Frequency-specific attentional modulation in human primary auditory cortex and midbrain

Lars Riecke a,*, Judith C. Peters a, b, Giancarlo Valente a, Benedikt A. Poser a, Valentin G. Kemper a, Elia Formisano a, Bettina Sorger a

a Department of Cognitive Neuroscience, Faculty of Psychology and Neuroscience, Maastricht University, Oxfordlaan 55, 6229, EV, Maastricht, The Netherlands
b Department of Vision and Cognition, Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences (KNAW), 1105, BA, Amsterdam, The Netherlands

A R T I C L E   I N F O

Keywords:
Auditory attention
Human primary auditory cortex
FMRI
Frequency tuning
Human inferior colliculus

A B S T R A C T

Paying selective attention to an audio frequency selectively enhances activity within primary auditory cortex (PAC) at the tonotopic site (frequency channel) representing that frequency. Animal PAC neurons achieve this ‘frequency-specific attentional spotlight’ by adapting their frequency tuning, yet comparable evidence in humans is scarce. Moreover, whether the spotlight operates in human midbrain is unknown.

To address these issues, we studied the spectral tuning of frequency channels in human PAC and inferior colliculus (IC), using 7-T functional magnetic resonance imaging (FMRI) and frequency mapping, while participants focused on different frequency-specific sounds. We found that shifts in frequency-specific attention alter the response gain, but not tuning profile, of PAC frequency channels. The gain modulation was strongest in low-frequency channels and varied near-monotonically across the tonotopic axis, giving rise to the attentional spotlight. We observed less prominent, non-tonotopic spatial patterns of attentional modulation in IC.

These results indicate that the frequency-specific attentional spotlight in human PAC as measured with FMRI arises primarily from tonotopic gain modulation, rather than adapted frequency tuning. Moreover, frequency-specific attentional modulation of afferent sound processing in human IC seems to be considerably weaker, suggesting that the spotlight diminishes toward this lower-order processing stage. Our study sheds light on how the human auditory pathway adapts to the different demands of selective hearing.

Introduction

Listening in natural auditory scenes typically involves multiple sound sources that simultaneously stimulate the ears. To effectively extract information about a given source of interest from the complex acoustic input, the auditory system relies on selective attention. This process involves the allocation of neural and cognitive resources to acoustic input features representing the target source. This prioritizes the processing of the aforementioned features and gives them a perceptual advantage over non-attended features. Which specific feature is selected depends on the relative saliency of the available features and the listener’s current goal: the listener’s focus of attention can be involuntarily attracted by highly salient acoustic input features (‘bottom-up’ or exogenous attention) or voluntarily allocated based on the listener’s current goals (‘top-down’ or endogenous attention). In everyday life, these two forces often compete: for example, when attempting to follow a conversation in a noisy environment, the listener pays endogenous selective attention to the voice of interest to overcome the distracting effect of irrelevant salient sounds (for reviews, see e.g. (Alain and Bernstein, 2008; Shinn-Cunningham, 2008)).

Human neuroimaging studies have shown that focusing endogenous selective auditory attention on a specific frequency in acoustic input (frequency-specific attention) selectively enhances the activity of neuronal populations in primary auditory cortex (PAC) that are tuned to this attended frequency, relative to populations tuned to the non-attended frequencies (Paltoglou et al., 2009; Da Costa et al., 2013; Oh et al., 2013; Riecke et al., 2017). Analogous to the visual system (Tootell et al., 1998; Brefczynski and DeYoe, 1999), this operation can be conceptualized as that of an ‘attentional spotlight’ that focuses afferent sound processing in a behaviorally relevant and frequency-specific manner so as to prioritize processing in primary cortical areas that are...
tuned to the frequency of interest. Despite the converging evidence for the existence of such a tonotopic attentional spotlight in human PAC, two important questions regarding the neural basis of endogenous frequency-specific auditory attention in the human brain remain unanswered.

First, does frequency-specific attention modulate the frequency tuning of human PAC? Animal electrophysiology studies have shown that PAC neurons implement an attentional spotlight mechanism by adapting their frequency response, i.e., their responsiveness to different audio frequencies (for a review, see (Fritz et al., 2007a)). More specifically, attention can modulate the overall strength of neurons’ frequency response (response gain) and/or the shape of this response (frequency tuning): Neurons whose preferred frequency lies closely to the attended frequency increase their responses to input (Atiani et al., 2009; Lakatos et al., 2013; O’Connell et al., 2014), compared with neurons tuned to more distant, unattended frequencies. In addition to this response gain, the ‘near’ neurons adapt (or ‘reshape’) their frequency tuning so as to increase and reduce their relative responsiveness to the attended and unattended frequencies, respectively (Fritz et al., 2007b; Atiani et al., 2009; David et al., 2012). In humans, it is still unclear whether attention affects response gain, frequency tuning, or both (Lee et al., 2014).

Invasive electrocorticography studies in neurological patients have found modulation of spectrotachymetric tuning in auditory cortex induced by variations in speech context and speaker-specific attention (Mesgarani and Chang, 2012; Holdgraf et al., 2016). Magnetoencephalography studies have observed attentional enhancement of tone-evoked long-latency auditory-evoked magnetic fields, which is thought to reflect modulation of neuronal-population frequency tuning (Kauramaki et al., 2007; Okamoto et al. 2007, 2009; Ahveninen et al., 2011). A recent functional magnetic resonance imaging (fMRI) study investigating the laminar organization of frequency tuning in human PAC finds that intermodal (auditory vs. visual) attention modulates the bandwidth of frequency tuning in the upper layers of PAC (De Martino et al., 2015). However, it remains unclear whether the attentional spotlight in the human brain emerges from modulation of response gain, frequency tuning, or both, as in animals.

Second, does frequency-specific attention modulate afferent sound processing in human auditory subcortical regions? Studies of frequency-specific attention have focused almost exclusively on cortical regions; however, recent electrophysiological recordings from the inferior colliculus (IC) of ferrets reveal attentional modulation of frequency tuning (Slee and David, 2015). This modulation is similar to that observed in PAC (Fritz et al., 2003; David et al., 2012), suggesting that frequency-specific attentional retuning in PAC is partially inherited to IC. Moreover, electric stimulation of PAC neurons has been shown to alter the frequency tuning of IC neurons (bats: (Zhang et al., 1997; Yan and Suga, 1998), mice: (Yan and Ehret, 2002), and gerbils (Sakai and Suga, 2001);) presumably via tonotopic corticofugal projections (cats: (Andersen et al., 1980; Winer et al., 1998); rats: (Herbert et al., 1991); ferret (Bajo et al., 2007)). In humans, findings of subcortical frequency-specific attentional modulation are relatively scarce (Galbraith et al., 1998; Maison et al., 2001; Lehmann and Schonwiesner, 2014; Yamagishi et al., 2016) and partially controversial (discussed in (Varghese et al., 2015)). Moreover, the locus of the reported subcortical attentional modulation remains unclear due to the limited spatial resolution of the applied methods (otoacoustic emission recordings and scalp electroencephalography).

In the present study, we addressed these unresolved questions by investigating frequency tuning in the human auditory pathway with ultra-high field fMRI during intramodal auditory attention tasks on target sounds with distinct frequencies. By measuring cortical and subcortical responses to various frequency-mapping stimuli under different states of endogenous frequency-specific attention, we could assess whether and how goal-directed frequency-specific attention shifts alter the frequency response of neuronal populations at different stages of the human auditory pathway. We demonstrate that the attentional spotlight in human PAC, as measured with fMRI, arises primarily from local alterations in overall response gain (i.e., elevation or reduction of the entire frequency response) rather than from frequency retuning (i.e., no shifts in characteristic frequencies or sharpening/widening of the frequency-tuning profile). We further provide results suggesting that frequency-specific attentional modulation can be observed with fMRI in human IC.

Materials and methods

Participants

Ten healthy volunteers (four females, ages: 21–39 years) participated in the study. They reported no history of hearing disorders, or vision or motor problems, and had normal hearing (defined as pure-tone hearing thresholds of less than 25 dB HL at 0.25, 0.5, 0.75, 1, 2, 3, 4, and 6 kHz), except for one participant (P7) who had slightly elevated threshold (30 dB HL) in one ear at 3 kHz. Another participant’s (P10) task performance was observed to approximate ceiling in the fMRI experiment (see Results). Excluding these participants’ data from the analysis did not alter the conclusions that can be drawn from the study. All participants gave written informed consent according to a protocol approved by the local research ethics committee (Ethical Review Committee Psychology and Neuroscience, Maastricht University) and received a monetary reimbursement for their participation.

Tasks

Auditory stimuli are described in Fig. 1A–B, and in more detail in the next section. The experimental manipulation of listeners’ focus of attention was implemented by varying the listener’s task. An active task served to direct frequency-specific attention to either a low-frequency auditory target stream (L-target) or a high-frequency target stream (H-target). A passive task served to either enable localization of functional brain regions of interest (see Definition of functional-anatomical regions of interest) or establish an attentional baseline. For the active task, listeners were instructed to attend to a pre-defined target stream (L-target or H-target) and count the number of gaps in it. For the passive task, they were instructed to ignore any sound and rest. We refer to these three different experimental conditions as ‘L-task’, ‘H-task’, and ‘No-task’ conditions; see Fig. 1B. Potential confounding of attention effects by physical stimulus variations was excluded by using the same set of stimuli for all task conditions.

Fig. 1C visualizes the design of the experimental trials. Trials lasted 13s and began with the simultaneous presentation of a visual cue (symbol ‘L’, ‘H’, or ‘+’, depending on the task) and an auditory stimulus, in the absence of fMRI scanning. After the auditory stimulus, a cluster of two consecutive fMRI volumes was acquired. Subsequently and only in the active tasks, listeners gave a behavioral response, and visual feedback was provided by displaying the correct response in either green or red color, depending on the correctness of the listener’s response. Stimulus presentation and fMRI were timed to capture the blood oxygenation level-dependent (BOLD) responses that were evoked primarily by the task and stimuli rather than by ambient scanner noise. This relative timing was estimated based on convolving the auditory stimulation protocol (task stimuli and scanner noise, modeled as boxcar waveforms) with a standard hemodynamic response function; for details see (Riecke et al., 2017). Occasional null trials served to establish a stable baseline. These trials involved no auditory stimulus but only the task cue for the upcoming block of trials. Participants were instructed to rest on these trials.

Stimuli

The stimulus design was inspired by a recent visual attention study, which mapped retinotopy in human primary visual cortex with fMRI.
Listeners’ task was varied across the three conditions (No-task, L-task, or H-task) shown on the horizontal axis. Each cell within the graph sketches sounds of a mapping band has a center frequency (CF) of 4.3 kHz, and the L-target contains two gaps. Bandwidth is expressed in units of ERBs (equivalent rectangular bands).

The present auditory stimuli while participants covertly attended to various peripheral target locations in visual space (Klein et al., 2014). The present auditory stimuli were composed of three simultaneous frequency bands: a specific mapping-frequency band and two target-frequency bands, all of which lasted 5.5 s; see Fig. 1A. The center frequency (CF) of the mapping band was varied across trials to enable the identification of frequency channels and the assessment of frequency tuning (frequency mapping). The CFs of the target bands were fixed above and below the edges of the mapped frequency range to provide listeners with distinct, frequency-specific targets on which they could focus their attention during frequency mapping. Exemplary stimuli are provided as audio files in the Supplementary Material.

**Mapping bands**

Mapping bands had a CF of 0.4, 0.7, 1.0, 1.5, 2.1, 3.0, 4.3, or 6.0 kHz (corresponding to a fixed CF-spacing of three equivalent rectangular bands [ERBs]). We refer to these eight different experimental conditions as ‘mapping frequencies’; see Fig. 1B. The bandwidth of each mapping band equaled two ERBs, thus adjacent bands were separated by a fixed spectral gap of one ERB. Mapping bands comprised a sequence of tone pairs pulsating at a fixed rate of 4 Hz, which has been shown to evoke strong BOLD responses in human PAC (Harms and Melcher, 2002; Hart et al., 2003). For each individual tone pair within the overall sequence, the inter-tone interval was 42 ms and the two tone frequencies were randomly chosen from each ERB flanking the current CF. Tones lasted 42 ms and had fixed amplitude. Mapping bands were matched for loudness (see Procedure) to promote uniform activation level across PAC.

**Target bands**

The low-frequency target (L-target) and high-frequency target (H-target) had a CF of 0.2 kHz and 8.3 kHz respectively. They comprised a 0.5ERB-wide band of spectrally uniform noise, which was generated by mixing tones with random phases, fixed amplitude, and fixed frequency spacing (0.25 Hz). To facilitate attentional selection of targets, the CFs of the noise bands were modulated (sinusoidal modulation; depth: ±0.25ERB, frequency: 1.2 Hz and 0.4 Hz, starting phase: 0° and 180°, for L-target and H-target respectively). Thus, the targets fluctuated within a 1-ERB band that was separated by 1.3 ERBs from the adjacent mapping band. To implement the gap detection task (see Tasks), a random number of maximally three gaps (duration: 313 ms) was inserted independently in each target band. Gaps could occur with a fixed probability of 50% within each third of the stimulus interval (sub-intervals), for each target band. The exact timing of gaps within these sub-intervals was random, with the constraint that they did not overlap with the edges of the sub-interval (its initial and final 375 ms) or any gap in the other target band.

**Mixing of mapping band and target bands**

To equalize task difficulty, the perceptual separability of targets from the distracting mapping bands was matched across the different stimuli and targets in two steps. First, spectrottemporal grouping cues were equalized across stimuli: increases in the spectral proximity of target band and mapping band were counteracted by decreasing the temporal coherence of these two bands. This was done by applying an amplitude modulation (a sinusoid with the same frequency as the tone-pair rate in the mapping band and fixed starting phase) to each target and varying its saliency (modulation depth: 100-0%) as a negative linear function of spectral proximity. For example, low-CF mapping bands would render the H-target more separable than the L-target due to their larger distance from the H-target; to compensate for this perceptual difference, the separability of the H-target was reduced by increasing its temporal coherence with the mapping band, i.e., by applying a more salient modulation to this target. Second, listeners fine-tuned the intensity of each target before the FMRI experiment (for details, see Procedure). The resulting target band intensities were substantially lower than the intensity of mapping bands, on average ~20 dB and ~35 dB for the L-target and H-target, respectively. Therefore, auditory-evoked brain responses were driven primarily by the mapping bands and potential distortion of frequency mapping by physical target variations, which were inevitably
induced by the difficulty-matching procedure, was reduced.

Onsets and offsets in the auditory stimulus components were ramped using 5-ms raised-cosine ramps, except for target gaps, which were adjoined by 30-ms ramps. The auditory stimuli were presented diotically via a soundcard (Sound Blaster X-Fi Xtreme Audio) with 16-bit resolution and 44.1-kHz sampling frequency at an overall sound pressure level of approximately 80 dB. Non-flatness of the frequency response of the MR-compatible audio system (Sensimetrics S14, Malden, MA, USA) in the auditory stimulus range (0.2–8.7 kHz) was corrected by applying the manufacturer's equalization filters. Stimulus presentation and button response collection were controlled using Presentation 17.1 software and trials were triggered by MR image acquisitions.

**Procedure**

The experimental procedure involved an initial training session and a subsequent FMRI session using the same MR-compatible audio system.

**Training session**

In the training session, participants first matched the loudness of the mapping bands to that of a fixed reference signal by adjusting the intensity of each band. Second, they were familiarized with the gap-detection task. Third, as mentioned above (see **Mixing of mapping band and target bands**), they adjusted the intensity of each target to the minimum level they deemed sufficient for just detecting the gaps. This was done in the presence of each the lowest and highest mapping band and repeated until listeners reported that task difficulty matched across all stimuli and targets. For intermediate mapping bands, the intensity of each target was set by interpolating logarithmically between the intensities obtained with the extreme mapping bands. Fourth, the obtained individual settings were applied in a subsequent training of the L-task and H-task, which lasted approximately 30min.

To enable a manipulation check, a control ‘?‐task’ was included that was identical to the L- and H-tasks, except that the visual cue was presented only after the auditory stimulus; thus listeners could not reliably focus their attention exclusively on the actual target as in the L- and H-tasks (Riecke et al., 2017). Impaired performance in this control task would indicate that listeners reliably paid attention exclusively to the specified target in the L-task and H-task conditions.

**FMRI session**

In the FMRI session, participants first verified whether task difficulty still matched across the individualized stimuli inside the scanner and further fine-tuned them if needed. Before each functional run, participants balanced loudness across the two ears if needed. In addition to the insert earphones, they wore ear muffs, which together attenuated ambient noise by approximately 45 dB. Participants’ heads were stabilized with cushions and they were instructed to keep as still as possible while being inside the scanner. They were further instructed to keep their eye gaze focused on the center of the screen throughout the experiment to reduce visual input variations. They gave behavioral responses by pressing one of four buttons with the four fingers of their right hand.

Each repetition (FMRI run) of the experiment comprised 48 task trials (two repetitions per mapping-frequency × task condition) and twelve null trials, lasting in total 13min. Trials of the same task condition were clustered into blocks of four, preceded by a null trial during which the task cue was presented to specify the upcoming task condition; see **Fig. 1D**. The twelve blocks and four task trials within each block were presented in random order; see **Fig. 1E**. Overall, an FMRI session comprised anatomical scans (overall duration: 8min) followed by five runs of the experiment (except for participants P1 and P4 who underwent four and three functional runs, respectively).

**Imaging settings**

Brain images were collected with a Siemens Magnetom 7T MRI scanner (Siemens Medical Systems, Erlangen, Germany) equipped with a 32-channel head coil (Nova Medical Inc., Wilmington, MA, USA). BOLD-signal changes were recorded using a modified gradient-echo echo-planar imaging (EPI) sequence at a nominal resolution of 1.5 × 1.5 × 1.5 mm3 (field of view: 201 × 201 mm2, 29 transverse slices without gap, echo time: 17 ms, repetition time: 1500 ms, readout bandwidth: 1776 Hz/pixel, in-plane GRAPPA acceleration factor: 2, nominal flip angle: 60°). Images were acquired such that a pair of two consecutive volumes was acquired on each trial followed by a 10-s interval without scanning (**Fig. 1D**), yielding two clustered volumes per every 11500 ms. This was done to increase sensitivity and enable the separation of auditory stimuli and scanner noise in terms of both perception and evoked BOLD responses (Hall et al., 1999). During each functional run, 120 vol were collected. Slices were positioned approximately orthogonally to the Sylvian fissure and in 45°-orientation relative to the longitudinal axis of the brainstem, so as to cover bilaterally the brain regions of interest (Heschl’s region and IC) and reduce cardiac-induced pulsatile motion artefacts in IC (Slabu, 2010).

Structural T1-weighted images optimized for gray-white matter contrast were obtained using a magnetization-prepared two rapid-acquisition gradient-echo pulse (MP2RAGE (Marques et al., 2010)) sequence at 0.7 × 0.7 × 0.7 mm3 resolution (field of view: 224 × 224 × 168 mm3, echo time: 2.47 ms, repetition time: 5000 ms, readout bandwidth: 250 Hz/pixel, GRAPPA acceleration factor: 3, inversion times: 900 ms and 2750 ms, flip angles: 5° and 3°). All functional and anatomical images were acquired in anterior-posterior phase encoding direction. To enable subsequent correction for echo-planar image distortions using TOPUP (Andersson et al., 2003), an additional pair of EPI volumes (correction volumes) with reversed phase-encoding direction was acquired between the second and third functional run.

**Functional data processing**

Data were analyzed using Brain Voyager QX 2.8 (BrainInnovation BV, Maastricht, The Netherlands) and Matlab (The Mathworks Inc., Natick, MA, USA) software. Preprocessing of functional images included the following steps: First, for each run, volume pairs were merged into single average volumes. This was done by decomposing the acquired series of volumes into a pair of two series (n = 1, 2) and computing their weighted average as described elsewhere (Riecke et al., 2017). All subsequent analyses were applied to the resulting average volumes. Second, head motion was corrected by co-registering all volumes with the last volume of the second functional run (target volume). Third, a map of estimated magnetic susceptibility-induced distortions was computed by combining the target volume with the acquired correction volume and then used to reduce distortions in all functional images. Fourth, images were co-registered with the individual’s structural images and spatially normalized to stereotactic Talairach space. Filth, slow signal fluctuations were removed by temporal high-pass filtering (cut-off: 3cycles/run) and the signal level was normalized to its time average. All aforementioned preparatory steps were applied on a voxel-by-voxel basis separately for each run and region of interest (ROI).

A supplementary analysis was applied to behaviorally weighted FMRI data. Data preparation involved extracting the FMRI-signal level from individual trials and normalizing it to its time average. Then, the weighted average of the resulting z-normalized single-trial FMRI signal was computed across trials, separately for L-task and H-task trials, as

\[ x_i = \frac{\sum_{n=1}^{N} (4 - e_i) x_{i,n}}{\sum_{n=1}^{N} (4 - e_i)} \]

where \( e_i \) corresponds to the absolute magnitude of the behavioral error (a value between 0 and 3; see section Behavioral data analysis) and \( x_{i,n} \) to the z-normalized single-trial FMRI signal observed on the \( n \)th trial. Thus, more weight was given to trials comprising better performance.
Anatomical data processing

Structural images were resampled to 0.5 × 0.5 × 0.5 mm³-resolution to fit one third of the resolution of the functional images. The resulting volume was spatially normalized to Talairach space. Gray-white matter borders were extracted, from which individual cortical surface meshes were reconstructed. All FMRI data analyses were conducted in volumetric space. Only for visualization purposes, functional maps were sampled using trilinear interpolation onto inflated versions of the reconstructed cortical-surface meshes.

Definition of functional-anatomical regions of interest

Individual ROIs were defined based on participants’ individual functional brain anatomy. This approach takes into account inter-individual functional-anatomical variations and thereby improves inter-subject alignment. First, areal borders were estimated using common macroanatomic criteria. The putative borders of PAC were manually defined on the inflated cortical surface meshes, based on Heschl’s gyrus (HG) and possible duplications of it. The anterior border was marked by the most anterior transverse temporal sulcus. For all other hemispheres, the posterior border of PAC was extended to include the medial two-third portion of HG as suggested by ex vivo macroanatomic findings (Rivier and Clarke, 1997; Hackett et al., 2001; Morosan et al., 2001; Wallace et al., 2002; Sweet et al., 2005). For all other hemispheres, the anterior border was marked by the most anterior transverse temporal sulcus. For cerebral hemispheres containing a single HG, PAC was defined as the medial two-third portion of that HG as suggested by ex vivo macroanatomic findings (Rivier and Clarke, 1997; Hackett et al., 2001; Morosan et al., 2001; Wallace et al., 2002; Sweet et al., 2005).

Frequency tuning analysis

Frequency tuning was assessed based on the frequency response, which we defined as the beta value as a function of mapping frequency. From this response, we further defined BF as the frequency that evoked the maximum (i.e., the CF of the mapping band that evoked the highest BOLD response, as described already above). Each of these measures was obtained separately for each task condition in order to extract a ‘passive’ frequency response (No-task) and two ‘active’ frequency responses (Low-task and High-task) as well as a ‘passive’ BF and two ‘active’ BFs. For the active tasks, also a ‘worst frequency’ (WF) was extracted, which we defined analogously to BF, but with the difference that this frequency evoked the minimum response. Effects of frequency-specific attention, which for simplicity we refer to as ‘attentional modulation’, were assessed by contrasting L-task vs. H-task. Thus, positive and negative values of this measure indicate relative enhancement induced by focusing attention on the low- and high-frequency target, respectively.

Multi-voxel pattern analysis

Frequency-specific attentional modulation patterns were assessed using multi-voxel pattern analysis (MVPA). First, the silent baseline (average across all null trials) was subtracted from the z-normalized, unweighted single-trial FMRI signal for each trial. Second, on each iteration of a leave-one-run-out procedure, single trials from L-task and H-task conditions from all but one experimental run were used for classifier training, i.e., for estimating the function that mapped the measured BOLD response patterns to the corresponding frequency-specific attention conditions. This was done within each ROI (i.e., no searchlight approach). Third, the trials from the omitted run were used for consecutive testing, i.e., to assess the generalization performance of the trained classifier. The omitted run varied across iterations. Fourth, classification accuracy was computed as the percentage of correctly classified test trials across all iterations. Finally, the chance level was estimated empirically as the median of 500 classification accuracy values obtained iteratively using the same procedure as above, but after permuting the task labels (L vs. H) on each iteration. A non-linear classifier (nearest-neighbor model, Statistics and Machine Learning Toolbox in Matlab) was used, which is especially suited for datasets comprising a small number of features like IC (Ouda et al., 2006). The classifier’s only free parameter (the number of neighbors) was optimized on each iteration of the training step using three-fold cross-validation to minimize prediction error.

Supplementary IC experiment

The main experiment was primarily designed to address our first research question, i.e., to assess frequency responses in PAC and detect their attentional modulation. Following up on our observation of comparatively weak attentional modulation in IC (see Results), we conducted a supplementary experiment with auditory-stimulus and FMRI...
settings adjusted to fit more our second research question (i.e., to detect subcortical attentional modulation).

Five normally-hearing, trained listeners participated in the supplementary experiment (one female, ages: 23–39 years; including two of the ten participants from the main experiment). Methods largely matched those applied in the main experiment, except for the following differences: Auditory stimuli were composed of two alternating, individually loudness-matched frequency bands within the pitch range, i.e., with a CF of 0.4 kHz (L-target) and 3.2 kHz (H-target) respectively. Each band comprised a common melody composed of a sequence of 29 pure-tone pairs. Overall tone pairs spanned a range of seven semitones and pulsed at a fixed rate of 5.3 Hz (Riecke et al., 2017). Within each pair, tones resembled the same note and pulsed at 27 Hz. Tone-pair omis-(gaps) were inserted in each band as described above. An exemplary stimulus is provided as audio file in the Supplementary Material. As before, these multi-band stimuli were presented in L-target and H-target conditions to enable extracting attentional modulation. The experimental design further included null trials and passive conditions during which only one of the two bands was presented in isolation; these conditions served to enable functionally define IC. In total, 60 trials were presented per experimental condition. Functional imaging data were acquired with a simultaneous-multi-slice gradient-echo EPI sequence (Setsompop et al., 2012) at a resolution of 1.1 × 1.1 × 1.1 mm³ (field of view: 201 × 201 mm², 50 slices, echo time: 21 ms, repetition time: 1500 ms, readout bandwidth: 1446 Hz/pixel, in-plane GRAPPA acceleration factor: 3, slice acceleration factor: 2, nominal flip angle: 50°) according to the sparse protocol described above. Structural images were acquired with an MP2RAGE sequence at 0.55 mm³ resolution.

Compared with the main experiment, the target bands in this supplementary experiment were expected to enhance attentional modulation in IC because of their higher intensity (due to their at least 20 dB higher sound level and the absence of any simultaneous band), higher modulation rate, and higher bandwidth—factors which have been shown to promote IC activation (Harms and Melcher, 2002; Hawley et al., 2005; Sigalovsky and Melcher, 2006; Overath et al., 2012). Moreover, the absence of any acoustic stimulus variation (i.e., no simultaneous mapping-frequency manipulation) and the increased spatial resolution of imaging were expected to facilitate the detection of finer-grained attentional modulation patterns.

Results

Behavioral results

Listeners’ behavioral performance in the FMRI session was accurate and stable across conditions. On average, they slightly underestimated/overestimated the number of gaps in the L-task/H-task, respectively (Fig. 2A), but the magnitude of these errors did not differ significantly from zero for any attention task (all P > 0.22) or across attention tasks (P = 0.19). Listeners on average responded correctly on 90.6% of the L-task trials and on 91.4% of the H-task trials, with no significant difference between these tasks in the FMRI session (F1,63 = 0.18, P = 0.68) or in the training session (F1,63 = 0.03, P = 0.88); see Fig. 2B. Accuracy in the L-task and H-task was generally better than in the ‘?task’ (all P < 10⁻³); see Fig. 2B. Accuracy further differed across mapping frequencies in the H-task (F2,63 = 3.46, P = 0.003), but not in the L-task (F2,63 = 1.11, P = 0.37); see Fig. 2C. The interaction of these factors was not significant (no mapping-frequency × target-frequency interaction: F2,63 = 1.71, P = 0.12); see Fig. 2D. The observation of no significant task difference or interaction indicates that task difficulty was similar across the different conditions. The significant difference with the ‘?task’ further suggest that listeners focused reliably on the cued target during the active tasks inside the FMRI scanner, indicating that our experimental manipulation of frequency-specific attention was successful.

In the supplementary IC experiment, behavioral accuracies were 96.9% (L-task) and 97.0% (H-task), without significant difference between these tasks as above (t4 = 0.01, P = 0.99; data not shown).

The group data (panels A–D) show mean ± sem across participants. The single-subject data (panel E, upper row) show mean ± sem across mapping frequencies. n.s. non-significant, *P < 0.05. The horizontal axis in panels B and C was slightly shifted for each data series to improve visibility.

FMRI results

Functional-anatomical definition of PAC and PAC frequency channels

Fig. 3A visualizes the location of PAC and its frequency channels, defined from functional-anatomical criteria, for each participant. Auditory stimuli in the passive task reliably activated most of the anatomically-defined PAC compared with silent null trials. These passively sound-activated PAC regions further exhibited spatial variations in BF that roughly agreed with tonotopic gradients (for reviews, see (Moerel et al., 2014; Saenz and Langers, 2014)), despite the fact that target bands were present during the BF mapping.

Data in panels B–G show mean ± sem across participants. Frequencies are expressed in units of kHz. BOLD response strength is expressed as beta value (arbitrary units). n.s. non-significant, * * * corrected P < 0.05, 0.0005. The horizontal axis in panels D and E was slightly shifted for each data series to improve visibility.

Fig. 3B shows the response characteristics of the functional-anatomical PAC during the passive task, averaged across participants. Frequency channels within PAC did not vary systematically with regard to their passive overall auditory-evoked response or number of voxels (no main effect of frequency channel in No-task condition on the BOLD response [F2,63 = 0.39, P = 0.91]; Fig. 3B left) or the number of voxels [F2,63 = 1.81, P = 0.10]. Similarly, the passive auditory-evoked response of the overall PAC varied only slightly across mapping-frequency conditions, exhibiting a weak skewness in PACs passive frequency response (no main effect of mapping frequency in No-task condition: F2,63 = 1.96, P = 0.07); see Fig. 3B right. These observations show that our loudness-matched mapping bands evoked approximately uniform
activation levels across tonotopic PAC (Fig. 3B left) and that overall PAC was approximately equally sensitive to these different bands (Fig. 3B right).

Fig. 3C shows PAC's passive frequency response separately for each frequency channel. While the observation of a marked single peak in each channel is trivial (because the channel had been defined based on this prior observation), a non-trivial observation is that the frequency response decreased near-monotonically beyond this channel BF, revealing band-pass frequency filtering in human PAC in the presence of the target bands.

Fig. 3A. The maps visualize functional-anatomical ROIs on inflated representations of each cerebral cortical hemisphere, for each participant. The upper rows show PAC (white patch) defined based on macroanatomic landmarks. The lower rows show BF maps obtained during passive auditory stimulation, confined to regions activated by this stimulation (P < 0.05) and with anatomically-defined PAC borders superimposed (black). BF maps obtained during the active tasks are shown in Supplementary Fig. 1.

Fig. 3B. The plots show the average activation level of PAC evoked by passive auditory stimulation. On the left, this measure is plotted as a function of frequency channel's BF to illustrate spatial variations in passive sound-evoked activation across PAC frequency channels. On the right, it is plotted as a function of mapping frequency to visualize overall PAC's passive frequency response. Numbers in the subtitle denote the overall volume of PAC.

Fig. 3C. The plots show frequency channels' passive frequency responses (i.e., data from panel B right, but stratified for channels). The title shows the passive BF of the channel and the subtitle denotes the channel's ROI volume. The activation level at BF is biased due to the averaging of data previously used for defining the channel. D-E. Analogous to panels B–C, except that data stem from L-task (red) and H-task (blue). Arrows highlight channels' active response at their passive BF (i.e., in the active mapping-frequency condition that matched the channel's passive BF). Unlike panel C, the activation level at BF here is not biased by prior channel definition. F-G. Analogous to panels D–E, except that the difference between the L-task vs H-task is shown to visualize effects of frequency-specific attention (attentional modulation).
Effects of frequency-specific attention on PAC

Fig. 3D–E are analogous to Fig. 3B–C, but show data separately for each active task (L-task and H-task). Fig. 3F–G shows the difference between these tasks (L-task minus H-task) to visualize our main measure: attentional modulation. As predicted by the ‘attentional spotlight’ model, PAC’s spatial activation profile varied across L-task vs. H-task. Frequency-specific attention altered the relative overall response strength of PAC frequency channels with remarkably high reliability across participants (frequency-channel × target-frequency interaction: $F_{7,63} = 10.0$, $P = 10^{-2}$); see Fig. 3F left (‘Channel BF’). This spatially-selective attentional enhancement exhibited a gradual change in sign across frequency channels, indicating qualitatively a relative overall response enhancement in low- and high-frequency channels by low- and high-frequency attention, respectively. In contradiction with the original ‘attentional retuning’ hypothesis, frequency-specific attention did not reliably alter or ‘reshape’ the frequency response of the overall PAC; see Fig. 3F right (‘Mapping freq.’): both the height and the shape of this response remained essentially unchanged during L-task vs. H-task (no main effect of target frequency: $F_{1,9} = 1.43$, $P = 0.26$; no mapping-frequency × target-frequency interaction: $F_{7,63} = 0.99$, $P = 0.45$). Similar observations were made when comparing each active task with the passive task (no main effect of task; L-task: $F_{1,9} = 1.39$, $P = 0.27$, H-task: $F_{1,9} = 0.54$, $P = 0.48$; no mapping-frequency × task interaction; L-task: $F_{7,63} = 0.57$, $P = 0.78$, H-task: $F_{7,63} = 0.74$, $P = 0.64$), which was done for a purely anatomically defined PAC to avoid circularity in the analysis. This latter observation is in line with previous human FMRI studies comparing auditory attention with passive stimulation (Paltoglou et al., 2009) or visual attention (Petkov et al., 2004; Woods et al., 2009). In sum, these results confirm the existence of the tonotopic attentional spotlight in human PAC (Fig. 3F left), as observed in previous studies that did not investigate frequency responses (Paltoglou et al., 2009; Da Costa et al., 2013; Oh et al., 2013; Riecke et al., 2017). However, they do not provide FMRI evidence for an attention-induced reshaping of frequency tuning in human PAC (Fig. 3F right).

Effects of frequency-specific attention on frequency channels within PAC

We did not statistically compare the frequency response of each frequency channel between active vs. passive tasks (Fritz et al., 2003; Atiani et al., 2009), because this comparison would be biased by the fact that we had already defined the channels based on the passive-condition data (Fig. 3C). Instead, we focused on the frequency response during the active tasks; see Fig. 3E: this ‘active’ frequency response varied significantly across frequency channels (interaction mapping-frequency × frequency-channel; $F_{49,441} = 15.3$, $P = 10^{-16}$) in a similar way as the passive frequency response (Fig. 3C), an observation that we elaborate on in the next section. As predicted by the ‘spotlight’ model and the whole-PAC results reported above, frequency-specific attention altered the average response of PAC frequency channels: overall response-gain modulation reached or approached statistical significance specifically in the three lowest frequency channels (0.4–1.0 kHz, post-hoc tests: $t > 2.26$, corrected $P < 0.07$; see Fig. 3F left), revealing a slight asymmetry along the tonotopic axis. In further contradiction with the ‘retuning’ hypothesis, frequency-specific attention did not alter much the shape of the frequency response of any frequency channel (no mapping-frequency × target-frequency interaction: all $F_{7,63} < 0.55$, corrected $P > 0.7$; see Fig. 3G).

These observations indicate that the observed tonotopic modulation pattern in PAC (Fig. 3F left) emerged primarily from modulation of frequency channel’s overall response gain rather than modulation of their frequency tuning (Fig. 3G) as measured with FMRI. Put differently, frequency-specific attention modulated the overall responsiveness of PAC frequency channels and this modulation depended primarily on the channel’s characteristic frequency (i.e., its tonotopic location) not the channel’s specific input (i.e., the mapping frequency).

Characteristics of frequency-specific attentional modulation in PAC

To further characterize the observed attentional modulation of response gain of PAC frequency channels, we replotted these channels’ responses to a specific mapping frequency (i.e., the frequency nearest to the target band, the frequency farthest away from the target band, or the channels’ respective BF) as a function of the channel’s passive BF. Fig. 4A shows that attentional modulation of channels’ response to the lowest mapping frequency (nearest to the L-target) varied across channels in a slightly asymmetric tonotopic modulation pattern that was similar to the pattern observed for the overall channel response (cf. Fig. 3F left). As shown in Fig. 4B, attentional modulation of channels’ response to the highest mapping frequency (nearest to the H-target) also exhibited such an across-channel pattern, although with a shallower slope. The latter resulted in exclusively positive values, indicating that the observed modulation was strongly driven by attention to the L-target. Fig. 4C shows that attentional modulation of channels’ BF-response (i.e., the channel’s response in the mapping-frequency condition that matched the channel’s passive BF) exhibited an across-channel pattern that resembled more the (non-significant) frequency response shape changes observed in overall PAC (cf. Fig. 3F right; average correlation coefficient: 0.77 ± 0.04).

Together, these observations indicate that the tonotopic response-gain modulation observed in the overall PAC emerged from modulation of frequency channels’ responsiveness to a range of frequencies (including the lowest and highest ones), not specifically the BF or the frequency nearest to the target. This observed lack of input-frequency specificity corroborates the above notion that attentional modulation of frequency channels in PAC depends primarily on the channel’s characteristic frequency (‘spotlight’ model), not its specific input (original ‘retuning’ hypothesis).

All data show mean ± sem across participants. Frequencies are expressed in units of kHz. n.s. non-significant, ***$P < 0.0005$. The
horizontal axis in panels D and F was slightly shifted for each data series to improve visibility.

In another attempt to detect the originally hypothesized attentional retuning, we assessed effects of frequency-specific attention on frequency channels’ BF and WF (see Materials and Methods). As suggested by our channel frequency-response results (see preceding section), the active BF, which we defined from the maximum of the active overall frequency response (Fig. 3E), increased significantly and monotonously across frequency channels (main effect of frequency channel: F_7,9 = 15.0, P < 10^-10), exhibiting a pattern resembling the passive channel BF; see Figs. 4D and 3A vs. Supplementary Fig. 1. More importantly, and again in contradiction with the ‘retuning’ hypothesis, frequency-specific attention did not modulate this pattern (no target-frequency x frequency-channel interaction: F_7,63 = 0.92, P = 0.50; see Fig. 4E) or the active BF itself (no main effect of target frequency: F_1,9 = 1.58, P = 0.24). Similar outcomes were obtained for the WF; see Fig. 4F–G (main effect of frequency channel: F_7,9 = 11.9, P < 10^-8; no target-frequency x frequency-channel interaction: F_7,63 = 1.59, P = 0.15; no main effect of target frequency: F_1,9 = 0.004, P = 0.95).

Together, these results show that PAC frequency channels’ most and least characteristic frequencies remained relatively stable under frequency-specific attention. This indicates that these channels were reproducible across task conditions and again provides no fMRI evidence for attentional modulation of frequency tuning in human PAC.

**Anatomical definition of IC and IC-frequency channels**

Fig. 5 shows results from the main experiment for IC, in a similar format as the PAC results in Fig. 3. The anatomically defined IC exhibited spatial variations in BF (De Martino et al., 2013); see Fig. 5A. However, in some participants, this region was not reliably activated in every mapping-frequency condition, hampering the definition of a complete set of frequency channels. To circumvent this problem, which affected only our analysis of channels’ frequency responses (not MVPA; see next section), we defined IC based on anatomical criteria alone (i.e., without prior functional voxel selection) for this specific analysis. The anatomically-defined IC exhibited passive sound-evoked responses reliably across participants (Fig. 5B). These responses were significantly above the silent baseline, yet much weaker than in PAC. Similar to PAC, these passive sound stimulation-evoked activation levels were relatively stable across frequency channels in IC (no main effect of frequency channel in No-task: F_7,63 = 1.25, P = 0.29; Fig. 5B left) and the passive frequency response was nearly flat (no main effect of mapping frequency in No-task: F_7,63 = 0.82, P = 0.58; Fig. 5B right). The passive frequency response of IC-frequency channels to our auditory stimulus set is illustrated in Fig. 5C.

**Frequency-specific attentional modulation of spatially-distributed activation patterns in IC**

The overall response of IC was slightly stronger in active vs. passive tasks, but this difference was not statistically significant (t_9 = 0.30, P = 0.38). Applying the previous analyses for frequency-specific attention effects (see sections 3.2.2 and 3.2.3) to the IC dataset yielded no significant result (all P > 0.05); see Fig. 5D–G. Thus our univariate analyses did not reveal any indication for a tonotopic spotlight or frequency retuning in human IC.

An alternative approach for assessing frequency-specific attentional modulation within a given ROI is based on the multivariate analysis of spatially distributed response patterns. In contrast to conventional univariate (voxel-by-voxel) analyses, this MVPA approach does not rely on the definition of a BF or frequency channel but takes into account inter-voxel dependencies, making it sensitive for finer-grained attentional modulation patterns. Fig. 6A shows results obtained from applying MVPA...
to the functional-anatomically defined IC in our main experiment. The trained classifier could decode the listener's focus of attention from the BOLD response patterns in IC measured during the active tasks with limited success: permutation testing revealed a significant outcome for one listener (P6: \( P < 0.014 \)) and on average, the observed classification accuracy was higher than chance level (52.3 ± 1.4% vs. 50.0 ± 0.06%), a small difference that was not statistically significant (\( t_{12} = 1.61, P = 0.071 \)). Fig. 6B shows matching results obtained analogously from the supplementary IC experiment (52.7 ± 2.8% vs. 49.9 ± 1.6%, \( t_{12} = 0.95, P = 0.20 \)). Pooled analysis of both datasets revealed that classification accuracy was significantly above chance (52.6 ± 1.4% vs. 50.0 ± 0.06%, \( t_{18} = 1.88, P = 0.042 \)). These data suggest that spatially-distributed patterns of frequency-specific attentional modulation may be observed with FMRI in human IC.

To validate our classification approach, we applied the same analyses to the PAC dataset. As shown in Fig. 6C–D, this revealed much clearer, significant attentional modulation patterns (main experiment: 54.1 ± 1.4% vs. 49.9 ± 0.1%, \( t_{9} = 2.96, P = 0.0080 \); supplementary IC experiment: 57.3 ± 2.3% vs. 49.9 ± 0.2%, \( t_{9} = 3.25, P = 0.016 \)), confirming our BF-based results (Fig. 3F left) and previous human FMRI classification results (Riecke et al., 2017).

Overall, our observations in PAC and IC were stable across various analyses: the conclusions that can be drawn from our study did not change when applying the ROI analyses separately to each cerebral hemisphere (all \( P > 0.05 \)), when applying behavioral weights to the single-trial FMRI signal level (see Functional data processing), when normalizing the active frequency response to the passive frequency response (e.g. (Fritz et al., 2003)), when applying more liberal functional-anatomical PAC definition criteria (e.g., larger anatomical masks or no functional mask), or when applying an alternative, regression-based frequency-response analysis that aims to separate gain from shape changes (for details, see Atiani et al., 2009).

**Discussion**

Our results demonstrate that the spotlight of frequency-specific attention, as measured with FMRI, emerges primarily from modulation of overall response gain, rather than from frequency retuning, or frequency channels in human PAC. They further suggest that frequency-specific attention may alter afferent sound processing in human IC.

**Frequency-specific attentional modulation of response gain in human PAC**

Our analysis of spatial activation profiles confirmed that shifts in the focus of frequency-specific attention selectively alter spatial (tonotopic) activation patterns in human PAC, such that voxels with characteristic frequencies near the attended frequency are enhanced relative to voxels with characteristic frequencies far away from the attended frequency. This result confirms previous findings of the tonotopic attentional spotlight in PAC by human FMRI studies that did not specifically assess frequency responses (Paltoglou et al., 2009; Da Costa et al., 2013; Oh et al., 2013; Riecke et al., 2017). It further reveals that this phenomenon can be observed reliably even during simultaneous frequency mapping. Crucially, this latter aspect of our study enabled us to investigate frequency responses. This revealed that human PAC achieves the attentional spotlight by applying an overall gain to the outputs of specific frequency channels, i.e., a response modulation that depends on the channel’s characteristic frequency, not the channel’s specific input. The relative ‘flatness’ of this response gain implies that the originally hypothesized frequency retuning (i.e., reshaping of the frequency response induced by, e.g., shifts in channels’ most or least characteristic frequencies or sharpening/widening of the tuning profile) seems to contribute relatively little to the attentional spotlight in human PAC, although this interpretation requires some cautionary remarks that we make in the next section.

Animal electrophysiological studies have indicated that gain modulation in PAC neurons can arise from synaptic depression (Rabinowitz et al., 2011) and temporal shifts in excitability (Lakatos et al., 2013; Reig et al., 2015) especially in lower PAC layers (O’Connell et al., 2014), and may be triggered by top-down signals originating in frontal cortex (Fritz et al., 2010). These mechanisms probably also contributed to our human PAC findings, which may be verified in future studies using invasive single-unit recordings in neurosurgical patients (e.g. (Howard et al., 2000; Mukamel et al., 2005; Bitterman et al., 2008)).

Our observation that gain modulation is strongest in frequency channels with low characteristic frequencies (0.4–1.0 kHz; Fig. 3F left)

---

**Fig. 6.** Classification results. A. MVPA results in IC from the main experiment. The bar plot on the left shows the average activation level of functional-anatomically defined IC, for each active task (L-task: red, H-task: blue) pooled across mapping frequencies. The title denotes the average number of voxels in the ROI. Data show mean ± sem across participants. The plot in the middle shows classification accuracy for each participant (circles), i.e., how well the trained classifier could decode the listener’s focus of attention from the BOLD response patterns measured in IC during the active tasks. The gray horizontal line illustrates the empirically derived chance level (shaded area: SD across permutations). \( P < 0.05 \) (permutation testing). The plot on the right shows classification accuracy averaged across participants, with the gray line illustrating the average chance level (line width: sem across participants). B. Analogous to panel A, but for data from the supplementary experiment. C–D. Analogous to panels A–B, but for data from PAC for reference.
and strongly driven by selective attention to low frequencies (Fig. 4A–B) suggests that frequency-specific attentional modulation is asymmetric across the tonotopic axis of PAC. Although frequency discrimination worsens at high frequencies (Sek and Moore, 1995), the observed asymmetry cannot be attributed to our individualized auditory stimuli, the task difficulty, or the sensitivity of the passive PAC in our study because none of these aspects varied appreciably across the attention tasks or the tonotopic axis. Instead, it might reflect a contribution from what we refer to as ‘harmonic attentional modulation’. More specifically, frequency-specific attention possibly also modulated neuronal populations tuned to the higher harmonics of the attended frequency in human PAC (Fritz et al., 2007b; Borra et al., 2013). Because our frequency-mapping range covered harmonics of only the L-target (0.4, 0.6, …kHz), but not of the H-target (8.3, 16.6, …kHz), such putative harmonic modulations could only be observable in our L-task.

To explore this idea, we applied our BF-channel analyses to peripheral spike activity evoked by our auditory stimuli, which we estimated with a computational model of the peripheral auditory system (Meddis et al., 2013) incorporating eight frequency channels (centered on our mapping frequencies) and a multiplicative harmonic attentional gain (i.e., a scaling of each channel’s output by its response to the attended target alone). These simulations indeed revealed harmonic responses patterns evoked by the L-target, but not the H-target, within the sampled range of the tonotopic axis. In particular, applying the same analysis for frequency-specific attentional modulation (see FMRI results in section 3.2.2) showed a low-frequency bias across this tonotopic range (Fig. 7). This indicates indeed that harmonic attentional modulation can in principle account for part of the asymmetry in our PAC results.

Fig. 7. Frequency-specific attentional modulation of simulated peripheral auditory responses.

Analogously to the FMRI results shown in Fig. 3F, the plot shows attentional modulation of peripheral auditory responses as a function of frequency channel (dark magenta, scaled to match the scale of the FMRI results). Spike rates were estimated using an eight-channel computational model of the peripheral auditory system. Harmonic attentional gain was modeled by multiplying each channel’s output by its response to the attended target stream alone. The same BF-channel analysis as for the FMRI data was applied to the spike rate data. The resulting spatial pattern of attentional modulation shows a slight asymmetry across the tonotopic axis: the majority of frequency channels, specifically those up to 3 kHz, show stronger enhancement during low-vs. high-frequency attention. The FMRI results are plotted in brighter color for reference. The horizontal axis was slightly shifted for each data series to improve visibility.

Detailed observations of attentional modulation of frequency responses in auditory cortex have been limited so far mostly to animals (see Introduction). In humans, MEG studies have provided rather indirect and coarse insights into this topic (Kauramaki et al., 2007; Okamoto et al., 2007, 2009; Ahveninen et al., 2011). More detailed invasive electrocorticography studies have associated modulation of spectrotemporal tuning in human auditory cortex with selective listening to speech (Mesgarani and Chang, 2012; Holdgraf et al., 2016) but did not investigate frequency-specific attention. An FMRI study of the laminar organization of frequency tuning in human PAC observed that intermodal attention alters the width of the frequency response of upper PAC layers (De Martino et al., 2015). However, attentional targets were confounded by the mapping stimuli, leaving relative contributions of the attentional spotlight, response gain, and frequency tuning unclear. Our study extends the previous findings by showing that the tonotopic attentional spotlight in human PAC, as measured with FMRI, emerges primarily from modulation of response gain rather than from frequency retuning (but see cautionary remarks in next section).

Does frequency-specific attention modulate frequency tuning in human PAC?

Our analysis of PAC frequency channels’ response to frequencies near or far from the target bands (Fig. 4A–B) revealed that attentional modulation of channel’s responsiveness to non-BFs is sufficient to give rise to the attentional spotlight in human PAC. However, we found no statistical evidence for frequency retuning: the overall shape of the frequency response, as well as the most and least characteristic frequencies of frequency channels, did not alter systematically across attentional states (Figs. 3G, 4E and 4G). This null result of frequency retuning could indicate that the frequency-tuning properties of PAC remain more stable under frequency-specific attention in humans than was expected based on animal findings (Jarрамillo and Zador, 2011; Massoudi et al., 2013). Indeed, this interpretation would fit with the fact that frequency tuning in PAC and electrically induced modulation thereof vary across species (Suga and Ma, 2003; Bitterman et al., 2008). However, an equally possible interpretation is that attentional frequency retuning exists in human PAC as suggested by animal models, but we were unable to detect it due the following potential methodological limitations:

Firstly, the signal-to-noise ratio in our frequency-response analysis was possibly insufficient, considering that this analysis required stratifying the data according to mapping frequencies (i.e., trials), not according to channels (i.e., voxels, as in our more successful spatial activation-profile analysis). A general lack of statistical power can be excluded because our spatial activation-profile analysis replicates several FMRI findings in PAC (Petkov et al., 2004; Paltoglou et al., 2009; Woods et al., 2009; Da Costa et al., 2013; Oh et al., 2013; Riecke et al., 2017).

Secondly, the spectral resolution of our frequency mapping was one ERB1; thus presumed finer-grained spectral modulations probably averaged out in our measures. The fact that our target frequencies fell outside the mapped frequency range unlikely caused our null result, given that attentional gain modulation and frequency retuning in animals are not limited to PAC neurons with BFs equal to the target frequency, but extend to neurons tuned to nearby frequencies (Atiani et al., 2009).

Thirdly, the spatial resolution of our FMRI measurements was 1.1–1.5 mm1, thus finer-grained spatial variations, especially putative local heterogeneities in frequency tuning (Castro and Kandler, 2010), averaged out within voxels. It further cannot be excluded that our functionally-anatomically defined PAC lacked or contained a fraction of primary or non-primary tissue, respectively, as there currently exists no unequivocal definition of human primary AC in vivo (Moerel et al., 2014). However, our PAC definition unlikely caused our null result, given that our results replicate several PAC findings (see above).

Finally, our measure of brain activity—the BOLD response—merges neural excitation and inhibition (Logothetis et al., 2001; Logothetis, 2008); thus their putative contributions to frequency retuning in PAC or cortical feedback to IC (Yan and Suga, 1998; Weinberger, 2004; Oht and Scheich, 2005; Galindo-Leon et al., 2009) probably averaged out in our measures.

Therefore, attentional frequency retuning may be observed in future human studies using methods that enable to resolve finer-grained spectral/spatial modulations and disambiguates potential contributions of neural excitation and inhibition, e.g., invasive neuroelectric recordings. Given the aforementioned potential limitations, our findings cannot be straightforwardly compared to related animal electrophysiology findings (Fritz et al. 2003, 2005, 2007b; David et al., 2012; Slee and David, 2015). As noted above, our spectral and spatial resolutions were much coarser.
and we used a fundamentally different measure of neural activity (for further discussion of FMRI-electrophysiology differences, see e.g. (Mukamel et al., 2005; Langers et al., 2012; Overath et al., 2012)). Consequently, caution must be exercised when attempting to relate our macroscopic measures of frequency channels, frequency tuning, and response gain to those obtained from invasive electrophysiology. Moreover, most animal studies investigated auditory attention vs. passive stimulation, leaving unclear to what extent the reported effects reflect task performance-induced changes in the animals’ arousal. Finally, target and mapping stimuli were presented consecutively, whereas in our study they were presented simultaneously (i.e., as an auditory scene). This latter difference seems to facilitate the observation of gain modulation: the only animal study that also presented simultaneous target/mapping stimuli also found prominent gain modulations in PAC, in addition to the previously observed frequency retuning (Atiani et al., 2009); this possibly reflects contributions from additional (mapping stimuli-related) neural suppression induced by additional streaming-related processes. Despite these differences, our human FMRI findings converge with animal electrophysiology findings showing that frequency-specific attention in auditory scenes induces frequency-specific response gain especially in ‘near’ neuronal populations in PAC.

**Does frequency-specific attention modulate response patterns in human IC?**

Our pattern-analysis results indicate that frequency-specific attentional modulation can be observed in the human PAC and to a much lesser extent in the human IC. The trained classifier could decode listeners’ focus of frequency-specific attention from spatially-distributed, sound-evoked activation patterns, but with limited success in IC. The observed decoding accuracies were rather modest, yet pooled across the two experiments, the average accuracy was reliably above chance, which suggests that sound-evoked activation patterns in IC may carry at least some information about the focus of frequency-specific attention. Therefore, it is likely that endogenous frequency-specific attention can influence afferent sound processing in human IC, probably via top-down attentional signals fed from PAC to IC.

This tentative interpretation fits with previous human studies that have used otoacoustic emission recordings and scalp electroencephalography to observe frequency-specific attentional modulation in subcortical stages as early as the cochlea (Galbraith et al., 1998; Maison et al., 2001; Lehmann and Schonwiesner, 2014; Yamagishi et al., 2016). Our study extends these findings by attempting to specify the subcortical locus of attentional modulations in IC, which could not be previously achieved due to the limited spatial detail of the employed non-invasive compound measures of neuroelectric activity. Our study further extends animal electrophysiology findings of frequency-specific attentional modulation in IC (Slee and David, 2015) to humans. The frequency-response changes in IC observed in that animal study arose from modulation of response gain rather than from frequency retuning, and this attentional response-gain modulation was exclusively suppressive. Otherwise it was similar to that observed in animal PAC (Fritz et al., 2003; David et al., 2012).

In contrast to the aforementioned animal IC study and our PAC findings, we found no evidence for a frequency channel-specific attentional response gain in IC. Although our pattern analysis suggests that spatial patterns of frequency-specific attentional modulation may exist in human IC, the observed lack of a univariate effect in IC indicates that this presumed modulation is weaker or less frequency channel-specific than in PAC. This further suggests that the operation of the attentional spotlight diminishes along the efferent cortico-collicular pathway—an idea that remains to be validated in future research. Human IC is a small structure (average volume in our study: 303 mm$^3$) and attentional modulation is inhomogeneous across central vs. peripheral IC nuclei in ferrets (Slee and David, 2015), which may explain why we observed only subtle attentional modulations with pattern analysis and none with univariate BF-based analyses. Moreover, as described in the previous section, finer-grained within-voxel spatial variations of attentional enhancement and suppression probably averaged out in our FMRI measurements. In particular, we could not reliably define a complete set of reliably activated frequency channels in all participants, although we observed spatial variations in BF and activation patterns that were on average approximately flat across BF channels (Fig. 5B). This hints at a partial lack of activation level, which could be reduced in future FMRI studies by using naturalistic sounds that have already proven effective for activating tonotopic frequency channels in human IC (De Martino et al., 2013). Cardiac gating of FMRI may help to further unmask attentional modulations from cardiac-induced pulsatile motion, which is more prominent in IC than PAC (Guimaraes et al., 1998). Our method could not correct for such motion artefacts; however, we oriented FMRI slices so as to reduce these putative artefacts and could reliably observe sound-evoked IC activation in all task conditions when pooling mapping-frequency conditions and channels for our pattern analysis (Fig. 5B–D, left).

Importantly, attentional modulation in our study was driven purely by endogenous frequency-specific attention, i.e., in the absence of physical stimulus variations. Human FMRI studies of interaural attention (Rinne et al., 2008) and perceptual switches in auditory streaming (Schadwinkel and Gutschalk, 2011) found similar top-down attentional modulation in the absence of physical stimulus variations in human IC, but based on listeners’ perceived location of the auditory targets. Given the prominent role of IC in spatial processing (King et al., 2001), these spatial cues have been suggested to be particularly effective in producing attentional modulation in human IC (Rinne et al., 2008). Our study suggests that sound processing in human IC may also be modulated by endogenous attention to spectral cues alone.

**Conclusion**

We demonstrated that the spotlight of frequency-specific attention in human PAC as measured with FMRI emerges primarily from tonotopic modulation of neuronal population-level response gain. Contributions from fine-grained frequency retuning in human PAC (induced by, e.g., neural excitation and inhibition) remain a possibility to be investigated with methods that enable assessing frequency tuning with higher spatial and spectral resolution. Our study further indicates that endogenous frequency-specific attention induces comparatively subtle response-pattern modulations in human IC, suggesting that the operation of the attentional spotlight may diminish from PAC toward hierarchically lower auditory processing stages. In sum, frequency-specific neuronal population responses in the human auditory pathway appear not to be fixed, but adapt constantly to meet changes in listening demands. Noteworthy, our study reveals that frequency tuning in human PAC can be reliably assessed even during simultaneous presentation of task-relevant, frequency mapping-unrelated auditory stimuli—a methodological advancement that opens up new possibilities for future FMRI studies investigating attentional modulation of auditory feature tuning.

**Author contributions**

L.R. and B.S. conceived the project. L.R., B.S., and J.C.P. conducted the experiments. G.V., B.P., and V.G.K. provided tools for data acquisition and analysis. L.R., B.S., and G.V. analyzed the data. L.R. wrote the manuscript and made the figures. B.S., J.C.P., E.F., B.P., V.G.K., and G.V. commented on drafts.

**Acknowledgments**

This work was supported by the Universiteitsfonds Limburg (SWOL Brains Unlimited Pioneer grant S.2014.15), the Maastricht Brain Imaging Center, and the National Institutes of Health (NIH grant R01MH11444). The authors declare no conflict of interest.
Saenz, M., Langers, D.R., 2014. Tonotopic mapping of human auditory cortex. Hear Res. 307, 42–52.
Sakai, M., Suga, N., 2001. Plasticity of the cochleotopic (frequency) map in specialized and nonspecialized auditory cortices. Proc. Natl. Acad. Sci. U. S. A. 98, 3507–3512.
Schadwinkel, S., Gutschalk, A., 2011. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. J. Neurophysiol. 105, 1977–1985.
Sek, A., Moore, B.C., 1995. Frequency discrimination as a function of frequency, measured in several ways. J. Acoust. Soc. Am. 97, 2479–2486.
Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson Med. 67, 1210–1224.
Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182–186.
Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182–186.
Sakai, M., Suga, N., 2001. Plasticity of the cochleotopic (frequency) map in specialized and nonspecialized auditory cortices. Proc. Natl. Acad. Sci. U. S. A. 98, 3507–3512.
Schadwinkel, S., Gutschalk, A., 2011. Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. J. Neurophysiol. 105, 1977–1985.
Sek, A., Moore, B.C., 1995. Frequency discrimination as a function of frequency, measured in several ways. J. Acoust. Soc. Am. 97, 2479–2486.
Setsompop, K., Gagoski, B.A., Polimeni, J.R., Witzel, T., Wedeen, V.J., Wald, L.L., 2012. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. Magn. Reson Med. 67, 1210–1224.
Shinn-Cunningham, B.G., 2008. Object-based auditory and visual attention. Trends Cogn. Sci. 12, 182–186.
Sigalovsky, I.S., Melcher, J.R., 2006. Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. Hear Res. 215, 67–76.
Slabu, L.M., 2010. The effect of slice orientation on auditory fMRI at the level of the brainstem. Brain Topogr. 23, 301–310.
Slee, S.J., David, S.V., 2015. Rapid task-related plasticity of spectrototemporal receptive fields in the auditory midbrain. J. Neurosci. 35, 13090–13102.
Suga, N., Ma, X., 2003. Multiparametric corticofugal modulation and plasticity in the auditory system. Nat. Rev. Neurosci. 4, 783–794.
Sweet, R.A., Dorph-Petersen, K.A., Lewis, D.A., 2005. Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. J. Comp. Neurol. 491, 270–289.