

Active Sound Localization Sharpens Spatial Tuning in Human Primary Auditory Cortex

Kiki van der Heijden,¹  Josef P. Rauschecker,²  Elia Formisano,^{1,3} Giancarlo Valente,¹ and Beatrice de Gelder^{1,4}

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

²Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007,

³Maastricht Center for Systems Biology, Maastricht University, 6200 MD, Maastricht, The Netherlands, and ⁴Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, United Kingdom

Spatial hearing sensitivity in humans is dynamic and task-dependent, but the mechanisms in human auditory cortex that enable dynamic sound location encoding remain unclear. Using functional magnetic resonance imaging (fMRI), we assessed how active behavior affects encoding of sound location (azimuth) in primary auditory cortical areas and planum temporale (PT). According to the hierarchical model of auditory processing and cortical functional specialization, PT is implicated in sound location (“where”) processing. Yet, our results show that spatial tuning profiles in primary auditory cortical areas (left primary core and right caudo-medial belt) sharpened during a sound localization (“where”) task compared with a sound identification (“what”) task. In contrast, spatial tuning in PT was sharp but did not vary with task performance. We further applied a population pattern decoder to the measured fMRI activity patterns, which confirmed the task-dependent effects in the left core: sound location estimates from fMRI patterns measured during active sound localization were most accurate. In PT, decoding accuracy was not modulated by task performance. These results indicate that changes of population activity in human primary auditory areas reflect dynamic and task-dependent processing of sound location. As such, our findings suggest that the hierarchical model of auditory processing may need to be revised to include an interaction between primary and functionally specialized areas depending on behavioral requirements.

Key words: cortical functional specialization; fMRI; human auditory cortex; sound localization

Significance Statement

According to a purely hierarchical view, cortical auditory processing consists of a series of analysis stages from sensory (acoustic) processing in primary auditory cortex to specialized processing in higher-order areas. Posterior-dorsal cortical auditory areas, planum temporale (PT) in humans, are considered to be functionally specialized for spatial processing. However, this model is based mostly on passive listening studies. Our results provide compelling evidence that active behavior (sound localization) sharpens spatial selectivity in primary auditory cortex, whereas spatial tuning in functionally specialized areas (PT) is narrow but task-invariant. These findings suggest that the hierarchical view of cortical functional specialization needs to be extended: our data indicate that active behavior involves feedback projections from higher-order regions to primary auditory cortex.

Introduction

Sound localization is a crucial component of mammalian hearing. In the mammalian auditory cortex, neural activity in poste-

rior areas is modulated by sound location more than in primary and anterior areas. These spatially-sensitive areas include the caudo-medial (CM) and caudo-lateral belt areas (CL) in nonhuman primates (Tian et al., 2001), the posterior auditory field (Harrington et al., 2008) and dorsal zone in cats (Stecker and

Received March 1, 2018; revised July 9, 2018; accepted July 19, 2018.

Author contributions: K.v.d.H. wrote the first draft of the paper; K.v.d.H., J.P.R., E.F., G.V., and B.d.G. edited the paper; K.v.d.H., J.P.R., E.F., and B.d.G. designed research; K.v.d.H. performed research; K.v.d.H., E.F., and G.V. analyzed data; K.v.d.H., J.P.R., E.F., and B.d.G. wrote the paper.

This work was supported by a European Research Council Grant under the European Union Seventh Framework Programme for Research 2007–2013 (Grant 295673), a European Union's Horizon 2020 Research and Innovation Programme Grant 645553, ICT DANCE (IA, 2015–2017; B.d.G.), a PIRE Grant from the U.S. National Science Foundation (OISE-0730255; J.P.R.), NIH Grants R01EY018923 and R01DC014989 (J.P.R.); partial support from the Technische Universität München Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme Grant 291763 (J.P.R.), and NWO Vici-Grant 453-12-002 and the Dutch

Province of Limburg (E.F.), and funding for a research exchange from the Erasmus Mundus Auditory Cognitive Neuroscience Network (K.H.).

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Beatrice de Gelder, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6200 MD, Maastricht, The Netherlands. E-mail: b.degelder@maastrichtuniversity.nl.

DOI:10.1523/JNEUROSCI.0587-18.2018

Copyright © 2018 the authors 0270-6474/18/388574-14\$15.00/0

Middlebrooks, 2003; Stecker et al., 2005; Lomber and Malhotra, 2008), and the planum temporale (PT) in humans (Warren and Griffiths, 2003; Brunetti et al., 2005; Deouell et al., 2007; van der Zwaag et al., 2011; Derey et al., 2016; McLaughlin et al., 2016). For this reason, cortical processing of sound location is presumably taking place in a functionally specialized, posterior-dorsal “where” stream (Rauschecker and Tian, 2000; Tian et al., 2001; Arnott et al., 2004; Rauschecker and Scott, 2009).

Behavioral evidence from psychophysical studies shows that auditory spatial sensitivity in humans is dynamic. For example, an auditory target is processed faster when auditory spatial attention is focused at the location of the target (Spence and Driver, 1994; Mondor and Zatorre, 1995; Rorden and Driver, 2001). A recent study investigating the neural mechanisms underlying this dynamic spatial sensitivity in cats identified the primary auditory cortex (A1) as a potential locus for such dynamic sound location processing. (Lee and Middlebrooks, 2011). In humans, a recent study reported a region in posterior auditory cortex that exhibited a differential level of activation based on task performance, but no task modulation of selectivity to interaural level differences (ILD) or interaural time differences (ITD) across the entire auditory cortex (Higgins et al., 2017). However, it is presently not clear whether task performance results in sharpening of spatial tuning within distinct regions of the human auditory cortex, and whether this sharpening occurs preferentially in functionally specialized “where” regions (i.e., PT) or also affects A1.

Moreover, the effects of task performance on the cortical encoding of sound location are not yet known. The computational mechanisms underlying cortical sound location encoding are still a matter of debate, and prior studies assessing the validity of these computational mechanisms have not addressed possible effects of task performance (McAlpine et al., 2001; Stecker and Middlebrooks, 2003; Harper and McAlpine, 2004; Stecker et al., 2005; King et al., 2007; Miller and Recanzone, 2009; Day and Delgutte, 2013; Derey et al., 2016; Ortiz-Rios et al., 2017).

Here we measured with functional magnetic resonance imaging (fMRI) the neuronal population responses to different sound azimuth positions in the human auditory core, lateral belt areas, and PT, while participants performed different behavioral tasks. We then evaluated the spatial selectivity of neuronal populations within these areas across task conditions. Additionally, we applied a modified version of a maximum-likelihood population-pattern decoder previously used to decode sound location from neural spike rates (Jazayeri and Movshon, 2006; Miller and Recanzone, 2009; Day and Delgutte, 2013) to assess whether sound location encoding in fMRI activity patterns in human auditory cortex within and across hemispheres is modulated by task performance. Our results provide new insights into the dynamic nature of sound location encoding in human A1. In particular, in agreement with “reverse hierarchy” (Ahissar et al., 2009) and “recurrent processing” models (Lamme and Roelfsema, 2000; Bullier, 2001), our data suggest that behavior (sound localization) is enabled by feedback from functionally specialized areas to A1.

Materials and Methods

Participants

Thirteen human volunteers gave informed consent to participate in the experiment. Data of two participants were excluded from the analysis due to insufficient data quality as a consequence of excessive motion and participant fatigue. Data of the remaining 11 participants (mean age = 28.9 years, SD = 11.7 year, 7 females) are presented here. Participants reported no history of neurological disorders. We assessed hearing levels with pure-tone thresholds (0.5, 1, 2, 4, 8 kHz) using an Oscilla SM910 Screening Audiometer. Hearing thresholds did not exceed 25 dB for any

of the frequencies tested. The institutional review board of Georgetown University granted approval for the study.

Stimuli

Stimuli consisted of amplitude-modulated (AM) white noise clips (*probe* sounds: duration = 1200 ms) and click trains (*target* sounds: click rate = 200 Hz, duration = 1200 ms). Probe and target sounds were created with MATLAB (MathWorks). Stimuli were presented at 1 of 7 locations (-90° , -60° , -30° , 0° , $+30^\circ$, 60° , and $+90^\circ$; Fig. 1A).

All stimuli were spatialized by making subject-specific binaural recordings (Derey et al., 2016). During the binaural-recording session, participants sat in a chair in the center of a production studio (internal volume = 66 m³; walls and ceiling consisted of gypsum board covered with fabric, the floor consisted of concrete covered with a carpet) with binaural microphones placed in their ear canals (OKM II Classic Microphone, Soundman). A loudspeaker positioned at zero elevation in the far field (distance to subject = 1.3 m) presented sounds at each of the locations (Fig. 1A). This procedure resulted in stimuli with a clear spatial percept based on available ILD, ITD, and spectral cues (Fig. 1C,D).

Each stimulus was prefiltered with headphone equalization filters provided by the manufacturer of the MRI-compatible earbuds used in the present study (Sensimetrics S14). The headphone equalization filters ensure a flat frequency response at the level of the earbuds and remove headphone-induced phase offsets between the earbuds.

For the tonotopy measurements, we used amplitude-modulated pure tones (rate of modulation = 10 Hz, full-depth modulation, 800 ms duration). Pure tones were centered on eight center frequencies (0.18, 0.30, 0.51, 0.86, 1.46, 2.48, 4.19, 7.09 kHz) with a slight variation of ± 0.1 octave to prevent habituation (De Martino et al., 2013). Stimuli for the tonotopy measurements were prefiltered with the headphone equalization filters as well.

Experimental design

Participants listened to probe trials in three behavioral conditions: *passive listening*, *sound identification*, and *sound localization*. Probe trials consisted of five repetitions of a probe sound clip (duration = 1200 ms) at the same location. Sound clips were presented in silent gaps (1.4 s) in between fMRI acquisition periods (2 s; see Data acquisition), resulting in a total duration of 17 s per trial (5 stimulus repetitions in silent gaps of 1.4 s plus 5 fMRI data acquisition periods of 2 s; Fig. 1B). In the active listening conditions only, participants also listened to target trials. Specifically, in the sound identification condition, target trials had a similar structure (i.e., 5 repetitions at the same azimuthal location), yet the fourth or the fifth repetition of the probe sounds (AM white noise) was replaced by a deviant target sound (click train) at the same location (Fig. 1B). In the sound localization condition, target trials had a similar structure as well, but the fourth or the fifth repetition of the probe sound (AM white noise) was replaced by a probe sound at a deviant azimuth location. For example, the first four stimuli were presented at -90° and the fifth stimulus at $+30^\circ$ (Fig. 1B).

During fMRI acquisition, trials were grouped by task (passive listening, sound identification, sound localization) in a block. In each block, probe trials were presented once at each azimuth location and were separated by an intertrial interval of 12.2 s (for detailed information, see Data acquisition). The order of azimuth locations was randomized within a block. Thus, for passive listening, a block consisted of seven probe trials, one at each azimuth location. For the active tasks, sound localization and sound identification, a block also contained two target trials (equivalent to $\sim 22\%$ of the total number of trials) in addition to the seven probe trials. The order of target and probe trials was randomized within a block.

Each participant performed one block of each task per run of fMRI acquisition. Thus, one run consisted of three blocks corresponding to the three behavioral task conditions. At the start of each task block, a short audio clip of a voice informed participants of the task at hand: “sound location”, “sound identity”, or “passive listening”. In the passive listening condition, participants listened to the sounds without making a response. In the sound identification condition, participants pressed a button immediately upon detection of a target sound within a target trial

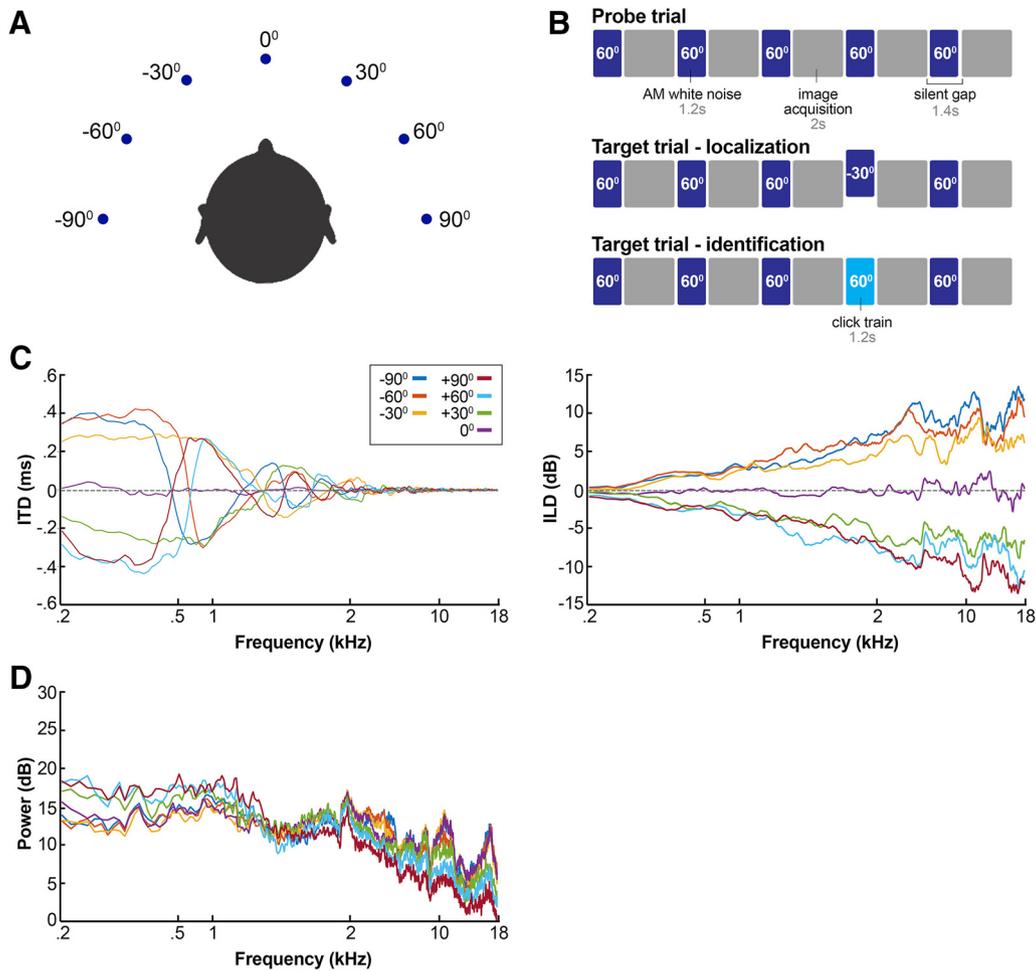


Figure 1. Stimuli. **A**, Azimuth locations at which sound sources were presented. **B**, Example of a probe trial (top), a target trial for the sound localization task (middle), and a target trial for the sound identification task (bottom). A probe trial consisted of a block of five stimulus presentations at one azimuth location. In the sound localization task, the target trial consisted of five stimulus presentations as well, yet for the fourth (depicted here) or fifth repetition the azimuth location was changed. For target trials in the sound identification condition, azimuth location remained constant across the five stimulus repetitions but the fourth or fifth repetition was replaced by a deviant click train. **C**, Lines reflect the ITD (left) and ILD (right) for stimuli at a specific sound azimuth position, averaged across the binaural recordings of all participants. ILD was computed as the arithmetic difference in power (measured as root mean square) between the left and right channel of each binaural recording. To compute ITD, we first computed the interaural phase difference, which we subsequently converted to time differences. **D**, Plotted is the power spectrum of the left channel of the binaural recordings (i.e., the left ear) at specific azimuth positions, averaged across all participants. The difference in power in specific frequency bands dependent on sound azimuth location illustrates the availability of spectral cues in the recordings. Colors similar to **C**.

(i.e., the click train). In the sound localization condition, participants pressed a button immediately upon detecting a location switch within a target trial.

The order of blocks was randomized and counterbalanced across participants. In total, participants completed four runs of the main experiment (~10 min each) in the MRI scanner. This resulted in four probe trial repetitions per azimuth location per task condition. Only probe trials were included in the subsequent analyses (see Data analysis).

Before the fMRI measurements, participants performed a short practice session to get familiar with the tasks and with the MRI environment. This also enabled participants to get accustomed to the auditory spatial percept in a supine frame of reference (due to the supine position required by the MRI scanner). The practice session consisted of passive presentation of the probe stimuli at each location as well as short task blocks of the sound localization and the sound identification task, in which one target trial was presented per task block.

Finally, the scan session was concluded with two runs of tonotopy measurements (~7.5 min each). For this experiment, participants listened passively to blocks of AM pure tones in the MRI scanner. Each block was repeated twice per run, resulting in four repetitions per center frequency. The order of frequency blocks was randomized (De Martino et al., 2013).

Data acquisition

Data were acquired with a Siemens TIM Trio 3-tesla MRI scanner at the Center for Functional and Molecular Imaging at Georgetown University. For the main experiment, blood oxygenation level-dependent (BOLD) signals were measured with a T_2^* -weighted echoplanar imaging (EPI) sequence covering the temporal cortex and parts of the occipital, parietal, and frontal cortex [echo time (TE) = 30 ms; repetition time (TR) = 3400 ms; flip angle = 90°; number of slices = 32; voxel size = 2 mm³ isotropic]. Image acquisition was clustered [acquisition time (TA) = 2000 ms], and binaural recordings were presented in silent gaps (duration = 1400 ms) between subsequent volume acquisitions through MR-compatible insert earphones (Sensimetrics S14) with sound-attenuating foam ear tips (>29 dB attenuation). One sound was presented per TR. Trials (i.e., 5 stimulus repetitions per azimuth location corresponding to 5 TRs, 17 s duration) were separated by three volumes in which no sound was presented (that is, 12.2 s silence) to allow the BOLD signal to return to baseline before the onset of the next trial.

We also acquired a high resolution anatomical image of the whole brain with a MPRAGE T1-weighted sequence (TE = 2.13 ms; TR = 2400 ms; voxel size = 1 mm³ isotropic). For the tonotopic measurements we also used a sparse T_2^* -weighted EPI sequence to measure the BOLD signal, covering mainly the temporal cortex (TE = 30 ms; TR = 2600 ms;

TA = 1600 ms; silent gap = 1000 ms; flip angle = 90°; number of slices = 25; voxel size = 2 mm³ isotropic). In each run, AM pure tones were presented in the silent intervals between subsequent volume acquisitions in blocks of six repetitions per center frequency (15.6 s). Blocks were separated by 12 s of silence (4 volumes).

Statistical analysis

Data preprocessing. Functional and anatomical data were analyzed using BrainVoyager QX (Brain Innovation), and customized MATLAB code. Preprocessing of functional images included motion correction (trilinear/sinc interpolation, we used the first run of first volume as reference volume for aligning), slice scan time correction (sinc interpolation), linear drifts removal, temporal high pass filtering (threshold = 7 cycles per run), and mild spatial smoothing (3 mm kernel). Functional images were coregistered to the anatomical T1-weighted image and transformed to 3D Talairach space (Tournoux and Talairach, 1988). Gray–white matter boundaries were defined with the BrainVoyager QX automatic segmentation procedure and manually improved when necessary.

Group analyses were performed in surface space to ensure optimal alignment of the auditory cortex across participants. To this end, we applied cortex-based alignment (CBA) to the surface reconstruction of each participant (Goebel et al., 2006) with the additional constraint of an anatomical definition of Heschl's gyrus (HG; Kim et al., 2000; Morosan et al., 2001). High-resolution surface mesh time courses were created by sampling and averaging for each point on the surface (that is, each vertex) the values from –1 mm below the gray–white matter boundary up to 2 mm in the gray matter toward the pial surface.

Univariate analysis of the processing of spatialized sounds. To test for the general response to presentation of spatialized sounds, we estimated a random effects general linear model (RFX GLM) with a predictor for sound presentation including all probe trials (regardless of azimuth location or behavioral task condition). Target trials were modeled with a separate predictor and not included in the contrast.

Response azimuth functions. We constructed a response azimuth function (RAF) for each auditory responsive voxel [individual subject GLM with one predictor per sound azimuth location per task condition and excluding target trials, contrast auditory stimuli > baseline, $q(\text{FDR}) < (0.05)$]. RAFs consisted of location-specific β values estimated with a GLM with one predictor per sound location per task. RAFs were mildly smoothed with a moving average window of three points [weights (0.2, 0.6, 0.2)]. A peak response was defined as a response at 75% or more of the maximum β value in the RAF (Stecker and Middlebrooks, 2003; Stecker et al., 2005; Derey et al., 2016). Each peak was described as a vector with length = β and angle = azimuth position. The vector sum then consisted of the summation of these individual vectors.

We considered a voxel to be spatially selective if the BOLD response was modulated by sound azimuth position, as reflected in the RAF, such that at least one and maximally three adjacent azimuth positions elicited a peak response. A voxel that exhibited a peak response to more than three adjacent azimuth positions was considered omni-responsive and therefore nonselective. Voxels that exhibited a peak response to two or more separate azimuth locations were also considered nonselective.

The tuning width of spatially selective voxels was quantified as the equivalent rectangular receptive field (ERRF) width (Lee and Middlebrooks, 2011). The ERRF is equal to the ratio between the amplitude of the peak response (that is, the β value at the preferred location), and the integral of the RAF. Although this measure does not provide an absolute measure of spatial selectivity, it enables the comparison of spatial selectivity across conditions, areas, and participants. Given that the rostral belt areas were not extensively activated, we focused this analysis on the caudal belt areas CM and CL.

Response sharpening versus response gain. We tested whether sharpening of spatial tuning resulted from *BOLD response gain* (that is, an increase of the BOLD response at the voxel's preferred location), *BOLD response sharpening* (a decrease in the BOLD response at the voxel's least preferred location), or a combination of the two. For this comparison, we defined the voxel's best location as the location with the highest β value in the task-independent RAF, that is, the average RAF across the two active task conditions. Similarly, we considered the least-preferred location the

azimuth location with the lowest β value in the average RAF (Lee and Middlebrooks, 2011, 2013).

Decoding sound azimuth position from fMRI activity patterns. To decode sound location, we applied a population-pattern decoder to the measured fMRI activity patterns in two regions of interest: the core region and PT. We selected these regions based on prior research in animals indicating A1 as a potential locus for dynamic spatial sensitivity (Lee and Middlebrooks, 2011) and prior neuroimaging research in humans illustrating the role of PT in spatial auditory processing in the human brain (Warren and Griffiths, 2003; Brunetti et al., 2005; Deouell et al., 2007; van der Zwaag et al., 2011; Derey et al., 2016; McLaughlin et al., 2016).

In general, the decoder—a modified version of a pattern decoder introduced to decode sensory information from neural spike rate patterns (Jazayeri and Movshon, 2006; Miller and Recanzone, 2009; Day and Delgutte, 2013)—computes the log-likelihood that a sound at a given azimuth location elicited the observed fMRI activity pattern. In particular, we computed for each voxel the log-likelihood that a stimulus at a particular azimuth location induced the observed BOLD response. The population log-likelihood then consists of the sum of the log-likelihoods across all voxels (Fig. 2).

Specifically, for each cortical area, we selected those voxels that responded to sounds (GLM sound > baseline, $p < 0.005$ uncorrected) and exhibited a spatially selective response (see previous section). Next, we estimated for each subject a GLM per functional data run with one predictor per azimuth position per task. This resulted in four β estimates per azimuth position, equivalent to the four functional runs. Beta estimates were normalized between 0 and 1 across the seven azimuth positions within each run. For each stimulus azimuth position, we then computed the log-likelihood that the observed BOLD response (β_i in the voxel under consideration was elicited by the presentation of a sound at that location. Assuming that the observed BOLD response β_i of voxel i for a given azimuth position θ_0 is normally distributed with mean $\mu_{0,i}$ and SD $\sigma_{0,i}$, the log-likelihood of the observation can be computed as follows:

$$\log L_i(\theta_0) = \frac{(\beta_i - \mu_{0,i})^2}{2\sigma_{0,i}^2} - \frac{1}{2} \log(\sigma_{0,i}^2) - \frac{1}{2} \log(2\pi).$$

The estimation was performed using cross-validation: we considered three runs to estimate the mean $\mu_{0,i}$ and SD $\sigma_{0,i}$ of a given voxel and azimuth position, and we used the left-out run to calculate the log-likelihood. The procedure was repeated for all the possible train-test combinations. Due to the limited amount of available data (1 trial per run), the estimation of the parameters was done using the β values of the selected voxel, as well as the six neighboring voxels, that is, those voxels sharing a side with the relevant voxel. Consequently, the number of data points to estimate $\mu_{0,i}$ and $\sigma_{0,i}$ was 21 (3 functional runs multiplied with 7 voxels). The test data β_i is the β estimate for this voxel for this azimuth position in the run that was left out. Assuming conditional (i.e., within each azimuth position) independence between different voxels, the population response was then computed as the sum of log likelihood of all voxels in the cortical area (N):

$$\log L(\theta_0) = \sum_{i=1}^N \log L_i(\theta_0).$$

In the test run, we predicted the sound azimuth location of a new, unseen sound, by selecting the location with the highest log-likelihood. This is equivalent to using a probabilistic classifier based on the posterior probability of azimuth location given the observed data, when class prior is uniform across all sound locations. Reported absolute errors are the average across the four train-test estimations. Statistical comparisons of absolute error across cortical areas and tasks were made with Wilcoxon signed rank tests (one-tailed) and corrected for multiple comparisons with the false discovery rate [$q(\text{FDR}) < 0.05$] unless mentioned differently. We determined the chance level of absolute error per azimuth position with permutation testing. Specifically, within each run we permuted β estimates randomly across the seven azimuth locations and for all voxels independently. We then applied the maximum likelihood de-

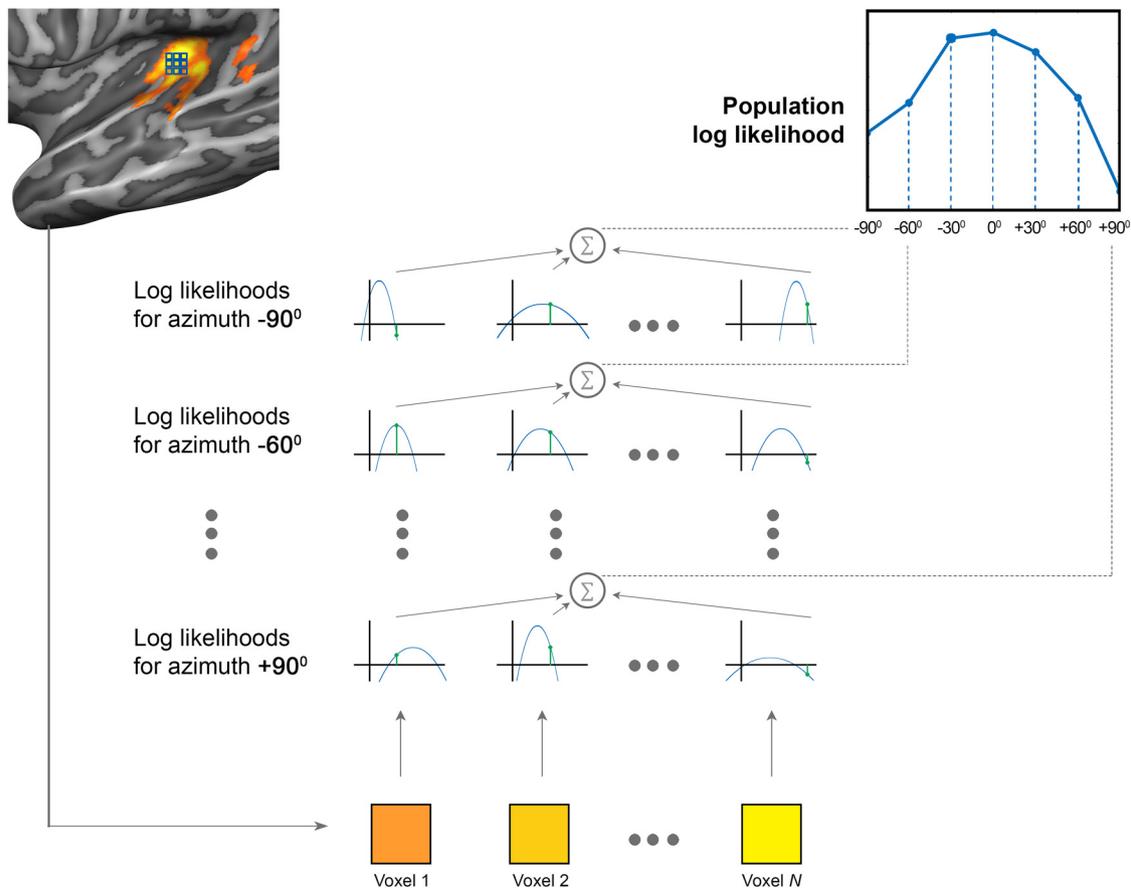


Figure 2. Estimating sound azimuth location with a maximum-likelihood population pattern decoder. Bottom row shows fMRI response (β value) to a sound presentation for individual voxels, with warmer colors (orange) indicating a weaker response (lower β value) and brighter colors (yellow) indicating a stronger response (higher β value). Small graphs show the log-likelihood function for each voxel for a given sound azimuth location (rows), with the fMRI response strength (β value) on the x -axis, and the log-likelihood on the y -axis. Large graph on the right shows the resulting population log-likelihood function, which is the sum of the log-likelihood functions of the individual voxels at each location.

coder to the permuted data. This procedure was repeated 1500 times per subject. Chance level of absolute error was computed as the mean absolute error across permutations.

Finally, we applied the population pattern decoder to data from both hemispheres simultaneously. In particular, we randomly sampled half of the voxels in the left hemisphere and half of the voxels in the right hemisphere. This procedure ensured that the number of data points used for the maximum likelihood estimation was equal when the decoder operated on data from two hemispheres versus data from a single hemisphere. We repeated the random sampling procedure 200 times per subject and computed absolute error as the average across samples. To determine the chance level for the population decoder operating on data from the two hemispheres, we applied a similar permutation procedure as described above. However, due to the interaction of the computationally intensive procedure of repeating the random sampling of half of the voxels in each hemisphere as well as the permutations, we limited the calculation to 30 random samples with 10 permutations each. Chance level of absolute error was computed as the average absolute error across samples and permutations.

Parcellation of the auditory cortex. To divide the auditory cortex into core, belt regions, and PT, we combined maps of frequency preference (tonotopy) and frequency selectivity. To construct these maps, we first estimated a voxel's frequency tuning profile by estimating GLM with one predictor per center frequency for each auditory active voxel (assessed with a GLM contrasting auditory stimulation > baseline, liberal threshold of $p < 0.05$ uncorrected). We inferred a voxel's preferred frequency (PF) from the frequency tuning profile. That is, a voxel's PF was defined as the frequency with the highest β value in the tuning profile (after z -normalizing across voxels). We then created tonotopic maps on the

cortical surface by color-coding the PF of all auditory responsive voxels in a blue (high-frequency) to red (low-frequency) color scale.

Next we estimated the frequency selectivity of a voxel by computing a frequency selectivity index (FSI). This index expresses the ratio between the peak β value (that is, the β corresponding to the PF) and the area under the frequency-tuning curve (the integral):

$$FSI = \frac{\int \text{frequency tuning curve}}{\beta_{PF}}$$

Then, similar to Moerel et al. (2012), we defined the tuning width (TW) of a voxel as follows:

$$TW = PF/f_2 - f_1,$$

where $(f_2 - f_1)$ is the frequency range in hertz corresponding to the FSI. As such, TW is high for voxels with a narrow tuning profile and small for voxels with a broad tuning profile. We color-coded the TW on the cortical sheet in a yellow (broad tuning) to purple (narrow tuning) color scale.

Finally, we used these maps to parcellate the auditory cortex following criteria based on the tonotopic organization described by Moerel et al. (2012) (Figure 3). Specifically, Moerel et al. (2012) identify the core region as a region overlapping with HG that is narrowly tuned to frequency and encompasses two mirror-symmetric tonotopic gradients (Formisano et al., 2003; Moerel et al., 2014; Leaver and Rauschecker, 2016). This core region is flanked by broadly tuned regions both anteriorly (overlapping with the first transverse sulcus and planum polare in

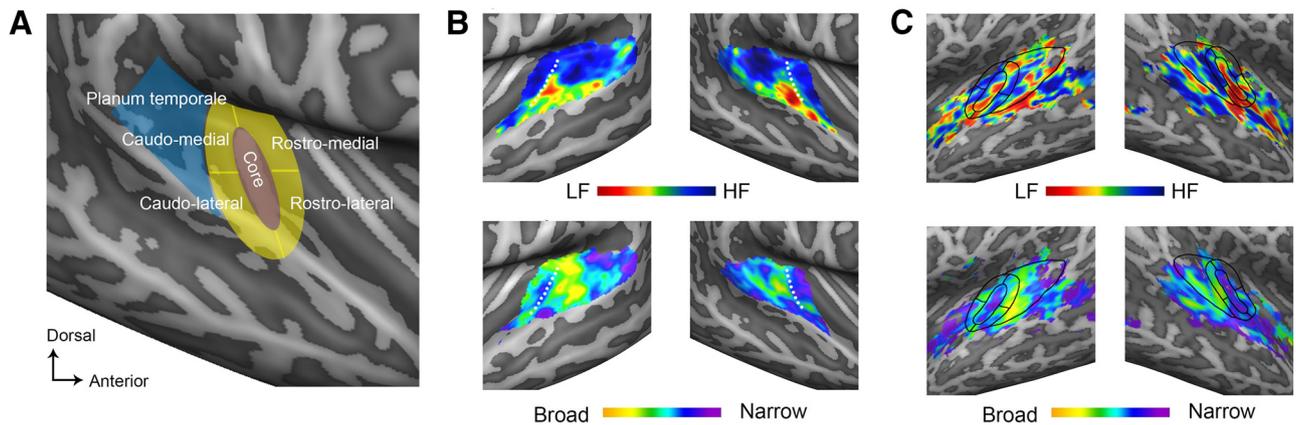


Figure 3. Parcellation of the human auditory cortex. **A**, The figure shows an enlarged view of the superior temporal plane in the right hemisphere, with a schematic overview of the parcellation used in the present study overlaid on top. **B**, Left and right superior temporal plane of a representative participant with the group map of frequency preference overlaid (top row; warm colors indicate a maximum response to low frequencies, cold colors to high frequencies), and frequency selectivity (bottom row; orange to green colors indicate broad tuning, blue to purple colors indicate progressively sharper tuning). **C**, Similar to **A** but displaying maps for a single representative participant.

Table 1. Number of auditory responsive voxels per cortical area

Cortical area	Average number of voxels (SD)	
	Left hemisphere	Right hemisphere
Core	160.0 (24.2)	138.9 (19.1)
CM	86.6 (14.9)	83.6 (12.9)
CL	73.9 (15.5)	68.3 (12.2)
RM	56.0 (15.0)	51.6 (15.7)
RL	57.8 (12.3)	44.6 (16.7)
PT	271.3 (49.4)	238.4 (48.0)

general), and posteriorly [coinciding with Heschl's sulcus (HS)]. Here we defined these broadly tuned bands as the rostral and caudal belt respectively (Fig. 3). We then evenly divided both the caudal and the rostral belt into medial and lateral parts, resulting in four belt areas: CM, CL, rostromedial (RM), and rostromedial (RL; Rauschecker et al., 1995; Kaas and Hackett, 2000). Finally, in line with Moerel et al. (2012) and the anatomical definition of PT provided by Kim et al. (2000), we defined the remaining posterior part of the superior temporal plane as PT. This region was bordered anteriorly by the caudal belt (overlapping largely with HS), medially by the insular cortex, and laterally by the superior temporal gyrus.

Note that two participants did not show extensive activation in the auditory cortex for the contrast auditory stimulation > baseline as a result of excessive movement during the tonotopy measurements (possibly due to participant fatigue). We parcellated the auditory cortex of these two participants based on anatomical criteria, resulting in areas that were similar in size and location to those of the other participants. Specifically, the core region was identified as approximately two-thirds of HG (starting from the medial border; Moerel et al., 2012, 2014). The caudal belt was defined by HS, bordered posteriorly by PT (Kim et al., 2000). The rostral belt was defined as anteriorly to HG, mainly overlapping with the first transverse sulcus, as the mirror image of the caudal belt. The rostral and caudal belt regions were evenly split into a lateral and medial part.

Maps of cortical auditory areas constructed in surface space were projected back into volume space. In subsequent analyses, we included for each area the voxels that responded to sounds (as established with a GLM, contrast auditory stimulation > baseline, liberal threshold of $p < 0.005$ uncorrected; Table 1).

Results

Behavioral task performance

Behavioral accuracy in the MRI scanner was high for both active tasks. The average hit rate for the sound localization task was 94.3% (SD: 15.2%), and for the sound identification task 90.9%

(SD: 12.6%). There was no difference in mean accuracy between tasks (paired samples t test, $t_{(10)} = 0.607$, $p = 0.557$).

Univariate analysis of the processing of spatialized sounds in human auditory cortex

RFX GLM contrasting *auditory stimulation > baseline* showed increases in BOLD signal in primary and secondary auditory cortices in response to the probe trials (corrected for multiple comparisons with the FDR, $q < 0.05$; (Benjamini and Hochberg, 1995). Activated areas included HG, HS, PT, and to a lesser extent, the first transverse sulcus and other parts of the planum polare. To investigate differences in the overall level of activation elicited by the three task conditions, we computed several balanced contrast maps. However, none of these contrasts revealed different activation levels between task conditions, either at a stringent threshold (FDR, $q < 0.05$) or at a more liberal threshold ($p < 0.005$ uncorrected), indicating that the overall BOLD signal amplitude in the auditory cortex was similar across tasks.

Parcellating the human auditory cortex

In agreement with prior tonotopic mapping studies (Wessinger et al., 2001; Formisano et al., 2003; Talavage et al., 2004; Striemi-Amit et al., 2011; Da Costa et al., 2011; Moerel et al., 2012; Leaver and Rauschecker, 2016), cortical maps of frequency preference revealed a region tuned to low frequencies overlapping partly with HG which was bordered anterolaterally and posteromedially by regions responding maximally to high frequencies (Fig. 3). Further, similar to Moerel et al. (2012) we observed a narrowly tuned region overlapping with (or in close vicinity to) HG in the frequency selectivity maps of most participants. This region was flanked by areas with broad frequency selectivity profiles (Fig. 3). We combined these maps of frequency preference and selectivity and derived an operational definition of the core region, the belt regions (Rauschecker et al., 1995) for original definitions in macaque auditory cortex), and PT (Fig. 3; see Materials and Methods).

Spatial selectivity in human auditory cortex is higher in posterior, higher-order regions than in primary regions

To start, we examined general differences in the presence of spatially selective voxels between cortical areas, i.e., interarea differences regardless of behavioral demands. The results show that the average proportion of auditory responsive voxels that was spa-

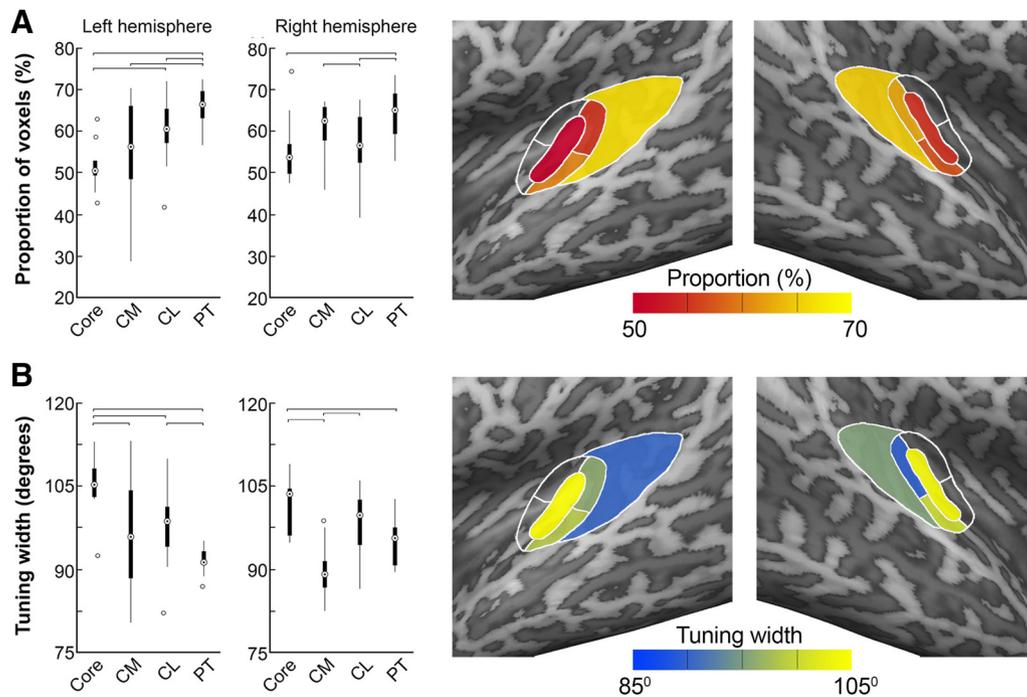


Figure 4. Spatial selectivity across auditory cortical areas in humans. **A**, Boxplots show, for each cortical area, the distribution of the proportion of spatially selective voxels across participants (averaged across task conditions). **B**, Boxplots reflect the distribution of relative spatial tuning width (ERRF width, averaged across task conditions) across participants. The central circle of a box indicates the median of the distribution, the edges the 25th and 75th percentiles, and lines the full range of values. Circles represent outliers. Horizontal lines indicate a significant difference between areas at $p < 0.05$, FDR corrected for multiple comparisons at $q < 0.05$.

Table 2. Differences in the proportion of spatially selective voxels and tuning width between cortical auditory areas

Condition	Left hemisphere								Right hemisphere							
	Friedman test			Median				Friedman test			Median					
	χ^2	df	p	Core	CM	CL	PT	χ^2	df	p	Core	CM	CL	PT		
Proportion spatially selective voxels	Passive	18.6	3	0.0003*	51.2	63.6	65.6	68.8	7.1	3	0.067	58.3	62.2	55.9	66.1	
	Identification	18.6	3	0.0003*	50.0	58.8	61.2	66.4	10.1	3	0.018*	51.3	60.7	60.8	63.3	
	Localization	9.4	3	0.024*	52.3	55.3	54.1	63.9	7.9	3	0.048	56.2	59.3	58.2	64.7	
	Average	19.5	3	0.0002*	50.6	56.3	60.6	66.5	12.4	3	0.0062*	53.8	62.6	56.6	65.2	
Tuning width	Passive	15.9	3	0.0012*	108.3	97.7	96.8	90.7	17.9	3	0.0005*	102.0	90.5	100.6	94.0	
	Identification	15.6	3	0.0014*	108.8	95.6	95.5	91.8	7.4	3	0.058	103.1	91.2	97.4	97.3	
	Localization	10.1	3	0.018*	104.5	98.8	99.2	92.8	15.4	3	0.0015*	102.0	85.0	100.4	95.1	
	Average	16.7	3	0.0008*	105.3	95.9	98.7	91.2	18.4	3	0.0004*	101.6	89.1	98.6	95.7	

Statistical results of Friedman tests compare the proportion of spatially selective voxels (top) and relative tuning width (bottom) between cortical auditory areas, as well as the median value for each cortical area.

* p value significant after FDR correction for multiple comparisons ($q < 0.05$). Proportions are displayed in percentages, tuning width in degrees.

tially selective (averaged across task conditions) varied across cortical regions in the left hemisphere (Fig. 4A), as well as in the right hemisphere (Fig. 4A). In particular, in the left hemisphere, PT contained relatively more spatially selective voxels than the core, CM, and CL. The proportion of selective voxels was also higher in left CL than in the left core (see Table 2 and Table 3). In the right hemisphere, PT contained a higher proportion of selective voxels than the core and CL as well, and the proportion of spatially selective voxels was higher in CM than in CL (Table 2 and Table 3).

We also assessed spatial selectivity by investigating the relative tuning width of spatially selective voxels within an area. For this measure of spatial selectivity, we observed an anterior to posterior (rostral-to-caudal) increase of spatial selectivity as well, both in the left hemisphere and right hemisphere (Table 2; Fig. 4B). Specifically, in the left hemisphere, spatial tuning width was broader in the core than in PT, CM, and CL. Finally, spatial

tuning width was narrower in PT than in CL (Table 3; Fig. 4B, left). In the right hemisphere, there was also a difference in spatial tuning width between PT and the core. However, in this hemisphere spatial tuning was sharpest in CM: there was a significant difference between CM and the core, and between CM and CL (Table 3; Fig. 4B).

Next, we investigated cortical inter-area differences in spatial selectivity per behavioral task condition. This revealed that there were differences in the proportion of spatially selective voxels across areas in all behavioral conditions (Table 2). Specifically, *post hoc* comparisons revealed that the rostral-to-caudal increase in the proportion of spatially selective voxels was present in all behavioral conditions in the left hemisphere. That is in each condition, there were more spatially selective voxels in PT than in the core and in CM. Further, in the *passive* listening and sound identification conditions, but not in the sound localization condition, there were more spatially selective voxels in PT than in CL. In the

Table 3. Statistical results (p values) of *post hoc* pairwise comparisons of the proportion of spatially selective voxels (top) and tuning width (bottom) between cortical auditory regions (two-sided Wilcoxon signed rank tests)

	Left hemisphere			Right hemisphere		
	Core	CM	CL	Core	CM	CL
Proportion of spatially selective voxels						
Passive						
CM	0.320			n.a.		
CL	0.001*	0.206		n.a.	n.a.	
PT	0.001*	0.024*	0.320	n.a.	n.a.	n.a.
Identification						
CM	0.206			0.007*		
CL	0.005*	0.175		0.577	0.042	
PT	0.001*	0.002*	0.054	0.010*	0.966	0.054
Localization						
CM	0.638			n.a.		
CL	0.577	0.700		n.a.	n.a.	
PT	0.002*	0.010*	0.032	n.a.	n.a.	n.a.
Average						
CM	0.320			0.042		
CL	0.002*	0.175		0.700	0.014*	
PT	0.001*	0.005*	0.014*	0.007*	0.465	0.010*
Tuning width						
Passive						
CM	0.042			0.001*		
CL	0.005*	0.765		0.320	0.003*	
PT	0.001*	0.083	0.024*	0.019*	0.148	0.067
Identification						
CM	0.007*			n.a.		
CL	0.010*	0.638		n.a.	n.a.	
PT	0.001*	0.278	0.032*	n.a.	n.a.	n.a.
Localization						
CM	0.067			0.005*		
CL	0.083	0.638		0.465	0.005*	
PT	0.002*	0.465	0.067	0.042	0.019*	0.413
Average						
CM	0.019*			0.002*		
CL	0.005*	0.700		0.148	0.001*	
PT	0.001*	0.123	0.032*	0.019*	0.054	0.365

*Significant after FDR correction at $q < 0.05$. Note that we only performed *post hoc* pairwise comparisons for those behavioral conditions that showed significant inter-area differences as assessed with the Friedman tests displayed in Table 2.

right hemisphere, we observed significant inter-area differences in the proportion of spatially selective voxels in the sound identification condition only. Similar trends were present for the passive listening and sound localization conditions, but these just failed to reach statistical significance (Table 2). *Post hoc* pairwise comparisons for the sound identification condition (Table 3) indicate that there are significantly more spatially selective voxels in PT as well as in CM, compared with the core region (Fig. 5).

We also observed inter-area differences in relative tuning width per behavioral task condition in the left hemisphere. That is, there were significant inter-area differences in all behavioral conditions (Table 2), and in all conditions spatial tuning was sharper in PT than in the core region (see results of *post hoc* pairwise comparisons in Table 3). In addition, spatial tuning in PT was sharper than CL in the passive listening and sound identification condition. Spatial tuning was also sharper in CL than in the core during the passive listening and sound localization condition. In the right hemisphere, we observed inter-area differences in the passive listening and sound localization condition (a similar pattern was observed in the sound identification condition, but this just failed to reach statistical significance; Table

2). *Post hoc* pairwise comparisons show that during passive listening, spatial tuning was sharper in PT than in the core region. In addition, spatial tuning was sharper in CM than in either the core region and CL. Also during active sound localization, spatial tuning in CM was sharper than in the core and CL, and even PT (Table 3; Fig. 5).

Task-modulations of spatial selectivity within cortical auditory regions

We then examined, for each cortical area, the effect of task performance on spatial selectivity. There were no differences in the proportion of auditory responsive voxels that were spatially selective across task conditions: none of the cortical regions showed an increase or decrease in the proportion of spatially selective voxels based on task performance (one-tailed Wilcoxon signed rank tests, all $p > 0.05$; Fig. 5A). However, spatial tuning was sharper in the localization condition compared with the sound identification condition in the left core region [median identification condition = 108.8°, median localization condition = 104.5°, one-tailed Wilcoxon signed rank test, $p = 0.001$, $q(\text{FDR}) < 0.05$], and in right CM [median identification condition = 91.2°, median localization condition = 85.0°, $p = 0.003$, $q(\text{FDR}) < 0.05$; Fig. 5B]. Figure 5C shows the population RAFs, which also reflect the sharpening of spatial selectivity in the left core and right CM during active sound localization.

Next, we investigated the mechanism underlying the observed sharpening of spatial tuning in the left core and right CM during the sound localization condition. Specifically, we evaluated whether the change in spatial tuning between the two active task conditions resulted from response gain (that is, an increase of the BOLD response amplitude at the voxel's preferred location), response sharpening (a decrease of the BOLD response at the voxel's nonpreferred location), or a combination of these processes. For this comparison, we defined the voxel's preferred location as the sound azimuth location with the maximum β value in the task-independent RAF (i.e., the average RAF across the two active task conditions). Similarly, we defined the nonpreferred location as the sound azimuth location with the minimum β value in the average RAF (Lee and Middlebrooks, 2011, 2013).

In both cortical areas, the BOLD response at the preferred location was similar for the two active task conditions, while the BOLD response at nonpreferred locations was lower in the sound localization than in the sound identification condition. Specifically, Figure 6 shows that the β values for the preferred location were similar for both active task conditions [reflected by the clustering of β values around the diagonal; median β left core in sound identification (sound localization) condition = 0.39 (0.40); median β right CM in sound identification (sound localization) condition = 0.27 (0.30); Wilcoxon signed rank tests for differences between task conditions, $p > 0.05$]. In contrast, the BOLD response at nonpreferred locations was lower in the sound localization than in the sound identification condition [most β values are below the diagonal; median β left core in sound identification (sound localization) condition = 0.13 (−0.04); median β right CM in sound identification (sound localization) condition = 0.04 (−0.11); Wilcoxon signed rank tests; left core: $p = 0.002$; right CM: $p = 0.014$; $q(\text{FDR}) < 0.05$]. Thus, sharpening of spatial tuning during active sound localization was mainly the result of a decrease of BOLD signal amplitude at nonpreferred locations, that is, response sharpening.

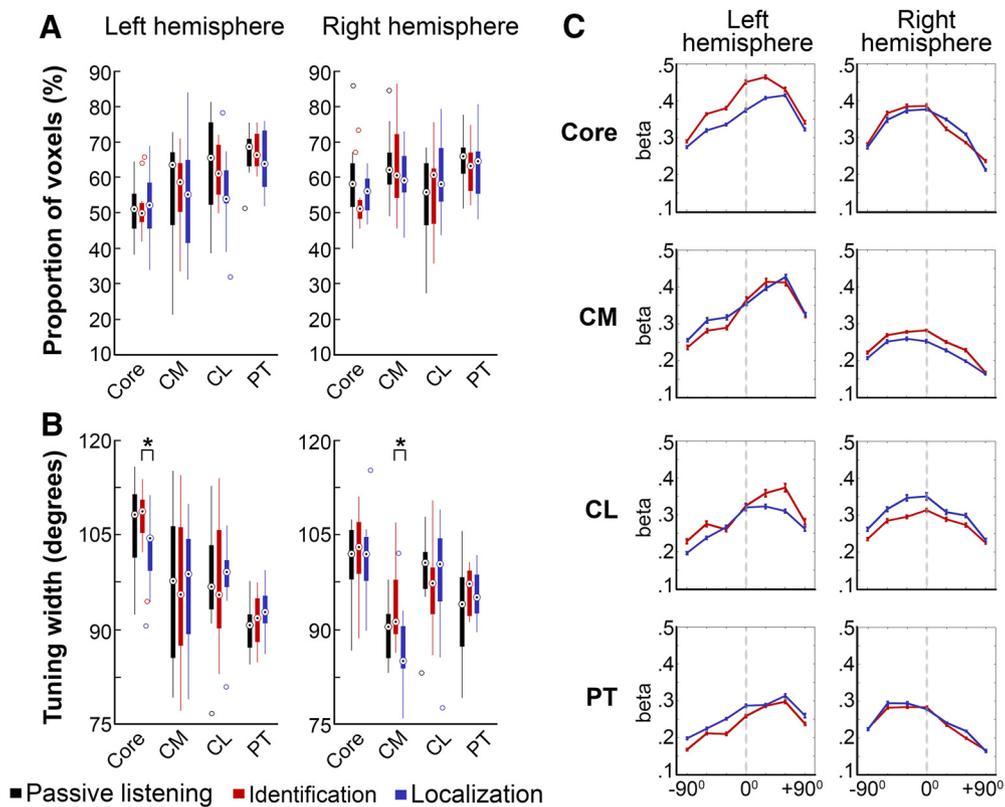


Figure 5. Task modulations of spatial selectivity in human auditory cortex. **A**, Boxplots show for each task condition the distribution of the proportion of voxels that exhibit a spatially selective response across participants. Black boxes indicate the passive listening condition, red boxes the sound identification condition, and blue boxes the sound localization condition. **B**, Boxplots reflect the distribution of relative spatial tuning width (ERRF width) across participants for each area and task condition. Colors similar to **A**. The central circle of a box indicates the median of the distribution, the edges the 25th and 75th percentiles, and lines the full range of values. Circles represent outliers. Horizontal lines with asterisks indicate a significant difference between areas at $p < 0.05$, FDR corrected for multiple comparisons at $q < 0.05$. **C**, Population RAFs are plotted for the spatially selective voxels within an area for the two active task conditions. RAFs are averaged across participants; blue lines indicate the sound identification condition, red lines the sound localization condition.

Decoding sound azimuth location from fMRI population activity patterns

Next we evaluated whether the encoding of sound azimuth in fMRI activity patterns in the core and in PT varies with behavioral task requirements. Specifically, we applied a population-pattern decoder based on maximum likelihood estimation to the measured fMRI responses to the probe sounds in the sound identification and sound localization condition (see Materials and Methods). Figure 7 shows for each cortical area and task condition the absolute error of the population pattern decoder as a function of sound azimuth location. There was no difference in decoding performance between ipsilateral and contralateral locations: a comparison of the average absolute error between hemifields (i.e., the average absolute error across -30° , -60° , and -90° , versus the average across $+30^\circ$, $+60^\circ$, and $+90^\circ$) did not yield significant results either for the core or for PT, in any behavioral task condition [two-sided Wilcoxon signed rank test per area and task condition, FDR corrected for multiple comparisons, all $q(\text{FDR}) \geq 0.05$].

For the purpose of statistical comparisons between cortical areas and behavioral task conditions, we computed the average

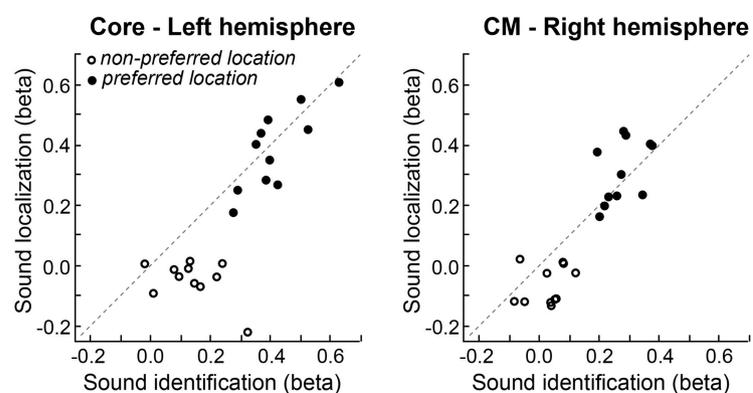


Figure 6. Sharper spatial selectivity during active sound localization is a result of response sharpening. Scatterplots show for each participant the average β value across voxels that exhibited sharper spatial selectivity (i.e., a decrease in ERRF width of 15% or more) during the sound localization condition (y -axis) than sound identification condition (x -axis) at the preferred (filled circles) and non-preferred location (open circles) for the left core region (left) and right CM (right). Circles below the diagonal reference line reflect a decrease in β value in the sound localization condition.

absolute error across azimuth positions for each area and task condition. Figure 7B shows that the population pattern decoder performed better than chance level in the left and right core in the sound localization condition. That is, in these areas and task conditions the absolute error was significantly lower than chance [one-sided Wilcoxon signed rank test, FDR corrected for multiple corrections; median absolute error sound localization condition left core = 61.1° , right core = 62.1° , chance error = 68.6° ,

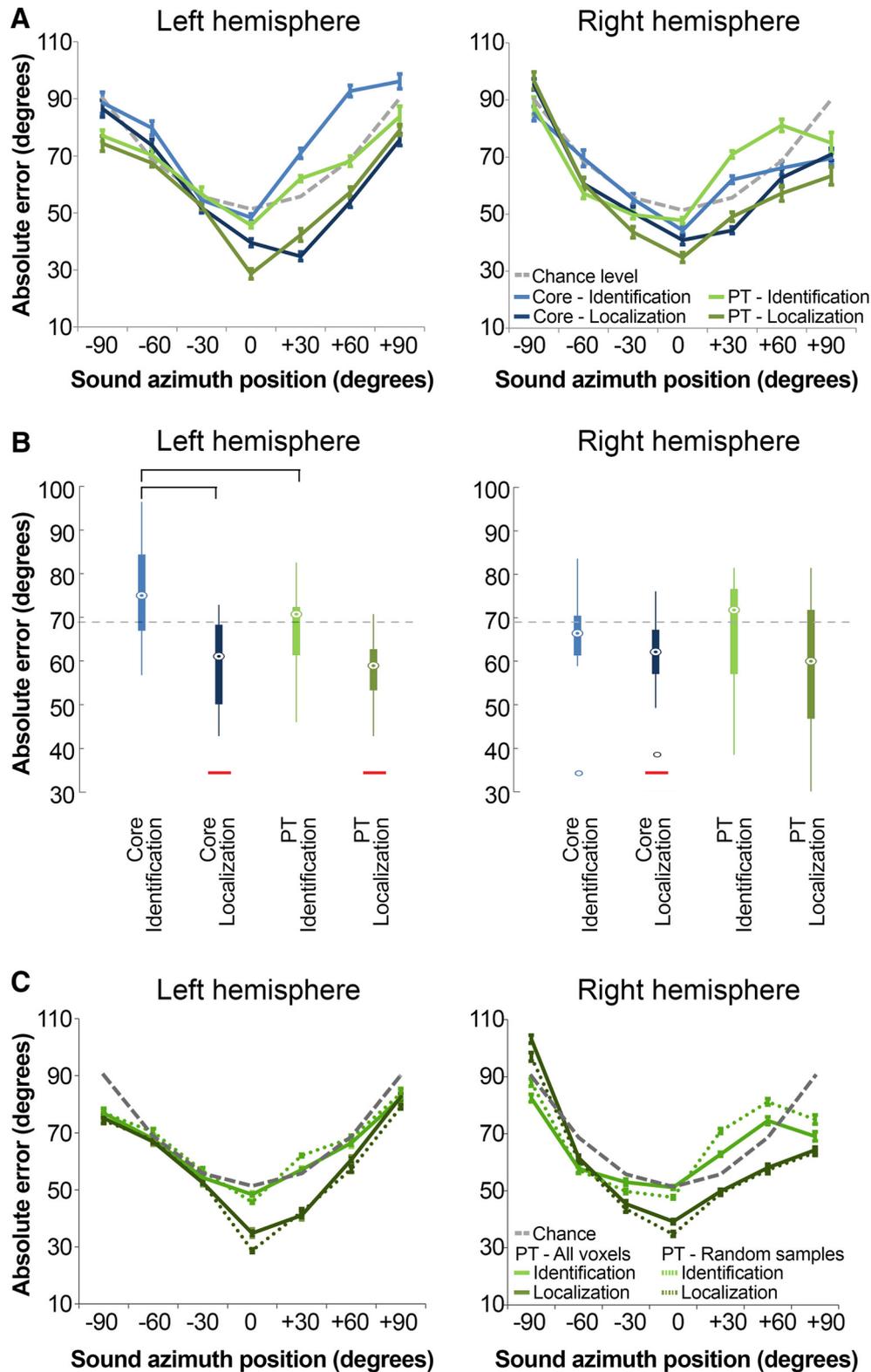


Figure 7. Decoding sound azimuth from population pattern activity in the core region and PT during a sound identification (“what”) and a sound localization (“where”) task. **A**, Lines reflect the average absolute error of the sound azimuth estimate resulting from the population pattern decoder (*y*-axis) as a function of actual sound azimuth (*x*-axis) for a particular cortical area and task condition. Light blue lines, Core region during sound identification task; dark blue lines, core region during sound localization task; light green lines, PT during sound identification task; dark green lines, PT during sound localization task. Error bars reflect the SEM. **B**, Boxplots of the absolute error of the sound azimuth estimates averaged across the seven sound azimuth positions. Colors similar to **A**. Horizontal black lines at the top of the figure indicate a significant difference in prediction error between cortical areas or task conditions [$p < 0.05$, $q(\text{FDR}) < 0.05$]. Horizontal red lines at the bottom of the figure indicate that the absolute error is below chance level [$p < 0.05$, $q(\text{FDR}) < 0.05$]. **C**, Lines reflect the performance of the population pattern decoder for PT controlled for the number of voxels. Similar to **A**, lines reflect the average absolute error. Solid lines are identical to those for area PT depicted in **A**. Dashed lines show the average absolute error across random samples (200 iterations) of voxels in PT. Specifically, for each participant we sampled a number of voxels from PT equal to the number of voxels included in the analysis for the core. Error bars reflect the SEM.

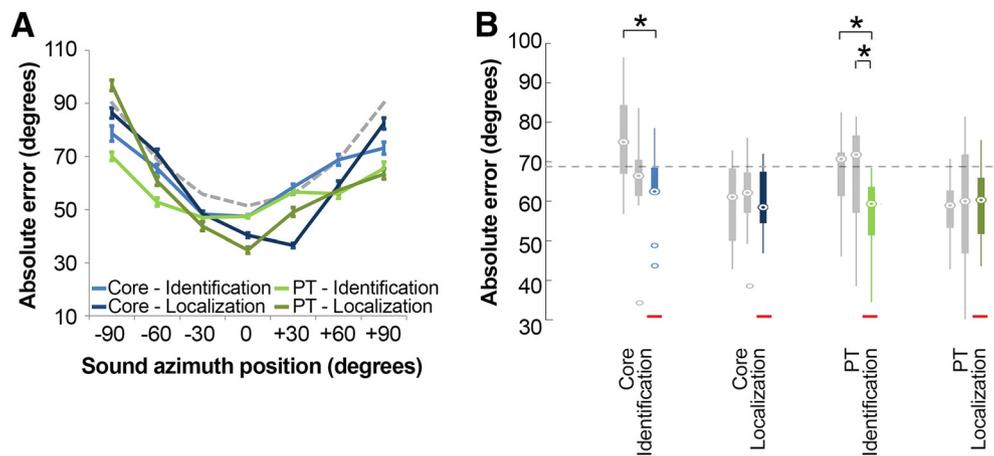


Figure 8. Decoding sound azimuth from population pattern activity across two hemispheres. **A**, Lines reflect the average absolute error of the sound azimuth estimate resulting from the population pattern decoder (y -axis) as a function of actual sound azimuth (x -axis) for a particular cortical area and task condition. Light blue lines: core region during sound identification task. Dark blue lines, Core region during sound localization task; light green lines, PT during sound identification task; dark green lines, PT during sound localization task. Error bars reflect the SEM. **B**, Boxplots of the absolute error of the sound azimuth estimates averaged across the seven sound azimuth positions. Colors similar to **A**. Gray boxes are identical to the boxes shown in Figure 7 and show the absolute error for the left hemisphere only (left-most gray box) and for the right hemisphere only (right-most gray box) for comparison. Horizontal black lines with asterisks at the top of the figure indicate a significant difference in prediction error between cortical areas or task conditions [$p < 0.05$, $q(\text{FDR}) < 0.05$]. Horizontal red lines at the bottom of the figure indicate that the absolute error is below chance level [$p < 0.05$, $q(\text{FDR}) < 0.05$].

$p = 0.009$ for both regions, $q(\text{FDR}) < 0.05$]. Chance level was computed with a permutation testing procedure in which we randomly scrambled the RAFs of each participant (1500 iterations). In left PT, the pattern decoder also performed better than chance in the localization condition [median absolute error left PT = 58.9° , $p = 9.8\text{E}-4$, $q(\text{FDR}) < 0.05$]. Similarly, in right PT the pattern decoder performed marginally better than chance in the localization condition [median absolute error right PT = 60.0° , $p = 0.051$, $q(\text{FDR}) = 0.076$]. However, in the sound identification condition the absolute error was larger than chance level in all cortical areas (median absolute error for the sound identification condition per area: left core = 75.0° , right core = 66.4° , left PT = 70.7° , right PT = 71.8° , $p > 0.05$; Fig. 7B), indicating that the pattern decoder did not perform well for this behavioral condition.

We then tested for differences in sound location decoding performance for the probe sounds between task conditions, within each cortical area. This showed that the pattern decoder performed significantly better in the sound localization than in the sound identification condition in the left core region; that is, the absolute error was significantly lower [one-sided Wilcoxon signed rank test, FDR-corrected for multiple comparisons; $p = 0.003$, $q(\text{FDR}) < 0.05$; Fig. 7B]. In left PT we observed a similar task effect, but this did not reach statistical significance [$p = 0.04$, $q(\text{FDR}) = 0.1$]. Figure 7A shows that the absolute error decreased especially at the midline and in contralateral space (0° to $+90^\circ$) for both the core and PT in the left hemisphere. There was no significant effect of task in the right core or in right PT ($p > 0.05$; Fig. 7). For the right core, this may be a consequence of the relatively high performance of the pattern decoder in the sound identification condition. In particular, sound azimuth location estimates were significantly more accurate in the right, than in the left core in the sound identification condition [two-sided Wilcoxon signed rank test; $p = 0.022$, $q(\text{FDR}) < 0.05$], but not in the sound localization condition ($p > 0.05$; Fig. 7B).

We also tested for each task condition whether there was a difference in decoding accuracy between cortical areas. In the left hemisphere, the absolute error was lower in PT than in the core

region in the sound identification condition [$p = 0.0098$, $q(\text{FDR}) < 0.05$] but not in the sound localization condition ($p > 0.05$). Figure 7A shows that the inter-area difference in the sound identification condition was mainly a result of lower absolute errors in PT in peripheral space. In the right hemisphere, there was no significant difference between the core and PT either in the sound identification condition ($p > 0.05$) or in the sound localization condition ($p > 0.05$). Note that the lower absolute error observed in left PT was not a consequence of a larger number of voxels in this cortical region: the inter-area effect persisted even if the number of voxels in PT included in the analysis was matched to the number of voxels in the core region (see Materials and Methods; Fig. 7C).

Finally, we applied the maximum-likelihood decoder to the fMRI activity patterns of the left and right hemisphere together: we provided the data of both hemispheres combined as input for the pattern decoder. Note that to ensure that the number of voxels on which the pattern decoder operates does not influence the sound location estimates, we randomly sampled half of the voxels in the relevant region within a hemisphere and combined this with a random sample of half of the voxels in the other hemisphere. This procedure was repeated 200 times, and we computed the absolute error of the two-hemisphere decoder as the average absolute error across those 200 iterations.

Figure 8 shows that combining the activity patterns in the two hemispheres resulted in lower absolute errors when decoding azimuth position for probe sounds in the sound identification, but not for probe sounds in the sound localization condition. Specifically, absolute error scores were lower than chance level in the sound identification condition in both the core and in PT [median absolute error core = 62.4° , median absolute error PT = 59.3° , chance error = 68.8° , $p = 0.03$ and $p = 0.009$ respectively, $q(\text{FDR}) < 0.05$]. In addition, the absolute error in PT was lower for the combined data than for either the left PT only [$p = 0.016$, $q(\text{FDR}) < 0.05$], or the right PT only [$p = 0.003$, $q(\text{FDR}) < 0.05$]. Inspecting absolute error as a function of sound azimuth location (Fig. 8A), shows that combining the data of left and right PT resulted in lower absolute error scores mainly in the periphery

(-90° , -60° , $+60^\circ$, and $+90^\circ$). In contrast, for the core the combination of the data of the left and right hemisphere resulted in more accurate azimuth estimates compared with the left core ($p = 0.002$), but not compared with the right core ($p > 0.05$). Further, the absolute error as a function of sound azimuth position (Fig. 8A) shows that the absolute errors resulting from the combined data were similar to those resulting from the decoder operating on the right core only. This indicates that the azimuth estimates resulting from the pattern decoder operating on the core in two hemispheres are driven by the activity patterns in the right core, rather than showing an improvement larger than the available information in either core.

Discussion

The major findings of the present study are that spatial selectivity of the left primary auditory core cortex and right area CM are dynamic and dependent on behavioral requirements, that fMRI activity patterns in the left core carry more information on sound azimuth location when participants engage in a sound-localization task (compared with a task unrelated to sound localization), and that integrating fMRI activity patterns measured during a “what” task, but not during a “where” task, across bilateral PT results in more accurate sound azimuth location estimates than in either left or right PT separately. Together, these results highlight the adaptive potential of spatial tuning in the A1 based on behavioral demands. A possible mechanism for the observed task-modulation of spatial sensitivity in A1 is the feedback from functionally specialized regions (PT) to this cortical area. Specifically, such feedback connections from higher-order to primary regions may be modulated by behavioral requirements to enable dynamic spatial sensitivity in the latter. Finally, these findings provide new insights into models of sound location encoding in unilateral and bilateral human auditory cortex.

Dynamic spatial tuning in human auditory cortex

Posterior auditory cortical regions are thought to be part of a functionally specialized stream for sound location processing in animals (Tian et al., 2001; Stecker and Middlebrooks, 2003; Stecker et al., 2005; Harrington et al., 2008; Lomber and Malhotra, 2008) and humans (Alain et al., 2001; Arnott et al., 2004; Brunetti et al., 2005; Ahveninen et al., 2006; Deouell et al., 2007; Derey et al., 2016). Although we replicate these inter-area differences in spatial selectivity between primary core and higher-order areas, and specifically the advantage of caudal belt regions, that have been reported previously for passive listening or non-spatial task conditions, we also show that these differences are reduced in the left core and right CM when humans engage in an active sound localization task. Thus, our findings indicate that, depending on the behavioral requirements, primary auditory areas may contribute to sound location processing as well.

Such task-dependent modulations of spatial sensitivity have not previously been observed in humans. Zimmer and Macaluso (2005) reported a relationship between the level of activity in posterior auditory regions and successful sound localization, but did not investigate cortical spatial selectivity. Further, a recent neuroimaging study in humans did not report a modulation of either ILD or ITD selectivity based on task performance (Higgins et al., 2017). Yet, in the latter study, the authors considered binaural cue response functions averaged across all auditory responsive voxels within the auditory cortex, which may have diluted the results. That is, our analyses show that task modulations of spatial selectivity are localized specifically in the left core and right CM.

Our findings in human auditory cortex are compatible with

animal studies showing that the performance of both spatial and non-spatial tasks affects neuronal receptive fields in A1 (Fritz et al., 2003; Otazu et al., 2009; Lee and Middlebrooks, 2011). One hypothesis is that higher-order, functionally specialized cortical areas, such as PT, modulate spatial tuning in A1 via back-projections. In particular, our data are compatible with theoretical frameworks of sensory processing such as the reverse hierarchy (Ahissar et al., 2009) and recurrent processing models (Lamme and Roelfsema, 2000; Bullier, 2001). Similar to visual cortex, the auditory cortex is characterized by dense reciprocal connections between primary and higher-order cortical areas (Kaas and Hackett, 2000; Lee and Winer, 2011). Lateral prefrontal cortex (PFC) may mediate such feedback processing: lateral PFC is known to project back to early regions of the lateral auditory belt (Romanski et al., 1999) and has been implicated in a two-stage model of categorization of sounds (Jiang et al., 2018).

Differences in sound location processing between the left and right auditory pathway

In humans, lesion and functional imaging studies suggest that the right (sub)cortical pathway may contain a representation of the entire acoustic azimuth, while in the left (sub)cortical pathway the representation of the contralateral acoustic azimuth is thought to be pre-dominant (Zatorre and Penhune, 2001; Krumbholz et al., 2005; Spierer et al., 2009; Briley et al., 2013; Higgins et al., 2017). Differential spatial processing between the left and right auditory pathway has also been observed in several animal species. For instance, Day and Delgutte (2013) observed in rabbit inferior colliculus a gradient of deteriorating sound location decoding accuracy from locations at the midline toward the periphery. In contrast, in monkeys, Miller and Recanzone (2009) observed in area A1 and CL most accurate sound location decoding results in contralateral space, with low decoding accuracies at the midline and especially in ipsilateral space: the magnitude of sound location estimation errors in the ipsilateral hemifield and around the midline was distinctly higher than the errors observed in the present study. Only in area R were decoding errors lower around the midline than in either ipsilateral or contralateral space. Here we did not observe a difference in location decoding accuracy between ipsilateral and contralateral space either for the left or right auditory cortex. Yet, our results did reflect sharper spatial tuning in the right than left core when the task was unrelated to sound location (the “what” task), which may be a reflection of the hypothesized right dominance for human spatial hearing. Future research with noninvasive lesion techniques in humans combined with advanced neuroimaging and computational modeling studies is required to elucidate these potential differences between the left and right human auditory pathway.

Integrating information on sound azimuth location across hemispheres

Our results show that the integration of sound location processing across the two hemispheres may be task dependent. Specifically, location estimates based on fMRI activity patterns in bilateral PT were more accurate than those based on either left or right PT independently for the task condition unrelated to sound localization (“what” task), although, this bilateral advantage was not present during active localization (“where” task). For the core region, we also observed a bilateral advantage for the “what” task compared with the left core separately, but not for the right core. This suggests that the bilateral advantage, is merely a reflection of the more accurate decoding obtained for the right core in itself.

Similar to PT, no bilateral decoding improvement was observed during active sound localization for the core region. Thus, fMRI activity patterns in left and right PT, and possibly in the left and right core, contain complementary information on sound azimuth location when participants are not engaged in active sound localization, resulting in better location estimates when the information in the two hemispheres is combined. In contrast, information in the two hemispheres appears to be overlapping during active sound localization, such that combining the information across the hemispheres appears to be redundant during this behavioral condition.

This may be explained by a task-dependent strength of functional callosal connections. In particular, in macaques there are major interhemispheric connections both between the left and right core, and between left and right parabelt (Kaas and Hackett, 2000). If similar callosal connections between bilateral primary and higher-order auditory cortices exist in humans, it is conceivable that during active sound localization the functional connectivity between left and right PT increases compared with during nonlocalization tasks. As a consequence, spatial processing in left PT may modulate spatial processing in right PT during active localization (and vice versa), whereas spatial information in left and right PT is relatively independent, and thus complementary during non-spatial tasks. Alternatively, corticofugal projections (Winer and Schreiner, 2005) may strengthen during active sound localization, and thereby indirectly modulate sound location processing in the contralateral hemisphere.

The observed task-dependency of bilateral integration of information is also of interest for the ongoing debate about the computational mechanisms underlying sound location processing in mammals. In particular, models for neural population coding of sound azimuth location have received wide attention in recent years, including population coding within a single hemisphere (Miller and Recanzone, 2009; unilateral population coding: Day and Delgutte, 2013), *unilateral opponent population coding* based on two oppositely tuned channels within a single hemisphere (i.e., an ipsilaterally and a contralaterally tuned channel; Stecker et al., 2005), and *bilateral opponent population coding* based on combining the sound azimuth information of contralaterally tuned channels in each hemisphere (McAlpine et al., 2001; Derey et al., 2016; Ortiz-Rios et al., 2017). Our current results suggest that the degree to which information is combined across hemispheres may be dependent on behavioral requirements, indicating that unilateral and bilateral models of sound location encoding may not be mutually exclusive.

References

- Ahissar M, Nahum M, Nelken I, Hochstein S (2009) Reverse hierarchies and sensory learning. *Philos Trans R Soc Lond B Biol Sci* 364:285–299. [CrossRef Medline](#)
- Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci U S A* 103:14608–14613. [CrossRef Medline](#)
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) “What” and “where” in the human auditory system. *Proc Natl Acad Sci U S A* 98:12301–12306. [CrossRef Medline](#)
- Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. *Neuroimage* 22:401–408. [CrossRef Medline](#)
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 57:289–300.
- Briley PM, Kitterick PT, Summerfield AQ (2013) Evidence for opponent process analysis of sound source location in humans. *J Assoc Res Otolaryngol* 14:83–101. [CrossRef Medline](#)
- Brunetti M, Belardinelli P, Caulo M, Del Gratta C, Della Penna S, Ferretti A, Lucci G, Moretti A, Pizzella V, Tartaro A, Torquati K, Olivetti Belardinelli M, Romani GL (2005) Human brain activation during passive listening to sounds from different locations: an fMRI and MEG study. *Hum Brain Mapp* 26:251–261. [CrossRef Medline](#)
- Bullier J (2001) Integrated model of visual processing. *Brain Res Rev* 36:96–107. [CrossRef Medline](#)
- Da Costa S, van der Zwaag W, Marques JP, Frackowiak RS, Clarke S, Saenz M (2011) Human primary auditory cortex follows the shape of Heschl’s gyrus. *J Neurosci* 31:14067–14075. [CrossRef Medline](#)
- Day ML, Delgutte B (2013) Decoding sound source location and separation using neural population activity patterns. *J Neurosci* 33:15837–15847. [CrossRef Medline](#)
- De Martino F, Moerel M, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E (2013) Spatial organization of frequency preference and selectivity in the human inferior colliculus. *Nat Commun* 4:1386. [CrossRef Medline](#)
- Deouell LY, Heller AS, Malach R, D’Esposito M, Knight RT (2007) Cerebral responses to change in spatial location of unattended sounds. *Neuron* 55:985–996. [CrossRef Medline](#)
- Derey K, Valente G, de Gelder B, Formisano E (2016) Opponent coding of sound location (azimuth) in planum temporale is robust to sound-level variations. *Cereb Cortex* 26:450–464. [CrossRef Medline](#)
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859–869. [CrossRef Medline](#)
- Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6:1216–1223. [CrossRef Medline](#)
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: from single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27:392–401. [CrossRef Medline](#)
- Harper NS, McAlpine D (2004) Optimal neural population coding of an auditory spatial cue. *Nature* 430:682–686. [CrossRef Medline](#)
- Harrington IA, Stecker GC, Macpherson EA, Middlebrooks JC (2008) Spatial sensitivity of neurons in the anterior, posterior, and primary fields of cat auditory cortex. *Hear Res* 240:22–41. [CrossRef Medline](#)
- Higgins NC, McLaughlin SA, Rinne T, Stecker GC (2017) Evidence for cue-independent spatial representation in the human auditory cortex during active listening. *Proc Natl Acad Sci U S A* 114: E7602–E7611. [CrossRef Medline](#)
- Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. *Nat Neurosci* 9:690–696. [CrossRef Medline](#)
- Jiang X, Chevillet MA, Rauschecker JP, Riesenhuber M (2018) Training humans to categorize monkey calls: auditory feature and category selective neural tuning changes. *Neuron* 98:405–416.e4. [CrossRef Medline](#)
- Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A* 97:11793–11799. [CrossRef Medline](#)
- Kim JJ, Crespo-Facorro B, Andreasen NC, O’Leary DS, Zhang B, Harris G, Magnotta VA (2000) An MRI-based parcellation method for the temporal lobe. *Neuroimage* 11:271–288. [CrossRef Medline](#)
- King AJ, Bajo VM, Bizley JK, Campbell RA, Nodal FR, Schulz AL, Schnupp JW (2007) Physiological and behavioral studies of spatial coding in the auditory cortex. *Hear Res* 229:106–115. [CrossRef Medline](#)
- Krumbholz K, Schönwiesner M, von Cramon DY, Rübsem R, Shah NJ, Zilles K, Fink GR (2005) Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. *Cereb Cortex* 15:317–324. [CrossRef Medline](#)
- Lamme VA, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci* 23:571–579. [CrossRef Medline](#)
- Leaver AM, Rauschecker JP (2016) Functional topography of human auditory cortex. *J Neurosci* 36:1416–1428. [CrossRef Medline](#)
- Lee CC, Middlebrooks JC (2011) Auditory cortex spatial sensitivity sharpens during task performance. *Nat Neurosci* 14:108–114. [CrossRef Medline](#)
- Lee CC, Middlebrooks JC (2013) Specialization for sound localization in

- fields A1, DZ, and PAF of cat auditory cortex. *J Assoc Res Otolaryngol* 14:61–82. [CrossRef Medline](#)
- Lee CC, Winer JA (2011) A synthesis of auditory cortical connections: thalamocortical, commissural and corticocortical systems. In: *The auditory cortex*, pp 147–170. New York: Springer.
- Lomber SG, Malhotra S (2008) Double dissociation of “what” and “where” processing in auditory cortex. *Nat Neurosci* 11:609–616. [CrossRef Medline](#)
- McAlpine D, Jiang D, Palmer AR (2001) A neural code for low-frequency sound localization in mammals. *Nat Neurosci* 4:396–401. [CrossRef Medline](#)
- McLaughlin SA, Higgins NC, Stecker GC (2016) Tuning to binaural cues in human auditory cortex. *J Assoc Res Otolaryngol* 17:37–53. [CrossRef Medline](#)
- Miller LM, Recanzone GH (2009) Populations of auditory cortical neurons can accurately encode acoustic space across stimulus intensity. *Proc Natl Acad Sci U S A* 106:5931–5935. [CrossRef Medline](#)
- Moerel M, De Martino F, Formisano E (2012) Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci* 32:14205–14216. [CrossRef Medline](#)
- Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. *Front Neurosci* 8:225. [CrossRef Medline](#)
- Mondor TA, Zatorre RJ (1995) Shifting and focusing auditory spatial attention. *J Exp Psychol Hum Percept Perform* 21:387–409. [CrossRef Medline](#)
- Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K (2001) Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* 13:684–701. [CrossRef Medline](#)
- Ortiz-Rios M, Azevedo FAC, Kuśmierk P, Balla DZ, Munk MH, Keliris GA, Logothetis NK, Rauschecker JP (2017) Widespread and opponent fMRI signals represent sound location in macaque auditory cortex. *Neuron* 93:971–983. [e974. CrossRef Medline](#)
- Otazu GH, Tai LH, Yang Y, Zador AM (2009) Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* 12:646–654. [CrossRef Medline](#)
- Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724. [CrossRef Medline](#)
- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800–11806. [CrossRef Medline](#)
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111–114. [CrossRef Medline](#)
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131–1136. [CrossRef Medline](#)
- Rorden C, Driver J (2001) Spatial deployment of attention within and across hemifields in an auditory task. *Exp Brain Res* 137:487–496. [CrossRef Medline](#)
- Spence CJ, Driver J (1994) Covert spatial orienting in audition: exogenous and endogenous mechanisms. *J Exp Psychol Hum Percept Perform* 20:555. [CrossRef](#)
- Spierer L, Bellmann-Thiran A, Maeder P, Murray MM, Clarke S (2009) Hemispheric competence for auditory spatial representation. *Brain* 132:1953–1966. [CrossRef Medline](#)
- Stecker GC, Middlebrooks JC (2003) Distributed coding of sound locations in the auditory cortex. *Biol Cybern* 89:341–349. [CrossRef Medline](#)
- Stecker GC, Harrington IA, Middlebrooks JC (2005) Location coding by opponent neural populations in the auditory cortex. *PLoS Biol* 3:e78. [CrossRef Medline](#)
- Striem-Amit E, Hertz U, Amedi A (2011) Extensive cochleotopic mapping of human auditory cortical fields obtained with phase-encoding fMRI. *PLoS One* 6:e17832. [CrossRef Medline](#)
- Talavage TM, Sereno MI, Melcher JR, Ledden PJ, Rosen BR, Dale AM (2004) Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *J Neurophysiol* 91:1282–1296. [CrossRef Medline](#)
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292:290–293. [CrossRef Medline](#)
- Tournoux T, Talairach J (1988) *Co-planar stereotaxic atlas of the human brain*. Stuttgart, Germany: Theime.
- van der Zwaag W, Gentile G, Gruetter R, Spierer L, Clarke S (2011) Where sound position influences sound object representations: a 7-T fMRI study. *Neuroimage* 54:1803–1811. [CrossRef Medline](#)
- Warren JD, Griffiths TD (2003) Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *J Neurosci* 23:5799–5804. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Winer JA, Schreiner CE (2005) The central auditory system: a functional analysis. In: *The inferior colliculus*, pp 1–68. New York: Springer.
- Zatorre RJ, Penhune VB (2001) Spatial localization after excision of human auditory cortex. *J Neurosci* 21:6321–6328. [CrossRef Medline](#)
- Zimmer U, Macaluso E (2005) High binaural coherence determines successful sound localization and increased activity in posterior auditory areas. *Neuron* 47:893–905. [CrossRef Medline](#)