

Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex

Citation for published version (APA):

Riecke, L., Peters, J. C., Valente, G., Kemper, V. G., Formisano, E., & Sorger, B. (2017). Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. *Cerebral Cortex*, 27(5), 3002-3014. <https://doi.org/10.1093/cercor/bhw160>

Document status and date:

Published: 01/05/2017

DOI:

[10.1093/cercor/bhw160](https://doi.org/10.1093/cercor/bhw160)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

ORIGINAL ARTICLE

Frequency-Selective Attention in Auditory Scenes Recruits Frequency Representations Throughout Human Superior Temporal Cortex

Lars Riecke¹, Judith C. Peters^{1,2}, Giancarlo Valente¹, Valentin G. Kemper¹, Elia Formisano¹ and Bettina Sorger¹

¹Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, 6229 EV Maastricht, The Netherlands and ²Netherlands Institute for Neuroscience, Institute of the Royal Netherlands Academy of Arts and Sciences (KNAW), 1105 BA Amsterdam, The Netherlands

Address correspondence to Lars Riecke, Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. Email: l.riecke@maastrichtuniversity.nl

Abstract

A sound of interest may be tracked amid other salient sounds by focusing attention on its characteristic features including its frequency. Functional magnetic resonance imaging findings have indicated that frequency representations in human primary auditory cortex (AC) contribute to this feat. However, attentional modulations were examined at relatively low spatial and spectral resolutions, and frequency-selective contributions outside the primary AC could not be established. To address these issues, we compared blood oxygenation level-dependent (BOLD) responses in the superior temporal cortex of human listeners while they identified single frequencies versus listened selectively for various frequencies within a multifrequency scene. Using best-frequency mapping, we observed that the detailed spatial layout of attention-induced BOLD response enhancements in primary AC follows the tonotopy of stimulus-driven frequency representations—analogue to the “spotlight” of attention enhancing visuospatial representations in retinotopic visual cortex. Moreover, using an algorithm trained to discriminate stimulus-driven frequency representations, we could successfully decode the focus of frequency-selective attention from listeners’ BOLD response patterns in nonprimary AC. Our results indicate that the human brain facilitates selective listening to a frequency of interest in a scene by reinforcing the fine-grained activity pattern throughout the entire superior temporal cortex that would be evoked if that frequency was present alone.

Key words: attention, auditory cortex, frequency, MVPA, tonotopy

Introduction

Identifying a sound of interest amid other salient sounds benefits from paying selective attention to one or more distinctive features of that sound including its frequency content or rhythm. When listeners are cued to a specific audio frequency, they become most sensitive to that frequency and neighboring frequencies, suggesting that frequency-selective attention involves the operation of a “listening band,” that is, an attentional bandpass filter in the auditory system that is centered on the attended

frequency (Scharf et al. 1987). Consistent with the general idea of an attentional filter, numerous neuroscientific studies have shown that selective attention to a target sound may enhance that sound’s representation in human auditory cortex (AC) (Hall et al. 2000; Rinne et al. 2005; Ahveninen et al. 2006; Degerman et al. 2006; Bidet-Caulet, Fischer, Bauchet, et al. 2007; Bidet-Caulet, Fischer, Besle, et al. 2007; Krumbholz et al. 2007; Altmann et al. 2008; Elhilali et al. 2009; Xiang et al. 2010; Besle et al. 2011; Mesgarani and Chang 2012; Nourski et al. 2015). However,

comparatively, little is known about the specific role of frequency representations in the human brain for frequency-selective attention in auditory scenes, that is, situations where a particular target sound needs to be separated from and tracked amid concurrent sounds based on its frequency content.

So far, 2 human studies have used functional magnetic resonance imaging (fMRI) to assess cortical frequency representations noninvasively at the neuronal population level and found frequency-selective attentional enhancement, that is, that selective auditory attention enhances blood oxygenation level-dependent (BOLD) responses in primary AC regions representing the attended sound's frequency (Paltoglou et al. 2009; Da Costa et al. 2013), consistent with fMRI findings on pitch imagery (Oh et al. 2013). Magnetoencephalography studies further showed that attention to a tone in noise may enhance long-latency auditory-evoked magnetic fields originating from the superior temporal cortex in a manner thought to reflect adaptation of population-level frequency tuning (Kauramaki et al. 2007; Okamoto et al. 2007, 2009; Ahveninen et al. 2011). Together with more detailed animal invasive electrophysiology studies showing frequency-selective attentional modulation of primary AC neurons (Fritz, Elhilali, David, et al. 2007; Fritz, Elhilali, Shamma, 2007; Lakatos et al. 2013), these findings indicate that frequency-selective attention in a scene involves a filtering mechanism that operates selectively on primary AC frequency representations.

A limitation of the previous human studies is that the key features of the observed attentional modulation, that is, its spatial and spectral distribution, could not be studied in detail because of methodological limitations: Analyses were performed on cortical responses averaged across voxel populations within primary AC, providing only coarse, aggregated information on the spatial pattern of the attentional modulations. In addition, auditory tasks were limited to only 2 attentional targets (frequency bands), providing rather limited information on the spectral specificity of the attentional modulations. Finally, evidence for a frequency-selective attentional mechanism in human nonprimary AC could not be obtained, although neuroimaging studies observed frequency representations in this region (Talavage et al. 2004; Humphries et al. 2010; Woods et al. 2010; Striem-Amit et al. 2011; Moerel et al. 2012; Nourski et al. 2014). Considering that listeners in the 2 aforementioned fMRI studies could identify the target sound based on not only its frequency, but also its location or rhythm, and that other fMRI studies failed to observe frequency-selective attentional enhancement in primary AC (Petkov et al. 2004; Woods et al. 2009), there is need for further characterization of frequency-selective attentional modulation in the human superior temporal cortex.

To address this point, we studied in 2 fMRI experiments the link between frequency-selective attention and frequency representations (as assessed by the BOLD response) at multiple stages of cortical sound processing (primary and nonprimary AC) and across multiple attentional states (attention to one of 3 simultaneously presented frequency bands or attention away from these bands). We first mapped out cortical BOLD responses evoked by auditory stimuli containing single frequency bands or by the listener's paying selective attention to these specific bands in a schematic triple-band scene containing no spatial or rhythmic cues. We then compared the resulting bottom-up stimulus-driven and purely top-down attention-driven response maps to characterize in detail (i.e., across individual cortical voxels and across the 3 target frequencies) the putative spatial and spectral patterns of attentional modulation in primary AC and nonprimary AC.

We observed that attentional modulation 1) closely follows the tonotopic layout of stimulus-driven BOLD responses in

primary AC, 2) reshapes the frequency response of primary AC in a manner consistent with a bandpass filter, and 3) modulates (or enforces) spatially distributed frequency representations, not only in primary AC, but also in nonprimary AC. While our first 2 findings (1 and 2) characterize the previously reported frequency-selective attentional modulation in primary AC (Paltoglou et al. 2009; Da Costa et al. 2013) with increased spatial and spectral detail, our third finding (3) provides novel evidence for the existence of frequency-selective attentional modulation in auditory areas beyond the primary AC. We argue that selective attention to a particular frequency in a scene operates in human superior temporal cortex by (1 and 2) directing a "spotlight" toward the tonotopic primary AC region tuned to the attended frequency and (3) overlaying higher order AC areas with a frequency-specific, non-tonotopic, spatially distributed pattern of modulation.

Materials and Methods

Participants and Stimuli

Fifteen healthy paid volunteers (8 females, ages: 21–42 years) participated in the fMRI study after providing their written informed consent. Auditory stimuli are described in Figure 1A and provided as exemplary [Supplementary Audio Files](#). They were presented diotically via a soundcard (Sound Blaster X-Fi Xtreme Audio) with 16-bit resolution and 44.1 kHz sampling frequency. For details, see [Supplementary Materials and Methods](#).

Design and Tasks

Figure 1 further visualizes the experimental designs (Fig. 1B,C) and trial design (Fig. 1D). The "stimulation experiment" was used to identify sensory frequency representations, whereas the "attention experiment" served to assess frequency-specific attentional modulations. In the stimulation experiment, listeners were stimulated with one of 3 frequency bands (Fig. 1B) and judged the overall pitch of the melody as "low," "middle," or "high." Conversely, in the attention experiment (Fig. 1C), the triple-band stimulus (the sum of the individual bands from the stimulation experiment, Fig. 1A) was presented and kept constant across attention conditions while variations were induced exclusively in the listener's focus of frequency-selective attention.

The manipulation of frequency-selective attention was implemented using detection tasks requiring the listener to pay attention to either a specific frequency band (auditory attention task) or away from auditory input (visual control task). For the auditory attention task, listeners were instructed to attend exclusively to the frequency band that was specified by a visual cue (a letter) and to report via button press whether a gap was present or absent in that band. Gaps were introduced on half of the trials of each condition by omitting the fourth last tone from a given band, with equal probability across bands (Fig. 1A exemplifies a gap in the low band). For the acoustically matched visual control task, listeners were instructed to ignore the auditory stimulus, track covertly the position of the visual cue, which was moving horizontally throughout the stimulus interval (Fig. 1D), and report the side (left or right) to which the cue's largest displacement occurred relative to a central fixation cross. The purpose of this control task was to establish a baseline for auditory attention by drawing participants' attention away from the auditory stimulation while keeping them in a stable alert state.

Participants were instructed to keep their gaze focused on a fixation cross throughout the experiment to reduce visual input

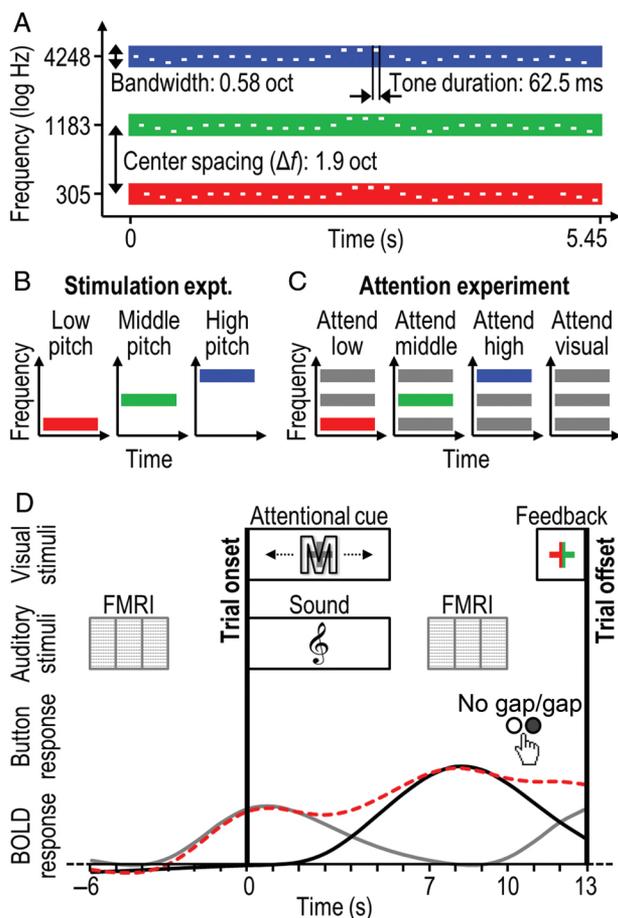


Figure 1. Design of stimuli, experiments, and trials. (A) A schematic spectrogram of an auditory stimulus from the attention experiment. A melody was transposed to 3 nonoverlapping frequency ranges labeled as “low,” “middle,” and “high.” Acoustic parameters were set to support attentional selection of the frequency bands based on their frequency alone. Auditory stimulus examples are provided as [Supplementary audio files](#). (B) The different conditions in the stimulation experiment. The center frequency of the auditory stimulus was varied (low, middle, or high) while participants judged its pitch. (C) The design of the attention experiment. In contrast to the stimulation experiment, the auditory stimulus contained multiple alternating bands and remained unchanged, and only the frequency band on which participants performed an attention-demanding task was varied (low, middle, or high; or none for the control task). (D) Two consecutive trials from the attention experiment. Task trials began with the simultaneous presentation of a fixation cross, visual attentional cue, and auditory stimulus. The cue—a letter that was horizontally relocated throughout the stimulus interval—specified the attention condition: Participants tracked either a specific frequency band (letter L/M/H) to detect a gap in it or the cue itself (letter V) to detect the side of the fixation cross to which its largest displacement occurred (control task). Subsequently, 3 fMRI volumes were acquired, participants responded, and visual feedback on response correctness was provided. Trials in the stimulation experiment were designed identically, except that no attentional cue was presented and 3 response buttons were used. BOLD response simulations (red curve; see [Supplementary Materials and Methods](#)) suggested that our measurements captured primarily BOLD responses evoked by the task and stimuli (black curve) rather than scanner noise (gray curve).

differences across conditions. In addition, on null trials, on which no auditory stimulus was presented, participants were instructed to rest and perform no task. Stimulus presentation and button response collection were controlled using Presentation 17.1 software, and trials were triggered by MR image acquisitions. For details, see [Supplementary Materials and Methods](#).

Procedures

All experimental procedures were approved by the local ethics committee (*Ethische Commissie Psychologie*) of the Faculty of Psychology and Neuroscience of Maastricht University. In an initial training session, participants adjusted the relative sound levels of the frequency bands in an attempt to match the difficulty of the auditory attention task across frequency bands, given a fixed peak intensity of the overall stimulus. The obtained individual settings were applied in the subsequent fMRI session, which is described next. Each repetition (fMRI run) of the stimulation experiment comprised 10 trials per condition (low, middle, high) and 10 null trials in random order (overall duration: 8.9 min). For the attention experiment, each run contained 12 trials per condition (low, middle, high, visual) and 8 null trials (overall duration: 12.4 min). Furthermore, trials of the same attention condition were clustered into blocks of 6 preceded by a null trial during which the attentional cue was presented to specify the upcoming attention condition. Blocks and trials within blocks were presented in random order. Overall, an fMRI session comprised an anatomical scan (duration: 5 min) followed by 3 runs of each experiment in alternating order, starting with the attention experiment. After the fMRI session, participants gave a written report of the mental strategies they followed during the gap detection task. For details, see [Supplementary Materials and Methods](#).

Imaging

Brain images were collected with a Siemens Prisma^{fit} 3 T MRI scanner (Siemens Medical Systems, Erlangen, Germany) equipped with a 64-channel receive head-neck coil. BOLD signal changes were recorded using a gradient-echo echo-planar imaging sequence (echo time: 28 ms, acquisition time: 1000 ms, flip angle: 70°, bandwidth: 1786 Hz/pixel, field of view: 200 × 200 mm², matrix size: 100 × 100, GRAPPA acceleration factor: 2, slice thickness: 2 mm, resulting in a voxel size of 2 × 2 × 2 mm³). Images were acquired according to a sparse clustered protocol (irregular repetition time: 1000, 1000, 11 000 ms, etc.); thus, 3 consecutive functional volumes were acquired on each trial followed by a long silent interval (Fig. 1D), enabling one to separate the auditory stimuli from scanner noise (Hall et al. 1999) as in previous human fMRI tonotopy studies (Langers et al. 2014). During each functional run, 120 volumes (stimulation experiment) or 168 volumes (attention experiment) were collected, in addition to 3 dummy volumes that were not analyzed. Each volume comprised 16 axial continuous slices (i.e., no inter-slice gap) centered on the Sylvian fissures and covering the superior temporal cortices (Fig. 3A). Structural T₁-weighted images optimized for gray-white matter contrast were obtained using a magnetization-prepared rapid-acquisition gradient-echo pulse sequence (voxel size: 1 × 1 × 1 mm³, echo time: 2.21 ms, repetition time: 2250 ms, inversion time: 900 ms).

Statistical Analysis

All measures described below were obtained separately for each subject and then submitted to second-level (random-effects) group analyses using nonparametric statistical tests (Friedman ANOVAs and Wilcoxon signed-rank tests based on the χ^2 statistic and *W*-statistic, respectively) and a significance criterion $\alpha = 0.05$. The family-wise error rate was controlled at α where indicated using the Bonferroni correction.

Data Analysis

The sensitivity index, d' , was obtained by extracting hit rates and false alarm rates while correcting for possible ceiling cases

(Brown and White 2005) and then subtracting the z-transformed corrected rates (Macmillan and Creelman 1991). Trials of the gap detection task on which listeners made an error were considered as attentional lapses and excluded from the fMRI data analysis (on average $9.9 \pm 1.7\%$ of the 108 auditory attention trials; for low, middle, and high frequency: 2.2 ± 0.6 , 4.2 ± 0.8 , and $3.5 \pm 0.7\%$, respectively). fMRI data preparation involved application of standard preprocessing steps and weighted averaging of the 3 functional volumes acquired within each trial; for details, see [Supplementary Materials and Methods](#). From the structural MR images, gray–white matter borders were extracted, from which individual cortical surfaces were reconstructed.

Definition of Regions of Interest

Individual regions of interest (ROIs) were defined within each cerebral hemisphere by partially inflating the reconstructed cortical surfaces and drawing manually on them the putative areal borders of primary AC and superior temporal cortex. These borders were estimated by authors B.S. and L.R. based on common macroanatomic criteria (gyri, sulci); for details, see [Supplementary Materials and Methods](#) and Figure 3B and see [Supplementary Figure 1](#). Subsequent analyses were applied separately to each of the defined ROIs.

An advantage of our primary AC definition is that it takes into account individuals' cortical anatomy in the region of Heschl's gyri, which is highly variable across both subjects and hemispheres (Campain and Minckler 1976). To compare our approach with a more standardized approach based on a probabilistic cytoarchitectonic atlas (Morosan et al. 2001), the cytoarchitectonically defined primary area Te1 (assembled areas Te10, Te11, Te12) was warped into Talairach space as the collected MR images, a probability threshold was set (Rademacher et al. 2001), and Te1 voxels resembling the individual's temporal cortex were extracted (i.e., voxels resembling frontal cortex and white matter were discarded; these voxels were abundant depending on the individual and probability threshold set). Next, for each ROI, the degree to which the ROI separated the retained portion of Te1 (target voxels) from the rest of the superior temporal cortex (nontarget voxels) was quantified using d' . ROI voxels overlapping with target voxels were considered as hits, whereas ROI voxels overlapping with nontarget voxels were considered as false alarms. The d' values associated with the 2 ROIs were then compared statistically. This was done for various probability thresholds (Fig. 3C).

Best-Frequency Mapping

In the first 2 analyses, frequency representations were assessed using an approach based on best-frequency (BF) mapping (Formisano et al. 2003); see Figures 4A and 5A. Sensory BF maps were obtained by first extracting maps of the sound-evoked BOLD response (obtained by comparing with null trials) for each stimulation experiment condition and then labeling (or color coding, for visualization purposes) each voxel according to its BF, that is, the condition that evoked the strongest response. Attentional modulation was assessed from the attention experiment as the ratio between the BOLD response in the given auditory attention condition and the BOLD response in the visual control condition, expressed on a dB scale. Thus, positive values of this quantity reflect BOLD response enhancement, whereas negative values reflect BOLD response attenuation, relative to nonauditory (i.e., visual) attention in the absence of auditory stimulus changes.

The aim of the first analysis was to characterize the detailed topography of attentional modulation in the superior temporal

cortex (Fig. 4A top). Analogous to the sensory BF maps described above, attended BF maps were obtained by extracting cortical maps of attentional modulation for each auditory attention condition and then labeling each voxel according to its BF, that is, the condition that evoked the strongest positive attentional modulation. To assess the congruency between sensory frequency representations and attentional modulations, a map-similarity index was defined. First, the proportion of voxels showing the same BF in both maps was extracted (referred to as observed proportion). Secondly, a null proportion resembling chance level was extracted after permuting (i.e., distorting the topography of) the sensory BF map; this step was repeated 5000 times. Thirdly, the similarity index was computed as the observed proportion minus the average of the permutation-based null proportions. Finally, this index was compared statistically against zero, thereby testing the congruency between the sensory and attended BF maps. These steps were applied separately to each ROI.

BOLD Response Modulation Analysis

The aim of the second analysis was to characterize the effect of attentional modulation on the shape of the frequency response of the superior temporal cortex, that is, the strength, nature (enhancement vs. attenuation), and frequency selectivity of attentional modulation in this region (Fig. 5A). First, spatially averaged, sensory BF channels were defined from the obtained sensory BF map by pooling voxels exhibiting the same BF. Secondly, for each of the obtained BF channels (or “sub-ROIs”) and each auditory attention condition, attentional modulation was extracted. Thirdly, BF-channel \times auditory attention condition combinations were averaged that were associated with the same spectral distance, a measure describing the mismatch between the channel's BF and the attended frequency (Atiani et al. 2009). Finally, frequency selectivity was tested by assessing attentional modulation as a function of spectral distance, that is, whether attentional modulation scaled with how well the channel's BF matched the attended frequency. These steps were applied separately to each ROI.

Multivoxel Pattern Analysis

In the third analysis, the congruency between sensory and attended frequency representations was further assessed with multivoxel pattern analysis (MVPA), an approach that does not rely on BF alone and is more sensitive than conventional univariate (voxel-by-voxel) analyses as it takes into account intervoxel dependencies (Mur et al. 2009). Single-trial data from the stimulation experiment were used for classifier training, that is, for estimating the function that mapped the BOLD response patterns to the corresponding frequency-specific conditions. Single-trial data from the attention experiment was used for consecutive testing, that is, to assess the generalization performance of the trained classifier. In addition to these between-experiment classifications, conventional within-experiment classifications were conducted based on a leave-one-run-out procedure. The chance level was estimated empirically as the median of 200 accuracy values obtained iteratively after shuffling the frequency labels on each iteration. The classifier used was a support vector machine with a linear kernel and a penalty parameter $C = 1$ (Chang and Lin 2011). MVPA was applied to the whole set of voxels within each ROI (i.e., no search-light approach).

Outcomes from the decoding were summarized as the percentage of correctly decoded test trials (an overall accuracy measure) and d' (a decision criterion-free discrimination performance measure). Contingency tables (Fig. 6C) were built by counting for

each auditory attention condition (rows) the number of trials for which the classifier reported a specific frequency (columns). For display purposes only, this number was expressed as a percentage of the row sum (i.e., of the total number of trials in the attention condition) to correct for potentially unbalanced trial numbers induced by rejecting trials containing attentional lapses. d' was computed for each frequency by first extracting the table column associated with that frequency and then taking the cell on the table diagonal as the number of hits and the sum of the 2 other cells within that column as the number of false alarms. Overall, d' was computed from the sums of the frequency-specific hits and false alarms.

Control Experiment Outside MR Scanner

A behavioral control experiment and fMRI debriefings were conducted with the aim to rule out the possibility that participants achieved their good performance in the attention experiment during fMRI by using mental strategies other than paying sustained selective attention to the cued frequency band (e.g., by restricting their attention only to the end of the stimuli [Thompson et al. 2011]). We recruited 10 control participants who matched our fMRI participants in terms of age and gender as much as possible. They underwent the same initial training session, but then performed the gap detection task in 2 blocks of 54 trials, with 1 block involving no attentional cue. Blocks and trials within blocks were presented in random order. Statistical group comparisons were done using the Wilcoxon rank-sum tests based on the W -statistic.

Results

Behavioral Results

Figure 2A shows behavioral results from the stimulation experiment: Listeners were able to judge pitch almost perfectly (overall accuracy: $98.0 \pm 0.4\%$, d' : 4.52 ± 0.1), with no significant difference between frequency bands [$\chi^2(2) = 3.62$, $P = 0.16$]. Figure 2B shows results from the attention experiment: Listeners were able to detect the gap with high accuracy in all bands (overall accuracy: $90.0 \pm 1.8\%$, d' : 2.34 ± 0.2), with higher sensitivity for the flanker bands than the middle band (low vs. middle: $W = 117$, $z = 3.24$, corrected $P = 0.0036$; middle vs. high: $W = 19$, $z = -2.33$, corrected $P = 0.060$). Overall performance differed substantially across

the 2 experiments (difference in overall d' : 2.18 ± 0.22 , $W = 120$, $z = 3.38$, $P = 0.00036$), indicating that the task in the attention experiment was much more demanding.

To corroborate that listeners in the attention experiment indeed paid selective attention to the frequency band specified by the attentional cue, we repeated this experiment outside the MR scanner in another group of listeners while varying the availability of the cue. Consistent with outcomes from uncertain-frequency detection experiments (Creelman 1960), results in Figure 2C show that when listeners were kept naïve about the target frequency band by not providing them with an attentional cue, they were overall much less able to detect a gap in that band ($W = 55$, $z = 2.75$, corrected $P = 0.0089$; comparison with the fMRI group: $W = 63$, $z = 3.14$, corrected $P = 0.0051$). These results, together with listeners' written introspective reports from debriefings, imply that they achieved their good performance in the fMRI experiment by exploiting the attentional cue and paying sustained selective attention to the target frequency band. In the control task, detection performance was $75.8 \pm 2.9\%$, suggesting that participants paid attention to the visual stimuli in this condition.

Definition of Primary and Nonprimary Auditory Cortical Regions

To enable investigation of attentional modulation at distinct auditory cortical processing stages, we first defined primary AC as the middle and medial portions of Heschl's gyrus (hereafter referred to as PAC), and nonprimary AC as *plana polare et temporale*, that is, as the superior temporal cortex excluding PAC (hereafter referred to as STC-); see Materials and Methods. The resulting 2 nonoverlapping ROIs are shown in Figure 3B and Supplementary Figure 1. Figure 3C illustrates the degree to which each ROI separated the cytoarchitectonically defined probabilistic primary area Te1 (Morosan et al. 2001) from the nonprimary superior temporal cortex for a range of probability thresholds. The latter parameter quantifies the reliability of area Te1 to resemble primary AC across individual brains, with larger values indicating a more confined and more reliable area Te1 (Rademacher et al. 2001). Results show that PAC was a significantly better identifier of Te1 than STC—at all probability thresholds (all corrected $P < 0.0034$), suggesting that the obtained ROIs indeed resembled distinct processing stages.

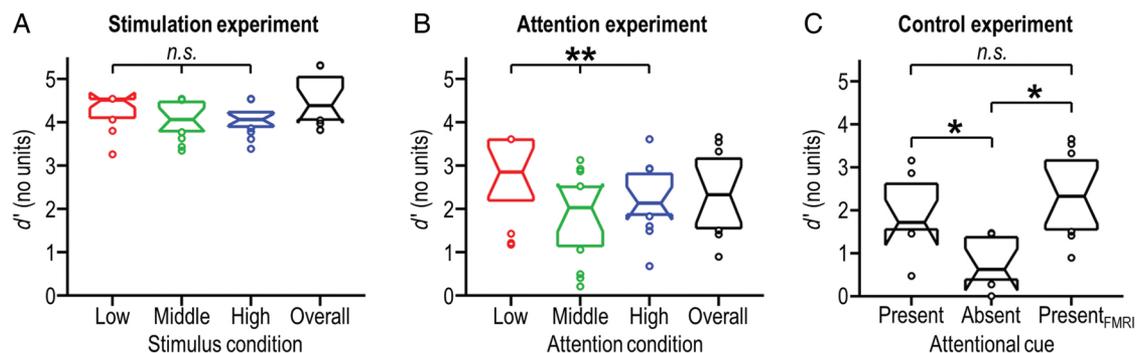


Figure 2. Behavioral results. The boxplots show listeners' average pitch judgment performance (A) and gap detection (B) performance inside the MR scanner, for each frequency band (red, green, blue) and pooled across bands (black). (C) Overall gap detection performance outside the MR scanner when the attentional cue was present or absent, and for reference inside the scanner when the cue was present (same as B rightmost). These results indicate that listeners could effortlessly identify pitch in the stimulation experiment and paid selective attention to the frequency band that was specified by the attentional cue in the attention experiment. * $P < 0.01$, ** $P < 0.001$, n.s., nonsignificant. Horizontal lines depict lower quartile, median, and upper quartile (from bottom to top); circles represent individual points outside this range; and notches indicate 95% confidence intervals about the median.

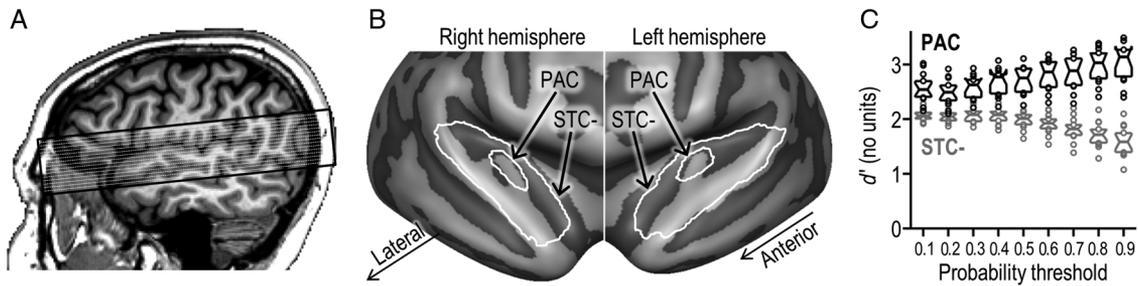


Figure 3. Definition of primary and nonprimary auditory cortical regions. The rectangle in (A) illustrates the positioning of the fMRI slices on an anatomical image from 1 participant. The white outlines in (B) illustrate, for each cerebral hemisphere, the locations of the 2 nonoverlapping ROIs (small region: PAC, large region excluding PAC: STC-) on an inflated representation of the cortical gray–white matter boundary averaged across participants. Single-subject data are provided in [Supplementary Figure 1](#). The boxplots in (C) show that on average, PAC (black) was a significantly better identifier of the cytoarchitecturally defined probabilistic primary area Te1 (Morosan et al. 2001) than STC- (gray), irrespective of the probability threshold used to define Te1. This indicates that PAC was more representative of primary AC than STC-.

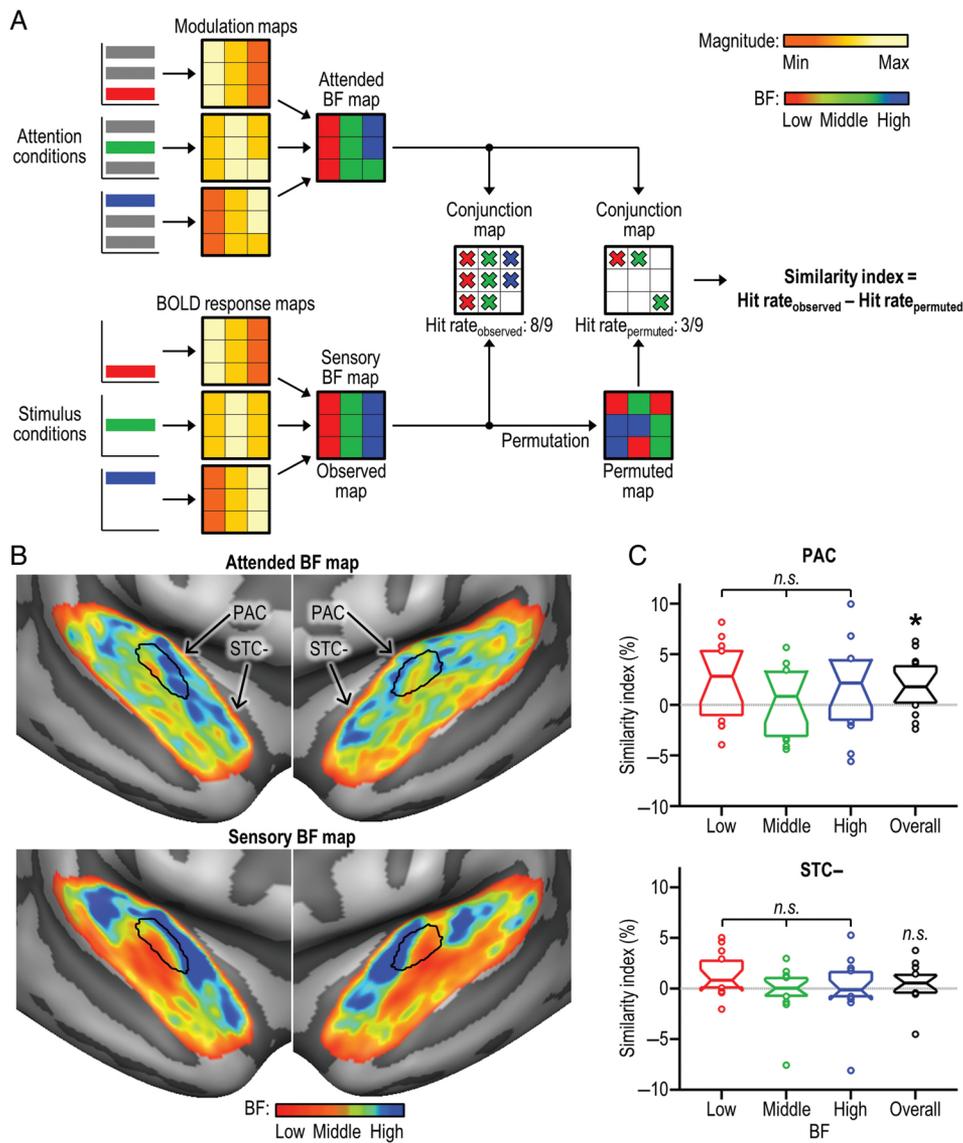


Figure 4. Topography of BOLD response enhancement induced by frequency-selective attention. (A) Our approach for investigating the detailed topography of attentional enhancement induced by frequency-selective attention. (B and C) Outcomes averaged across participants. Single-subject data are provided in [Supplementary Figure 2](#). (B) (Upper part) The topography of attentional BOLD response enhancement in the left and right superior temporal cortex. Analogously, (B) (lower part) illustrates the topography of observed conventional sensory BF representations. In both cerebral hemispheres, attentional modulation exhibited a topography that matched well the tonotopic organization of AC identified by the sensory BF representations (compare upper and lower parts of B). Boxplots in (C) that the similarity index (a measure of overlap between 2 BF maps, explained in A) was significantly larger than zero in PAC (upper plot) but not in STC- (lower plot), indicating that the topographies of attentional enhancement and sensory BF representation in primary AC were congruent well above chance. These results show that BOLD response enhancement induced by frequency-selective attention follows the detailed tonotopic layout of stimulus frequency-driven BOLD responses in primary AC. * $P < 0.05$, n.s., nonsignificant.

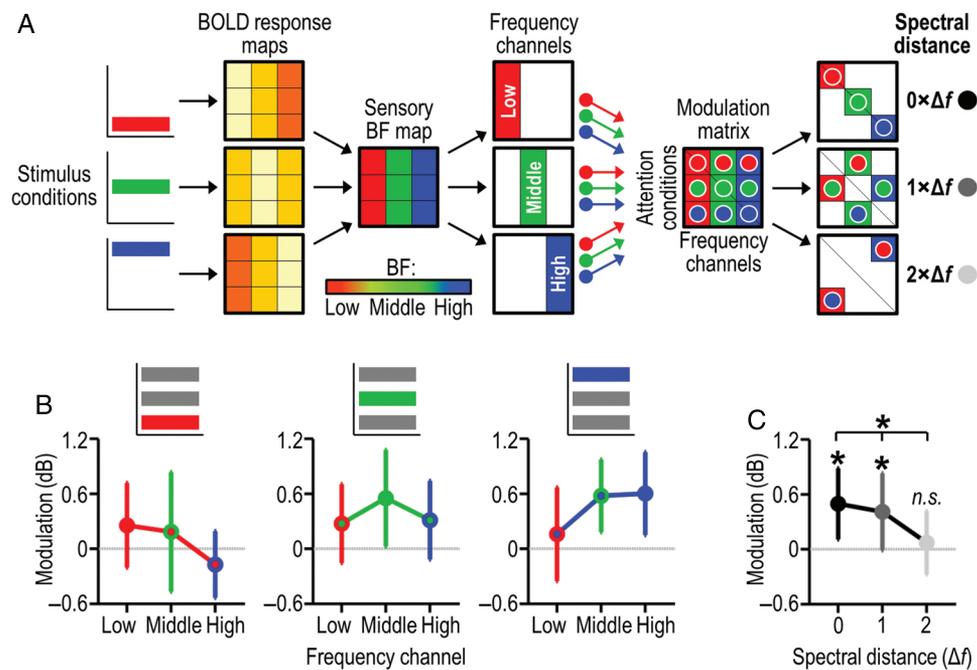


Figure 5. Frequency selectivity of attentional BOLD response modulation. (A) Our approach for investigating the nature, strength, and frequency selectivity of attentional modulation in sensory BF channels. (B and C) Outcomes (median \pm SEM) averaged across participants for region PAC. Outcomes for STC- are shown in [Supplementary Figure 3](#). In (B), attentional BOLD response modulation is plotted as a function of the BF channel (a voxel population responding preferentially and uniformly to a specific sensory frequency), separately for each auditory attention condition (i.e., for each row of the modulation matrix in A). In (C), attentional modulation is plotted as a function of spectral distance, a measure quantifying how much the channel's BF matches the attended frequency. These results show a frequency-specific reshaping of the frequency response of primary AC by frequency-selective attention, revealing frequency-selective BOLD response enhancements of up to approximately 0.6 dB and no significant suppression relative to nonauditory attention. * $P < 0.05$, n.s., nonsignificant.

Topography of BOLD Response Enhancement Induced by Frequency-Selective Attention

Having established a primary region (PAC) and a predominantly nonprimary region (STC-), we next conducted 3 analyses aimed to characterize the link between frequency-selective attention and sensory frequency representations at each putative processing stage.

In the first analysis, we obtained the detailed topography of subject-specific cortical BOLD response enhancements induced by frequency-selective attention and tested whether this topography matched the sensory-driven tonotopic layout of superior temporal cortex (Fig. 4A and Materials and Methods). As described in the Materials and Methods section, we obtained BF maps from the attention experiment to visualize spatial variations in maximum attentional enhancement that were induced by shifts in frequency-selective attention, not acoustic input changes. Figure 4B (top) shows the resulting attended BF maps averaged across listeners; for single-subject data, see [Supplementary Figure 2](#). The maps revealed multiple delimited regions: patches neighboring PAC in the anterior and posterior direction (in *plana polare et temporale*) typically exhibited the strongest modulation when the listener focused on the high frequency, whereas patches neighboring PAC in the lateral direction (anterolateral Heschl's gyrus and superior temporal gyrus) showed the strongest modulation for attention to the lower frequencies. This average spatial layout was visible in both cerebral hemispheres. Overall, it was similar to the topography of conventional sensory BF maps obtained from the stimulation experiment (Fig. 4B bottom), which revealed a clear tonotopic organization in PAC and nearby regions as expected (for a review, see [Saenz and Langers 2014](#)).

To validate this informal observation, we assessed the congruency of attended and sensory BF maps within each ROI using the similarity index (Fig. 4A and Materials and Methods). We observed that the average proportion of PAC voxels exhibiting a congruent BF was $34.8 \pm 0.74\%$, whereas for STC- voxels, this proportion was $33.2 \pm 0.54\%$. Results in Figure 4C show that across listeners, the similarity index was indeed reliably larger than zero, indicating above-chance topographic congruency in PAC (upper plot: similarity index $1.83 \pm 0.67\%$, $W = 99$, $z = 2.19$, $P = 0.014$), but not in STC- (lower plot: similarity index $0.43 \pm 0.48\%$, $W = 83$, $z = 1.28$, $P = 0.10$), with no significant difference between ROIs ($W = 87$, $z = 1.51$, $P = 0.066$) or between BF regions within each ROI [PAC: $\chi^2(2) = 3.73$, $P = 0.15$; STC-: $\chi^2(2) = 3.60$, $P = 0.17$]. In sum, these spatial results show that BOLD response enhancement induced by frequency-selective attention in a scene follows the detailed tonotopic layout of sensory frequency-evoked BOLD responses in primary AC.

Frequency Selectivity of Attentional BOLD Response Modulation

While the results so far detail the spatial cortical layout of attentional enhancement induced by frequency-selective attention in a scene, they provide only limited information about the effect of paying selective attention to (or away from) a particular frequency band on the shape of the cortical frequency response. To address this issue, we assessed in the second analysis attentional modulation within spatially averaged BF channels (Fig. 5A and Materials and Methods) as a function of the attended frequency and the channel's BF.

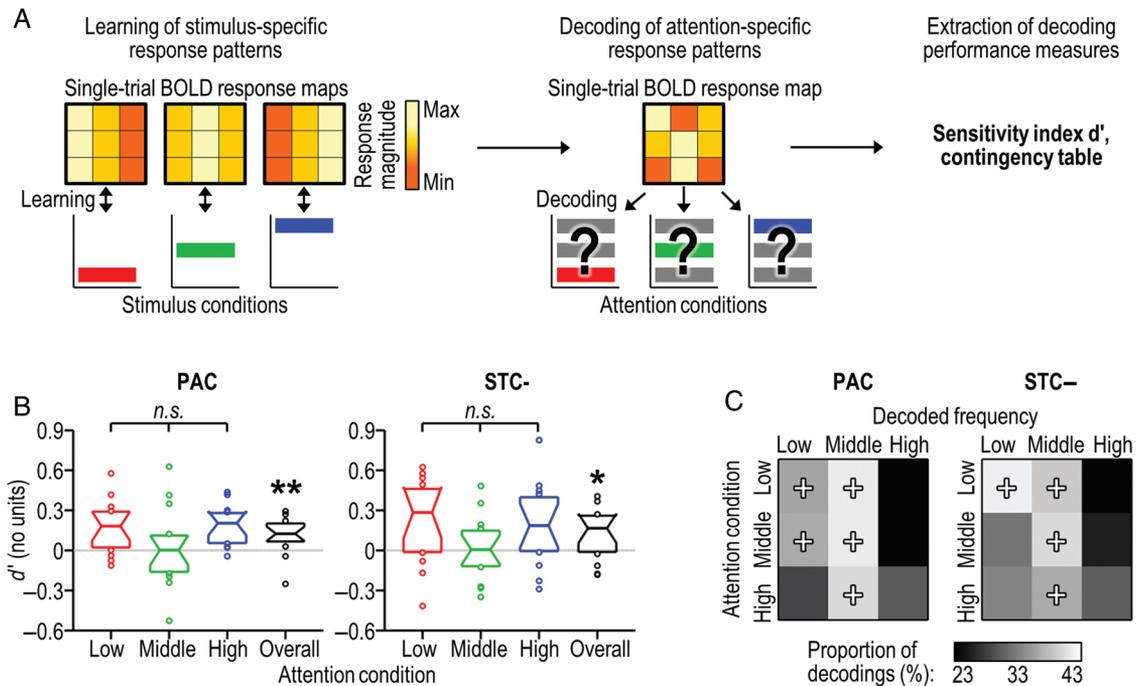


Figure 6. Informational content of cortical frequency representations. (A) Our MVPA approach for investigating the cortical encoding of the listener's focus of frequency-selective attention based on spatially distributed sensory frequency representations. (B and C) Outcomes averaged across participants for both ROIs (left: PAC, right: STC-). The boxplots in (B) summarize the classifier's unbiased decoding performance as assessed with d' for each attended frequency band (red, green, blue) and pooled across bands (black), relative to the chance level (gray horizontal line). These results show that, after learning to identify sensory frequency representations (i.e., the BOLD response patterns evoked by the different sensory frequencies), the classifier could successfully decode the flanker frequency that the listener was paying attention to, based on the accompanying BOLD response patterns in each PAC and STC-. The contingency tables in (C) illustrate for each auditory attention condition (rows) the proportion of trials (median across participants) for which the classifier decoded a specific frequency (columns) from PAC (left table) and STC- (right table), using a grayscale. Cells where the proportion of decodings was above chance (>33.33%) are indicated by crosses. The tables reveal that the decoding was biased toward the middle band. In sum, these results indicate that sensory frequency representations in primary AC and nonprimary AC, as assessed with fMRI and MVPA, encode the frequency that the listener is paying selective attention to. * $P < 0.05$, ** $P < 0.005$, n.s., nonsignificant.

Results in Figure 5B show that auditory attention-induced modulations of BF channels in PAC were predominantly positive, indicating BOLD response enhancement rather than attenuation, relative to nonauditory (visual) attention. The frequency response of PAC (i.e., the function relating attentional modulation to the channel's BF) revealed a characteristic peak for the low, middle, and high BF channel under auditory attention to the low, middle, or high frequency band (respectively), indicating a frequency-selective attentional reshaping of the frequency response of human primary AC as observed previously based on a smaller number of target frequencies (Da Costa et al. 2013).

To verify this informal observation, we assessed attentional modulation as a function of spectral distance, a measure quantifying the mismatch between the channel's BF and the attended frequency (Fig. 5A and Materials and Methods). Results in Figure 5C show that spectral distance had a systematic effect on attentional modulation [$\chi^2(2) = 10.13$, $P = 0.0063$]: channels tuned to the attended frequency underwent the strongest modulation, on average 0.58 dB. Modulation of these "attended channels" was significantly larger than zero ($0 \times \Delta f$: $W = 91$, $z = 1.73$, $P = 0.042$), and a similar effect was found for the adjacent channels ($1 \times \Delta f$: $W = 90$, $z = 1.68$, $P = 0.047$), but no such effect was found for more distant channels ($2 \times \Delta f$: $W = 68$, $z = 0.43$, $P = 0.34$). This frequency selectivity of attentional modulation was observed exclusively in PAC, not in STC- (see Supplementary Fig. 3A and B and Results). In sum, these spectral results show that frequency-selective attentional reshaping of the frequency response of human primary AC involves selective enhancement

of attended frequency-evoked BOLD responses by up to approximately 0.6 dB, but no significant suppression of unattended frequency-evoked BOLD responses, relative to nonauditory (visual) attention. This pattern is consistent with the operation of an attentional bandpass filter in human primary AC (see Introduction) that uses amplification rather than attenuation.

Informational Content of Cortical Frequency Representations

While the results so far characterize the link between frequency-selective attention in a scene and sensory frequency representations in primary AC, they allow no firm conclusion regarding nonprimary AC, probably because the applied BF approach failed to capture the more complex neuronal frequency tuning commonly observed in this region (Moerel et al. 2012; Nourski et al. 2014; Santoro et al. 2014) (see Supplementary Fig. 3C and Results). To circumvent this issue, we assessed frequency representations in our third analysis with MVPA, an approach that does not depend on single BFs alone (Fig. 6A and Materials and Methods). We first trained a classifier to discriminate sensory frequency representations, defined here as the spatially distributed BOLD response patterns that we observed in the different stimulation experiment conditions. We then tested whether the trained classifier could successfully decode the listener's focus of frequency-selective attention from attended frequency representations, defined here as the patterns observed in the different auditory attention conditions of the attention experiment.

Results from this between-experiment classification show that the classifier could successfully decode the attended frequency from the measured BOLD response patterns, not only in PAC (which was expected based on our BF-based results), but also in STC-. Averaged across listeners, the observed decoding accuracy was $36.36 \pm 0.87\%$ (PAC) and $37.37 \pm 1.26\%$ (STC-). As in many previous MVPA studies, these overall accuracies differed much less from the chance level (33.33%) compared with listeners' behavioral performance; yet above-chance decoding performance could be reliably observed across listeners: Statistical group results in Figure 6B show that the classifier's unbiased decoding performance was significantly above chance in each ROI (PAC: overall d' : $0.11 \pm 0.04 > 0$, $W = 106$, $z = 2.58$, $P = 0.0049$; STC-: overall d' : $0.13 \pm 0.05 > 0$, $W = 99$, $z = 2.18$, $P = 0.014$) with no significant difference between ROIs ($W = 50$, $z = -0.19$, $P = 0.57$). Closer inspection of decoding of individual frequency bands revealed that these overall above-chance results were dominated by the flanker bands (low or high band), that is, the classifier was less sensitive (it made more false alarms and/or misses) for the middle band than the flanker bands. Although this difference in unbiased decoding performance did not reach statistical significance [PAC: $\chi^2(2) = 5.73$, $P = 0.057$; STC-: $\chi^2(2) = 2.53$, $P = 0.28$], further analysis of decoding accuracies (Fig. 6C) revealed that the classifier misclassified selective attention to the flanker bands significantly more often as attention to the adjacent (middle) band than as attention to the other flanker band in each ROI (PAC: difference = 23.0 ± 4.4 trials, $W = 102$, $z = 3.077$, $P = 0.0010$; STC-: difference = 18.2 ± 4.1 trials, $W = 102$, $z = 3.077$, $P = 0.0010$). This indicates that the classifier was biased, that is, it had a strong tendency to report specifically the middle band, implying that its low sensitivity to this band (see unbiased decoding performance) reflects primarily an inflated number of false alarms.

In sum, these results show that frequency-selective attention in a scene induces spatially distributed BOLD response patterns congruent with those evoked by the corresponding sensory frequency alone, not only in primary AC, but also in nonprimary AC. They further indicate that more distant attended auditory frequencies are associated with more discriminable cortical representations, as observed in previous classification studies on spatially distributed sensory frequency representations in human superior temporal cortex (Nourski et al. 2014; Schonwiesner et al. 2015). Noteworthy, we also observed significant above-chance decoding performance in conventional within-experiment classifications (i.e., when data used for training and decoding were limited to a single experiment; see [Supplementary Results](#)), underscoring the validity of our decoding approach.

Discussion

Our fMRI results characterize the link between frequency-selective attention in an auditory scene and frequency representations throughout the human superior temporal cortex. First, the spatial results show that frequency-selective attentional BOLD response enhancement follows the tonotopic organization of sensory frequency representations in human primary AC. Secondly, the spectral results underscore that frequency-selective attentional BOLD response modulation reshapes the frequency response of primary AC analogous to an attentional bandpass filter primarily based on amplification. Finally, the decoding results reveal that frequency-selective attentional BOLD response modulation enforces frequency representations not only in primary AC, but also in higher auditory areas.

Tonotopy of Frequency-Selective Attentional Enhancement

By investigating attentional modulation throughout the entire superior temporal cortex, across individual voxels, and across 3 target frequencies, we were able to show that attentional modulation follows the detailed tonotopy of human primary AC. This finding is consistent with previous animal electrophysiology findings (Atiani et al. 2009) and extends human fMRI studies observing this spatial congruency with lower spatial detail (i.e., within large pooled voxel populations) (Paltoglou et al. 2009; Da Costa et al. 2013; Oh et al. 2013). Our observation of cortical tonotopy of attentional modulation has at least 3 implications: First, it extends animal findings showing that frequency-selective attention modulates frequency-specific AC neuronal populations and their receptive fields (Fritz, Elhilali, David, et al. 2007; Fritz, Elhilali, Shamma, 2007; Lakatos et al. 2013; O'Connell et al. 2014) to the human AC (Da Costa et al. 2013). Secondly and relatedly, it supports the more general notion that feature-selective auditory attention modulates the representation of attended sensory features in human AC (Ahveninen et al. 2006; Krumbholz et al. 2007; Altmann et al. 2008; Paltoglou et al. 2009; Woods et al. 2009; Da Costa et al. 2013; Oh et al. 2013; Lee et al. 2014) rather than constituting an independent processing stage (Näätänen 1992; Petkov et al. 2004). Finally, it indicates that frequency-selective attention may act as a spatial filter in early processing stages in human AC, that is, as an attention-controlled, large "spotlight" directed toward the specific tonotopic frequency region that is relevant for the current task, as proposed earlier based on animal findings (Fritz et al. 2010). This idea parallels fMRI findings on visuospatial attention in retinotopic maps of the primary visual cortex (Tootell et al. 1998; Brefczynski and DeYoe 1999), suggesting a generic mechanism for feature-selective attention in human primary topographic cortex.

Using a voxel-wise BF approach in nonprimary AC, we initially failed to replicate our spatial result in primary AC, although visual inspection suggested some congruency not only in our primary region but also in immediately neighboring regions. Based on supplementary frequency-selectivity analyses (see [Supplementary Fig. 3C and Results](#)), this initial null result can be attributed to a failure of our BF approach to capture frequency representations in nonprimary AC. Neuronal populations in this higher order region typically exhibit more complex frequency tuning curves (Moerel et al. 2012; Nourski et al. 2014; Santoro et al. 2014) that may have captured more than single "best" frequency band in our attention experiment, a point on which we elaborate below.

Attentional Bandpass Filtering in Human Primary AC

Together with previous studies in animals (Fritz et al. 2003, 2005; Fritz, Elhilali, David, et al. 2007; Fritz, Elhilali, Shamma, 2007; Atiani et al. 2009; Lakatos et al. 2013; O'Connell et al. 2014) and humans (Da Costa et al. 2013), our results on attentional reshaping provide converging support for the existence of an attentional bandpass filter that operates on tonotopic frequency representations at a primary cortical processing stage or possibly earlier stages (Suga et al. 1997; Perrot et al. 2006). Compared with related human fMRI studies, our spectral results build on a larger sample and a more rigorous approach that prevents potentially confounding effects of attention to irrelevant sound features or scanner noise. Moreover, our results reveal the frequency response of the proposed filter based on a larger number of target frequencies and attentional states, highlighting that attentional filtering of

sound-evoked BOLD responses in human primary AC can take the form of a bandpass filter (rather than a low- or high-pass filter alone, as was shown before) that operates primarily based on amplification.

Our bandpass filter interpretation builds on our observation that attentional modulation is a function of spectral distance in primary AC: representations of frequencies that were closer to the attended frequency underwent stronger attentional BOLD response enhancement, as observed in animal electrophysiology work (Atiani et al. 2009). Noteworthy, we observed enhancement of frequency representations as far as 1.9 octaves away from the attended frequency, implying that the proposed filter integrates several critical bands. This fairly broad tuning could reflect that listeners occasionally failed to completely ignore the outputs of peripheral auditory filters centered off the target frequency (Moore et al. 1996) when attempting to segregate the target frequency. While we analyzed only trials with correct responses, such presumed failures might have resulted in attentional comodulation of nontarget frequency representations (Alain and Arnott 2000; Shinn-Cunningham 2008), especially for the frequencies immediately neighboring the target frequency, because perceptual segregation of more nearby frequencies tends to build up more slowly (Bregman 1978; Carlyon et al. 2001). In this view, the observed frequency-selective attentional BOLD response enhancement could hint to a neural basis for auditory stream segregation based on spectral cues (Wilson et al. 2007; Hill et al. 2011; Wiegand and Gutschalk 2012).

Indeed, the attentional modulation patterns observed here and elsewhere (Bidet-Caulet, Fischer, Bauchet, et al. 2007; Bidet-Caulet, Fischer, Besle, et al. 2007; Xiang et al. 2010) fit the notion of a neural contrast-matched filter in human AC. According to this model originally derived from animal studies (Fritz et al. 2003, 2005; Fritz, Elhilali, David, et al. 2007; Fritz, Elhilali, Shamma, 2007), feature-selective attention in a scene operates by enhancing the cortical contrast between the attended feature and unattended features to facilitate perceptual segregation. More recent findings have indicated that this contrast may arise from a counter-phase oscillatory entrainment across differently tuned primary AC regions, that is, representations of an attended/unattended frequency may synchronize/desynchronize their neuronal excitability with the rhythm of the attended sound (Lakatos et al. 2013; O'Connell et al. 2014; Kayser et al. 2015).

Our observation of a general BOLD response amplification in primary AC induced by auditory attention versus visual attention agrees with previous human neuroimaging findings (Woodruff et al. 1996; Grady et al. 1997; Jancke et al. 1999; Johnson and Zatorre 2005; Rinne et al. 2005; Degerman et al. 2006; Paltoglou et al. 2009) and motivates our interpretation that attentional bandpass filtering in this region relies mostly on amplification of sound-evoked BOLD responses, rather than attenuation. However, as discussed below (section Limitations), our fMRI results do not preclude that frequency-selective attention can involve suppression of neural responses to unattended frequencies (lateral inhibition or sideband suppression) when, for example, compared with passive listening or assessed with electrophysiology (Bidet-Caulet, Fischer, Bauchet, et al. 2007; Bidet-Caulet, Fischer, Besle, et al. 2007; Besle et al. 2011; Lakatos et al. 2013; O'Connell et al. 2014).

Frequency-Selective Attention Recruits Sensory Frequency Representations in Nonprimary AC

By exploiting the high sensitivity of MVPA, we were able to establish a link between frequency-selective attention and (spatially

distributed) frequency representations in human nonprimary AC. Similar to our spatial results reported above, our fMRI decoding results show that frequency-selective attention in a scene overlays AC with a frequency-specific modulation pattern, that is, it induces a (cortically distributed) BOLD response modulation pattern that matches the response pattern evoked by the attended sensory frequency alone. Importantly, our decoding approach enabled us to extend this outcome to nonprimary AC and thus emphasize that frequency representations at each processing stage of AC (primary and nonprimary) can encode the focus of frequency-selective attention. This novel evidence for the existence of frequency-selective attention modulation of frequency representations in human nonprimary AC lines up with previous human findings on feature-selective attentional modulation (Ahveninen et al. 2006; Krumbholz et al. 2007; Altmann et al. 2008) and spatially distributed sound/task representations (Formisano et al. 2008; Staeren et al. 2009; Bonte et al. 2014) in nonprimary AC.

Besides this conceptual advance, our study provides a possible methodological contribution to future studies of sensory feature representations and feature-selective attention. Our decoding approach enables not only to capture sensory feature representations arising from neuronal populations with complex (multi-peaked) feature tuning curves, but also to detect with high sensitivity whether these representations are modulated under feature-selective attention. This may render our approach an attractive alternative to conventional “best” (maximum) response-based approaches that fail to capture multi-peaked feature tuning curves.

Limitations

Our interpretation requires a few cautionary remarks. First, while we succeeded in visualizing and characterizing large-scale cortical patterns of attentional modulation based on the BOLD response, we cannot disambiguate whether these modulations reflect increased neuronal gain (Hillyard et al. 1973; Woldorff et al. 1993), increased neuronal selectivity (Fritz et al. 2003, 2005; Fritz, Elhilali, David, et al. 2007; Fritz, Elhilali, Shamma, 2007; Jaramillo and Zador 2011; David et al. 2012), neuronal oscillatory phase entrainment (Besle et al. 2011; Lakatos et al. 2013; O'Connell et al. 2014), or combinations thereof (Kauramaki et al. 2007; Okamoto et al. 2007, 2009; Ahveninen et al. 2011), because the measured BOLD signals likely reflect a mixture of excitatory and inhibitory neuronal activity (Logothetis and Wandell 2004). Further experimentation using our paradigm with more direct measures of neural activity will be necessary to resolve questions regarding underlying neuronal mechanisms. Moreover, although we compared individual ROIs with a standardized, cytoarchitecturally defined probabilistic map of primary AC, we cannot exclude that these ROIs contained a fraction of nonprimary tissue and primary tissue, respectively, as there exists currently no unequivocal definition of human primary AC in vivo (Moerel et al. 2014), or that the observed effects originate from lower level structures. In future studies, inclusion of additional individual anatomical constraints based on myeloarchitecture mapping and subcortical structures in the imaging volume (Dick et al. 2012; De Martino et al. 2015) may help reduce these ambiguities.

Conclusion

Our fMRI study shows that goal-oriented frequency-selective attention in an auditory scene may operate by enforcing the fine-grained large-scale spatial activity pattern that would be evoked

throughout the entire superior temporal cortex by the attended frequency if that frequency was present alone. Spatially distributed frequency representations at multiple (tonotopic and nontotopic) AC processing stages appear to play pivotal roles in this regard, contributing to auditory phenomena such as selective perception of audio frequencies, the “listening band,” and perceptual segregation of sounds in auditory scenes. An interesting question for further research is how primary and nonprimary frequency representations are modulated under selective attention to more complex sounds, such as speech, that contain multiple salient frequencies and preferentially activate nonprimary AC regions.

Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Funding

This work was supported by the Netherlands Organization for Scientific Research (Veni grant 451-11-014 to L.R.).

Notes

We thank Fren Smulders for useful discussions, Johannes Mehrer for help with the behavioral training, and three anonymous reviewers for help with improving the manuscript. *Conflict of Interest*: None declared.

References

- Ahveninen J, Hamalainen M, Jaaskelainen IP, Ahlfors SP, Huang S, Lin FH, Raji T, Sams M, Vasios CE, Belliveau JW. 2011. Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc Natl Acad Sci USA*. 108:4182–4187.
- Ahveninen J, Jaaskelainen IP, Raji T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, et al. 2006. Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci USA*. 103:14608–14613.
- Alain C, Arnott SR. 2000. Selectively attending to auditory objects. *Front Biosci*. 5:D202–D212.
- Altmann CF, Henning M, Doring MK, Kaiser J. 2008. Effects of feature-selective attention on auditory pattern and location processing. *Neuroimage*. 41:69–79.
- Atiani S, Elhilali M, David SV, Fritz JB, Shamma SA. 2009. Task difficulty and performance induce diverse adaptive patterns in gain and shape of primary auditory cortical receptive fields. *Neuron*. 61:467–480.
- Besle J, Schevon CA, Mehta AD, Lakatos P, Goodman RR, McKhann GM, Emerson RG, Schroeder CE. 2011. Tuning of the human neocortex to the temporal dynamics of attended events. *J Neurosci*. 31:3176–3185.
- Bidet-Caulet A, Fischer C, Bauchet F, Aguera PE, Bertrand O. 2007. Neural substrate of concurrent sound perception: direct electrophysiological recordings from human auditory cortex. *Front Hum Neurosci*. 1:5.
- Bidet-Caulet A, Fischer C, Besle J, Aguera PE, Giard MH, Bertrand O. 2007. Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J Neurosci*. 27:9252–9261.
- Bonte M, Hausfeld L, Scharke W, Valente G, Formisano E. 2014. Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J Neurosci*. 34:4548–4557.
- Brefczynski JA, DeYoe EA. 1999. A physiological correlate of the “spotlight” of visual attention. *Nat Neurosci*. 2:370–374.
- Bregman AS. 1978. Auditory streaming is cumulative. *J Exp Psychol Hum Percept Perform*. 4:380–387.
- Brown GS, White KG. 2005. The optimal correction for estimating extreme discriminability. *Behav Res Methods*. 37:436–449.
- Campain R, Minckler J. 1976. A note on the gross configurations of the human auditory cortex. *Brain Lang*. 3:318–323.
- Carlyon RP, Cusack R, Foxton JM, Robertson IH. 2001. Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform*. 27:115–127.
- Chang C, Lin C. 2011. LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol*. 2:1–39.
- Creelman CD. 1960. Detection of signals of uncertain frequency. *J Acoust Soc Am*. 32:805–810.
- Da Costa S, van der Zwaag W, Miller LM, Clarke S, Saenz M. 2013. Tuning in to sound: frequency-selective attentional filter in human primary auditory cortex. *J Neurosci*. 33:1858–1863.
- David SV, Fritz JB, Shamma SA. 2012. Task reward structure shapes rapid receptive field plasticity in auditory cortex. *Proc Natl Acad Sci USA*. 109:2144–2149.
- De Martino F, Moerel M, Xu J, van de Moortele PF, Ugurbil K, Goebel R, Yacoub E, Formisano E. 2015. High-resolution mapping of myeloarchitecture in vivo: localization of auditory areas in the human brain. *Cereb Cortex*. 25:3394–3405.
- Degerman A, Rinne T, Salmi J, Salonen O, Alho K. 2006. Selective attention to sound location or pitch studied with fMRI. *Brain Res*. 1077:123–134.
- Dick F, Tierney AT, Lutti A, Josephs O, Sereno MI, Weiskopf N. 2012. In vivo functional and myeloarchitectonic mapping of human primary auditory areas. *J Neurosci*. 32:16095–16105.
- Elhilali M, Xiang J, Shamma SA, Simon JZ. 2009. Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biol*. 7:e1000129.
- Formisano E, De Martino F, Bonte M, Goebel R. 2008. “Who” is saying “what”? Brain-based decoding of human voice and speech. *Science*. 322:970–973.
- 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.
- Fritz J, Shamma S, Elhilali M, Klein D. 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*. 6:1216–1223.
- Fritz JB, David SV, Radtke-Schuller S, Yin P, Shamma SA. 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. *Nat Neurosci*. 13:1011–1019.
- Fritz JB, Elhilali M, David SV, Shamma SA. 2007. Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear Res*. 229:186–203.
- Fritz JB, Elhilali M, Shamma SA. 2007. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J Neurophysiol*. 98:2337–2346.
- Fritz JB, Elhilali M, Shamma SA. 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci*. 25:7623–7635.
- Grady CL, Van Meter JW, Maisog JM, Pietrini P, Krasuski J, Rauschecker JP. 1997. Attention-related modulation of activity

- in primary and secondary auditory cortex. *Neuroreport*. 8:2511–2516.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW. 1999. “Sparse” temporal sampling in auditory fMRI. *Hum Brain Mapp*. 7:213–223.
- Hall DA, Haggard MP, Akeroyd MA, Summerfield AQ, Palmer AR, Elliott MR, Bowtell RW. 2000. Modulation and task effects in auditory processing measured using fMRI. *Hum Brain Mapp*. 10:107–119.
- Hill KT, Bishop CW, Yadav D, Miller LM. 2011. Pattern of BOLD signal in auditory cortex relates acoustic response to perceptual streaming. *BMC Neurosci*. 12:85.
- Hillyard SA, Hink RF, Schwent VL, Picton TW. 1973. Electrical signs of selective attention in the human brain. *Science*. 182:177–180.
- Humphries C, Liebenthal E, Binder JR. 2010. Tonotopic organization of human auditory cortex. *Neuroimage*. 50:1202–1211.
- Jancke L, Mirzazade S, Shah NJ. 1999. Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci Lett*. 266:125–128.
- Jaramillo S, Zador AM. 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat Neurosci*. 14:246–251.
- Johnson JA, Zatorre RJ. 2005. Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. *Cereb Cortex*. 15:1609–1620.
- Kauramaki J, Jaaskelainen IP, Sams M. 2007. Selective attention increases both gain and feature selectivity of the human auditory cortex. *PLoS One*. 2:e909.
- Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S. 2015. Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. *J Neurosci*. 35:7750–7762.
- Krumbholz K, Eickhoff SB, Fink GR. 2007. Feature- and object-based attentional modulation in the human auditory “where” pathway. *J Cogn Neurosci*. 19:1721–1733.
- Lakatos P, Musacchia G, O’Connell MN, Falchier AY, Javitt DC, Schroeder CE. 2013. The spectrotemporal filter mechanism of auditory selective attention. *Neuron*. 77:750–761.
- Langers DR, Sanchez-Panchuelo RM, Francis ST, Krumbholz K, Hall DA. 2014. Neuroimaging paradigms for tonotopic mapping (II): the influence of acquisition protocol. *Neuroimage*. 100:663–675.
- Lee AK, Larson E, Maddox RK, Shinn-Cunningham BG. 2014. Using neuroimaging to understand the cortical mechanisms of auditory selective attention. *Hear Res*. 307:111–120.
- Logothetis NK, Wandell BA. 2004. Interpreting the BOLD signal. *Annu Rev Physiol*. 66:735–769.
- Macmillan NA, Creelman CD. 1991. *Detection theory: a user’s guide*. Cambridge: Cambridge UP.
- Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*. 485:233–236.
- Moerel M, De Martino F, Formisano E. 2014. An anatomical and functional topography of human auditory cortical areas. *Front Neurosci*. 8:225.
- Moerel M, De Martino F, Formisano E. 2012. Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J Neurosci*. 32:14205–14216.
- Moore BC, Hafter ER, Glasberg BR. 1996. The probe-signal method and auditory-filter shape: results from normal- and hearing-impaired subjects. *J Acoust Soc Am*. 99:542–552.
- 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.
- Mur M, Bandettini PA, Kriegeskorte N. 2009. Revealing representational content with pattern-information fMRI—an introductory guide. *Soc Cogn Affect Neurosci*. 4:101–109.
- Näätänen R. 1992. *Attention and brain function*. Hillsdale (NJ): Lawrence Erlbaum.
- Nourski KV, Steinschneider M, Oya H, Kawasaki H, Howard MA III. 2015. Modulation of response patterns in human auditory cortex during a target detection task: an intracranial electrophysiology study. *Int J Psychophysiol*. 95:191–201.
- Nourski KV, Steinschneider M, Oya H, Kawasaki H, Jones RD, Howard MA. 2014. Spectral organization of the human lateral superior temporal gyrus revealed by intracranial recordings. *Cereb Cortex*. 24:340–352.
- O’Connell MN, Barczak A, Schroeder CE, Lakatos P. 2014. Layer specific sharpening of frequency tuning by selective attention in primary auditory cortex. *J Neurosci*. 34:16496–16508.
- Oh J, Kwon JH, Yang PS, Jeong J. 2013. Auditory imagery modulates frequency-specific areas in the human auditory cortex. *J Cogn Neurosci*. 25:175–187.
- Okamoto H, Stracke H, Wolters CH, Schmael F, Pantev C. 2007. Attention improves population-level frequency tuning in human auditory cortex. *J Neurosci*. 27:10383–10390.
- Okamoto H, Stracke H, Zwislerood P, Roberts LE, Pantev C. 2009. Frequency-specific modulation of population-level frequency tuning in human auditory cortex. *BMC Neurosci*. 10:1.
- Paltoglou AE, Sumner CJ, Hall DA. 2009. Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. *Hear Res*. 257:106–118.
- Perrot X, Ryvlin P, Isnard J, Guenot M, Catenoix H, Fischer C, Mauguere F, Collet L. 2006. Evidence for corticofugal modulation of peripheral auditory activity in humans. *Cereb Cortex*. 16:941–948.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL. 2004. Attentional modulation of human auditory cortex. *Nat Neurosci*. 7:658–663.
- 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.
- Rinne T, Pekkola J, Degerman A, Autti T, Jaaskelainen IP, Sams M, Alho K. 2005. Modulation of auditory cortex activation by sound presentation rate and attention. *Hum Brain Mapp*. 26:94–99.
- Saenz M, Langers DR. 2014. Tonotopic mapping of human auditory cortex. *Hear Res*. 307:42–52.
- Santoro R, Moerel M, De Martino F, Goebel R, Uğurbil K, Yacoub E, Formisano E. 2014. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol*. 10:e1003412.
- Scharf B, Quigley S, Aoki C, Peachey N, Reeves A. 1987. Focused auditory attention and frequency selectivity. *Percept Psychophys*. 42:215–223.
- 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.
- Shinn-Cunningham BG. 2008. Object-based auditory and visual attention. *Trends Cogn Sci*. 12:182–186.

- 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.
- 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.
- Suga N, Zhang Y, Yan J. 1997. Sharpening of frequency tuning by inhibition in the thalamic auditory nucleus of the mustached bat. *J Neurophysiol.* 77:2098–2114.
- 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.
- Thompson SK, Carlyon RP, Cusack R. 2011. An objective measurement of the build-up of auditory streaming and of its modulation by attention. *J Exp Psychol Hum Percept Perform.* 37:1253–1262.
- Tootell RB, Hadjikhani N, Hall EK, Marrett S, Vanduffel W, Vaughan JT, Dale AM. 1998. The retinotopy of visual spatial attention. *Neuron.* 21:1409–1422.
- Wiegand K, Gutschalk A. 2012. Correlates of perceptual awareness in human primary auditory cortex revealed by an informational masking experiment. *Neuroimage.* 61:62–69.
- Wilson EC, Melcher JR, Micheyl C, Gutschalk A, Oxenham AJ. 2007. Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J Neurophysiol.* 97:2230–2238.
- Woldorff MG, Gallen CC, Hampson SA, Hillyard SA, Pantev C, Sobel D, Bloom FE. 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc Natl Acad Sci USA.* 90:8722–8726.
- Woodruff PW, Benson RR, Bandettini PA, Kwong KK, Howard RJ, Talavage T, Belliveau J, Rosen BR. 1996. Modulation of auditory and visual cortex by selective attention is modality-dependent. *Neuroreport.* 7:1909–1913.
- Woods DL, Herron TJ, Cate AD, Yund EW, Stecker GC, Rinne T, Kang X. 2010. Functional properties of human auditory cortical fields. *Front Syst Neurosci.* 4:155.
- Woods 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.
- Xiang J, Simon J, Elhilali M. 2010. Competing streams at the cocktail party: exploring the mechanisms of attention and temporal integration. *J Neurosci.* 30:12084–12093.