Behavioral/Cognitive

Mapping Frequency-Specific Tone Predictions in the Human Auditory Cortex at High Spatial Resolution

Eva Berlot,[1] Elia Formisano,[1,2] and Federico De Martino[1,3]

[1]Department of Cognitive Neuroscience, [2]Maastricht Centre for Systems Biology, Maastricht University, Maastricht, The Netherlands, and [3]Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota

Auditory inputs reaching our ears are often incomplete, but our brains nevertheless transform them into rich and complete perceptual phenomena such as meaningful conversations or pleasurable music. It has been hypothesized that our brains extract regularities in inputs, which enables us to predict the upcoming stimuli, leading to efficient sensory processing. However, it is unclear whether tone predictions are encoded with similar specificity as perceived signals. Here, we used high-field fMRI to investigate whether human auditory regions encode one of the most defining characteristics of auditory perception: the frequency of predicted tones. Two pairs of tone sequences were presented in ascending or descending directions, with the last tone omitted in half of the trials. Every pair of incomplete sequences contained identical sounds, but was associated with different expectations about the last tone (a high- or low-frequency target). This allowed us to disambiguate predictive signaling from sensory-driven processing. We recorded fMRI responses from eight female participants during passive listening to complete and incomplete sequences. Inspection of specificity and spatial patterns of responses revealed that target frequencies were encoded similarly during their presentations, as well as during omissions, suggesting frequency-specific encoding of predicted tones in the auditory cortex (AC). Importantly, frequency specificity of predictive signaling was observed already at the earliest levels of auditory cortical hierarchy: in the primary AC. Our findings provide evidence for content-specific predictive processing starting at the earliest cortical levels.

Key words: auditory cortex; fMRI; predictions; predictive processing

---

Significance Statement

Given the abundance of sensory information around us in any given moment, it has been proposed that our brain uses contextual information to prioritize and form predictions about incoming signals. However, there remains a surprising lack of understanding of the specificity and content of such prediction signaling; for example, whether a predicted tone is encoded with similar specificity as a perceived tone. Here, we show that early auditory regions encode the frequency of a tone that is predicted yet omitted. Our findings contribute to the understanding of how expectations shape sound processing in the human auditory cortex and provide further insights into how contextual information influences computations in neuronal circuits.

Introduction

An astounding amount of information reaches our ears through incessant streams of sound waves. Received inputs are often incomplete, but our brains nevertheless transform them into rich and complete perceptual phenomena such as meaningful conver-

sations or pleasurable music. It has been hypothesized that this efficiency in sensory processing is achieved due to the highly regular structure of the environment (Nelken, 2012), which allows the brain to form predictions of upcoming signals. The discrepancy between predictions and received sensory input is used to refine future expectations. Several models describe how expectations influence sensory processing (Helmholtz, 1860; MacKay, 1957; Gregory, 1980; Rao and Ballard, 1999; Friston, 2005; Bastos et al., 2012). The proposals vary widely, but they all emphasize the critical role of predictions in sensory processing.

Experiments investigating predictive processing induce regularity in the presented stimuli, which is subsequently violated, most often by presenting a different, unexpected stimulus. Such rule violations lead to distinct brain responses with some being explained by passive stimulus-specific adaptation mechanisms
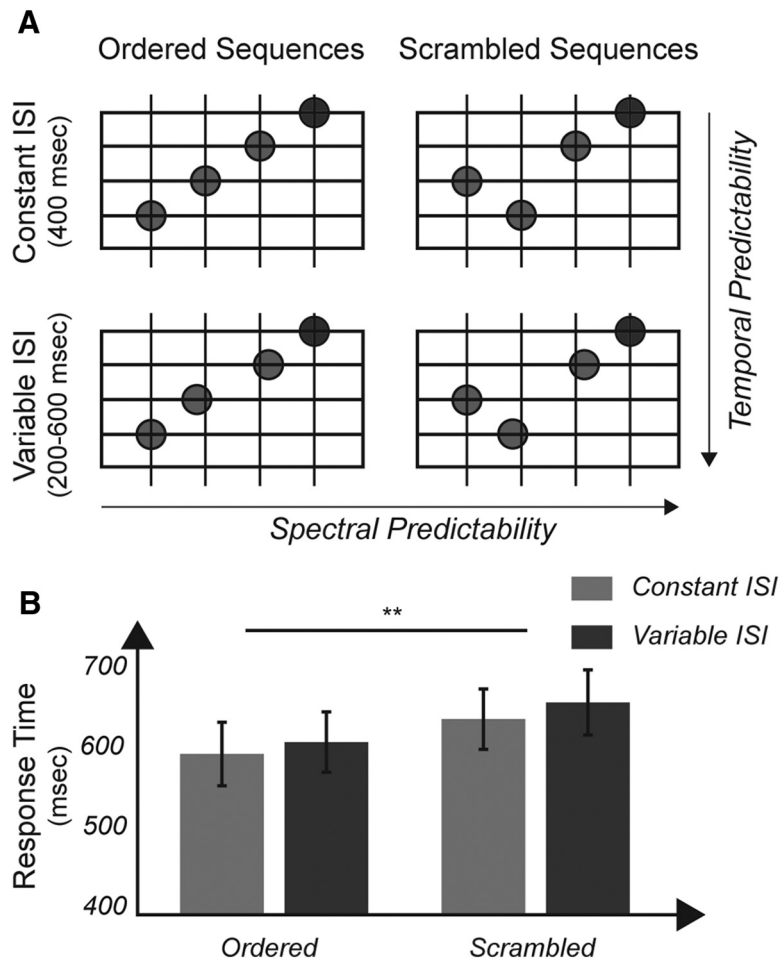
A

Figure 1. Sequences with regularity in both frequency and temporal domain are behaviorally most predictable. **A**, Sequences in the psychophysical paradigm. Frequency expectations were modulated by presenting tones in an orderly (left) or scrambled (right) way with the target frequency always remaining as the final, fourth tone (blue circle). Temporal expectations were manipulated by inserting either regular (top) or irregular (bottom) ISIs between tones. The target could either be the lowest or the highest frequency in the presented sequence and the participants' task was to distinguish between the two cases and indicate whether the target was the highest or the lowest tone as quickly as possible. **B**, Ordered sequences were associated with significantly faster reaction times (~40 ms) than scrambled sequences ($F_{(1,14)} = 51.76, p = 3.031e^{-9}$). Temporally regular sequences produced, on average, 20 ms faster responses than sequences with irregular ISIs ($F_{(1,14)} = 3.60, p = 0.078$). There was no interaction between spectral and temporal regularities ($F_{(1,14)} = 0.40, p = 0.54$). Because ordered sequences with regular ISIs were responded to the fastest, we chose ascending and descending sequences with constant ISIs as stimuli for the neuroimaging experiment. **Significance below $p < 0.01$.

plete sequences (Bendixen et al., 2009; Chouiter et al., 2015). However, the restrictions in the spatial resolution of EEG limited their assessment on how precisely tone characteristics are encoded.

Tone frequency is one of the most defining characteristics of auditory perception and a prominent feature of the auditory pathway (Kaas and Hackett, 1998; Formisano et al., 2003; Moerel et al., 2014). To better understand the importance of tone frequency as an organizing acoustic feature, it is important to determine how auditory responses are organized across a range of different acoustic contexts. An important question is whether auditory predictions are frequency specific and follow the characteristic tonotopic organization across the AC. Although auditory predictions have hitherto not been investigated in such spatial detail, research in vision neuroscience has indicated that fMRI provides the spatial resolution to examine the detailed functional organization of predictions in humans *in vivo* (Smith and Muckli, 2010; Kok et al., 2014; Muckli et al., 2015).

The present study investigated frequency specificity of auditory expectations in the human AC. We examined brain activity to tone omissions in predictable sequences using high-field fMRI. Auditory sequences consisted of four tones, with the last tone omitted in half of the trials. The predictability of the sequences used was assessed using psychophysics. We show that the AC encodes the frequency of the expected tone during its omission. Frequency-specific predictive responses were observed across all auditory cortical regions, including the earliest levels of the primary AC (PAC). We propose that frequency regularities in auditory stimuli are extracted already at the level of the PAC (or earlier) and are used to predict the upcoming tones.

(Ulanovsky et al., 2003) and others with more active prediction mechanisms (mismatch negativity) (Näätänen et al., 1978). Such responses indicate rule extraction, but the presence of unexpected stimulus makes it is difficult to distinguish whether they are driven by an erroneous prediction signal or a pure detection of a novel stimulus. More conclusive evidence stems from paradigms that only omit (rather than replace) predictable stimuli. Such studies have consistently reported increased auditory cortex (AC) activity during omissions compared with repetitions of expected stimuli, lending support for accounts of predictive signaling (Todorovic et al., 2011; Bendixen et al., 2012; Rohrmeier and Koelsch, 2012; Nazimek et al., 2013; SanMiguel et al., 2013; Chouiter et al., 2015). However, most studies use single tone repetitions, which limits the conclusions that can be drawn because increased activity could reflect release of adaptation from previously presented tone rather than prediction of an upcoming tone. A recent study using rule-based sequences of different tones demonstrated that omitted tones evoke similar responses as com-

Materials and Methods

Construction of sequences with predictable targets. Tone sequences were composed of four harmonically related tones, specifically the tonal centers and dominant intervals, presented in an ascending or descending manner (e.g., g1–c2–g2–*c3* or g2–c2–g1–*c1*; in Hertz: 247-330-494-*659* Hz or 494-330-247-*165* Hz) and with equal interstimulus intervals (ISIs). Regularities in both spectral (i.e., ascending/descending sequences) and temporal (regular ISIs) domains were chosen because our psychophysical experiment before scanning demonstrated that orderly sequences with regular ISIs created the strongest expectations (for details, see Fig. 1).
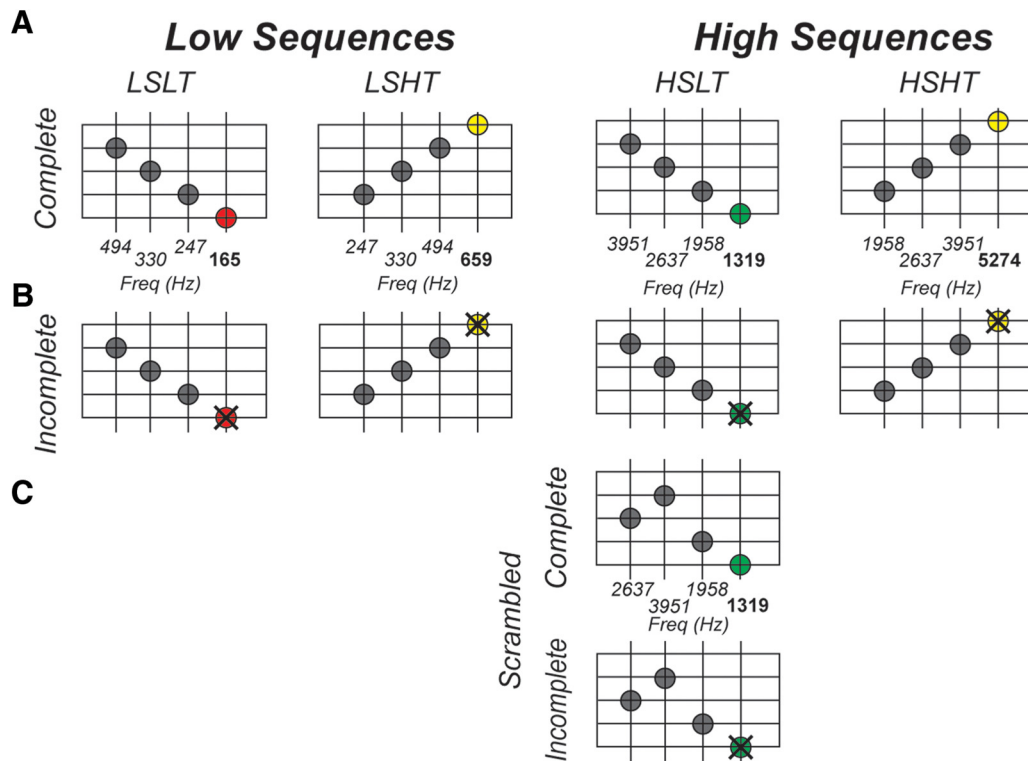
Figure 2. Sequences in the neuroimaging experiment. One pair of sequences was in the low-frequency range (first and second column from the left) and the second pair was in the high-frequency range (third and fourth column from the left). For half of the trials, the sequences were complete (*A*), whereas for the other half, the targets were omitted (*B*). The three tones preceding the targets were identical for both sequence pairs; therefore, the omission trials contained identical tones for every sequence pair. The third and target tones of the scrambled condition were the same as for the descending sequence in high-frequency range (i.e., HSLT sequence), whereas the first and second tones were reversed in order (*C*).

We used a total of four sequences, with the first pair consisting of the ascending and descending sequence in the lower frequency range and the second pair representing the ascending and descending sequence in the higher frequency range (Fig. 2A). The last tone in a sequence, the target, corresponded to the highest (e.g., *c3*/659 Hz) or the lowest (e.g., *c1*/165 Hz) tonal centers in ascending and descending sequences, respectively. The three tones preceding the target were identical for each sequence pair (e.g., *g1-c2-g2*/Hz: 247–330-494). The lowest target frequency corresponded to the target in the descending low sequence (164.8 Hz), followed by the target in the ascending low sequence (659.3 Hz), then the descending high sequence target (1318 Hz), and the highest target frequency was presented in the ascending high sequence (5274 Hz). The four sequences will be referred to as the low sequence with the low or high target (LSLT and LSHT, respectively) and the high sequence with the low or high target (HSLT and HSHT). To isolate expectations about the target from the presented tones, targets were omitted in half of the trials (i.e., incomplete sequences). Importantly, tones presented during incomplete sequences were identical for each pair of sequences (e.g., incomplete LSLT and incomplete LSHT) and differed only in the temporal order of presented sounds (Fig. 2B).

Five participants were additionally presented with a scrambled version of the HSLT sequence in which the order of the first two HSLT tones was reversed while the third and target tones remained the same (Fig. 2C). Note that, for the scrambled condition, we created both complete and incomplete sequences just like ordered sequences. The scrambled sequences (complete and incomplete) were presented the same number of times as each of the orderly sequences. This scrambled condition enabled us to determine whether predictive responses of the target are truly driven by the expectations formed by the first three sequential tones. Target expectation is reduced in scrambled sequences (see our psychophysics results in Fig. 1B), so we expected to see attenuated responses in voxels responding to HSLT targets in scrambled compared to ordered conditions (i.e., incomplete HSLT ordered greater than incomplete HSLT scrambled). Both the presented tones and the spectral proximity of

the third presented tone to the target were identical in both scrambled and ordered incomplete HSLT sequences, thereby controlling for any other potential confounding factors.

To control for attention during the experiment, participants were asked to respond to catch trials in which one of the presented tones contained a gap of 30 ms. These represented 5% of all trials and were excluded from the subsequent fMRI analyses.

Tonotopic localizer. To map the tonotopic organization in the AC, a frequency localizer was performed. In the localizer, we presented nine center frequencies (104.5, 164.8, 331.8, 659.3, 938.4, 1318, 1988, 5274, and 6313 Hz) in blocks of three stimuli each. Each block consisted of three tones centered on each of the center frequencies (i.e., center frequency ± 0.05 octaves). Tones were amplitude modulated (5 Hz, modulation depth of 0.1) and presented for 800 ms.

All of the sounds (both in the sequences and in the tonotopic localizer) were sampled at 44.1 kHz ramped with a 10 ms rise and 10 ms fade-out times. The stimuli were created in MATLAB (The MathWorks) and the paradigms were programmed in PsychoPy (Peirce, 2007). Before starting the experiments, the sound intensity of stimuli was adjusted individually for each participant to equalize the perceived loudness of different tones.

MRI. Participants underwent scanning in a 7 T scanner (Siemens). Anatomical T_1-weighted (T_1w) images were obtained using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (voxel size = 0.6 × 0.6 × 0.6 mm; TR = 3100 ms; TI = 1500 ms; TE = 2.52 ms). Proton-density-weighted (PDw) images were additionally acquired to correct for field inhomogeneities (Van de Moortele et al., 2009) (voxel size = 0.6 × 0.6 × 0.6 mm). Gradient-echo echoplanar imaging was used to obtain T_2*-weighted functional data. Functional scans with the opposite phase-encoding polarities were acquired to correct for geometric distortions (Andersson et al., 2003).

Data for the tonotopic localizer were acquired following a block design. Acquisition parameters were as follows: TR = 2600 ms; TA = 1200 ms; TE = 19 ms; number of slices = 44, GRAPPA acceleration ×3, multiband factor = 2; voxel size = 1.2 × 1.2 × 1.2 mm, silent gap = 1400
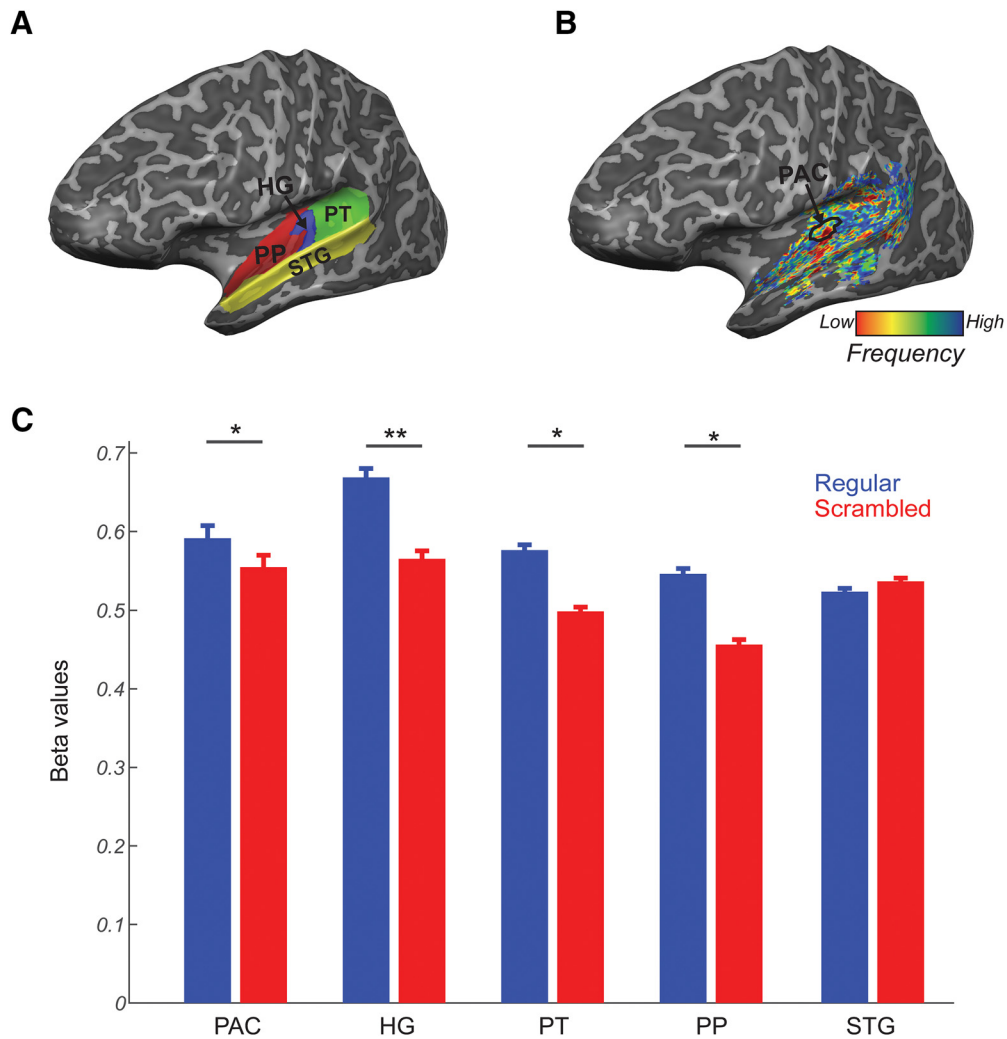
Figure 3. Regions of interest (ROIs). **A**, Anatomically defined ROIs: HG (blue), PT (green), PP (red), and STG (yellow). **B**, Functionally defined PAC based on the tonotopic gradient. **C**, Predictive responses for regular and scrambled incomplete sequences. Responses in voxels encoding the HSLT target frequency were stronger when the omitted tone was preceded by a regular sequence compared with when the tones were scrambled. *Significance below $p < 0.05$; **significance below $p < 0.01$.

ms. Tones centered around the same center frequency were grouped into blocks of three TRs with each tone occurring during the silent gap. The blocks were separated from each other by 13 s (i.e., five TRs) of silence. Localizer mapping was split into two runs and took 15 min with every center frequency presented five times.

The main experiment followed a slow event-related (sparse) design with a TR of 12 s and a TA of 1400 ms (gap = 10.6 s). All other acquisition parameters were identical to the localizer runs. Tone sequences were presented 5 or 6 s before each data acquisition point. Every run contained 32 tone sequence presentations and lasted for ∼9 min. The whole experiment consisted of nine runs. Each scanning session lasted for 2.5 h.

Data preprocessing. Functional and anatomical images were analyzed in BrainVoyager QX (Brain Innovations) and using custom MATLAB scripts. The ratio between anatomical T₁w images and PDw images was computed to obtain unbiased anatomical images (Van de Moortele et al., 2009). The unbiased anatomical data were normalized in Talairach space (Talairach and Tournoux, 1988) and resampled (with sinc interpolation) to a resolution of 0.5 mm isotropic. A surface reconstruction of each individual hemisphere was obtained by segmenting the gray–white matter boundary. Preprocessing of functional data consisted of slice-scan-time correction (with sinc interpolation), temporal high-pass filtering (removing drifts of five cycles or fewer per run), 3D motion correction (with trilinear/sinc interpolation and aligning each volume to the first volume of functional run 1), and temporal smoothing (two consecutive data points). Geometric distortions were corrected using scans with the

opposite phase encoding polarities in BrainVoyager's plugin COPE. Functional data were coregistered to the anatomical data and projected to the normalized space at a resolution of 1 mm isotropic. All statistical computations were performed on a single-subject level using a general linear model with a predictor for each center frequency for the tonotopic localizer and for every type of sequence for the experimental runs. Predictors were convolved with a standard two-gamma hemodynamic response function, peaking at 5 s after stimulus onset. By fitting the general linear model with the described predictors, we obtained for each voxel the response strength (i.e., β values) to every presented stimulus (i.e., frequency for the tonotopic localizer and every sequence type in the experimental runs).

Selecting ROIs. Based on macro-anatomical landmarks (sulci and gyri) and following the definition reported in Kim et al. (2000), the temporal lobe of each subject was divided into four ROIs in each hemisphere: Heschl's gyrus (HG), planum temporale (PT), planum polare (PP), and superior temporal gyrus (STG) (Fig. 3A). The core of the AC (PAC) was defined functionally as the high-low-high gradient on the tonotopic map, proceeding from posterior and medial parts of the HG to the anterior and lateral parts of HG (based on Moerel et al., 2014; Fig. 3B). We combined data for each ROI across the two hemispheres for all statistical analyses performed.

Mapping frequency and sequence preferences. Statistical activation maps were computed for the localizer runs and for the experimental runs. We restricted our analyses to voxels that were within the anatomically de-fined ROIs and showed a significant response (false discovery rate cor-
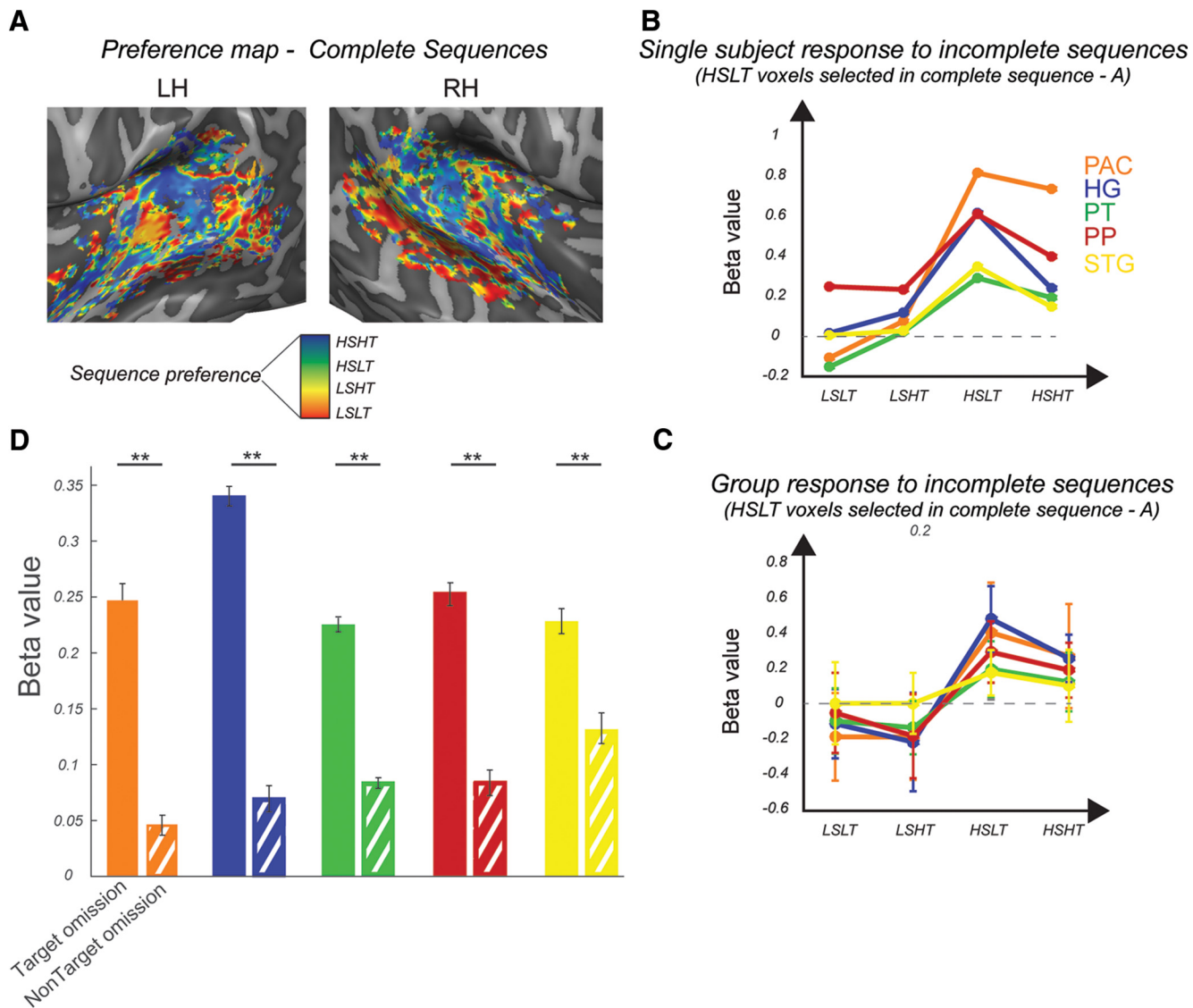
Figure 4. Strength of responses is preserved for targets in incomplete sequences. ***A***, Best-sequence map to presentation of complete sequences in a representative subject. Each voxel was tagged with the sequence that elicited its highest response and the color was chosen correspondingly. ***B***, Strength of responses to incomplete sequence in voxels that respond strongest to the HSLT complete sequences in ***A*** across ROIs in a representative subject and group (***C***). In both cases, the preference is preserved during omissions. ***D***, Strength of responses to incomplete TOs and NTOs. Voxels respond more strongly to presentation of the incomplete part of their complete preferred sequence than a different incomplete sequence. **significance below $p < 0.01$.

rected for multiple comparisons; $q = 0.05$) to sounds presented in the localizer run. Tonotopic maps were obtained using "best frequency mapping" (Formisano et al., 2003), in which we determined the tone frequency that elicited the highest fMRI response in every AC voxel and color coded the voxel accordingly. Using the localizer data, we obtained tonotopic maps using all nine center frequencies. Additionally, we restricted the analysis to only the four center frequencies corresponding to the target tones. Similarly, maps of sequence preference were obtained by color coding each voxel according to the sequence eliciting the strongest response (i.e., splitting all voxels into LSLT-, LSHT-, HSLT-, or HSHT-preferring voxels). This was performed separately for complete sequences and sequences with target omissions (incomplete sequences). Sequence preference and tonotopic maps were projected to the inflated hemispheric surfaces and their similarities were characterized by computing spatial correlations. Note that, before best frequency mapping, the voxelwise distribution of responses to all stimuli (e.g., the nine center frequencies in the tonotopic localizer, the four complete sequences, or the four omission sequences in the main experiment) was spatially normalized (z-scored) to remove any bias in the responses to one of the stimuli (Formisano et al., 2003).

Characterizing responses to omissions. We examined the specificity of prediction signaling in several ways considering both the average response strength in ROIs and the spatial pattern of responses across all voxels in ROIs.

First, to establish that predictive responses to the target were truly driven by the expectations formed by the first three sequential tones, we compared (one-way ANOVA) the strength of every ROI's response to incomplete orderly and scrambled HSLT sequences in HSLT voxels as defined on the basis of the response to complete sequences. This analysis was performed using the data of five subjects in which the scrambled control sequences were presented.

Second, using data from all our participants, we examined the frequency specificity of the prediction signal. This analysis was performed separately for each anatomical ROI. We defined groups of voxels based on response to complete sequences (i.e., the complete sequence evoking the highest response, resulting in LSLT, LSHT, HSLT, and HSHT voxels). In each functionally defined group of voxels, we averaged the β weights estimated in response to the presentation of incomplete sequences, and divided the incomplete sequences into target omissions (TOs), which represent the omission sequence corresponding to the voxels' preferred complete
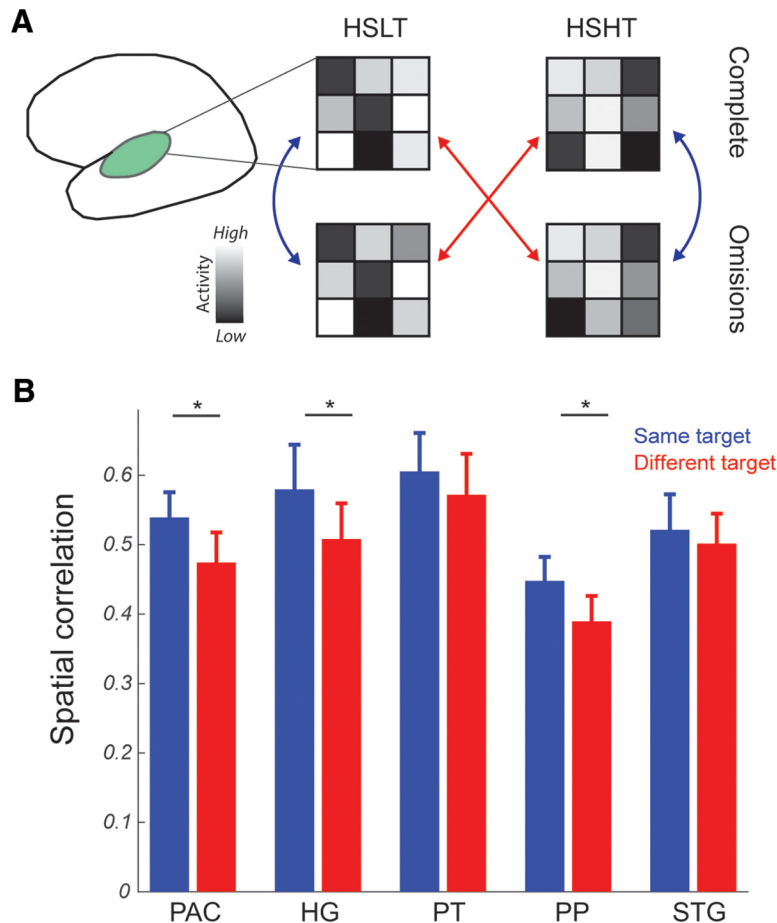
Figure 5. Spatial correspondence for sequences with the same predicted or perceived target. ***A***, Schematic depiction of the performed analyses. We compared patterns of voxels activation (i.e., β values) for complete and incomplete sequences with the same target (i.e., blue arrows) or across different targets (e.g., complete HSLT and incomplete HSHT target; red arrows). ***B***, Sequences with the same target demonstrated higher spatial correlation than sequences with different targets in the PAC, HG, and PP. *Significance below $p < 0.05$.

sequence, and nontarget omissions (NTOs), which constitute the other three incomplete sequences. For instance, for voxels responding the strongest to HSLT complete sequences, HSLT incomplete sequences are the TOs, whereas LSLT, LSHT, and HSHT omissions are grouped into NTOs. Responses to TOs and NTOs for all functionally defined groups of voxels (LSLT, LSHT, HSLT, and HSHT voxels) were averaged and compared in each anatomical ROI.

We also investigated whether the spatial patterns activation evoked by incomplete sequences correspond to the patterns of activation in complete sequences across in all anatomical ROIs. Specifically, we examined the spatial correlation between the activation patterns of matching pairs of complete/incomplete sequences (i.e., sequences with the same target such as HSLT complete and incomplete sequences) and compared it with the correlations between the activation patterns of nonmatching pairs of incomplete/complete sequences (e.g., HSLT complete and HSHT incomplete sequence). Finally, after having considered the spatial pattern of responses for each sequence individually, we examined the correlation between tonotopic maps obtained with standard tonotopic mapping (i.e., presenting only the target frequencies) and the sequence (i.e., complete and incomplete) preference maps.

Results

Regular sequences elicit prediction signaling

We first investigated whether expectations elicited by regular sequences create predictive responses in auditory areas by comparing responses to incomplete ordered and scrambled sequences (with the same frequency content). We hypothesized that if expectation drives responses in auditory cortical areas, then or-

dered incomplete HSLT sequences would elicit more activation than the incomplete scrambled sequence. As predicted, there was a significant main effect of sequence ($F_{(1,4)} = 8.213$, $p = 0.045$), with ordered HSLT sequences eliciting higher responses than scrambled sequences (Fig. 3C). There was no effect of ROI ($F_{(4,16)} = 1.07$, $p = 0.40$). Responses to orderly sequences were significantly stronger than to scrambled sequences in primary auditory regions (HG and PAC), PT, and PP, but the STG responses did not show a significant difference between the two conditions. However, the interaction between the ROI and sequence was not significant ($F_{(4,16)} = 1.65$, $p = 0.21$). Altogether, this suggests that regular sequences contain high target expectancy that elicits prediction signaling across the whole AC.

Omission responses contain information about the expected target

Next, we investigated whether this increased response for ordered incomplete sequences contains information about the frequency of the omitted target. This was done in two ways: (1) by comparing the strength of responses elicited to different incomplete sequences in voxel groups defined based on their responses to complete sequences and (2) by examining the spatial correspondence of complete targets with their omission counterparts. All tests were performed at the group level across all participants ($N = 8$).

Figure 4B depicts the colored map obtained from responses to the complete sequences of a representative subject. In voxels that prefer the HSLT complete sequences, sequence preference was maintained when the targets were omitted (i.e., incomplete HSLT sequences evoke the strongest responses) both for an individual subject (Fig. 4B) and on the group level (Fig. 4C). To summarize across all targets and sequences, we collapsed all incomplete sequences into TOs and NTOs (Fig. 4D) and performed an ROI × target type (TO vs NTOs) ANOVA. The main effect of target was significant ($F_{(1,7)} = 42.79$, $p = 3.21e^{-4}$). The effect of the region was not significant ($F_{(4,28)} = 0.78$, $p = 0.55$), whereas the interaction between the two factors was significant ($F_{(4,28)} = 7.11$, $p = 4.40e^{-4}$). *Post hoc* t tests revealed a highly significant effect, with TOs eliciting higher responses than NTOs in all ROIs (PAC: $t_{(7)} = 3.98$, $p = 0.0026$, HG: $t_{(7)} = 7.93$, $p = 4.81e^{-5}$, PT: $t_{(7)} = 6.36$, $p = 1.9e^{-4}$, PP: $t_{(7)} = 5.34$, $p = 5.41e^{-4}$, STG: $t_{(7)} = 5.69$, $p = 3.72e^{-4}$). To test whether the effect of target type (TO vs NTO) varies across regions, we performed an ANOVA on a difference index between TO and NTOs. The main effect of region was significant ($F_{(4,28)} = 4.125$, $p = 9.4e^{-3}$). *Post hoc* t tests (corrected for Bonferroni) revealed that it was particularly the difference between the HG and the STG, with the HG having a significantly stronger effect of target type than STG, that was driving the effect ($t_{(7)} = 5.44$, $p = 0.003$). Overall, these analyses demonstrate responses that across the whole AC are stronger to the presentation of the incomplete counterpart of voxels' preferred complete se-
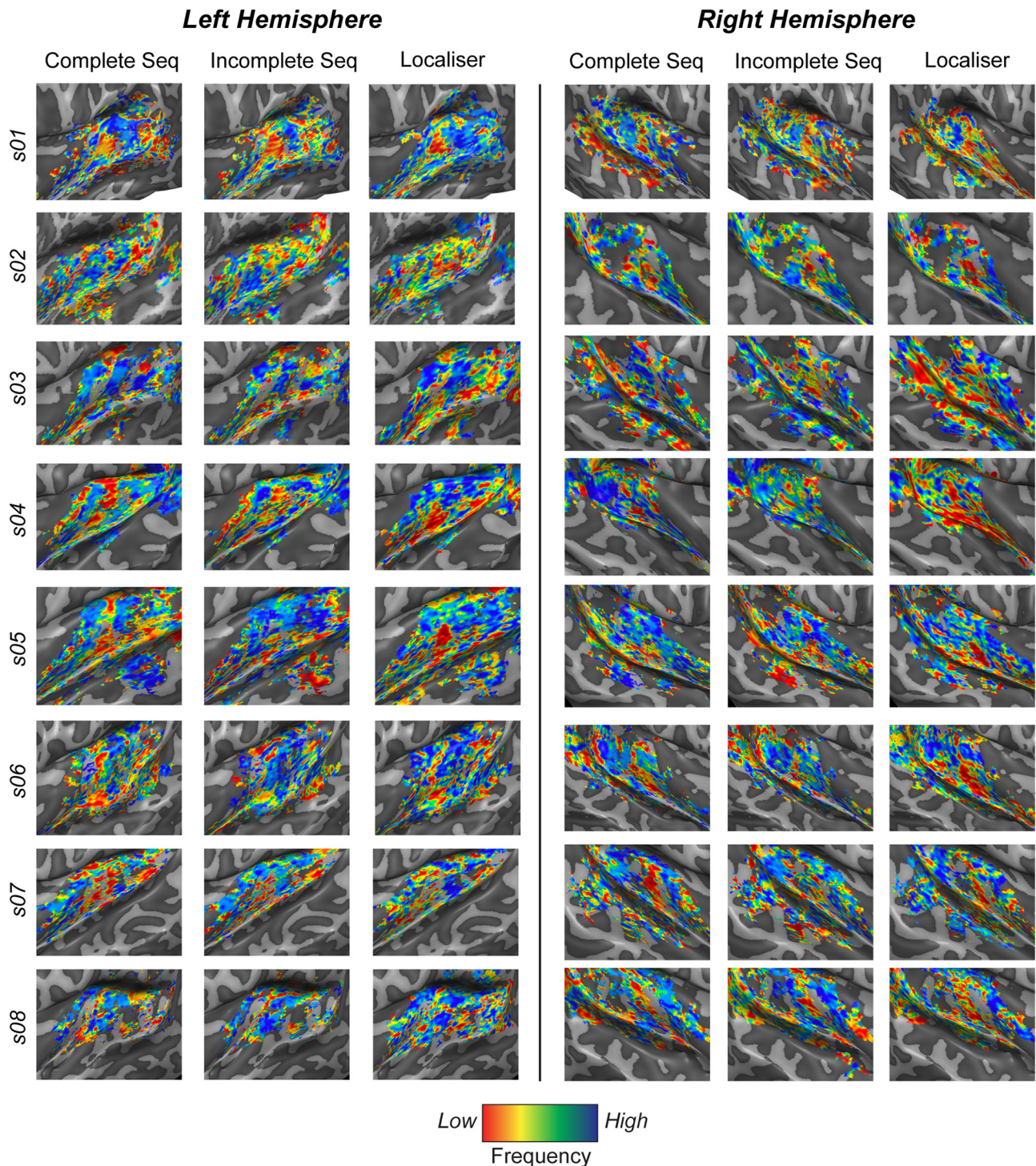
Figure 6. Tonotopic maps reconstructed from the activity to sequence presentation (Complete Seq, Omissions) and localizer frequencies (Localizer) for all subjects. Sequence preference maps were reconstructed by color coding the voxels independently of the complete and omission sequences that evoked the highest responses. Color codes were assigned depending on the frequency of the target in the sequence. The best-frequency maps were obtained by color coding each voxel depending on which of the four localizer frequencies elicited the highest voxel activation. Obtained surface maps are highly similar and reflect the characteristic tonotopic gradients in the AC.

quence compared with the nonpreferred incomplete sequence, especially in more primary regions.

To ensure that the main effect of sequence in ANOVA was not driven by the difference in the spectral range of the sequences (i.e., low sequences LSLT and LSHT vs high sequences HSLT and HSHT), we split the sequences into low and high pairs and repeated the ANOVA analysis. This is a more stringent analy-

sis because the two pairs differ only in the frequency of the (omitted) target, whereas all presented tones are the same. The main effect of sequence remained significant in both cases (low: $F_{(1,7)} = 22.59, p = 0.0021$; high: $F_{(1,7)} = 39.00, p = 4e^{-4}$). This argues against the possible explanation of spectral context and in favor of explanation that the prediction of the target alone modulates activation in the AC.

Importantly, because each of the sequence pairs (e.g., HSLT and HSHT) contained identical physical stimuli in the incomplete sequences, but the sequences differed in the direction of sound presentation (e.g., descending for HSLT and ascending for HSHT), the increased responses to TOs compared with NTOs could be explained by a preference for the direction of the presented sounds rather than differences in target expectation. To investigate this, incomplete sequences were defined as sequences with preferred and nonpreferred directions. For example, for a voxel with the strongest responses to the HSLT sequence, the descending sequences (LSLT and HSLT) were considered to be the preferred direction sequences, whereas the ascending sequences (LSHT and HSHT) were the nonpreferred direction sequences. Comparison of β values between the preferred and nonpreferred directions revealed that none of the ROIs was sensitive to the direction of the presented sequence.

We also examined the selectivity of responses to incomplete sequences by correlating activity patterns evoked by a given complete sequence with the activity pattern of corresponding incomplete sequence (e.g., Fig. 5A, blue arrows) and compared the correlation value with the spatial correlation between nonmatching complete–incomplete sequences (e.g., Fig. 5A, red arrows). As predicted, a sequence (same/different target) $\times$ ROI ANOVA revealed a significant effect of sequence ($F_{(1,7)} = 3.788$, $p = 0.046$), demonstrating that the spatial correspondence of activity patterns is higher when the same specific target frequency is perceived/expected than when the target frequency differs. There was no significant effect of ROI ($F_{(4,28)} = 2.047$, $p = 0.1149$), but there was a significant interaction between the two factors ($F_{(4,28)} = 4.125$, $p = 0.0094$). *Post hoc t* tests on correlation values in each region separately showed that the matching complete–omission sequence corresponded significantly more than the nonmatching pair in the PAC ($t_{(7)} = 2.274$, $p = 0.029$), HG ($t_{(7)} = 2.462$, $p = 0.022$), and PP ($t_{(7)} = 2.009$, $p = 0.042$) (Fig. 5B). Altogether, these results suggest that the spatial distribution of voxel responses in complete and incomplete sequences is influenced by the target, which is either perceived or just expected.

Prediction signaling is tonotopic

From the previous analysis, it follows that the predictive responses to omissions of targets should respect the tonotopic organization of the AC. We tested for this hypothesis by comparing the surface maps of sequence preference for complete and incomplete sequences with the tonotopic maps obtained with standard tonotopic mapping. Importantly, all three maps revealed the typical mirror-reversed topographic organization in the AC in all subjects (Fig. 6). The correlation between pairs of surface maps was calculated separately for each subject and separately for each ROI. Because ROIs differ in the number of vertices on the cortical surface, we calculated correlations by randomly resampling each ROI map using a subset of vertices (i.e., 50 vertices less than the smallest number of vertices across ROIs). The average of all 10,000 correlations was *z*-transformed and the group average correlation values are reported in Table 1. The average correlation (Pearson's *r*) between complete sequence map and tonotopy across all ROIs was 0.19. The correlation between tonotopic map and the map of best omitted sequence preference was 0.16. The significance of correlation values was determined separately for each ROI using one-sample permutation testing by performing all possible sign switches of the individual correlation values. The correspondence between pairs of maps was significant in all regions and for both combinations, with the exception of the PAC for the correspondence between tonotopy and omission sequence maps,

Table 1. Similarities between sequence preference maps and tonotopic maps across ROIs

	PAC	HG	PT	PP	STG
Complete tonotopy	0.22*	0.30*	0.13*	0.19*	0.10*
Omission tonotopy	0.18	0.29*	0.13*	0.12*	0.09*

The significance of the observed correlation value was tested at the group level using permutations (i.e., by changing the sign of the individual correlation values and counting the number of times that the actual group mean exceeded the group mean obtained in all possible permutations).

*$p < 0.05$.

which was probably due to higher intersubject variability. The correspondence between omission and tonotopic maps is even more remarkable considering how drastically different the stimuli in the two task are; that is, single-amplitude-modulated tones in the localizer versus the same tone frequency omitted in harmonic sequences. This makes the correspondence between them a more powerful indication of a tonotopic encoding of the target frequency.

Overall, our results on average responses and spatial patterns results suggest that predictive coding of the expected target is encoded in the neuronal populations corresponding to the frequency of the expected tone.

Discussion

We investigated whether the human AC forms frequency-specific tone predictions by examining the fMRI activity elicited in response to sequences with a tone omission. Most studies so far have employed single tone repetitions, but we used sequences with different tones, which allowed us to inspect the frequency specificity of predictive signaling.

We first investigated whether responses are higher when the expectation over the target tone is stronger. We compared responses to ordered and scrambled incomplete sequences since our behavioral paradigm indicated that orderly sequences increase expectation of the target frequency compared to scrambled sequences. We observed stronger fMRI responses to presentations of orderly incomplete sequences than scrambled sequences, indicating that AC encodes the expectation of the target frequency. These results agree with previous behavioral work demonstrating regular sequences inducing higher expectation than scrambled ones (Lange, 2009), as well as reports on increased AC activity in cases of stronger expectation or regularity violation (Todorovic et al., 2011; Nazimek et al., 2013; SanMiguel et al., 2013).

Next, we tested the specificity of prediction signaling. We demonstrated that the response to incomplete sequences that predict a target is higher than the response to incomplete sequences which predict a nontarget (where the target was defined on the basis of the response to complete sequences). We additionally refuted a possible alternative explanation of voxel responses reflecting the direction of the presented tones in the incomplete sequences. We also examined the specificity of prediction signaling, demonstrating that the spatial patterns of responses of incomplete sequences is more similar to the pattern elicited by the complete sequence which share the same target than complete sequences with a different target. These analyses indicate that, in AC, incomplete sequences are processed similarly to complete sequences that share the same target, analogous to the "filling-in" illusion, which has also been shown in an auditory EEG study (Chouiter et al., 2015). Importantly, our results did not only demonstrate that voxel preferences for the complete and incomplete sequences correspond highly.

Finally, we showed that prediction signaling respects the tonotopic organization of the AC by demonstrating that the pattern of responses elicited by best-sequence maps (with missing targets) corresponds to the pattern (i.e., tonotopy) obtained with a tono-

topic localizer, a result that is similar to visual studies on retino-topic content of predictions (Smith and Muckli, 2010; Kok et al., 2014; Muckli et al., 2015).

Altogether, our results present evidence for tonotopic encoding of target prediction in incomplete sequences across the human AC.

Recent studies have shown that, similar to our frequency-specific result in omission trials, the response to multiple auditory stimuli also demonstrates frequency specificity for the stimulus that is in the attentional spotlight (Da Costa et al., 2013; Riecke et al., 2017). It might be that the mechanisms of frequency-specific predictions and auditory attention have shared substrates.

An important question for future research is how predictions are propagated across the auditory hierarchy. One option is that higher cortical areas signal the anticipated tone to the tonotopi-cally relevant regions of primary areas (as suggested by Bastos et al., 2012). Another possibility is that predictions are constructed in a feedforward manner, with primary areas extracting regularities in basic stimulus properties (e.g., tone frequencies), whereas higher areas build on these codes and construct more abstract representations (Nelken, 2004). The current study cannot distinguish these two possibilities. Other techniques using higher temporal resolution (e.g., electrophysiology) or examining activations at various cortical depths could provide valuable insights into the interplay between perceptual and predictive signal propagation.

Another interesting question is how predictive coding is shaped while learning the statistical regularities. The stimuli used in our paradigm contained strong regularities that are enforced through lifetime exposure to sounds in our environments. Another option for inspecting learned sequencing operations is using an implicit sequence learning (an "artificial grammar"). It has been shown recently that violations of regularities modulate neural oscillatory coupling in both monkeys and humans in the PAC (Kikuchi et al., 2017). Although it is insightful to determine how learned relationships modulate sensory processing, it would be very informative to study how predictive processing develops with learning.

In summary, our results corroborate the notion that the complexity of responses in primary sensory areas stretches beyond simple stimulus detection. Low-level areas are sensitive to environmental regularities and, by anticipating upcoming signals, this enables efficient sensory processing. Our paradigm used simple stimuli, but it allows for conceptual extension to more complex processes such as speech and music. The finding of content-specific predictions may suggest that these processes are required for our brain to process sounds encountered in our everyday lives successfully.

References

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870–888. CrossRef Medline

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ (2012) Canonical microcircuits for predictive coding. Neuron 76:695–711. CrossRef Medline

Bendixen A, Schröger E, Winkler I (2009) I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. J Neurosci 29:8447–8451. CrossRef Medline

Bendixen A, SanMiguel I, Schröger E (2012) Early electrophysiological indicators for predictive processing in audition: a review. Int J Psychophysiol 83:120–131. CrossRef Medline

Chouiter L, Tzovara A, Dieguez S, Annoni JM, Magezi D, De Lucia M, Spierer L (2015) Experience-based auditory predictions modulate brain activity to silence as do real sounds. J Cogn Neurosci 27:1968–1980. CrossRef Medline

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. CrossRef Medline

Formisano E, Kim DS, Di Salle F, van de Moortele P, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. Neuron 40:859–869. CrossRef Medline

Friston K (2005) A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360:815–836. CrossRef Medline

Gregory RL (1980) Perceptions as hypotheses. Philos Trans R Soc Lond B Biol Sci 290:181–197. CrossRef Medline

Helmholtz H (1860) Theorie der Luftschwingungen in Röhren mit offenen Enden. J für die reine und angewandte Mathematik 57:1–72.

Kaas JH, Hackett TA (1998) Subdivisions of auditory cortex and levels of processing in primates. Audiol Neurootol 3:73–85. CrossRef Medline

Kikuchi Y, Attaheri A, Wilson B, Rhone AE, Nourski KV, Gander PE, Kovach CK, Kawasaki H, Griffiths TD, Howard MA 3rd, Petkov CI (2017) Sequence learning modulates neural responses and oscillatory coupling in human and monkey auditory cortex. PLoS Biol 15:e2000219. CrossRef Medline

Kim JJ, Crespo-Facorro B, Andreasen NC, O'Leary DS, Zhang B, Harris G, Magnotta VA (2000) An MRI-based parcellation method for the temporal lobe. Neuroimage 11:271–288. CrossRef Medline

Kok P, de Lange FP (2014) Shape perception simultaneously up-and down-regulates neural activity in the primary visual cortex. Curr Biol 24:1531–1535. CrossRef Medline

Lange K (2009) Brain correlates of early auditory processing are attenuated by expectations for time and pitch. Brain Cogn 69:127–137. CrossRef Medline

MacKay DM (1957) Complementary descriptions. Mind 66:390–394.

Moerel M, De Martino F, Formisano E (2014) An anatomical and functional topography of human auditory cortical areas. Front Neurosci 8:225. CrossRef Medline

Muckli L, De Martino F, Vizioli L, Petro LS, Smith FW, Ugurbil K, Goebel R, Yacoub E (2015) Contextual feedback to superficial layers of V1. Curr Biol 25:2690–2695. CrossRef Medline

Näätänen R, Gaillard AW, Mäntysalo S (1978) Early selective attention reinterpreted. Acta Psychologica 42:313–329. CrossRef Medline

Nazimek JM, Hunter MD, Hoskin R, Wilkinson I, Woodruff PW (2013) Neural basis of auditory expectation within temporal cortex. Neuropsychologia 51:2245–2250. CrossRef Medline

Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. Curr Opin Neurobiol 14:474–480. CrossRef Medline

Nelken I (2012) Predictive information processing in the brain: the neural perspective. Int J Psychophysiol 83:253–255. CrossRef Medline

Peirce JW (2007) PsychoPy: psychophysics software in Python. J Neurosci Methods 162:8–13. CrossRef Medline

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat Neurosci 2:79–87. CrossRef Medline

Riecke L, Peters JC, Valente G, Kemper VG, Formisano E, Sorger B (2017) Frequency-selective attention in auditory scenes recruits frequency representations throughout human superior temporal cortex. Cereb Cortex 27:3002–3014. CrossRef Medline

Rohrmeier MA, Koelsch S (2012) Predictive information processing in music cognition: a critical review. Int J Psychophysiol 83:164–175. CrossRef Medline

SanMiguel I, Widmann A, Bendixen A, Trujillo-Barreto N, Schröger E (2013) Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. J Neurosci 33:8633–8639. CrossRef Medline

Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. Proc Natl Acad Sci U S A 107:20099–20103. CrossRef Medline

Talairach K, Tournoux P (1988) Co-planar stereotaxic atlas of the human brain. New York: Thieme.

Todorovic A, van Ede F, Maris E, de Lange FP (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. J Neurosci 31:9118–9123. CrossRef Medline

Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. Nat Neurosci 6:391–398. CrossRef Medline

Van de Moortele PF, Auerbach EJ, Olman C, Yacoub E, Uğurbil K, Moeller S (2009) T1 weighted brain images at 7 tesla unbiased for proton density, T2 contrast and RF coil receive B1 sensitivity with simultaneous vessel visualization. Neuroimage 46:432–446. CrossRef Medline