The Encoding of Sound Source Elevation in the Human Auditory Cortex

Régis Trapeau^{1,2} and Marc Schönwiesner^{1,2,3,4}

¹International Laboratory for Brain, Music and Sound Research, Université de Montréal, Department of Psychology, Outremont, Quebec H2V 4P3, Canada, ²Centre for Research on Brain, Language and Music, McGill University, Montreal, Quebec H3A 0G4, Canada, ³Department of Neurology and Neurosurgery, Faculty of Medicine, McGill University, Montreal, Quebec H3A 0G4, Canada, and ⁴Institute of Biology, University of Leipzig, 04103 Leipzig, Germany

Spatial hearing is a crucial capacity of the auditory system. While the encoding of horizontal sound direction has been extensively studied, very little is known about the representation of vertical sound direction in the auditory cortex. Using high-resolution fMRI, we measured voxelwise sound elevation tuning curves in human auditory cortex and show that sound elevation is represented by broad tuning functions preferring lower elevations as well as secondary narrow tuning functions preferring individual elevation directions. We changed the ear shape of participants (male and female) with silicone molds for several days. This manipulation reduced or abolished the ability to discriminate sound elevation and flattened cortical tuning curves. Tuning curves recovered their original shape as participants adapted to the modified ears and regained elevation perception over time. These findings suggest that the elevation tuning observed in low-level auditory cortex did not arise from the physical features of the stimuli but is contingent on experience with spectral cues and covaries with the change in perception. One explanation for this observation may be that the tuning in low-level auditory cortex underlies the subjective perception of sound elevation.

Key words: auditory cortex; fMRI; plasticity; sound elevation; spatial hearing

Significance Statement

This study addresses two fundamental questions about the brain representation of sensory stimuli: how the vertical spatial axis of auditory space is represented in the auditory cortex and whether low-level sensory cortex represents physical stimulus features or subjective perceptual attributes. Using high-resolution fMRI, we show that vertical sound direction is represented by broad tuning functions preferring lower elevations as well as secondary narrow tuning functions preferring individual elevation directions. In addition, we demonstrate that the shape of these tuning functions is contingent on experience with spectral cues and covaries with the change in perception, which may indicate that the tuning functions in low-level auditory cortex underlie the perceived elevation of a sound source.

Introduction

Spatial hearing is the capacity of the auditory system to infer the location of a sound source from the complex acoustic signals that reach both ears. This ability is crucial for an efficient interaction with the environment because it guides attention (Broadbent, 1954; Scharf, 1998) and improves the detection, segregation, and

recognition of sounds (Dirks and Wilson, 1969; Bregman, 1994; Roman et al., 2003). Sound localization is achieved by extracting spatial cues from the acoustic signal that arise from the position and shape of the two ears. The separation of the two ears produces interaural time (ITD) and interaural level differences, which enable sound localization on the horizontal plane. The direction-dependent filtering of the pinnae and the upper body generates spectral cues, which allow sound localization on the vertical plane as well as disambiguating sounds that originate from the front and the back of the listener (Wightman and Kistler, 1989; Blauert, 1997).

The cortical encoding of the location of sound sources has been mostly studied on the horizontal plane. There is a large body of evidence from both animal and human models that horizontal sound direction is represented in the auditory cortex by a rate code and two opponent neural populations, each one broadly tuned to the contralateral hemifield (Recanzone et al., 2000;

Received Sept. 4, 2017; revised Feb. 11, 2018; accepted Feb. 14, 2018.

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

This work was supported by Canadian Institutes of Health Research Operating Grant MOP-133716 to M.S. We thank Vincent Migneron-Foisy for collecting most of the data.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Régis Trapeau, International Laboratory for Brain, Music and Sound Research, Université de Montréal, Department of Psychology, Pavillon 1420 boul. Mont-Royal, Outremont, Quebec H2V 4P3, Canada. E-mail: regis.trapeau@umontreal.ca.

DOI:10.1523/JNEUROSCI.2530-17.2018

Copyright © 2018 the authors 0270-6474/18/383252-13\$15.00/0

Figure 1. Silicone mold and array of loudspeakers. **A**, Example of a silicone mold in a participant's right ear. **B**, Spherical array of loudspeakers used for behavioral testing.

Stecker et al., 2005; Werner-Reiss and Groh, 2008; Salminen et al., 2009; Magezi and Krumbholz, 2010). However, single-unit recordings in animals often show a greater diversity of spatial sensitivity than can be captured by far-field measures, such as EEG and MEG or by data pooled over many single-unit recordings (Xu et al., 1998; Mscis-Flogel et al., 2005; Woods et al., 2006; Bizley et al., 2007; Middlebrooks and Bremen, 2013). These studies have shown sound elevation sensitivity in neurons in primary and higher-level auditory cortical areas, but it is still unclear how the auditory cortex represents elevation. One fMRI study in humans (Zhang et al., 2015) presented stimuli at different elevations, computed an up-versus-down contrast, but did not construct tuning curves. Their results suggested a nontopographic, distributed, representation of elevation in auditory cortex. Other neuroimaging attempts to study elevation coding were limited by poor behavioral discrimination of the different elevations presented through headphones (Fujiki et al., 2002; Lewald et al., 2008).

We aimed to reveal the encoding of sound elevation in the human auditory cortex by extracting voxelwise elevation tuning curves and to identify the relationship between these tuning curves and elevation perception by manipulating the acoustic cues for elevation perception. In a first fMRI session, sound stimuli recorded through the participants' unmodified ears were presented. These stimuli included the participants' native spectral cues and allowed us to measure encoding of sound elevation under natural conditions. In two subsequent sessions, stimuli carrying modified spectral cues were presented. Hofman et al. (1998) showed that adult humans can adapt to modified spectral cues. Audio-sensory-motor training accelerates this adaptation (Parseihian and Katz, 2012; Carlile et al., 2014; Trapeau et al., 2016). We modified spectral cues by fitting silicone earmolds to the outer ears of our participants. Participants wore these earmolds for 1 week and received daily training to allow them to adapt to the modified cues. Stimuli recorded through the modified ears were presented in two identical fMRI sessions: one before and one after this adaptation. Any change in tuning between the two sessions is therefore due to the recovery of elevation perception achieved during adaptation, rather than physical features of the spectral cues, which allows us to directly link auditory cortex tuning to elevation perception.

Materials and Methods

Participants

Sixteen volunteers took part in the experiment after having provided informed consent. The sample size was based on data from a previous

study (Trapeau et al., 2016). One participant did not complete the experiment due to a computer error during the first scanning session. The 15 remaining participants (9 male, 6 female) were between 22 and 41 years of age (mean $\pm$ SD, 26 ± 4.7 years). They were right-handed as assessed by a questionnaire adapted from the Edinburgh Handedness Inventory (Oldfield, 1971), had no history of hearing disorder or neurological disease, and had normal or corrected-to-normal vision. Participants had hearing thresholds of ≤ 15 dB HL, for octave frequencies between 0.125 and 8 kHz. They had unblocked ear canals, as determined by a nondiagnostic otoscopy. The experimental procedures conformed to the World Medical Association's Declaration of Helsinki and were approved by McGill University's Research Ethics Board.

Procedure overview

Each participant completed three fMRI sessions. In each session, participants listened passively to individual binaural recordings of sounds emanating from different elevations. Stimuli played in fMRI Session 1 were recorded from the participants' unmodified ears and thus included their native spectral cues. Stimuli played in Session 2 were recorded from participants' ears with added silicone earmolds. These earmolds were applied to the concha to modify spectral cues and consequently disrupt elevation perception. Session 3 was identical to Session 2 but took place after participants had worn the earmolds for 1 week. To track performance and accelerate the adaptation process, participants performed daily sound localization tests and training sessions in the free field and with earphones.

Apparatus

The molds were created by applying a fast-curing medical-grade silicone (SkinTite, Smooth-On) in a tapered layer of ~ 3 – 7 mm thickness on the cymba concha and the cavum concha of the external ear, keeping the ear canal unobstructed (Fig. 1A).

Behavioral testing, training, and binaural recordings were conducted in a hemi-anechoic room ($2.5 \times 5.5 \times 2.5$ m). Participants were seated in a comfortable chair with a neck rest, located in the center of a spherical array of loudspeakers (Orb Audio) with radius of 90 cm (Fig. 1B). A LED was mounted at the center of each loudspeaker. To minimize location cues caused by differences in loudspeaker transfer functions, we measured these functions using a robotic arm that pointed a Brüel and Kjær 1/2-inch probe microphone at each loudspeaker from the center of the sphere and designed inverse finite impulse response filters to equalize amplitude and frequency responses across loudspeakers.

An acoustically transparent black curtain was placed in front of the loudspeakers to avoid visual capture. A small label indicating the central location (0° in azimuth and elevation) was visible on the curtain and served as a fixation mark.

A laser pointer and an electromagnetic head-tracking sensor (Polhemus Fastrak) were attached to a headband worn by the participant. The laser light reflected from the curtain served as visual feedback for the head position. Real-time head position and orientation were used to calculate azimuth and elevation of pointed directions.

Individual binaural recordings were taken with miniature microphones (66AF31, Sonion). Sound stimuli in the fMRI sessions were delivered through MR-compatible S14 insert earphones (Sensimetrics). The same pair of earphones was used to present stimuli in the closed-field behavioral task. Frequency response equalization of the insert earphones was achieved using the custom-designed filters provided by Sensimetrics. No equalization was applied to the miniature microphones because their exact transfer function depends partly on the position in the ear canal and is difficult to correct. However, the probe microphones were audiological devices and had a sufficiently flat transfer function. The slight spectral coloration was not elevation-dependent and did not hinder elevation perception.

Sound localization tests, binaural recordings, and stimulus presentations during fMRI sessions were controlled with custom MATLAB scripts (The MathWorks). Stimuli and recordings were generated and processed digitally (48.8 kHz sampling rate, 24 bit amplitude resolution) using TDT System 3 hardware (Tucker Davis Technologies).

Detailed procedure

Earmolds were fitted to the participants' ears during their first visit. Fitting was done in silence, and the molds were removed immediately after the silicone had cured (4–5 min). Individual binaural recordings were then acquired using miniature microphones placed 2 mm inside of the entrance to the blocked ear canal. Participants were asked not to move during the recordings and keep the dot of the head-mounted laser on the fixation mark. Recordings were first acquired from free ears. Without removing the microphones, the earmolds were then inserted for a second recording and removed immediately afterward.

To minimize procedural learning during the experiment and to make sure that participants would have a good elevation perception of our stimuli before their first scan, participants completed three practice sessions. Each practice session took place on a different day, was done without molds, and consisted of a sound localization run, a training run, and a localization run again for both free-field and earphone stimuli. In a previous study using the same equipment, no procedural learning was found for this localization task (Trapeau et al., 2016).

The first fMRI scanning session was then performed, in which binaural recordings from free ears were presented. The second scanning session was performed on the next day. Participants wore the molds during this session and binaural recordings taken with molds were presented. Earmolds were inserted in silence right before the participants entered the scanner and were removed immediately after the end of the scanning sequence.

From the next day onwards, participants wore the molds continuously. They were informed that the earmolds would "slightly modify their sound perception" but did not know the specific effect of this modification. Once the molds were inserted, participants immediately performed two localization runs (one with earphones, the other in the free field), followed by two training runs (with earphones and free field). On each of the following 7 d, participants completed a sound localization run, a training run, and a localization run again for both free-field and earphone stimuli.

Participants were scanned for a third time after 7 d wearing the molds. The third scanning session was identical to the second one. Participants continued wearing the molds until they performed a last localization run with the molds, <24 h after the last scanning session.

Stimuli

All stimuli used in this experiment, including fMRI stimuli, consisted of broadband pulsed pink noise (25 ms pulses interleaved with 25 ms silence). Stimuli were presented at 55 dB SPL in all conditions to optimize elevation perception (Vliegen and van Opstal, 2004).

During the free-field sound localization task and training, the pulsed pink noise was presented directly through the loudspeakers. During the closed-field sound localization task and training, as well as during fMRI sessions, stimuli were presented via the same pair of MRI-compatible earphones (Sensimetrics S14). Stimuli presented via earphones were individual binaural recordings of the pulsed pink noise presented from loudspeakers at various locations. The locations of the stimuli presented in the different tasks are displayed in Figure 2.

Seven elevations were presented during the fMRI sessions (-45° to $+45^\circ$ in 15° steps). Because elevation perception is more accurate for azimuths away from the median plane (Wightman and Kistler, 1989; Makous and Middlebrooks, 1990; Carlile et al., 1997) and for sounds in the left hemifield (Burke et al., 1994), our fMRI stimuli were recorded at 22.5° azimuth in the left hemifield (Fig. 2B). To ensure identical binaural cues at each elevation, we matched ITD across elevations by shifting the left and right channels to match the ITD measured at 0° elevation (maximum ITD correction: $\sim 80 \mu s$), and we matched interaural level differences across elevations by normalizing broadband RMS amplitude of the left and right channel to the amplitudes measured at 0° elevation (maximum

Figure 2. Stimuli locations in the different tasks. **A**, Speaker arrangement during the free-field localization task. **B**, Stimulus locations during the closed-field localization task. **C**, Stimulus locations during the closed-field training.

Figure 3. Spectra and difference spectra of fMRI stimuli. Solid lines indicate the free ears condition. Dotted lines indicate the earmolds condition. **A**, Spectra of the fMRI stimuli at each elevation for one randomly picked participant. Spectra have been computed from KEMAR recordings of the stimuli delivered through the pair of Sensimetrics S14 earphones used during the experiment. **B**, Difference spectra at each elevation for the same participant. Each transfer function was divided by the average of the transfer functions across elevations. Black dashed lines indicate a 0 dB gain. Earmolds mostly changed the spectra for frequencies >6 kHz, whereas lower frequencies were practically unaffected.

interaural level differences correction: ~ 2 dB). The overall intensity at each elevation and for each set of recordings (free vs molds) was also normalized to the same RMS amplitude to remove potential intensity cues (maximum overall intensity correction: ~ 2 dB). The same treatment was applied to the recordings with earmolds. To check that earmolds did not introduce physiologically implausible distortions in the stimulus spectra, we plotted fMRI stimulus spectra and difference spectra for both the free ears and earmold conditions (Fig. 3A). Earmolds mostly changed the spectra for frequencies >6 kHz, whereas lower frequencies were practically unaffected. Removing the common, nondirectional, portion of this limited set of transfer functions helps to compare spectral differences across elevations (Fig. 3B). For a more detailed analysis of the acoustic properties of the earmolds, see Trapeau et al. (2016).

Figure 4. Sound localization performance. Left, Closed-field sound localization performance of a representative participant. Each dot indicates one response (perceived elevation vs target elevation). Red dots indicate targets at 22.5° azimuth in the left hemifield (locations presented in the scanner). Black and red lines indicate linear fits of the responses for all locations and for locations at 22.5° azimuth in the left hemifield, respectively. The EG is the slope of these regression lines. **A**, Baseline performance with free ears (practice day 3). **B**, First performance with molds (before adaptation). **C**, Last performance with molds (adaptation). **D**, Time course of EG averaged across all participants ($n = 15$), shown for the practice period (ears free), and for the adaptation period (with molds). Dashed lines indicate the free-field EG. Solid lines indicate the closed-field EG. Error bars indicate SE. EG was calculated from all target locations. Red diamond symbols represent the closed-field EG calculated using only the locations corresponding to the fMRI stimuli (i.e., seven elevations from -45° to 45° at 22.5° azimuth in the left hemifield).

The duration of the stimulus used in both sound localization tasks (free field and closed field), as well as in the closed-field training task, was 225 ms (i.e., 5 pulses). It was 125 ms (3 pulses) in the free-field training task and 4350 ms (87 pulses) during the fMRI sessions.

Localization tasks

Free-field localization task. The free-field localization task was similar to the one described by Trapeau et al. (2016). In each trial, a 225 ms pulsed pink noise was presented from 1 of 23 loudspeakers covering directions from -45° to 45° in azimuth and elevation (for speaker locations, see Fig. 2A). Each sound direction was repeated five times during a run, in pseudorandom order. Participants responded by turning their head and thus pointing the head-mounted laser toward the perceived location. At the beginning of a run, participants sat so that the head was centered in the loudspeaker array and the laser dot pointed to the central location. This initial head position was recorded and the participant had to return to this position with a tolerance of 2 cm in head location and 2° in pointed angle before each trial. If the head was correctly placed when the button was pressed, a stimulus was played from one of the speakers. If the head was misplaced, a brief warning tone was played from a speaker located above the participant's head (Az: 0° , El: 82.5°).

Closed-field localization task. The procedure in the closed-field localization task was the same as in free field, except that stimuli were presented via MRI-compatible earphones instead of loudspeakers. Stimuli were individual binaural recordings of the free-field stimulus presented from 31 different locations (for stimulus locations, see Fig. 2B).

Statistical analysis

Vertical localization performance was quantified by the elevation gain (EG), defined as the slope of a linear regression line of perceived versus physical elevations (Hofman et al., 1998). Perfect localization corresponds to an EG of 1, whereas random elevation responses result in an EG of 0.

Behavioral EG was calculated on all locations except when reporting the closed-field EG corresponding to a scanning session, where it was calculated using only the locations presented in the scanner (i.e., -45° to 45° elevations at 22.5° azimuth in the left hemifield) (Fig. 2B). No behav-

ioral EG was measured in the scanner; these measurements were taken before or after the scanning sessions, as indicated in the timeline in Figure 4D. The closed-field EG reported for the first fMRI session was thus computed from the data collected during practice day 3 in Figure 4D (using trials with stimuli at 22.5° azimuth). The closed-field EG reported for the second fMRI session (in which binaural recordings with molds were presented) was computed from the data collected during adaptation day 0, and the closed-field EG reported for the third fMRI session was computed as weighted average of data collected during adaptation days 7 and 8 (because comparable data were available from before and after the session). The average was weighted by the time separating the two localization tests from the scanning session. Throughout the analysis, paired comparisons were statistically assessed using Wilcoxon signed-rank tests, p values <0.05 were regarded as significant.

Training tasks

Both training tasks involved auditory and sensory-motor interactions, which has been shown to accelerate the adaptation process compared with tasks that comprise visual feedback only (Carlile et al., 2014).

Free-field training task. The free-field training sessions were 10 min long, and the procedure was inspired by Parsehian and Katz (2012). A continuous train of stimuli (each stimulus was 125 ms and consisted of 3 pulses

of pink noise) was presented from a random location in the frontal field (between $\pm 45^\circ$ elevation and $\pm 90^\circ$ azimuth, 61 loudspeakers). The interstimulus interval depended on the angular distance between the sound direction and the participant's head direction (measured continuously); the smaller the difference, the shorter the interstimulus interval, an effect similar to a Geiger counter. Participants were instructed to point their head toward the sound source location as fast as possible (no button needed to be pressed), which allowed active exploration of the auditory space and aids adaptation (Parsehian and Katz, 2012; Carlile et al., 2014). Once the participants kept their head in an area of 6° radius around the target location for 500 ms, the stimulus was considered found and the location was switched to a random new loudspeaker with the constraint that the new location was at least 45° away from the previous one. Participants were instructed to find as many locations as possible during the training run.

Closed-field training task. The closed-field training consisted of paired auditory and visual stimuli. The auditory stimuli were identical to the ones used in the closed-field localization task (one stimulus consisted of 5 pulses of pink noise and had a total duration of 225 ms) and were also presented through MRI-compatible earphones. A trial consisted in the presentation of five such stimuli (25 pulses in total) from a given location, each stimulus being separated by a 275 ms interstimulus interval. Trials were separated by 1 s pauses without any stimulation. Each trial corresponded to a pseudorandom location in the frontal field (Fig. 2C). The visual stimuli were LEDs blinking in synchrony with the acoustic stimuli and at the location corresponding to the sound source in the binaural recording in each stimulus presentation. Because binaural recordings were taken while participant's head pointed at the central location, participants were asked to keep their head (and the laser dot) pointed to this position during stimulus presentation. The task was automatically paused when the head position or orientation was incorrect. To introduce sensory-motor interactions, participants were asked to shift their gaze toward the blinking LED.

To ensure that participants attended the sound stimuli, they were asked to detect decreases in sound intensity that occurred infrequently (28%) and randomly in the third, fourth, or fifth noise burst (Recanzone,

1998). Participants indicated a detection by pressing a button, which paused the task. To add a sensory-motor component to the task, participants pointed their head toward the position of the detected deviant sound (no sound was played during head motion). Once their head direction was held for 500 ms in a 6° radius area around the target location, the target was considered found and feedback was given to the participant by briefly blinking the LED at the target location. The task was restarted when the head position returned at the central location.

During one training run, each location was presented seven times and a decrease in intensity occurred twice per location. One run took ~8 min to complete. The amount of decrease in intensity was adjusted by a 1-up 2-down staircase procedure so that participants achieved a hit rate of 70.71%. The initial intensity decrement was 10 dB, and participants typically ended between 2 and 4 dB.

fMRI data acquisition

Imaging protocol. Imaging was performed on a 3 T MRI scanner (Trio, Siemens Healthcare), equipped with a 12 channel matrix head-coil. For each participant and each session, a high-resolution ($1 \times 1 \times 1$ mm) whole-brain T1-weighted structural scan (MPRAGE sequence) was obtained for anatomical registration. Functional imaging was performed using an EPI sequence and sparse sampling (Hall et al., 1999) to minimize the effect of scanner noise artifacts (gradient echo; TR = 8.4 s, acquisition time = 1 s, echo time = 36 ms; flip angle = 90°). Each functional volume comprised 13 slices with an in-plane resolution of 1.5×1.5 mm and a thickness of 2.5 mm (FOV = 192 mm). The voxel volume was thus only 5.625 mm^3 , approximately one-fifth of a more commonly used $3 \times 3 \times 3$ mm voxel (27 mm^3). The slices were oriented parallel to the average angle of left and right lateral sulci (measured on the structural scan) to fully cover the superior temporal plane in both hemispheres. As a result, the functional volumes included Heschl's gyrus, planum temporale, planum polare, and the superior temporal gyrus and sulcus.

Procedure. During the fMRI measurements, the participants listened to the sound stimuli while watching a silent nature documentary. Participants were instructed to be mindful of the sound stimulation, but no task was performed in the scanner.

Each session was divided into two runs of 154 acquisitions, including one initial dummy acquisition, for a total of 308 acquisitions per session and per subject. Each acquisition was directly preceded by either a 4350-ms-long stimulus (positioned in the TR of 8.4 s so that the maximum response is measured by the subsequent volume acquisition) or silence. In a session, each of the 7 sound elevations was presented 34 times, interleaved with 68 silent trials in total. Spatial stimuli and silent trials were presented in a pseudorandom order, so that the angular difference between any two subsequent stimuli was $>30^\circ$ and silent trials were not presented in direct succession. Stimulus sequences in the three scanning sessions were identical for each participant but differed across participants.

fMRI data analysis. The functional data were preprocessed using the MINC software package (McConnell Brain Imaging Center, Montreal Neurological Institute, Montreal). Preprocessing included head motion correction and spatial smoothing with a 3 mm FWHM 3D Gaussian kernel. Each run was linearly registered to the structural scan of the first session and to the ICBM-152 template for group averaging. Group averaging was only used for visualization of the approximate location of active voxels at the group level, and to estimate tuning functions for regions outside of the auditory cortex. GLM estimates of responses to sound stimuli versus silence were computed using fMRISTAT (Worsley et al., 2002). A mask of sound-responsive voxels (all sounds vs silence and each elevation vs silence) in each session was computed for each participant, and all further analysis was restricted to voxels in the conjunction of the masks from the three sessions. Voxel significance was assessed by thresholding T maps at $p < 0.05$, corrected for multiple comparisons using Gaussian Random Field Theory (Worsley et al., 1996). Because our stimuli were presented in the left hemifield and due to the mostly contralateral representation of auditory space (Pavani et al., 2002; Krumbholz et al., 2005; Palomäki et al., 2005; Woods et al., 2009; Trapeau and Schönwiesner, 2015), we analyzed data only in the right hemisphere.

We computed voxelwise elevation tuning curves by plotting the response size in each elevation versus silence contrast. Because we were interested in the shape rather than the overall magnitude of the tuning curves, tuning curves were standardized by subtracting the mean and dividing by the SD (z scores). We determined the elevation preference of each voxel by computing three measures: the center of gravity of the voxel tuning curve (COG), the maximum of the tuning curve, and the slope of the linear regression of standardized activation versus standardized elevation (analogous to the behavioral EG, we refer to this slope as the tuning curve EG). The COG is a simple noise-robust nonparametric method (Schönwiesner et al., 2015), computed as follows:

$$\text{COG} = \frac{1}{n} \sum_{i=1}^n \frac{r_i}{n \cdot r} \cdot e_i,$$

where n is the number of presented elevations (7 in the current experiment), e_i is the elevation value, and r_i is the respective response amplitudes in percent signal change. r is the average response amplitude across elevations. The maximum of the tuning curve is less robust to noise but avoids compression at the highest and lowest elevations. It is therefore more appropriate when plotting sharp tuning functions across elevations and, after verifying that both methods yield similar results, the maximum was used in the analysis of sharp tuning. To avoid selection bias in the mean tuning curves, we selected voxels according to their COG, maximum, or EG in one of the two functional runs and extracted tuning curves from the other run (and inversely; cross-validation).

For all neuroimaging data averaged across participants, SE was obtained by bootstrapping the data 10000 times, using each participant's mean response.

fMRI measures a hemodynamic response, which fortunately is very tightly coupled to neural responses. The BOLD signal correlates directly with the sustained local field potential (and to a lesser degree with spiking activity) as recorded *in vivo* with microelectrodes during fMRI measurements (Logothetis et al., 2001; Ekstrom, 2010). This correlation has also been shown in low-level auditory cortex (Mukamel et al., 2005; Nir et al., 2007) and holds for awake behaving primates viewing visual stimuli (Goense and Logothetis, 2008). Sustained local field potentials reflect mostly intracortical processing. These data are the basis for inferences on neural tuning functions from fMRI recordings.

Results

Behavioral results

Participants had normal elevation perception with a mean EG of 0.75 ± 0.05 (mean $\pm$ SE) in the first free-field test. Figure 4A–C shows behavioral results from a representative participant. The three practice sessions slightly increased free-field EG (0.81 ± 0.04). The practice sessions were especially important for the closed-field condition because participants required training to perform as well in the less natural closed field as in the free field. The average closed-field EG in the first practice session was 0.42 ± 0.06 , which improved to free-field performance by the third session (0.81 ± 0.05). The poor results of the first session were mainly due to reduced externalization of the stimuli (as reported by the participants). The first practice session drastically helped to achieve an externalized perception of the stimuli and practically all of the increase in performance happened in the first practice session. Performance was stable from the second to the third practice session (Fig. 4D). According to participant reports, stimuli were mainly heard as coming from the earphones before the first training and became clearly externalized during training. The large and fast increase in performance is likely due to the acquisition of an externalized percept during the first training.

The insertion of the earmolds produced a marked reduction in the free-field EG (0.38 ± 0.05 ; one-tailed Wilcoxon signed-rank test: $p = 3 \times 10^{-5}$). The same reduced EG (0.34 ± 0.06 ; $p = 6 \times 10^{-5}$) was observed in the closed field when presenting

Figure 5. Mean elevation tuning curve of all sound-responsive voxels in the right hemisphere. **A–M**, Mean elevation tuning curves. Shaded areas represent the SE across participants ($n = 15$, except **F**: $n = 14$). **A**, Mean elevation tuning curve of all sound-responsive voxels (solid black line) and a linear regression of measured activity against stimulus direction (dashed line). The slope of the regression line corresponds to the EG of the mean tuning curve. **B–M**, Mean tuning curves (cross-validated) of voxels binned according to their tuning curve COG, EG, or maximum. The mean number of voxels across participants selected by cross-validation for COG, EG, and maximum is listed in each plot. In each bin, corresponding tuning curves were found in all participants, except **F** (COG = 30°; EG = 1), in which corresponding tuning curves were found in 14 of 15 participants. **B–F**, Black represents mean tuning curves of voxels by COG. Gray represents mean tuning curves of voxels selected by EG. **G–M**, Black represents mean tuning curves. Cross-validated activity did not peak at selected COGs, maxima (vertical dotted lines) or follow the selected EGs (angled dotted lines), but rather decreased with increasing elevation in all voxels. **N**, Distribution of tuning curve COG. **O**, Distribution of tuning curve EG.

individual binaural recordings taken with earmolds. After 7 d of wearing the molds continuously and undergoing short daily training sessions, the mean EG recovered to 0.6 ± 0.06 ($p = 4 \times 10^{-4}$) in the free field and to 0.52 ± 0.05 ($p = 1.5 \times 10^{-4}$) in the closed field (Fig. 4D). Individual differences in adaptation were large and ranged from full recovery to no adaptation.

Because in the fMRI scanner, sounds were presented via earphones from a subset of locations, we also calculated EG values in the closed field based only on those locations (seven vertical directions at 22.5° azimuth in the left hemifield). These EG values were comparable with the free- and closed-field values in the first session (0.86 ± 0.06) and showed a similar reduction (0.42 ± 0.07 ; $p = 2 \times 10^{-4}$) and recovery (0.56 ± 0.07 ; $p = 0.0177$) in the second and third scanning session (Fig. 4D, red diamonds).

Elevation tuning curves with free ears

Sound stimuli evoked significant activation in the auditory cortex of all participants in the three scanning sessions. We computed elevation tuning curves for all sound-responsive voxels and observed signal changes that covaried with elevation. The average tuning curve across voxels and participants showed a decrease of activation level with increasing elevation (Fig. 5A). This decrease was observed in all participants, although the shape and slope of the mean tuning curves differed between participants. The same response pattern was found in both hemispheres, but as expected, the activation was stronger and more widespread in the right hemisphere, contralateral to the sound sources, and we thus only report data from that hemisphere.

The average tuning curve indicated that the majority of voxels may represent increasing elevation as an approximately linear decrease in activity. To determine whether there were any subpopulations of voxels with different tuning, such as narrow tuning curves selective for different sound directions, we plotted the mean tuning curve across all voxels that preferred a given elevation (quantified as the COG and the maximum of the tuning curve) or showed similar EG. Tuning curves were cross-validated

to avoid selection bias. All resulting tuning curves had broad negative slopes (Fig. 5B–M). Cross-validation showed that single-voxel elevation preferences to mid or high elevations were not repeatable across runs and thus due to noise. This is also supported by the distributions of the COG and EG across all sound-responsive voxels and participants, which exhibited only one distinct population, centered (mode) at -15.12° COG and -0.84 EG (Fig. 5N, O). We also applied a dimensionality-reduction technique that excels at clustering high-dimensional data (t-SNE) (Van der Maaten and Hinton, 2008). Although this procedure accurately separated tuning curves of voxels tuned to the left and right hemifield from a previous study on horizontal sound directions, it did not identify distinct subgroups in the present elevation tuning curves.

The shape of the voxelwise elevation tuning curves could be related to elevation perception, or to acoustical properties of the stimuli that change with elevation, or to a mixture of both. To address this question, we first quantified systematic changes with elevation of spectral amplitude in different frequency bands of the stimuli presented in the MRI scanner. We computed a linear regression coefficient between elevation and the mean spectral amplitude per octave band, which results in a frequency-dependent “acoustic EG” (Fig. 6A). We found strong positive correlations in octave bands between 4 and 8 kHz. In these bands, spectral amplitude covaried linearly with elevation. The earmolds slightly shifted these positive correlations toward lower frequencies but did not reduce them. Negative covariations at low frequencies (<1 kHz) probably reflect elevation cues from torso and head and was unaffected by the earmolds.

Such acoustic changes might have contributed to the observed voxel tuning functions. To address this important caveat, we conducted a control experiment. Two pairs of participants performed an additional fMRI session in which we swapped the binaural recordings for each pair of participants. This allowed us to compare functional responses of a same group of participants to a same set of stimuli, in two different conditions: when stimuli matched participants’ spectral cues and led to full elevation perception (own ears); when stimuli were swapped between partic-

Figure 6. Acoustic EG and mean elevation tuning curves in control experiment. **A**, Frequency-dependent acoustic EG. Each point represents the mean slope across participants ($n = 15$, shaded areas indicate the SE across participants) of the linear regression between standardized elevations and the standardized spectral amplitudes of the difference spectrum at each elevation computed in a 1-octave window. The acoustic EG was computed separately for recordings with free ears (green line) and earmolds (orange line). Strong positive correlations are visible between 4 and 8 kHz. **B**, **C**, Mean elevation tuning curves of all sound-responsive voxels in the right hemisphere, averaged across the 3 control participants. **B**, For stimuli corresponding to the participants' own ears, the mean tuning curve displayed a linear decrease with increasing elevation. **C**, Mean tuning curves were no longer monotonic and much flatter when stimuli corresponding to someone else's ears were presented.

ipants and elevation perception was reduced (swapped ears). To avoid any learning of the swapped spectral cues before the scanning session, behavioral performance with swapped stimuli was assessed after the scanning session. All control participants showed reduced elevation perception with swapped stimuli ($EG = 0.38 \pm 0.04$). Their performance was higher for locations corresponding to stimuli presented in the scanner ($EG = 0.52 \pm 0.12$) than for other locations ($EG = 0.34 \pm 0.04$). One participant in particular had a very large EG of 0.82 for the locations presented in the scanner compared with other locations ($EG = 0.38$) and was excluded from the analysis. The mean EG for the remaining participants at the locations presented in the scanner was 0.42 ± 0.11 .

With stimuli corresponding to their own ears, elevation tuning curves of the control participants clearly displayed a linear decrease with increasing elevation (Fig. 6B). With swapped stimuli, the shape of the tuning curves did not exhibit this pattern (Fig. 6C). Mean tuning curves were no longer monotonic and were much flatter (tuning curve EG with own ears = -0.56 ± 0.03 ; tuning curve EG with swapped ears = -0.27 ± 0.07).

The better behavioral performance at locations previously heard in the scanner suggests that some adaptation occurred during the 45 min of passive listening in the scanner. We analyzed the time course of the tuning curve shape during the scanning session by extracting average tuning curves across auditory cortex in a moving time window of ~ 10 min (seven trials per elevation). Tuning curve EG increased over the course of the experiment from essentially zero at the beginning to ~ 0.3 for the final stimulus presentations (Fig. 7). This suggests that the remaining amount of tuning curve EG observed with the swapped recordings is due to adaptation during the measurement, and that the tuning curve EG is close to zero when elevation perception is entirely absent (i.e., at the beginning of the scanning session).

Because the prevalent shape of tuning in auditory cortex appeared to be a linear decrease with increasing elevation, we modeled a linear effect of elevation to highlight areas of auditory cortex exhibiting that tuning. We selected voxels that exhibited a significantly positive or negative effect in this contrast in at least one of the three sessions. At this stage of the analysis, 1 participant had to be excluded because of a malfunction of the earphones

Figure 7. Evolution of the mean elevation tuning curve and of the tuning curve EG with swapped stimuli. Mean tuning curves were extracted from the functional data of the control experiment in a moving time window of ~ 10 min (seven trials per elevation), yielding 22 windows for the 45 min scanning session. Top, Mean tuning curves at windows 3, 11, and 20. Bottom, Time course of the individual (gray lines) and mean (bold black line) tuning curve EG during the session. Tuning curve EG increased over the course of the experiment from essentially zero at the beginning to ~ 0.3 for the final stimulus presentations.

during the second scanning session. Ten of 14 participants displayed significant linear covariation of single-voxel activity and sound elevation (Fig. 8).

Activation decreased with increasing elevation in all significantly active voxels in this contrast. We found no voxels whose activity increased with elevation. The mean tuning of voxels in this contrast had the same shape as the one observed in all sound-responsive voxels but was less noisy.

Figure 8. Distribution of tuning curve EG across auditory cortex. Fixed-effects group average data (labeled Group) and individual data (labeled 1–10) are shown. Slices are oriented to run parallel to the lateral sulcus (see red line on inset schematic brain). Color code represents tuning curve EG of sound-responsive voxels. White outlines indicate voxels that showed a significant covariation with elevation in at least one of the three sessions.

Interindividual variability of the mapping of elevation preference in the auditory cortex was large, which was expected by the great anatomical variability of auditory cortices across individuals (Rademacher et al., 2001) and the functional variability reported for other types of mappings (e.g., Humphries et al., 2010; Griffiths and Hall, 2012; Trapeau and Schönwiesner, 2015).

There is a known relation between frequency and elevation such that sounds at higher elevations tend to contain more power at high frequencies (Blauert, 1997; Parise et al., 2014). We tested whether elevation preference correlated with frequency preference by comparing individual elevation preference maps and a published average map for frequency preference (Schönwiesner et al., 2015). The mean correlation between individual maps of elevation COG and the average map of frequency COG did not differ significantly from zero ($p = 0.56$, Wilcoxon signed rank test of Spearman correlation coefficients across participants). In addition, the mean frequency preference of voxels that showed a significant covariation with elevation (Fig. 8, white outline) did not differ significantly from the frequency preference of other sound-responsive voxels (sounds-vs-silence contrast, colored voxels in Fig. 8 excluding white outline; $p = 0.12$, Wilcoxon signed rank test of mean frequency tuning of the two groups of voxels across participants).

Effect of the molds and adaptation on elevation tuning

The above results demonstrate a relationship between physical sound position on the vertical axis and auditory cortex activity. Our manipulation of spectral cues and elevation perception allowed us to go one step further and to relate the shape of these tuning curves directly to elevation perception.

If voxel activity in low-level auditory cortex relates to the perception of sound source elevation, then we predict that our measure of elevation perception (behavioral EG) should correlate with the equivalent brain measure (tuning curve EG) across all participants and sessions. Behavioral EG correlated significantly with tuning curve EG (in voxels that showed a significant covariation with elevation, the contrast used in the previous section) across individuals and sessions (Fig. 9A; Spearman Correlation Coefficient: $r = -0.44$, $p = 0.019$), indicating that changes in tuning curves covary with changes in reported stimulus elevation across sessions. We did not find significant correlations between these measures within sessions or between behavioral adaptation and individual differences in neural tuning, perhaps due to insufficient within-session variance or statistical power.

On the basis of the measured tuning curves, we hypothesized that auditory cortex encodes the perception of vertical sound

location by a decrease in activity with increasing elevation. Under this hypothesis, we predict less informative (flatter and/or noisier) tuning curves when elevation perception is reduced by the earmolds, and a recovery of the tuning curve shape when participants adapt to the earmolds and regain elevation perception. Steepness of the tuning curves was quantified by their EG, and noisiness was quantified by the R^2 of the EG regression line, in voxels that showed a significant covariation with elevation. Results showed a significant flattening of the participants' mean tuning curve from the first to the second session (mean tuning curve EG from Session 1 to Session 2: 0.73 ± 0.03 to 0.49 ± 0.06 ; one-tailed Wilcoxon signed-rank test, $p = 0.0098$, $n = 10$). Noisiness also increased from Session 1 to Session 2, as shown by the reduction of R^2 (0.56 ± 0.04 for Session 1, 0.33 ± 0.03 for Session 2; $p = 0.0068$, $n = 10$).

To observe the effects of behavioral adaptation on the shape of elevation tuning curves, we then examined differences between Sessions 2 and 3. Two of our participants showed no behavioral adaptation to the earmolds and were therefore excluded from the following comparison (the behavioral EG across sessions of the remaining participants is shown in Fig. 9B). From Session 2 to Session 3, tuning curves changed in the predicted direction (i.e., the opposite direction of the changes observed between Sessions 1 and 2). Tuning curve slope increased (mean tuning curve EG from Session 2 to Session 3: 0.5 ± 0.07 to 0.67 ± 0.03 ; $p = 0.0273$, $n = 8$; Fig. 9C) and the curves were less noisy in Session 3 (R^2 from Session 2 to Session 3: 0.33 ± 0.04 to 0.46 ± 0.04 ; $p = 0.0195$, $n = 8$; Fig. 9D). These differences between the sessions are also visible in the mean tuning curves across participants. The mean tuning curve with molds (Session 2) was flatter and less monotonic than the one with native ears (Session 1), and recovered after adaptation to the molds (Session 3; Fig. 9E).

Secondary narrow tuning to elevation

Other features of the response, apart from the overall slope of the tuning functions, may change with sound elevation. To identify such changes, we subtracted the mean tuning function of each participant from all voxel tuning functions and examined the remainder separately in voxels that responded maximally to each of the seven elevations. Without subtracting the mean tuning, local maxima in the cross-validated tuning function are obscured by the overall decreasing tuning curve (Fig. 5G–M). Cross-validated mean tuning curves after subtracting the mean had narrow peaks at each selected maximum. Each of these peaks was significantly higher than the mean activation at other elevations (as tested by a one-tailed Wilcoxon signed rank test; Fig. 10A).

Figure 9. Behavioral and tuning curve results across sessions. **A**, Correlation between behavioral and tuning curve EG across all participants and sessions. The Spearman correlation coefficient (bottom right inset) was calculated after excluding two outlier data points (unfilled dots; coefficient including outliers: $r = -0.48$, $p = 0.008$). Black dots indicate results of Session 1. Light gray dots indicate results of Session 2. Dark gray dots indicate results of Session 3. Large dots indicate the mean of each session (outliers excluded). **A–D**, Data show clear covariation of reported elevation perception with mean voxel tuning curve slope. The position of Session 3 approximately halfway in between Sessions 1 and 2 in **D** suggests a proportional recovery of elevation perception and tuning curve shape during adaptation. **B**, Behavioral EG across participants who showed adaptation to the earmolds ($n = 8$). Error bars indicate the SE across participants. **C**, Mean tuning curve EG across the same participants. The absolute EG dropped in Session 2 (shallower tuning curves) and partially recovered in Session 3. **D**, Mean R^2 of the EG regression line across participants. R^2 dropped in Session 2 (tuning curves noisier) and partially recovered in Session 3. **E**, Mean elevation tuning curves. Tuning curves showed a steep decrease toward higher elevations with free ears (Session 1, black). Tuning curves flattened when molds were inserted (Session 2, light gray) but recovered after adaptation to the molds (Session 3, dark gray). Shaded areas represent the SE across participants.

Figure 10. Mean tuning curves after subtracting the global mean. Mean tuning curves of voxels binned according to their tuning curve maximum, after subtracting the global mean tuning of each participant. Shaded areas represent the SE across participants ($n = 15$). In each bin, corresponding tuning curves were found in all participants. To avoid selection bias, maxima were computed in one-half of the data, and mean tuning curves were calculated in the other half of the data (and inversely). Black represents mean tuning curves. Vertical dotted line indicates the bin maximum. The mean number of voxels across participants selected by cross-validation is listed in each plot. Whether the peaks at each selected elevation were significantly higher than the mean activation at other elevations was tested by a one-tailed Wilcoxon signed rank test as follows: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$. The peaks were significant at each selected elevation in Session 1 (**A**). Only two of seven peaks were significant in Session 2, after insertion of the molds (**B**). All but one peak were significant in Session 3, after adaptation to the molds (**C**).

The mean height of the peaks compared with other elevations corresponded to $\sim 20\%$ of the BOLD signal change of the global decreasing tuning curve. No topographical pattern or gradient was evident in the distribution of these peaks across the surface of the auditory cortex. If the peaks relate to elevation perception, the

loss in elevation perception caused by the earmolds should be reflected by a decrease in peak height from Session 1 to Session 2, and adaptation to the molds should be reflected by an increase from Session 2 to Session 3. This was observed. After insertion of the molds, narrow tuning was no longer evident (Fig. 10B), and

the height of the peaks decreased (mean standardized peak height from Session 1 to Session 2: 0.2 ± 0.03 to 0.1 ± 0.03 ; one-tailed Wilcoxon signed-rank test, $p = 0.015$, $n = 15$), and most of them were no longer significantly higher than the mean activity at other elevations. After adaptation, all peaks, except one, were significantly above the mean activity (Fig. 10C), and peak height increased significantly from Session 2 to Session 3 (mean standardized peak height: from 0.1 ± 0.03 to 0.19 ± 0.03 ; one-tailed Wilcoxon signed-rank test, $p = 0.011$, $n = 15$).

Discussion

Behavioral adaptation to modified spectral cues

Consistent with previous studies, the insertion of earmolds produced a marked decrease in elevation performance (Hofman et al., 1998; Carlile, 2014; Trapeau et al., 2016). Most participants recovered a portion of their elevation perception after wearing the molds during 7 d and taking part in daily training sessions. As observed in previous earmold adaptation studies (Hofman et al., 1998; Van Wanrooij and van Opstal, 2005; Carlile and Blackman, 2014; Trapeau et al., 2016), individual differences in adaptation were large and ranged from full recovery to no adaptation. Various factors might explain these differences, such as the individual capacity for plasticity or differences in lifestyle that may have led to different amounts of daily sound exposure or multisensory interactions (Javer and Schwarz, 1995; Carlile, 2014). In addition, the acoustical effects of the earmolds vary from one ear to another and acoustical differences between directional transfer functions with and without molds explain a portion of individual differences in the amount of adaptation (Trapeau et al., 2016).

Only closed-field stimuli can be presented in typical human neuroimaging setups, and the difficulty to achieve proper elevation perception for such stimuli hampered earlier attempts to measure elevation tuning (Fujiki et al., 2002; Lewald et al., 2008). To circumvent this problem, we trained participants in an audio-visual sensory-motor task using the same earphones as in the scanner, which allowed participants to achieve accurate perception of the fMRI stimuli.

Elevation tuning in auditory cortex

We measured fMRI responses to sounds perceived at different elevations and computed voxelwise elevation tuning curves. Elevation tuning curves were wide and approximated a linear decrease in activation with increasing elevation. Their slope was correlated with elevation perception as measured by behavioral responses: modifying spectral cues with earmolds, thereby reducing elevation perception, resulted in noisier and flatter tuning curves; changes reflecting the perceptual adaptation to the modified cues were observed from Session 2 to Session 3, which were identical in all aspects. After adaptation, the shape of the tuning curves reapproached the one measured with free ears. The difference in the tuning observed for physically identical stimuli in Sessions 2 and 3 indicates that voxel tuning functions depend on listening experience with spectral cues and thus do not solely represent physical features of the spectral cues. The change in tuning covaried with changes in elevation perception between the sessions, which may indicate that the tuning functions in low-level auditory cortex underlie the perceived elevation of a sound source. At the population level, cortical neurons may thus encode perceived sound source elevation using the slope of the tuning curves, as previously proposed for the coding of horizontal sound direction in the human auditory cortex (Salminen et al., 2009, 2010; Magezi and Krumbholz, 2010; Trapeau and Schönwiesner,

2015; McLaughlin et al., 2016). A potential neural mechanism for the observed voxel-level responses is a rate code.

Why would cortical neurons respond most strongly to lower elevations? In cats, an early stage of spectral cue processing is performed by notch-sensitive neurons of the dorsal cochlear nucleus (Young et al., 1992; Imig et al., 2000) and by their projections to Type O units of the central nucleus of the inferior colliculus (Davis et al., 2003). Deeper spectral notches at low elevations might result in increased sensitivity of these cells to low elevations and might have influenced the general cortical tuning preference for low rather than high elevations during evolution. Previous plots of directional transfer functions suggest deeper spectral notches at low elevations than at high elevations (Middlebrooks, 1997, 1999; Cheng and Wakefield, 1999; Raykar et al., 2005; Trapeau et al., 2016). The acoustic EG computed from our stimuli shows that low elevations contained deeper notches in frequency bands previously identified as crucial for elevation perception (Langendijk and Bronkhorst, 2002; Trapeau et al., 2016). These acoustic features could conceivably drive stronger responses at low elevations. However, such spectral features do not solely determine the shape of cortical elevation tuning curves: swapping stimuli between participants disrupted elevation perception and led to tuning curves that no longer exhibited a linear decrease with increasing elevation. Additionally, the adaptation to modified spectral cues showed that differences in perception of physically identical stimuli change the shape of tuning curves in low-level auditory cortex. These results suggest that the tuning curves observed in auditory cortex do not encode acoustical elevation cues, but the source elevation computed from these cues. This provides evidence for the notion that low-level auditory cortex represents not only physical stimulus features but also the subjective perception of these features (Staeren et al., 2009; Kilian-Hütten et al., 2011). A model of elevation processing posits that spectral profiles of an incoming sound are compared with templates for different elevations (Hofman and van Opstal, 1998). In terms of this model, auditory cortex activity may reflect the result of this comparison, and a correct mapping would enable accurate perception. In turn, experience may reshape this mapping or templates.

Obscured by the main broad tuning, a secondary narrow tuning was observed after subtracting each participant's mean tuning curve from the data. Voxels exhibiting significant secondary tuning peaks were found at each elevation, and the height of these peaks was approximately one-fifth of the size of the overall decrease with increasing elevation. With modified spectral cues, secondary peaks were initially almost absent, but reappeared after adaptation to the modified cues. Therefore, both broad tuning aspect and secondary narrow tuning aspect of the elevation tuning curves covary with elevation perception. This finding may indicate that sound elevation is encoded by two complementary tuning mechanisms. In gerbils, the representation of ITDs seems to be transformed from broadly tuned channels in the midbrain to a uniform sampling by sharply tuned responses in the primary auditory cortex (Belliveau et al., 2014), and a representation of ITDs involving multiple types of codes in human auditory cortex has been recently speculated about (McLaughlin et al., 2016). Two populations of neurons may be necessary for elevation encoding because pinna and head/torso contribute independent spectral cues in different frequency ranges. Pinna cues lie >4 kHz, whereas torso cues exist only <3 kHz (Asano et al., 1990; Algazi et al., 2001). Different optimal coding strategies could then arise from the different spectral cue frequency ranges (Harper and McAlpine, 2004).

Comprehensive data from animal models would aid in interpreting the voxel-level results in humans. No clear pattern for an overall elevation tuning preference was observed among the few elevation-sensitive neurons found in macaque primary auditory cortex and caudomedial field (Recanzone, 2000; Recanzone et al., 2000). One study in cats reported distributed elevation tuning preference across neurons in nontopographic auditory cortex, with some evidence for an overall preference for upper frontal elevations (Xu et al., 1998). However, elevation encoding may differ between cats and humans due to differences in body height, ear shape, and ear mobility.

We observed experience-dependent changes in auditory cortex, but this plasticity may emerge already in subcortical structures, where spectral cues are extracted. The corticofugal system has been suggested to play a role in the modification of subcortical sensory maps in response to sensory experience (Yan and Suga, 1998) and plays a critical role in experience-dependent auditory plasticity (Suga et al., 2002). Bajo et al. (2010) demonstrated that auditory corticocollicular lesions abolish experience-dependent recalibration of horizontal sound localization. The integrity of this descending pathway may be needed in the adaptation to modified spectral cues.

Limitations

The supine position in the scanner may have influenced our results. Body tilt affects horizontal sound localization judgments (Comalli and Altshuler, 1971; Lackner, 1974) and sound location is perceived in a world coordinate reference frame (Goossens and van Opstal, 1999; Vliegen et al., 2004). At the level of the auditory cortex, it is still unclear whether the representation is craniocentric (Altmann et al., 2009) or both craniocentric and allocentric (Schechtman et al., 2012; Town et al., 2017). In the latter case, subsets of neurons representing sound location in a world coordinate frame might have influenced the shape of the elevation tuning curves.

We acquired a total of 52 one-hour sessions of fMRI data from 16 participants, considerably exceeding the amount of data in a typical fMRI study. A higher number of participants would have been useful to identify sources of the interindividual variability in adaptation to our ear manipulation and the potential correlation between behavioral and brain data within sessions but was technically and economically not feasible.

Outlook

Our results suggest that sound elevation is represented in the auditory cortex by broad tuning functions preferring lower elevations as well as secondary narrow tuning functions preferring individual elevation directions. The shape of these tuning functions covaried with the perception of the stimuli, highlighting the link between low-level cortical activity and perception. Our results emphasize the complexity of cortical processing of sound location. A distinct neural population or tuning might be needed to represent sounds in the rear field. Regions involved in vertical and horizontal auditory space processing seem to largely overlap (compare with Trapeau and Schönwiesner, 2015), and there is growing evidence for an integrated cortical processing of localization cues (Pavani et al., 2002; Edmonds and Krumbholz, 2014). In animal models, single-unit recordings demonstrated that cortical neurons are sensitive to multiple localization cues (Brugge et al., 1994; Xu et al., 1998; Mrsic-Flogel et al., 2005). Horizontal and vertical auditory space might be represented by integrated codes, and future experiments should combine these dimensions.

References

- Algazi VR, Avendano C, Duda RO (2001) Elevation localization and head-related transfer function analysis at low frequencies. *J Acoust Soc Am* 109:1110–1122. [CrossRef Medline](#)
- Altmann CF, Wilczek E, Kaiser J (2009) Processing of auditory location changes after horizontal head rotation. *J Neurosci* 29:13074–13078. [CrossRef Medline](#)
- Asano F, Suzuki Y, Sone T (1990) Role of spectral cues in median plane localization. *J Acoust Soc Am* 88:159–168. [CrossRef Medline](#)
- Bajo VM, Nodal FR, Moore DR, King AJ (2010) The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci* 13:253–260. [CrossRef Medline](#)
- Belliveau LA, Lyamzin DR, Lesica NA (2014) The neural representation of interaural time differences in gerbils is transformed from midbrain to cortex. *J Neurosci* 34:16796–16808. [CrossRef Medline](#)
- Bizley JK, Nodal FR, Parsons CH, King AJ (2007) Role of auditory cortex in sound localization in the midsagittal plane. *J Neurophysiol* 98:1763–1774. [CrossRef Medline](#)
- Blauert J (1997) Spatial hearing: the psychophysics of human sound localization. Cambridge, MA: Massachusetts Institute of Technology.
- Bregman AS (1994) Auditory scene analysis: the perceptual organization of sound. Cambridge, MA: Massachusetts Institute of Technology.
- Broadbent DE (1954) The role of auditory localization in attention and memory span. *J Exp Psychol* 47:191–196. [CrossRef Medline](#)
- Brugge JF, Reale RA, Hind JE, Chan JC, Musicant AD, Poon PW (1994) Simulation of free-field sound sources and its application to studies of cortical mechanisms of sound localization in the cat. *Hear Res* 73:67–84. [CrossRef Medline](#)
- Burke KA, Letsos A, Butler RA (1994) Asymmetric performances in binaural localization of sound in space. *Neuropsychologia* 32:1409–1417. [CrossRef Medline](#)
- Carlile S (2014) The plastic ear and perceptual relearning in auditory spatial perception. *Front Neurosci* 8:237. [CrossRef Medline](#)
- Carlile S, Blackman T (2014) Relearning auditory spectral cues for locations inside and outside the visual field. *J Assoc Res Otolaryngol* 15:249–263. [CrossRef Medline](#)
- Carlile S, Leong P, Hyams S (1997) The nature and distribution of errors in sound localization by human listeners. *Hear Res* 114:179–196. [CrossRef Medline](#)
- Carlile S, Balachandrar K, Kelly H (2014) Accommodating to new ears: the effects of sensory and sensory-motor feedback. *J Acoust Soc Am* 135:2002–2011. [CrossRef Medline](#)
- Cheng CI, Wakefield GH (1999) Introduction to Head-Related Transfer Functions (HRTFs): representations of HRTFs in time, frequency, and space. In: *Audio Engineering Society*. <http://www.aes.org/e-lib/browse.cfm?elib=8154>. Accessed December 4, 2015.
- Comalli PE Jr, Altshuler MW (1971) Effect of body tilt on auditory localization. *Percept Mot Skills* 32:723–726. [CrossRef Medline](#)
- Davis KA, Ramachandran R, May BJ (2003) Auditory processing of spectral cues for sound localization in the inferior colliculus. *J Assoc Res Otolaryngol* 4:148–163. [CrossRef Medline](#)
- Dirks DD, Wilson RH (1969) The effect of spatially separated sound sources on speech intelligibility. *J Speech Hear Res* 12:5–38. [CrossRef Medline](#)
- Edmonds BA, Krumbholz K (2014) Are interaural time and level differences represented by independent or integrated codes in the human auditory cortex? *J Assoc Res Otolaryngol* 15:103–114. [CrossRef Medline](#)
- Ekstrom A (2010) How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain Res Rev* 62:233–244. [CrossRef Medline](#)
- Fujiki N, Riederer KA, Jousmäki V, Mäkelä JP, Hari R (2002) Human cortical representation of virtual auditory space: differences between sound azimuth and elevation. *Eur J Neurosci* 16:2207–2213. [CrossRef Medline](#)
- Goense JB, Logothetis NK (2008) Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr Biol* 18:631–640. [CrossRef Medline](#)
- Goossens HH, van Opstal AJ (1999) Influence of head position on the spatial representation of acoustic targets. *J Neurophysiol* 81:2720–2736. [CrossRef Medline](#)
- Griffiths TD, Hall DA (2012) Mapping pitch representation in neural ensembles with fMRI. *J Neurosci* 32:13343–13347. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Harper NS, McAlpine D (2004) Optimal neural population coding of an auditory spatial cue. *Nature* 430:682–686. [CrossRef Medline](#)
- Hofman PM, van Opstal AJ (1998) Spectro-temporal factors in two-dimensional human sound localization. *J Acoust Soc Am* 103:2634–2648. [CrossRef Medline](#)
- Hofman PM, Van Riswick JG, van Opstal AJ (1998) Relearning sound localization with new ears. *Nat Neurosci* 1:417–421. [CrossRef Medline](#)
- Humphries C, Liebethal E, Binder JR (2010) Tonotopic organization of human auditory cortex. *Neuroimage* 50:1202–1211. [CrossRef Medline](#)
- Imig TJ, Bibikov NG, Poirier P, Samson FK (2000) Directionality derived from pinna-cue spectral notches in cat dorsal cochlear nucleus. *J Neurophysiol* 83:907–925. [CrossRef Medline](#)
- Javer AR, Schwarz DW (1995) Plasticity in human directional hearing. *J Otolaryngol* 24:111–117. [Medline](#)
- Kilian-Hütten N, Valente G, Vroomen J, Formisano E (2011) Auditory cortex encodes the perceptual interpretation of ambiguous sound. *J Neurosci* 31:1715–1720. [CrossRef Medline](#)
- Krumbholz K, Schönwiesner M, von Cramon DY, Rübsamen 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](#)
- Lackner JR (1974) Changes in auditory localization during body tilt. *Acta Otolaryngol* 77:19–28. [CrossRef Medline](#)
- Langendijk EH, Bronkhorst AW (2002) Contribution of spectral cues to human sound localization. *J Acoust Soc Am* 112:1583–1596. [CrossRef Medline](#)
- Lewald J, Riederer KA, Lentz T, Meister IG (2008) Processing of sound location in human cortex. *Eur J Neurosci* 27:1261–1270. [CrossRef Medline](#)
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157. [CrossRef Medline](#)
- Magazi DA, Krumbholz K (2010) Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *J Neurophysiol* 104:1997–2007. [CrossRef Medline](#)
- Makous JC, Middlebrooks JC (1990) Two-dimensional sound localization by human listeners. *J Acoust Soc Am* 87:2188–2200. [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](#)
- Middlebrooks JC (1997) Spectral shape cues for sound localization. In: *Binaural and spatial hearing in real and virtual environments*, pp 77–97. Mahwah, NJ: Lawrence Erlbaum.
- Middlebrooks JC (1999) Individual differences in external-ear transfer functions reduced by scaling in frequency. *J Acoust Soc Am* 106:1480–1492. [CrossRef Medline](#)
- Middlebrooks JC, Bremen P (2013) Spatial stream segregation by auditory cortical neurons. *J Neurosci* 33:10986–11001. [CrossRef Medline](#)
- Mrsic-Flogel TD, King AJ, Schnupp JW (2005) Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. *J Neurophysiol* 93:3489–3503. [CrossRef Medline](#)
- Mukamel R, Gelbard H, Arieli A, Hasson U, Fried I, Malach R (2005) Coupling between neuronal firing, field potentials, and fMRI in human auditory cortex. *Science* 309:951–954. [CrossRef Medline](#)
- Nir Y, Fisch L, Mukamel R, Gelbard-Sagiv H, Arieli A, Fried I, Malach R (2007) Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. *Curr Biol* 17:1275–1285. [CrossRef Medline](#)
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9:97–113. [CrossRef Medline](#)
- Palomäki KJ, Tiitinen H, Mäkinen V, May PJ, Alku P (2005) Spatial processing in human auditory cortex: the effects of 3D, ITD, and ILD stimulation techniques. *Cogn Brain Res* 24:364–379. [CrossRef Medline](#)
- Parise CV, Knorre K, Ernst MO (2014) Natural auditory scene statistics shapes human spatial hearing. *Proc Natl Acad Sci U S A* 111:6104–6108. [CrossRef Medline](#)
- Parseihian G, Katz BF (2012) Rapid head-related transfer function adaptation using a virtual auditory environment. *J Acoust Soc Am* 131:2948–2957. [CrossRef Medline](#)
- Pavani F, Macaluso E, Warren JD, Driver J, Griffiths TD (2002) A common cortical substrate activated by horizontal and vertical sound movement in the human brain. *Curr Biol* 12:1584–1590. [CrossRef Medline](#)
- 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. [CrossRef Medline](#)
- Raykar VC, Duraiswami R, Yegnanarayana B (2005) Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *J Acoust Soc Am* 118:364–374. [CrossRef Medline](#)
- Recanzone GH (1998) Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc Natl Acad Sci U S A* 95:869–875. [CrossRef Medline](#)
- Recanzone GH (2000) Spatial processing in the auditory cortex of the macaque monkey. *Proc Natl Acad Sci U S A* 97:11829–11835. [CrossRef Medline](#)
- Recanzone GH, Guard DC, Phan ML, Su TK (2000) Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol* 83:2723–2739. [CrossRef Medline](#)
- Roman N, Wang D, Brown GJ (2003) Speech segregation based on sound localization. *J Acoust Soc Am* 114:2236–2252. [CrossRef Medline](#)
- Salminen NH, May PJ, Alku P, Tiitinen H (2009) A population rate code of auditory space in the human cortex. *PLoS One* 4:e7600. [CrossRef Medline](#)
- Salminen NH, Tiitinen H, Miettinen I, Alku P, May PJ (2010) Asymmetrical representation of auditory space in human cortex. *Brain Res* 1306:93–99. [CrossRef Medline](#)
- Scharf B (1998) Auditory attention: the psychoacoustical approach. In: *H. Pashler (Ed.), Attention*, pp 75–117. Hove, UK: Psychology Press.
- Schechtman E, Shrem T, Deouell LY (2012) Spatial localization of auditory stimuli in human auditory cortex is based on both head-independent and head-centered coordinate systems. *J Neurosci* 32:13501–13509. [CrossRef Medline](#)
- Schönwiesner M, Dechent P, Voit D, Petkov CI, Krumbholz K (2015) Parcellation of human and monkey core auditory cortex with fMRI pattern classification and objective detection of tonotopic gradient reversals. *Cereb Cortex* 25:3278–3289. [CrossRef Medline](#)
- Staeren N, Renvall H, De Martino F, Goebel R, Formisano E (2009) Sound categories are represented as distributed patterns in the human auditory cortex. *Curr Biol* 19:498–502. [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](#)
- Suga N, Xiao Z, Ma X, Ji W (2002) Plasticity and corticofugal modulation for hearing in adult animals. *Neuron* 36:9–18. [CrossRef Medline](#)
- Town SM, Brimijoin WO, Bizley JK (2017) Egocentric and allocentric representations in auditory cortex. *PLoS Biol* 15:e2001878. [CrossRef Medline](#)
- Trapeau R, Schönwiesner M (2015) Adaptation to shifted interaural time differences changes encoding of sound location in human auditory cortex. *Neuroimage* 118:26–38. [CrossRef Medline](#)
- Trapeau R, Aubrais V, Schönwiesner M (2016) Fast and persistent adaptation to new spectral cues for sound localization suggests a many-to-one mapping mechanism. *J Acoust Soc Am* 140:879–890. [CrossRef Medline](#)
- Van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Machine Learn Res* 9:85.
- Van Wanrooij MM, van Opstal AJ (2005) Relearning sound localization with a new ear. *J Neurosci* 25:5413–5424. [CrossRef Medline](#)
- Vliegen J, van Opstal AJ (2004) The influence of duration and level on human sound localization. *J Acoust Soc Am* 115:1705–1713. [CrossRef Medline](#)
- Vliegen J, Van Grootel TJ, van Opstal AJ (2004) Dynamic sound localization during rapid eye-head gaze shifts. *J Neurosci* 24:9291–9302. [CrossRef Medline](#)
- Werner-Reiss U, Groh JM (2008) A rate code for sound azimuth in monkey auditory cortex: implications for human neuroimaging studies. *J Neurosci* 28:3747–3758. [CrossRef Medline](#)
- Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening: II. Psychophysical validation. *J Acoust Soc Am* 85:868–878. [CrossRef Medline](#)
- Woods DL, Stecker GC, Rinne T, Herron TJ, Cate AD, Yund EW, Liao I,

- Kang X (2009) Functional maps of human auditory cortex: effects of acoustic features and attention. *PLoS One* 4:e5183. [CrossRef Medline](#)
- Woods TM, Lopez SE, Long JH, Rahman JE, Recanzone GH (2006) Effects of stimulus azimuth and intensity on the single-neuron activity in the auditory cortex of the alert macaque monkey. *J Neurophysiol* 96:3323–3337. [CrossRef Medline](#)
- Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC (1996) A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp* 4:58–73. [CrossRef Medline](#)
- Worsley KJ, Liao CH, Aston J, Petre V, Duncan GH, Morales F, Evans AC (2002) A general statistical analysis for fMRI data. *Neuroimage* 15:1–15. [CrossRef Medline](#)
- Xu L, Furukawa S, Middlebrooks JC (1998) Sensitivity to sound-source elevation in nontopographic auditory cortex. *J Neurophysiol* 80:882–894. [CrossRef Medline](#)
- Yan W, Suga N (1998) Corticofugal modulation of the midbrain frequency map in the bat auditory system. *Nat Neurosci* 1:54–58. [CrossRef Medline](#)
- Young ED, Spirou GA, Rice JJ, Voigt HF (1992) Neural organization and responses to complex stimuli in the dorsal cochlear nucleus. *Philos Trans R Soc Lond B Biol Sci* 336:407–413. [CrossRef Medline](#)
- Zhang X, Zhang Q, Hu X, Zhang B (2015) Neural representation of three-dimensional acoustic space in the human temporal lobe. *Front Hum Neurosci* 9:203. [CrossRef Medline](#)