate over the entire 2 s duration. The HNR values of the selected IRNs ranged from −3.5 to +25.2 dB HNR (grouped in increments of 3 dB HNR), with ±1.3 dB HNR average SD (range 0.3–7.9). For the animal vocalizations, we carefully selected those having a relatively stable pitch and cadence over time, ranging from −6.5 to +32.7 dB HNR with ±5.4 dB HNR average SD (range 0.8–10.9). The estimated pitches of the animal vocalizations (Fig. 1*f*) were also derived using a 75 Hz floor and 5 kHz ceiling (Praat software).

Care must be taken in applying the HNR calculation. We derived HNR values over a two second duration, which proved to be adequate for relatively continuous or temporally homogeneous sounds. However, the HNR estimate was sensitive to abrupt acoustic transitions, such as fricatives and plosives, because it relies on providing a good estimate of the fundamental frequency of the sound sample (Boersma, 1993; Riede et al., 2005). We found that for some sound stimuli, and some sound categories such as sounds produced by hand tools, it was difficult to derive reliable HNR estimates, especially when using long (2 s) duration sound samples. Thus, for many natural sound stimuli it may be more meaningful to examine shorter segments of time, characterizing discrete segments as the sound dynamically changes (Riede et al., 2001).

fMRI imaging paradigms. Each participant ($n = 16$) performed one to five different scanning paradigms (41 scanning sessions total). In all paradigms, we used a clustered acquisition design allowing sounds to be presented during scanner silence, and allowed a one-to-one correspondence between a stimulus presentation and a brain image acquisition (Edmister et al., 1999; Hall et al., 1999).

Paradigm 1: "Tonotopy" localizers. In one scanning session (12 scanning runs, ~8 min each; $n = 4$ participants), we randomly presented 15 repetitions of 12 test sounds and 120 silent events as a control. The test sounds included six PTs at 125, 250, 2000, 4000, 12,000 and 16,000 Hz, plus six corresponding versions of bandpass noise (BPN) stimuli having the same six center frequencies. The BPNs were generated from one white noise sample that was modified by seventh order Butterworth filters to yield ±1 octave bandwidths (Fig. 1*d*). The sound intensity of the PT and BPN stimuli had been assessed psychophysically before scanning by three participants and equated for perceived loudness (Fig. 1*e*). All stimuli consisted of five 400 ms bursts with 35 ms on/off ramps, spanning 2 s duration.

For purposes of a task, a second PT or BPN (2 s) was presented 200 ms after each respective PT or BPN test sound, having a lower, the same, or a higher center frequency. The task sounds spanned a gradient of ~3% difference at the lower and higher center frequencies and 0.5% difference at the middle center frequency ranges to match for approximate discrimination difficulty. During scanning, participants, with eyes closed, responded by three alternative forced choices (3AFC) as to whether the second sound was lower, the same, or higher in pitch, responding quickly before the second sound had stopped playing.

A multiple linear regression analysis modeled the contribution to the blood oxygen level-dependent (BOLD) signal time series data for each of the 6 PTs and 6 BPNs, plus and error term (see below, Image analysis). A winner-take-all algorithm identified voxels showing the greatest average BOLD signal magnitude responses, relative to silent events, to one of the three different frequency ranges presented, low (125 + 250 Hz, yellow), medium (2 + 4 kHz, orange), and high (12 + 16 kHz, red), separately for the PT and BPN stimuli. We then masked the winner-take-all map for significant activation to the PT or BPN tonotopy data separately for each individual at two conservative threshold settings ($p < 10^{-4}$ and $p < 10^{-6}$) and projected these data onto the cortical surface models for each individual (see below, Image analysis). The surface models were then highly inflated and unfolded to facilitate viewing of the functional data (unfolded flat maps not shown), and these were used to guide the generation of outlines around tonotopic progressions (for outlining criteria, see Results). For illustration purposes, individual cortical surface models of the left and right hemisphere were slightly inflated, smoothed, and cut away so as to reveal each individual's unique cortical geography along Heschl's complex, including Heschl's gyrus (HG) (or gyri in some individual hemispheres), planum polare, and planum temporale.

Paradigm 2: IRN HNR paradigm. For the IRN paradigm we randomly presented 180 pairs of IRN stimuli and 60 silent events (6 runs, ~7 min each, $n = 16$). The 60 IRN test stimuli included six pitches across ten 3 dB increments in HNR value, ranging from −3.6 to +25.2 dB (Fig. 1*f*). A second IRN "task" sound was presented 200 ms after the test sound, and included the 60 sounds together with two additional IRNs at −6 dB HNR and one at +27. The test and task IRN sound pairs had the same pitch, but had either higher, the same, or lower HNR value (ranging in difference from 0 to 5 dB HNR). Participants indicated whether the second sound was more tonal, the same, or more "noisy" than the first, responding (3AFC) before the second sound had stopped playing. A multiple linear regression analysis modeled the BOLD response using two terms plus an error term. The first term modeled variance in the time series data due to the presence of sound versus silent events. The second term assessed how much additional variance was accounted for by activity that linearly correlated with the HNR value of each sound (partial *F* statistic). HNR-sensitive regions were selected based on the second term in the model. Individual data sets were thresholded to $p < 0.01$, and whole-brain corrected for multiple comparisons using Monte Carlo randomization statistics (see below, Image analyses), yielding a whole-brain corrections of $\alpha < 0.05$. The IRN HNR-sensitive regions-of-interest (ROIs) were also separately modeled for sensitivity to the six different IRN pitches, using a second regression analysis similar to that described for paradigm 1.

Paradigm 3: Loudness biased IRN control paradigm. When assessed psychophysically in a sound isolation booth, the perceived loudness of the different IRN stimuli with differing pitches and HNR values proved difficult to precisely balance across individuals. Because increases in sound intensity have generally been reported to activate larger and/or varying extents of auditory cortex (Jäncke et al., 1998; Bilecen et al., 2002; Yetkin et al., 2004), a subset of participants ($n = 4$) also underwent a separate scan to directly test the effects of sound intensity versus HNR value of the IRN stimuli. In one condition, the 60 IRN stimuli (test and task sound pairs) were reverse-biased for sound intensity (supplemental Fig. 5*a*, available at www.jneurosci.org as supplemental material), applying a linear gradient from −5 dB to +5 dB average RMS power to the lower to higher HNR valued IRN stimuli in steps of 3 dB HNR. In a second condition, the opposite forward-bias with intensity was applied. Scanning parameters and the listening task were identical to those for IRN paradigm 2 (sometimes conducted during the same scanning session as paradigm 2), and six runs of each condition were randomly intermixed (12 or 18 runs, ~7 min each). Multiple linear regression analyses modeled sensitivity to HNR value, as described in paradigm 2, for each of the separate loudness conditions.

Paradigm 4: Animal vocalization HNR paradigm. We randomly presented 160 unique animal vocalizations, 120 IRNs (the above described 60 IRNs, presented twice), and 40 silent events (7 runs, ~7 min each) using the same scanning parameters as those for paradigms 2–3 ($n = 11$ of the 16 from paradigm 2, including the 4 participants from paradigm 1). For animal vocalization task sounds, the HNR values of the test sounds (original recordings) were modified by either adding white noise or by filtering out white noise (CoolEditPro v1.2 software). This allowed for the same 3AFC task as with the IRN paradigms, judging whether the task sound was more tonal, same, of noisier than the test sound. A multiple linear regression analysis included four terms: Two terms modeled variance due to the presence of vocalizations or IRN sounds versus silent events, respectively, while two additional terms assessed how much additional variance was accounted for by activity that linearly (positively or negatively) correlated with the HNR value of the vocalizations or IRN sounds. These latter two terms were used to generate HNR-sensitive ROIs, as described in paradigm 2

A *post hoc* nonlinear regression analysis was additionally used to model the response profile between BOLD signal and HNR value of the animal vocalizations (see Fig. 4, blue curves) using the following equation:

$$\mathrm{BOLD} = b_0 + g_0/(1 + g_1{}^{*}e(-g2^{*}\mathrm{HNR}))\,.$$

Although the coefficients in this equation (b_0, g_0, g_1, g_2) do not necessarily reflect any physiologically relevant measures, this nonlinear regression model was chosen as it could more closely fit the data (blue dots) and reflect biologically plausible floor and ceiling limits in BOLD signal "activation" levels than could a linear fit. This approach also had the advantage of being able to reveal an HNR range where the slope might be changing more rapidly.

Paradigm 5: Human vocalization HNR paradigm. For this paradigm, we included unique samples of (1) 60 human speech phrases (balanced male and female speakers) (2) 60 human nonverbal vocalizations and utterances (3) 60 animal vocalizations (a subset from paradigm 4), and (4) 60 IRNs (from paradigm 2), together with 60 silent events (8 runs, ~7 min each). Each sound category was matched for HNR value range (+3 to +27 dB HNR) and HNR mean (+11.6 dB HNR). Participants ($n = 6$; five from paradigm 4) performed a 2AFC task, indicating whether the sound stimulus was produced by a human or not. A multiple linear regression analysis modeled the contribution to the BOLD signal from each of the four categories of sound, each relative to responses to silent events as the baseline control.

Stimulus presentation. For all paradigms, the high-fidelity sound stimuli were delivered via a Windows PC computer with a sound card interface (CDX01, Digital Audio), a sound mixer (1642VLZ pro mixer, Mackie) and MR compatible electrostatic ear buds (STAX SRS-005 Earspeaker system; Stax), worn under sound attenuating ear muffs. Sound stimuli were presented at 80–83 dBC-weighted, as assessed at the time of scanning (Brüel & Kjær 2239A sound meter) using one of the IRN stimuli (1 kHz pitch, 11.3 dB HNR) as a "standard" loudness test stimulus. The sound delivery system imparted a 75 Hz high-pass filter (at rate of 18 dB/octave), and the ear buds exhibited a flat frequency response out to 20 kHz (±4 dB).

Image acquisition. Scanning was conducted with a 3 Tesla General Electric Horizon HD scanner equipped with a body gradient coil optimized to conduct whole-head, spiral imaging of BOLD signals (Glover and Law, 2001). For paradigms 2–5, a sound pair or silent event was presented every 10 s, and 4.4 s after onset of the test sound there followed the collection of BOLD signals from axial brain slices (28 spiral "in" and "out" images, with 1.87 × 1.87 × 2.00 mm^3 spatial resolution, echo time = 36 ms, repetition time = 10 s, 2.3 s slice package, field of view = 24 mm). The tonotopy localizer paradigms used a 12 s cycle to further minimize possible contamination of the sound frequencies emitted by the scanner itself. The presentation of each event was triggered by a TTL pulse from the MRI scanner. During every scanning session, T1-weighted anatomical MR images were collected using a spoiled GRASS pulse sequence; 1.2 mm slices, with 0.9375 × 0.9375 mm in-plane resolution.

Image analysis. Data were viewed and analyzed using Analysis of Functional NeuroImages (AFNI) (Cox, 1996) and related software plug-ins (http://afni.nimh.nih.gov/). BOLD data of each participant were converted to percentage signal changes relative to the mean of the responses to silent events on a voxel-wise basis for each scan run. For each paradigm, the scanning runs from a single session (6–18 scans) were concatenated into one time series. Brain volume images were motion corrected for global head translations and rotations, by reregistering them to the 20th volume of the scan closest in time to the anatomical image acquisition. Functional data (multiple regression coefficients) were thresholded based on partial *F* statistic fits to the regression models, and significantly activated voxels were overlaid onto anatomical images.

Using the public domain software package Caret (http://brainmap.wustl.edu), three-dimension cortical surface models were constructed from the anatomical images for several individuals (Van Essen et al., 2001; Van Essen, 2003), onto which the volumetric fMRI data were projected. For all paradigms, the combination of individual voxel probability threshold (partial *F* statistic, typically $p < 0.01$ or $p < 0.05$), a cluster size minimum (typically 9 or 50 voxels), and an estimate of signal variance correlation between neighboring voxels (filter width at half maximum of 2–4 mm) yielded the equivalent of a whole-brain corrected significance level of $\alpha < 0.05$ (AFNI plug-in AlphaSim).

For group-average analyses, each individual's anatomical and functional brain maps were transformed into the standardized Talairach (AFNI-tlrc) coordinate space. Functional data were spatially low-pass filtered (4 mm Gaussian filter), then merged volumetrically by combining coefficient values for each interpolated voxel across all participants. Combined data sets were subjected to *t* tests (typically $p < 0.05$), and to a cluster size minimum (typically 9 voxels).

Averaged cortical hemisphere surface models were derived from three of our participants, using Caret software, on which the group-averaged fMRI results were illustrated. Briefly, six geographical landmarks, including the ridge of the STG, central sulcus, Sylvian fissure, the corpus callosum (defining dorsal and ventral wall divisions), and calcarine sulcus of each hemisphere of each participant were used to guide surface deformations to render averaged cortical surface models. Portions of these data can be viewed at http://sumsdb.wustl.edu/sums/directory.do?id=6694031&dir_name=LEWIS_JN09, which contains a database of surface-related data from other brain mapping studies.

Results

The following progression of five experimental paradigms, using high spatial resolution fMRI (<2 mm^3 voxels), was designed to test for HNR-sensitive patches of auditory cortex in humans, using both artificially constructed IRNs and real-world recordings of animal vocalizations. This included identifying tonotopically organized cortices and regions sensitive to human vocalizations within individuals, allowing for a direct test of our proposed hierarchical model for processing vocalizations. To explore the possible behavioral significance that the HNR signal attribute might generally have in vocal communication across species, we further investigated the harmonic content of various "subcategories" of human and animal vocalizations to provide further context.

Estimated localizations of primary auditory cortices

Based on a cytoarchitectonic study (Rademacher et al., 2001), the location of primary auditory cortices (PAC) (including A1, R, and possibly a third subdivision), tends to overlap the medial two thirds of HG, although with considerable range in individual and hemispheric variability. Although the correspondence between functional estimates for PAC with histological and anatomical criteria remains to be resolved (Talavage et al., 2004), the identification of frequency-dependent response regions (FDRRs) allowed for more precise and direct localization of HNR-sensitive regions (addressed below) relative to tonotopically organized patches of auditory cortex within individual hemispheres. We identified the location of tonotopically organized cortices in a subset of our participants ($n = 4$, paradigm 1), using techniques similar to those described previously (Formisano et al., 2003).

We charted cortex sensitive to pure tones, and additionally to 1 octave bandpass noises, at low (Fig. 2,125 and 250 Hz, yellow), medium (2000 and 4000 Hz, orange), and high (12,000 and 16,000 Hz, red) center frequency ranges, wherein participants performed a three 3AFC tone or pitch discrimination task. In contrast to previous fMRI tonotopy mapping studies (Wessinger et al., 2001; Schönwiesner et al., 2002; Formisano et al., 2003; Talavage et al., 2004; Langers et al., 2007), we derived perimeter boundary outlines of FDRRs based on the presence of tonotopic gradients at conservative threshold settings, as illustrated for three representative individuals who participated in three or more of our paradigms (Fig. 2*a–c*, black outlines) (see Materials and Methods). The tonotopic subdivisions of the FDRRs were characterized by cortex that responded preferentially, but not exclusively, to particular pure tone frequency bands (Fig. 2*a*, histograms). Three criteria were used to define FDRR outlines. First, a red to orange to yellow contiguous progression, in any direction, had to be present along the individual's cortical surface model using either the pure tone data or a combination of pure tone and bandpass noise data. However, outlines only encircled the high threshold ($p < 10^{-6}$) pure tone data. Second, some of the FDRR progressions showed a mirror image organization with neighboring progressions, as reported previously in human and nonhuman animal studies (ibid). In those instances, activation gradients were divided approximately midway between the two FDRRs (Fig. 2*a*, left hemisphere midway along the yellow cortex). Third, FDRR progressions had to show continuity in both volumetric and surface projection maps to be included within an outline.

To our knowledge, this is the first fMRI study to chart the location of cortex sensitive to very high frequency tones (12,000 and 16,000 Hz, red). A right hemisphere bias for high frequencies (up to 14,000 Hz) and left hemisphere bias for low frequencies has been reported using auditory evoked potentials (Fujioka et al., 2002). However, the results of the present study demonstrated significant activation in both hemispheres to the high frequencies (red), which was even evident when examining responses to only the 16,000 Hz pure tones relative to silence (supplemental Fig. 2, available at www.jneurosci.org as supplemental material).

FDRR organizations defined by pure tones and bandpass noises were largely congruent with one another (Fig. 2*a*, top vs bottom panels), although the bandpass noises generally activated a greater expanse of auditory cortex, which may include "belt" regions as reported previously in human (Wessinger et al., 2001) and nonhuman primates (Rauschecker et al., 1995). Note, however, that the functionally defined FDRR outlines may not accurately reflect genuine boundaries between primary and nonprimary areas since they were dependent on relative threshold settings (Hall, 2005). Nonetheless, we could reliably reveal one to three FDRRs located along Heschl's gyrus in each hemisphere of each participant, thereby refining estimated locations of PACs, allowing for direct comparisons within individuals with the location of HNR-sensitive cortices, as addressed in the following section.

Iterated rippled noises reveal HNR-sensitive patches of cortex

Next, we investigated our hypothesis that portions of cortex outside of FDRRs would be characterized by activity that increased with increasing harmonic content (HNR value) of the sound stimuli, representing "intermediate" acoustic processing stages. We used IRN sounds because they could be systematically varied in HNR value, yet not be confounded by additional complex spectro-temporal signal attributes that are typically present in real-world sounds such as vocalizations. Sixty IRN stimuli were used, spanning 10 different ranges of HNR value for each of 6 different pitches (Fig. 1*f*, green). In contrast to previous studies using IRNs to study pitch depth or pitch salience (Griffiths et al., 1998; Hall et al., 2002; Patterson et al., 2002; Krumbholz et al., 2003; Penagos et al., 2004), we included a much broader range of effective HNR values (and pitches) that was more comparable with ranges observed with vocalization sounds. Participants heard sequential pairs of IRN stimuli and performed a 3AFC discrimination task indicating whether the second sound was more tonal, the same, or noisier than the first (n = 16, paradigm 2; see Materials and Methods).

Relative to silent events, IRN stimuli activated a broad expanse of auditory cortex, including the FDRRs (data not shown). More importantly, all 16 participants revealed multiple foci in auditory cortex characterized by increasing activity that showed a significant positive, linear correlation with parametric increase in HNR value of the IRN stimuli. All of the illustrated IRN HNR-sensitive ROIs showed significantly greater, positive BOLD signal activation relative to silent events (Fig. 3, error bars in charts). The topography of these regions was illustrated on cortical surface models generated for the same three individuals depicted previously (Fig. 3*a–c*, green).

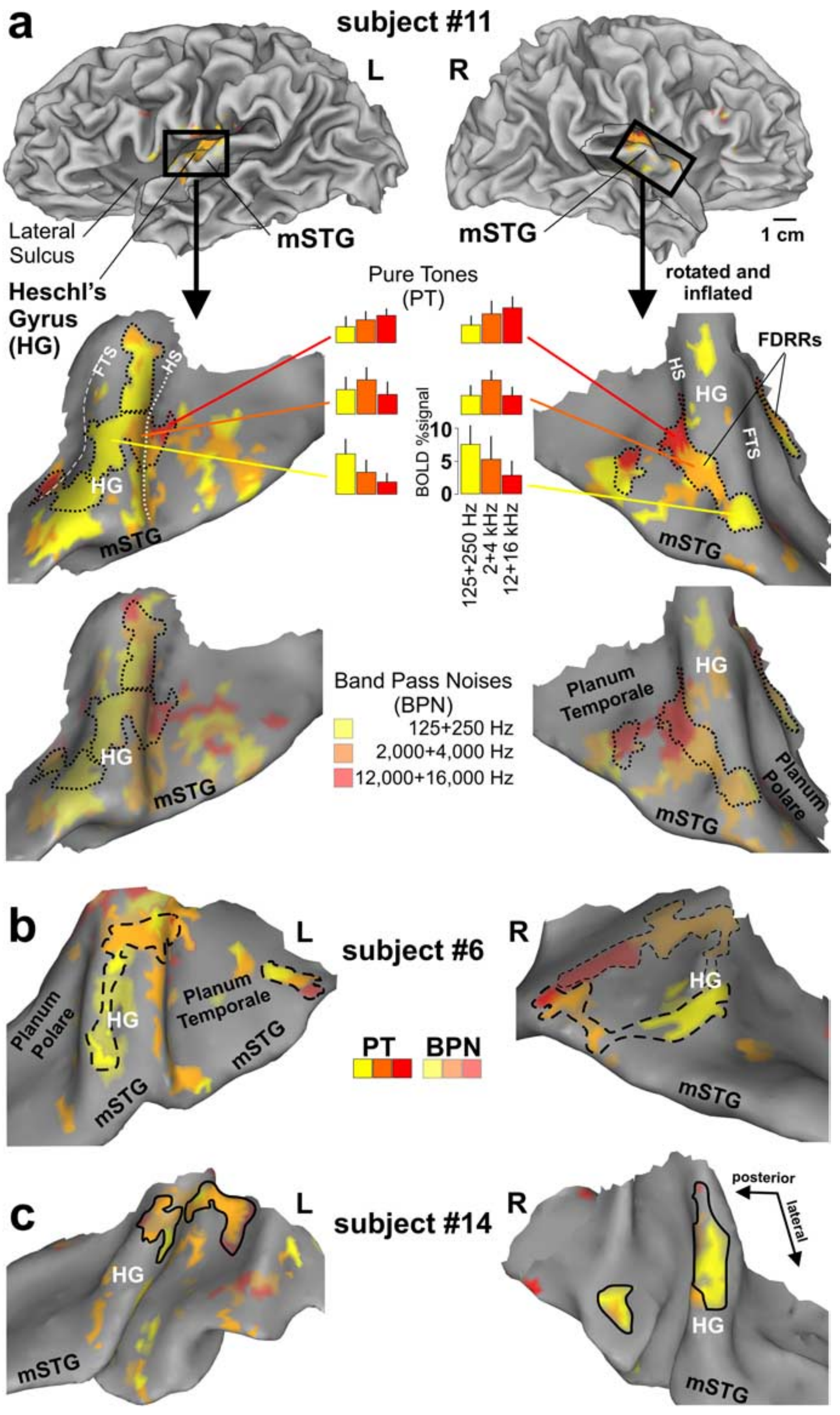

Figure 2. Functional localization of FDRRs in auditory cortex of three participants (***a–c***). Cortical hemisphere models of one participant (top) illustrate typical "cuts" (thin black outlines and black boxes) made to optimally view auditory cortex along the superior temporal plane and mSTG in this and subsequent figures. The cortical models of each hemisphere were slightly inflated and smoothed to facilitate viewing of Heschl's complex, including HG, Heschl's sulcus (HS) (white-dotted line), and the first transverse sulcus (FTS) (white-dashed line). The fainter dashed outline in ***b*** (right) depicts a prominent FDRR defined by the BPNs. The dotted, dashed, and solid black FDRR outlines distinguish these three representative individuals in this and subsequent figures. Refer to text for FDRR outlining criteria.

In general, the IRN HNR-sensitive foci showed a patchy distribution along much of Heschl's complex, the superior temporal plane, and in some hemispheres included cortex extending out to the mSTG. Within individuals, some of these foci partially overlapped portions of the outlined FDRRs. In these regions of overlap, the tonotopically organized frequency sensitive ranges were sometimes congruent with the pitch range of the IRN (supplemental Fig. 3, available at www.jneurosci.org as supplemental material), although the degree to which representations of periodicity pitch versus spectral pitch overlap remains a controversial

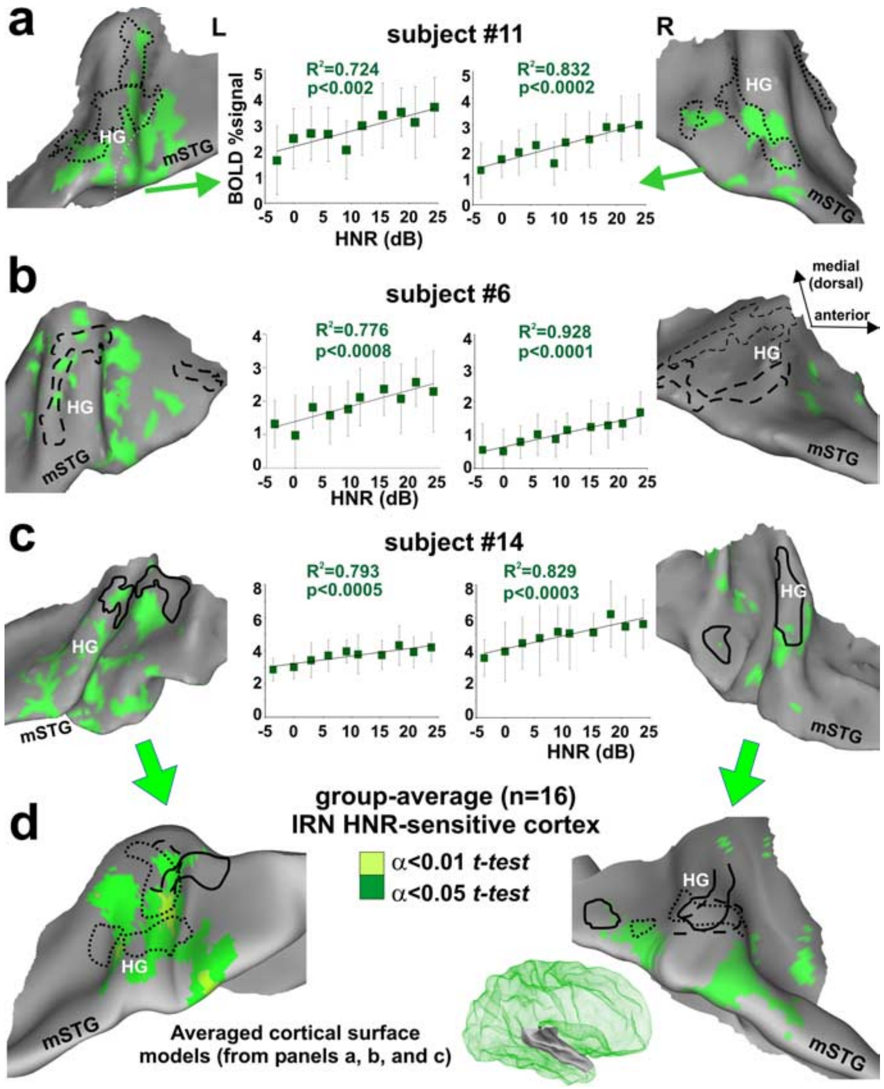

Figure 3. Cortex sensitive to the degree of harmonic structure of IRNs. ***a–c***, Individual data sets showing location of IRN HNR-sensitive cortical foci ($\alpha < 0.05$, corrected) relative to the location of FDRRs specific to each individual (dotted, dashed, and solid outlines from Fig. 2). Charts show the linear correlation between HNR value and BOLD activity (percentage signal change relative to silent events) combined across the multiple foci along Heschl's complex and the mSTG (mean + SD). The 180 IRN data points were binned at 3 dB HNR intervals for clarity. ***d***, Group-average overlap of HNR-sensitive cortex after thresholding each individual data set (individual $\alpha < 0.05$, and two *t* test levels, $\alpha < 0.05$ and $\alpha < 0.01$, corrected) and projected onto averaged brain surface models derived from these three participants (right hemisphere model shown in green mesh inset).

issue outside the scope of the present study (Langner, 1992; Jones, 2006). Nonetheless, these results show, at high spatial resolution within individuals, that substantial portions of IRN HNR-sensitive regions were located along and just outside the FDRRs.

Group-averaged IRN HNR-sensitive regions were projected onto an averaged cortical surface model (Fig. 3*d*) (see Materials and Methods). These results revealed a left hemisphere bias for IRN HNR-sensitive activation, evident as more significant and expansive areas (green and light green) involving portions of HG and cortex extending out to the mSTG. In contrast to previous studies that localized cortex sensitive to increasing pitch depth, or pitch strength, using rippled noises or other complex harmonic stimuli (Griffiths et al., 1998; Hall et al., 2002; Patterson et al., 2002; Penagos et al., 2004), the present results (1) indicated that the global HNR value of a sound represents a quantifiable acoustic signal attribute that is explicitly reflected in activation of human auditory cortex (2) demonstrated that IRN HNR-sensitive foci partially overlapped, but clearly did not completely overlap, with estimates of tonotopically organized cortices, suggestive of a hierarchical relationship, and (3) showed that there was a left hemisphere lateralization bias for HNR sensitivity, although non-natural and relatively acoustically "simple" sound stimuli were used.

Control conditions for IRN pitch and loudness

As control measures, we explicitly examined IRN pitch and perceived loudness (intensity) as variables that might affect the cortical activation patterns (Bilecen et al., 2002). A secondary analysis restricted to the IRN HNR-sensitive ROIs tested for linear correlations with increasing or decreasing IRN pitch sensitivity, and failed to show any significant correlations (supplemental Fig. 4, available at www.jneurosci.org as supplemental material).

To directly assess the effects of parametric increases or decreases in IRN stimulus intensity, a subset of the participants ($n = 4$, paradigm 3) were tested using IRN stimuli where the HNR values were forward- or reverse-biased with intensity (supplemental Fig. 5, available at www.jneurosci.org as supplemental material) (see Materials and Methods). Both forward- and reverse-biased IRN sounds yielded positive, linearly correlated activation foci that overlapped one another, demonstrating that the identification of IRN HNR-sensitive regions was not simply due to unintended differences in perceived loudness of the IRNs with different HNR values.

Animal vocalizations also reveal HNR-sensitive cortices

Next, we investigated whether we could reveal HNR-sensitive regions using recordings of natural animal vocalizations, and, if they existed, whether they overlapped with IRN HNR-sensitive regions. One possibility was that there might be a single HNR-sensitive processing "module" that would show HNR sensitivity independent of the type of sound presented. Alternatively, because animal vocalizations contain additional signal attributes statistically more similar to human vocalizations than to IRNs, HNR sensitivity using vocalizations might reveal additional or different foci along "higher-level" stages of auditory cortex, such as mSTG (Lewis et al., 2005; Altmann et al., 2007). As in the previous paradigm, we used a 3AFC harmonic discrimination task, and included IRNs and silent events as controls ($n = 11$, paradigm 4, see Materials and Methods).

Relative to IRNs, animal vocalizations activated a wider expanse of auditory cortex, and with greater intensity, including near maximal BOLD signal responses within the FDRRs and IRN HNR-sensitive regions (Fig. 4*d*, IRN foci charts; Fig. 5, histograms). Moreover, all participants revealed activation foci showing a significant positive, linear correlation with increase in HNR value of the animal vocalizations (Fig. 4*a–c*, blue cortex). Similar to the IRN HNR-sensitive regions (green cortex), these foci also showed a patchy distribution. However, within individuals there was only a moderate degree of overlap between IRN and animal vocalization HNR-sensitive regions (blue-green intermediate color) at these threshold settings, despite similarity in the range of HNR values used. Most vocalization HNR-sensitive regions were located further peripheral (lateral and medial) to the FDRRs and IRN HNR-sensitive regions, including regions along the mSTG in both hemispheres. Response profiles for nearly all animal vocalization HNR-sensitive ROIs (Fig. 4*a–c*, charts) revealed at least a trend for also showing positive, linear correlations with the HNR value of the IRN sound stimuli. However, the IRNs were generally less effective at driving activity in these regions, which is evident in all the charts (green lines) (Fig. 5, histograms). In some hemispheres, the animal vocalization data points resembled more of a negative exponential or sigmoid-shaped response curve. Thus, in addition to linear fits we also modeled these data using an exponential function (see Materials and Methods), thereby constructing a more biologically plausible activation profile that respected floor and ceiling limits in BOLD signal (e.g., Fig. 4*a*, right hemisphere; see supplemental Fig. 6, available at www.jneurosci.org as supplemental material, for additional individual charts).

Group-averaged data, similar to the individual data sets, demonstrated that the HNR-sensitive regions defined using animal vocalizations (Fig. 4*d*, blue), as opposed to using IRNs (green), were located further laterally, predominantly along the mSTG, with a strong left-lateralization. Moreover, animal vocalization HNR-sensitive regions in all participants showed greater response magnitudes than those defined using IRNs (Fig. 4*d*, charts). However, within the IRN HNR-sensitive ROIs (charts in green boxes) the linear correlations with animal vocalizations were relatively flat, appearing to have reached a ceiling plateau in both hemispheres. Within the animal vocalization HNR-sensitive ROIs (charts in blue boxes) both the IRN and animal vocalizations yielded positive, linear correlations with the HNR values, but the IRN data were of relatively lower response magnitudes and slightly less steep slopes, and thus tended to not meet statistical significance at our threshold settings.

In sum, these results revealed the existence of HNR-sensitive regions when using animal vocalizations and/or IRN stimuli, but the two respective activation patterns showed only a moderate degree of overlap (Fig. 4, blue vs green). The extent of overlap appeared to be in part due to floor and ceiling effects with the BOLD signal, in that the animal vocalizations, regardless of HNR value, lead to near maximal activation (green-boxed charts). However, other acoustic signal differences between vocalizations and IRNs are also likely to have contributed to the degree of overlap. Activation of the mSTG may have required sounds with more specific effective stimulus bandwidths, specific power spectral density distributions of different harmonic peaks (e.g., the $1/f^{\alpha}$-like power spectrum density in panels of Fig. 1*a* vs 1*b*; f = frequency, $1<\alpha<2$), and/or different specific frequencies, harmonics and subharmonics that are present in natural vocalizations but not IRNs (see Discussion). Nonetheless, these results demonstrated that there exists cortex, especially in the left hemisphere, that is generally sensitive to the degree of harmonic structure present in artificial sounds and real-world vocalizations.

HNR-sensitive regions lie between FDRRs and human voice-sensitive cortices

Our working model for HNR sensitivity, as representing intermediate processing stages, assumes that portions of auditory cortex of adult human listeners are optimally organized to process the signal attributes characteristic of human vocalizations and speech. Thus, as a critical comparison, we localized cortex sensitive to human nonverbal vocalizations (Hvocs) and to human speech (Speech) in a subset of the participants (n = 6, paradigm 5). In the same experimental session, we also presented animal vocalizations (Avocs), IRN stimuli, and silent events (see Materials and Methods). For this paradigm, all four sound categories (Speech, Hvocs, Avocs, and IRNs) had the same restricted range of HNR values (mean = +11.2, range +3 to +25 dB HNR), and participants indicated by 2AFC whether or not the sound was produced by a human (see Materials and Methods).

As expected, all four sound categories presented yielded significant activation throughout the FDRRs, IRN HNR-sensitive ROIs, and other portions of auditory cortex (data not shown) (although see group-average data in Fig. 5, histograms). More specifically, we charted the locations of foci that showed differential activation to one versus another category of sound in relation to the previously charted FDRRs and HNR-sensitive regions (Fig. 5, colored cortical maps). In particular, regions sensitive to human nonverbal vocalizations (pink) relative to animal vocalizations, speech (purple) relative to human nonverbal vocalizations, and regions preferential for animal vocalizations (light blue) relative to IRNs, were all superimposed onto averaged cortical surface maps. The HNR-sensitive regions (dark blue and green hues) and FDRRs (yellow and outlines) were those depicted previously (refer to Fig. 5 color key). Although the combined overlapping patterns of activation are complex, a clear progression of at least three tiers of activation was evident (Fig. 5*a*, rainbow-colored arrows). FDRRs (yellow and outlines, derived from Fig. 2) represented the first tier, and were located mostly along the medial two thirds of Heschl's gyri, consistent with probabilistic locations for primary auditory cortices (Rademacher et al., 2001). FDRRs were surrounded by, and partially overlapped with, HNR-sensitive regions defined using IRNs (green), and those regions were flanked laterally by HNR-sensitive regions defined using animal vocalizations (dark blue). Together, these HNR-sensitive regions were tentatively regarded as encompassing a second tier, although they may be comprised of multiple processing stages.

Regions preferential for processing human vocalizations comprised a third tier, which included cortex extending into the superior temporal sulcus (STS). This included patches of cortex preferential for speech (purple) relative to human nonverbal vocalizations, which were strongly lateralized to the left STS, consistent with earlier studies (Zatorre et al., 1992; Belin et al., 2000; Binder et al., 2000; Scott and Wise, 2003), and patches of cortex preferential for human nonverbal vocalizations (pink) relative to animal vocalizations, which were lateralized to the right hemisphere, also consistent with earlier studies (Belin et al., 2000, 2002).

Within all ROIs representative of these three tiers (Fig. 5, color-coded histograms), human vocalizations produced the greatest degree of activation, even within the IRN HNR-sensitive regions (green boxes). However, when progressing from IRN HNR-sensitive regions to animal vocalization HNR-sensitive re-

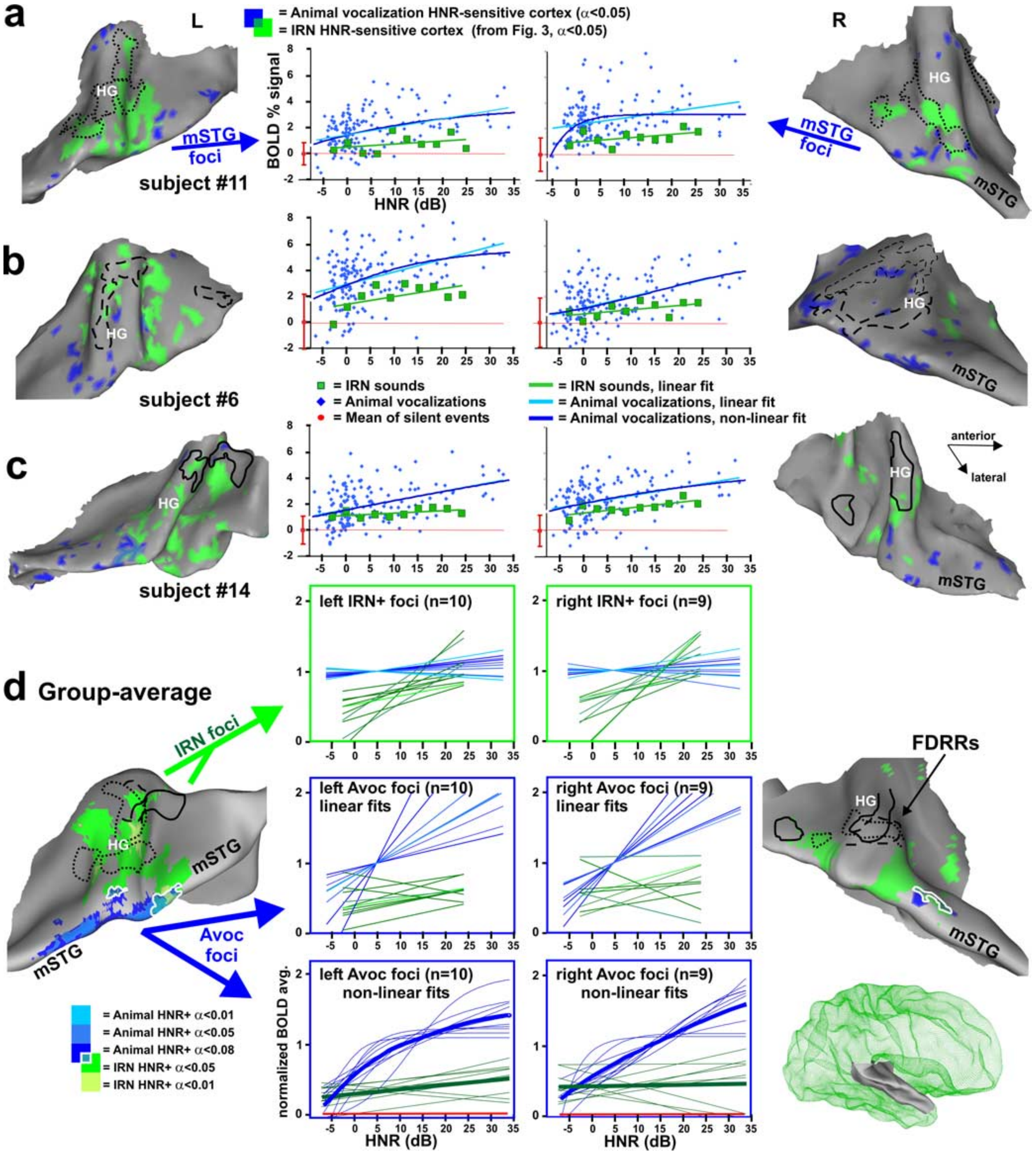

Figure 4. Cortex sensitive to the degree of harmonic structure of animal vocalizations. ***a–c***, Individual cortical maps illustrating animal vocalization HNR-sensitive cortex (blue), based on a linear regression model. IRN HNR-sensitive foci (green) and FDRR outlines (black) are from Figure 3. Charts show the relation between HNR value and BOLD signal from the animal vocalization foci (blue) and IRN HNR-sensitive foci (green). The IRN data depicted in the charts were the control stimuli from paradigm 4 (as opposed to the data from paradigm 3 in Fig. 3), allowing for a direct comparison of relative activation response magnitudes (BOLD signal). All data are in percentage BOLD signal change relative to the mean responses to silent events (red dot at zero, mean + SD). ***d***, Group-averaged maps of HNR-sensitive cortex to animal vocalizations ($n = 11$, blue: *t* tests, see color key) and to IRN stimuli (green, from Fig. 3*d*) on the averaged surface model from Figure 3. White outlines encircle regions of overlap between IRN and animal vocalization HNR-sensitive regions. In the charts, thin curves are those from different individuals, normalized to the mean BOLD response within each ROI defined by the animal vocalization data. Not all participants showed significant bilateral activation ($n = 10$ left, $n = 9$ right hemisphere). Thick curves show the respective response averages. Some hemispheres revealed foci showing a significant negative, linear correlation with HNR value of the IRN and/or animal vocalizations (data not shown). When present, these foci were typically located along the medial wall of the lateral sulcus, and were more commonly observed in the right hemisphere. However, these negatively correlated HNR-sensitive foci were not significant in the group-averaged data.

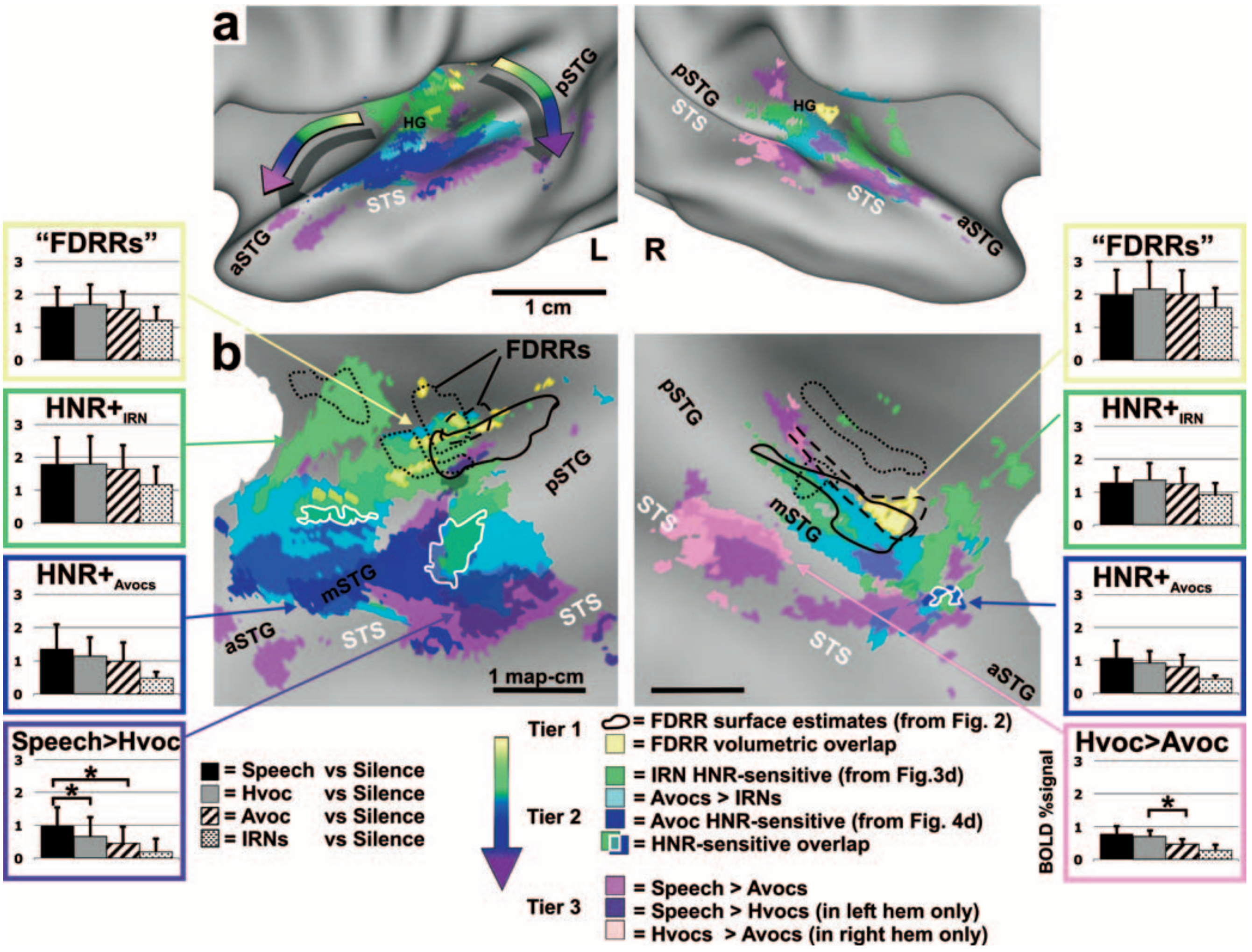

Figure 5. Location of HNR-sensitive cortices relative to human vocalization processing pathways and FDRRs. Data are illustrated on slightly inflated (***a***) and "flat map" (***b***) renderings of our averaged cortical surface models. Volumetric averages of FDRR (yellow) and volumetrically aligned FDRR boundary outlines (black) were derived from data in Figure 2. HNR-sensitive data are from Figure 4*d*. Data from paradigm 5 (Speech, Hvoc, Avoc, IRN) are all at $\alpha < 0.01$, corrected. Refer to key for color codes. Intermediate colors depict regions of overlap. The "rainbow" arrows in ***a*** depict two prominent progressions of processing tiers showing increasing specificity for the acoustic features present in human vocalizations. Overlap of IRN and animal vocalization HNR sensitivity are indicated (white outlines). Histograms from several ROIs show group-averaged response magnitudes (mean + SD) to each of the four sound categories used in paradigm 5 (refer to Results for other details).

gions to speech-sensitive regions, activation became significantly preferential for human vocalizations (e.g., purple and pink boxed histograms). This three-tiered spatial progression was generally consistent with proposed hierarchically organized pathways for processing conspecific vocalizations in both human (Binder et al., 2000; Davis and Johnsrude, 2003; Scott and Wise, 2003; Uppenkamp et al., 2006) and nonhuman primates (Rauschecker et al., 1995; Petkov et al., 2008), and with the identification of an auditory "what" stream for processing conspecific vocalizations and calls (Rauschecker et al., 1995; Wang, 2000).

Subcategories of vocalizations fall along an HNR continuum

Do the global HNR values of human or nonhuman animal vocal communication sounds have any behavioral relevance? We further sought to determine whether our approach of exploring global HNR values could be useful for further characterizing different subcategories of human and animal vocal communication sounds, concordant with ethological considerations in the evolution of vocal production (Wilden et al., 1998; Riede et al., 2005; Bass et al., 2008)

In addition to the vocalizations used in neuroimaging paradigms 4 and 5, we also derived HNR value ranges and means for several conceptually distinct subcategories of human communication sounds (see Materials and Methods). Indeed, various subcategories of vocalizations could be at least roughly organized along the HNR continuum (Fig. 6, colored ovals and boxes). In the lower HNR ranges, this included hisses and a subcategory that included growls, grunts, and groans, most of which are vocalizations associated with threat warnings or negative emotional valence. Whispered speech, as a subcategory, was also characterized by relatively low HNR values, consistent with its social function as an acoustic signal with a low transmission range and reduced speech perceptibility (Cirillo, 2004). At the other extreme, vocal singing and whistling sounds (although not produced by vibrating tissue folds) were characterized by significantly higher HNR values than those typical for conversational speech. We also derived HNR values of spoken phrase segments from adults (n = 10) when speaking in monologue to other adults versus when speaking to a realistic infant doll (Fig. 6, rectangles) (see Materials and Methods). Interestingly, in addition to generally increasing

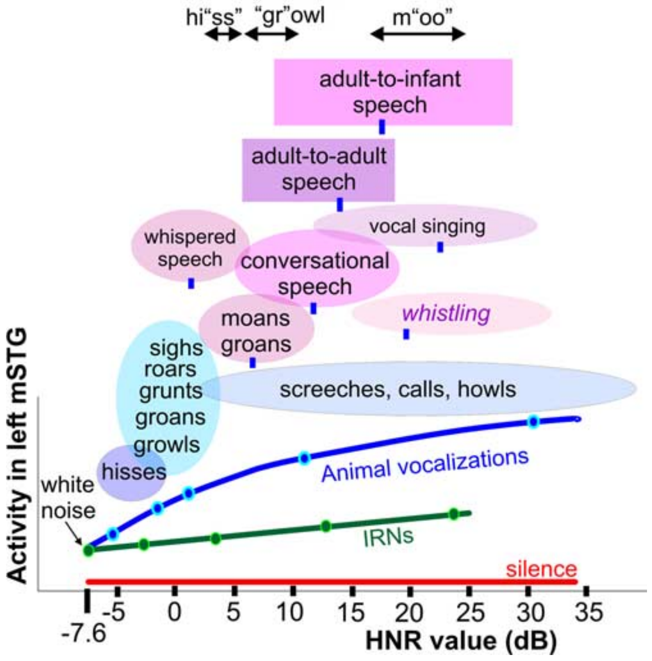

Figure 6. Typical HNR value ranges for various subcategories of vocalizations. Oval and box widths depict the minimum to maximum HNR values of the sounds we sampled, charted relative to the group-averaged HNR-sensitive response profile of the left mSTG (from Fig. 4*d*). Green and blue dots correspond to sound stimuli illustrated in Figure 1, *a* and *b*. Blue ovals depict subcategories of animal vocalizations explicitly tested in paradigm 4. Ovals and boxes with violet hues depict subcategories of human vocalizations (12–18 samples per category), and blue tick marks indicate the mean HNR value. For instance, conversational speech, including phrases explicitly tested in paradigm 5, had a mean of +12 dB HNR, within a range from approximately +5 to +20 dB HNR. Adult-to-adult speech (purple box; mean = +17.2 dB HNR) and adult-to-infant speech (violet box; mean = +14.0 dB HNR) produced by the same individual speakers were significantly different (*t* test, $p < 10^{-5}$). Stressed phonemes of three spoken onomatopoetic words depicting different classes of vocalizations are also indicated. Refer to Materials and Methods for other details.

in pitch, each participant’s voice was characterized by significantly greater harmonic structure when speaking to an infant.

Also noteworthy was that the vocalization subcategories tended to have onomatopoetic descriptors (in many languages), which when spoken stress phonemes that correlate with the HNR structure of the corresponding category of sound. For instance, we recorded phrases from multiple speakers and found the “ss” in “hissing” to be consistently lower in HNR value range than the “gr” in “growling”, which was lower than the “oo” in “mooing” (Fig. 6, top) (see Materials and Methods). Moreover, onomatopoetic words (in Japanese) have previously been associated with activation of the bilateral (left > right) STG/STS (Hashimoto et al., 2006), overlapping blue to violet/purple regions in Figure 5. Together, these results suggest that variations of harmonic structure during vocal production, by animals or humans, can be used to convey fundamentally different types of behaviorally relevant information.

Discussion

The main finding of the present study was that bilateral portions of Heschl’s gyri and mSTG (left > right) showed significant increases in activation to parametric increases in overall harmonic structure of either artificially constructed IRNs and/or natural animal vocalizations. Within individuals, these HNR-sensitive foci were situated between functionally defined primary auditory cortices and regions preferential for human vocalizations in both hemispheres, but with a significant left-lateralization. We propose that the explicit processing of harmonic content serves as an important bottom-up, second-order signal attribute in a hierarchical model of auditory processing, which are comprised of pathways optimized for extracting vocalizations. In particular, HNR-sensitive cortex may function as an integral component of computationally theorized spectro-temporal template staging, which serves as a basic neural mechanism for the segregation of acoustic events (Medvedev et al., 2002; Kumar et al., 2007). Thus, higher-order signal attributes, or primitives, that are characteristic of behaviorally relevant real-world sounds experienced by the listener may become encoded along intermediate processing stages leading to the formation of spectro-temporal templates, which dynamically develop to statistically reflect these acoustic structures. In the mature brain, matches between components of an incoming sound and these templates may subsequently convey information onto later processing stages to further group acoustic features, segment the sound, and ultimately lead to its identification, meaning or relevance.

However, why did the IRN and animal vocalization HNR-sensitive regions (Fig. 4, green vs blue foci) of auditory cortex not completely overlap to indicate a single, centralized stage of HNR processing? Our results were consistent with previous neuroimaging studies manipulating pitch salience or temporal regularity of IRNs or complex tones (cf. Figs. 3–5, green), all of which revealed bilateral activation along lateral portions of Heschl’s gyri and/or the STG (Griffiths et al., 1998; Patterson et al., 2002; Krumbholz et al., 2003; Penagos et al., 2004; Hall et al., 2005). HNR sensitivity for animal vocalizations may not have overlapped the entire IRN HNR-sensitive region because other features of animal vocalizations, regardless of their HNR value, contributed to the maximal or near maximal BOLD activation within both FDRRs and IRN HNR-sensitive locations (Fig. 5). As a result, animal vocalization HNR sensitivity may not have been detectable. Conversely, IRN HNR-sensitive regions may not overlap animal vocalization HNR-sensitive regions due to serial hierarchical processing of acoustic features. IRNs, with relatively simple harmonic structure (equal power at every integer harmonic), appeared to be effectively driving early stages of frequency combination-sensitive processing. However, the IRNs were less capable of significantly driving subsequent stages along the mSTG, and thus may have been effectively filtered out from the pathways we identified for processing vocalizations. The other signal attributes required to drive higher stages (mSTG and STS) presumably include more specific combinations and distributions of power of harmonic and subharmonic frequencies that more closely reflect the statistical structure of components characteristic of vocalizations (Darwin, 1984; Shannon et al., 1995; Giraud et al., 2000). The series of acoustic paradigms that we used at minimum serve to identify cortical regions for further study highlighting additional acoustic attributes. Although other higher-order signal attributes that would further test this model remain to be explored, the present data indicate that harmonic structure represents a major, quantifiable second-order attribute that can differentially drive intermediate processing stages of auditory cortex, consistent with a hierarchical stectro-temporal template model for sound processing.

The apparent hierarchical location of HNR-sensitive regions may be a corollary to the intermediate cortical stages of other sensory systems. For example, V2, V4 and TEO in human visual cortex (Kastner et al., 2000) and S2 in primate somatosensory cortex (Jiang et al., 1997) have "larger" and more complex receptive fields relative to their respective primary sensory areas, showing sensitivity to textures, shapes, and patterns leading to object segmentation. In all three modalities, these intermediate cortical stages may be integrating specific combinations (second-order features) of input energy across spatially organized maps corresponding to their respective sensory epithelia. In this regard, HNR-sensitive regions appear to represent cortical processing stages analogous to intermediate hierarchical stages in other sensory modalities, potentially reflecting a general processing mechanism of sensory cortex.

Cortical organization for processing different categories of real-world sounds

The present results supported and further extended our previous findings, in that the preferential activation of mSTG by animal vocalizations, compared with hand-tool sounds, was likely due to the greater degree of harmonic content in the vocalizations (Lewis et al., 2005, 2006). Thus, HNR-sensitive stages could be facilitating the processing of vocalizations as a distinct category of real-world sound. However, an auditory evoked potential study examining responses to sounds representative of living objects (which included vocalizations) versus man-made objects, both of which were explicitly matched overall in HNR values, reported a differential processing component between the two categories starting ~70 ms from the onset of sound (Murray et al., 2006). Thus, it is clear that complex signal attributes other than global HNR value are contributing grossly to early stages of sound categorization. Nonetheless, HNR sensitivity should be considered when exploring processing pathways for different categories of sound.

Human vocalizations, as a subcategory of sound distinct from animal vocalizations, are generally characterized by more idiosyncratic combinations of frequencies, specific relative power distributions, as well as other spectral and temporal attributes not taken into consideration here (Rosen, 1992; Shannon et al., 1995; Wilden et al., 1998; Belin et al., 2000, 2004; Cooke and Ellis, 2001). These other more subtle signal attribute differences appear to be necessary to evoke activation of the speech-sensitive regions we and others have observed along the STG/STS regions. Those regions are thought to represent subsequent hierarchical stages involved more with processing acoustic primitives or symbols just before extracting linguistic content (Binder et al., 1997; Cooke and Ellis, 2001; Scott and Wise, 2003; Price et al., 2005). Thus, the contributions of HNR relative to other higher-order signal attributes toward the processing of human vocalizations, as an apparently distinct subcategory of vocalizations, remains to be explored.

Relation of HNR sensitivity to speech processing

Evidence for the presence of spectral templates in humans has significant implications for advancing our understanding how one may process and learn to recognize sounds, including speech. In early development, experience with behaviorally relevant vocalizations produced by one's caretakers, and perhaps one's own voice, could help establish the receptive fields of auditory neurons to exhibit sensitivity to their specific frequency combinations, thereby reflecting the statistical distributions of harmonic structure of human (conspecific) vocalizations. These experiences and subsequent cortical encodings will be unique to each individual's listening experience. Large cortical ensembles of frequency combination-sensitive neurons may thus develop (Fig. 4*a–c*, HNR-sensitive patches specific to each individual) to comprise spectral and spectro-temporal templates, and these templates could serve as Bayesian-like networks to rapidly group or stream vocalizations from a person or sound-source (Medvedev et al., 2002; Kumar et al., 2007). As a side note, such principles have already been implemented in automated speech recognition algorithms, in the form of "weft-resynthesis" (Ellis, 1997), which may be an important biologically inspired mechanism for the future development of hearing devices optimized for amplifying speech sounds.

On a larger scale of auditory cortex, and common across individuals, a hierarchical organization appears to become further established. In our data, sounds containing increasing degrees of acoustic structure, defined here as becoming more characteristic of human vocalizations, preferentially recruited cortex extending out to the mSTG and STS in both hemispheres (Fig. 5, rainbow-colored progressions). However, the left hemisphere had more, and better organized, cortex devoted to HNR-sensitive processing, and also a stronger bias for processing human speech sounds (Binder et al., 2000; Boemio et al., 2005). Interestingly, at birth, humans are reported to already have a left hemisphere superiority for processing human linguistic stimuli (Peña et al., 2003). Thus, there may be a predisposition for the left hemisphere to process harmonic sounds, perhaps even being influenced by listening experiences *in utero*.

Interestingly, modifying one's voice to speak to infants, ostensibly to make them happy, was strongly associated with an increase in the harmonic structure of spoken words and phrases (Fig. 6, rectangles). This largely appeared to be due to the elongation of vowel sounds, accompanied by a decrease in noise and other "complicated" acoustic features. Although speculative, this could serve as a socially interactive mechanism to help train the auditory system of a developing infant to recognize and perceive the basic statistical structure of human vocalizations. He or she would then eventually learn to process more complex variations in spectral, temporal, and spectro-temporal structure that convey more specific and behaviorally relevant meaning or communicative content, such as with phonemes, words, prosody, and other basic units of vocal communication and language.

In sum, although the HNR value of a sound is by no means the only important acoustic signal attribute for processing real-world sounds, our results indicate that harmonic structure is parametrically reflected along human auditory cortical pathways for processing vocalizations. This attribute may serve as an integral component for hierarchical processing of sounds, notably including vocalizations as a distinct category of sound. Consequently, the HNR acoustic signal attribute should be considered when studying and distinguishing among neural pathways for processing and recognizing human vocalizations, auditory objects, and other "conceptually" distinct categories of real-world sounds.

References

Altmann CF, Doehrmann O, Kaiser J (2007) Selectivity for animal vocalizations in the human auditory cortex. Cereb Cortex 17:2601–2608.

Bass AH, Gilland EH, Baker R (2008) Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. Science 321:417–421.

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

Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Res Cogn Brain Res 13:17–26.

Belin P, Fecteau S, Bédard C (2004) Thinking the voice: neural correlates of voice perception. Trends Cogn Sci 8:129–135.

Bilecen D, Seifritz E, Scheffler K, Henning J, Schulte AC (2002) Amplitopicity of the human auditory cortex: an fMRI study. Neuroimage 17:710–718.

Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T (1997) Human brain language areas identified by functional magnetic resonance imaging. J Neurosci 17:353–362.

Binder JR, Frost JA, Hammeke TA, Bellgowan PS, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. Cereb Cortex 10:512–528.

Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat Neurosci 8:389–395.

Boersma P (1993) Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc Inst Phon Sci 17:97–110.

Cheang HS, Pell MD (2008) The sound of sarcasm. Speech commun 50:366–381.

Cirillo J (2004) Communication by unvoiced speech: the role of whispering. An Acad Bras Cienc 76:413–423.

Cooke M, Ellis DPW (2001) The auditory organization of speech and other sources in listeners and computational models. Speech commun 35:141–177.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Darwin CJ (1984) Perceiving vowels in the presence of another sound: constraints on formant perception. J Acoust Soc Am 76:1636–1647.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Edmister WB, Talavage TM, Ledden PJ, Weisskoff RM (1999) Improved auditory cortex imaging using clustered volume acquisitions. Hum Brain Mapp 7:89–97.

Ellis DPW (1997) The weft: a representation for periodic sounds. In: IEEE International Conference on Acoustics, Speech and Siginal Processing, pp 1307–1310.

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.

Fujioka T, Kakigi R, Gunji A, Takeshima Y (2002) The auditory evoked magnetic fields to very high frequency tones. Neuroscience 112:367–381.

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.

Glover GH, Law CS (2001) Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46:515–522.

Griffiths TD, Büchel C, Frackowiak RS, Patterson RD (1998) Analysis of temporal structure in sound by the human brain. Nat Neurosci 1:422–427.

Hall DA (2005) Representations of spectral coding in the human brain. Int Rev Neurobiol 70:331–369.

Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. Hum Brain Mapp 7:213–223.

Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. Cereb Cortex 12:140–149.

Hall DA, Barrett DJ, Akeroyd MA, Summerfield AQ (2005) Cortical representations of temporal structure in sound. J Neurophysiol 94:3181–3191.

Hashimoto T, Usui N, Taira M, Nose I, Haji T, Kojima S (2006) The neural mechanism associated with the processing of onomatopoeic sounds. Neuroimage 31:1762–1770.

Jäncke L, Shah NJ, Posse S, Grosse-Ryuken M, Müller-Gärtner HW (1998) Intensity coding of auditory stimuli: an fMRI study. Neuropsychologia 36:875–883.

Jiang W, Tremblay F, Chapman CE (1997) Neuronal encoding of texture changes in the primary and the secondary somatosensory cortical areas of monkeys during passive texture discrimination. J Neurophysiol 77:1656–1662.

Jones SJ (2006) Cortical processing of quasi-periodic versus random noise sounds. Hear Res 221:65–72.

Kastner S, De Weerd P, Ungerleider LG (2000) Texture segregation in the human visual cortex: A functional MRI study. J Neurophysiol 83:2453–2457.

Krumbholz K, Patterson RD, Seither-Preisler A, Lammertmann C, Lütkenhöner B (2003) Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. Cereb Cortex 13:765–772.

Kumar S, Stephan KE, Warren JD, Friston KJ, Griffiths TD (2007) Hierarchical processing of auditory objects in humans. PLoS Comput Biol 3:e100.

Langers DR, Backes WH, van Dijk P (2007) Representation of lateralization and tonotopy in primary versus secondary human auditory cortex. Neuroimage 34:264–273.

Langner G (1992) Periodicity coding in the auditory system. Hear Res 60:115–142.

Lewicki MS, Konishi M (1995) Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. Proc Natl Acad Sci U S A 92:5582–5586.

Lewis JW (2006) Cortical networks related to human use of tools. Neuroscientist 12:211–231.

Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. J Neurosci 25:5148–5158.

Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe EA (2006) Lefties get it "right" when hearing tool sounds. J Cogn Neurosci 18:1314–1330.

Medvedev AV, Kanwal JS (2004) Local field potentials and spiking activity in the primary auditory cortex in response to social calls. J Neurophysiol 92:52–65.

Medvedev AV, Chiao F, Kanwal JS (2002) Modeling complex tone perception: grouping harmonics with combination-sensitive neurons. Biol Cybern 86:497–505.

Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S (2006) Rapid brain discrimination of sounds of objects. J Neurosci 26:1293–1302.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

Peña M, Maki A, Kovacić D, Dehaene-Lambertz G, Koizumi H, Bouquet F, Mehler J (2003) Sounds and silence: an optical topography study of language recognition at birth. Proc Natl Acad Sci U S A 100:11702–11705.

Penagos H, Melcher JR, Oxenham AJ (2004) A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. J Neurosci 24:6810–6815.

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.

Price C, Thierry G, Griffiths T (2005) Speech-specific auditory processing: where is it? Trends Cogn Sci 9:271–276.

Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. Neuroimage 13:669–683.

Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. Science 268:111–114.

Riede T, Herzel H, Hammerschmidt K, Brunnberg L, Tembrock G (2001) The harmonic-to-noise ratio applied to dog barks. J Acoust Soc Am 110:2191–2197.

Riede T, Mitchell BR, Tokuda I, Owren MJ (2005) Characterizing noise in nonhuman vocalizations: acoustic analysis and human perception of barks by coyotes and dogs. J Acoust Soc Am 118:514–522.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philos Trans R Soc Lond B Biol Sci 336:367–373.

Schönwiesner M, von Cramon DY, Rübsamen R (2002) Is it tonotopy after all? Neuroimage 17:1144–1161.

Scott S, Wise R (2003) PET and fMRI studies of the neural basis of speech perception. Speech commun 41:23–34.

Scott SK (2005) Auditory processing–speech, space and auditory objects. Curr Opin Neurobiol 15:197–201.

Shama K, Krishna A, Cholayya NU (2007) Study of harmonics-to-noise ratio and critical-band energy spectrum of speech as acoustic indicators of laryngeal and voice pathology. EURASIP J Adv Signal Process 2007:85286 (1–9).

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Shofner WP (1999) Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: analysis of neural autocorrelograms. J Neurophysiol 81:2662–2674.

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.

Terhardt E (1974) Pitch, consonance, and harmony. J Acoust Soc Am 55:1061–1069.

Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD (2006) Locating the initial stages of speech-sound processing in human temporal cortex. Neuroimage 31:1284–1296.

Van Essen DC (2003) Organization of visual areas in macaque and human cerebral cortex. In: The visual neurosciences (Chalupa L, Werner JS, eds), pp 507–521. Cambridge, MA: MIT.

Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Inform Assoc 8:443–459.

Wang X (2000) On cortical coding of vocal communication sounds in primates. Proc Natl Acad Sci U S A 97:11843–11849.

Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. J Cogn Neurosci 13:1–7.

Wilden I, Herzel H, Peters G, Tembrock G (1998) Subharmonics, biphonation, and deterministic chaos in mammal vocalization. Bioacoustics 9:171–196.

Yetkin FZ, Roland PS, Christensen WF, Purdy PD (2004) Silent functional magnetic resonance imaging (FMRI) of tonotopicity and stimulus intensity coding in human primary auditory cortex. Laryngoscope 114:512–518.

Yost WA (1996) Pitch strength of iterated rippled noise. J Acoust Soc Am 100:3329–3335.

Zatorre RJ, Evans AC, Meyer E, Gjedde A (1992) Lateralization of phonetic and pitch discrimination in speech processing. Science 256:846–849.