Reexamining the Evidence for a Pitch-Sensitive Region: A Human fMRI Study Using Iterated Ripple Noise

Daphne Barker1,2, Christopher J. Plack2 and Deborah A. Hall1,3,4

1MRC Institute of Hearing Research, Nottingham, NG7 2RD, UK, 2Division of Human Communication and Deafness, The University of Manchester, Manchester, M13 9PL, UK, 3NIHR National Biomedical Research Unit in Hearing, Nottingham, NG1 5DU, UK and 4Division of Psychology, School of Psychological Sciences, Nottingham Trent University, Nottingham, NG1 4BU, UK

Keywords: Heschl’s gyrus, modulation, planum polare, planum temporale, salience

Introduction

Pitch is one of the primary auditory percepts. Pitch can be defined as the sensation whose variation is associated with musical melodies (Plack and Oxenham 2005) and is one of the most important perceptual features in music. Pitch also plays an important role in spoken language by providing lexical information in tonal languages and prosodic information in nontonal languages, and it is one of the main perceptual cues for segregating the sources of different concurrent sounds.

Despite a large body of research examining the neural correlates of pitch perception, debate continues as to whether there exists an area of auditory cortex that represents the percept itself more so than the physical attributes responsible for its creation. The same pitch can be elicited by sounds with different spectral and temporal characteristics, and this has led many researchers to postulate the existence of neurons responsive to the perceptual property “pitch.” It has been suggested that any potential pitch center should satisfy 4 criteria: 1) It should respond selectively to pitch compared with an appropriately matched noise. 2) Activity should still be present after the elimination of peripheral effects, such as cochlear distortions. 3) It should respond to all pitch-evoking stimuli, regardless of physical attributes. 4) Activity should increase with increasing pitch salience (Hall and Plack 2009).

Note that these criteria do not imply that the pitch center should respond exclusively to pitch.

A landmark primate study used single-unit extracellular recordings in the vicinity of primary auditory cortex to find such a region (Bendor and Wang 2005). This study identified a cluster of neurons in the anterolateral border of primary auditory cortex that met all 4 criteria for a pitch center. However, there are a number of problems involved in translating such results into the domain of human cognitive neuroscience. For example, Bendor and Wang (2005) were recording from a population of 131 individual units, only 51 of which exhibited a significant pitch response. In contrast, functional magnetic resonance imaging (fMRI) detects changes in blood oxygenation levels that occur as an indirect consequence of population neural activity. Hence, it is unclear that the same effects should necessarily be observed using the 2 different methods even if both species possess pitch sensitivity at the neuronal level.

Regions of pitch-related auditory activity are often identified by contrasting pitch-evoking stimuli with stimuli that are matched in terms of spectral content but do not evoke a pitch percept. One contrast of this type is that of an iterated ripple noise (IRN) pitch stimulus and a Gaussian noise control. IRN is created by generating a sample of Gaussian noise, imposing a delay to the noise, and adding (or subtracting) the delayed version back to (or from) the original (Yost 1996). The pitch sensation of IRN is determined by the reciprocal of the imposed delay. IRN can be high-pass or band-pass filtered so that it contains no perceptually resolvable spectral peaks at harmonic frequencies. Instead, the pitch percept is wholly determined by fast-rate temporal regularities in the stimulus. Another appeal is that pitch salience can be easily manipulated by changing the number of delay-and-add (or subtract) iterations. Increasing the number of iterations increases the salience of the pitch (Yost 1996). A physical correlate of pitch salience is the height of the first peak in the autocorrelation function, which increases with increasing iterations, and correlates strongly with the measured pitch salience (Yost 1996).

One of the earliest human neuroimaging studies of pitch used positron emission tomography to examine the effect of pitch salience for IRN by manipulating the number of delay-and-add iterations (Griffiths et al. 1998). The authors concluded that activation in an area of auditory cortex encompassing parts of lateral Heschl’s gyrus (HG) increased with increasing number of iterations. However, the effect seems to have been determined by the difference between the 0-iteration condition (i.e., Gaussian noise) and the pitch-evoking conditions (1, 2, 4, 8, and 16 iterations). In other words, it is not clear that there
would have been a significant linear relationship if the 0-
iteration noise had been excluded. A number of human
neuroimaging studies have since contrasted IRN with a spe-
cially matched noise control and have demonstrated significant
activation in lateral HG (e.g., Patterson et al. 2002; Hall et al.
2006; Hall and Plack 2009). Based on the animal and human data,
Bendor and Wang (2006) suggested that lateral HG is a good
candidate for a human pitch center.

At present, the evidence for lateral HG as a pitch center is somewhat mixed (Penagos et al. 2004; Chait et al. 2006; Hall and Plack 2007, 2009; García et al. 2010; Puschmann et al. 2010; Barker et al. 2011). When results using a wide range of pitch-evoking stimuli are more closely scrutinized, it appears that activity in lateral HG is not crucial for pitch coding. For example, Hall and Plack (2009) found that planum temporale (PT) was typically responsive to many different pitch-evoking stimuli, including tone-in-noise, wideband harmonic complex, and Huggins pitch. In contrast, lateral HG was found to respond no differently to these stimuli than to the spectrally matched noise control. However, consistent with the earlier results, lateral HG did respond significantly to 2 types of IRN stimulus compared with the corresponding spectrally matched noise control. Although the IRN-related response was highly consistent across listeners (>50%), the activation in PT produced by the other pitch-evoking stimuli was less so (<25%). This lack of consistency led the authors to conclude that no one region could reliably be assigned the label of "pitch center."

To explain the discrepancy in the spatial distribution and consistency of activity for the different pitch contrasts, Hall and Plack (2009) demonstrated that IRN contains acoustic features unrelated to pitch that are not present in the other pitch-evoking stimuli nor in the noise control. The iterative delay-and-
add process introduces unpredictable spectrotemporal varia-
tions that occur over a longer time scale (hundreds of milliseconds) than the temporal regularity responsible for pitch (tens of milliseconds). Increasing the number of delay-and-add iterations in IRN increases both the pitch salience and the depth of the modulations across time and frequency, hence increasing the perceptual salience of the modulations as well as the pitch salience. This finding supported a suggestion by de Cheveigne (2007) that the "spectral ripple" in IRN could set it apart from other pitch-evoking stimuli and that these additional features could explain the disparity in results from studies using IRN and those using different pitch-evoking stimuli. Further support for the suggestion that the robust activation seen for IRN might be a result of the modulations, rather than the pitch of IRN, was provided by results from a recent fMRI study (Schönwiesner and Zatorre 2009). These results indicated strong selectivity to specific properties of dynamic spectral ripples in HG and around Heschl's sulcus.

To investigate the possibility that fMRI effects attributed to pitch may instead be due to spectrotemporal fluctuations, we designed a novel type of stimulus that includes the slowly varying spectrotemporal fluctuations of IRN but does not include the fine temporal structure responsible for the pitch percept. We have called this new stimulus "no-pitch IRN" (IRNo). To verify that the spectrotemporal fluctuations are perceptible and that their salience is dependent on iterations, a psychophysical modulation discrimination task was performed with IRNo. In addition, a pitch discrimination task was performed with IRN (2, 4, 16, and 64 delay-and-add iterations) to confirm previous results showing that pitch salience depends on the number of iterations. To disentangle the potential cortical effects of pitch strength and spectro-
temporal fluctuations, auditory cortical responses to IRN and IRNo were measured using fMRI. For a pitch-specific response, we expect 1) a significant difference between IRN and noise but not between IRNo and noise, 2) a dependency of response on the number of iterations for IRN but not IRNo, and 3) a significant difference between corresponding IRN and IRNo conditions, at least for large numbers of iterations. For a modulation-specific response, we expect 1) a significant difference between IRN and noise and IRNo and noise, 2) a dependency on iterations for both IRN and IRNo, and 3) no significant difference between corresponding IRN and IRNo conditions.

Materials and Methods

Listeners

Sixteen listeners (11 males, 5 females; age range 20–47 years) with normal hearing (≥20 dB hearing level between 250 and 8 kHz) took part in both the psychophysical and fMRI testing. All listeners were right-handed (laterality index = 50, Oldfield 1971). Seven listeners were musically trained between grade 2 and grade 7 (7 ≤ 02, 07, 18, 19, 22, 23, and 25), while 5 others reported informal musical experience (self-taught/ungraded, 7 ≤ 05, 09, 16, 17, 21). Fourteen additional participants were included in the psychophysical testing for IRN and 10 additional participants for IRNo. These participants were recruited as part of 2 separate undergraduate projects, and all were students of the University of Nottingham, who gave written informed consent. None had a history of any neurological or hearing impairment. All listeners gave written informed consent, and the study was approved by the Medical School Research Ethics Committee, University of Nottingham.

Stimuli

Dioctic IRN stimuli were generated by a delay-and-add process performed on a Gaussian noise. The noise was band-pass filtered (1-2 kHz) to remove low-numbered harmonics that are resolved (i.e., separated out) by the peripheral auditory system. A delay of 10 ms was imposed before adding the delayed noise back to the original sample, generating a stimulus with a nominal fundamental frequency (f0) of 1000 Hz. This process was repeated 2, 4, 16, or 64 times using the output of the previous delay-and-add iteration as the input to the following delay-and-add iteration to create all 4 IRN stimulus conditions, each with a pitch corresponding to a 100-Hz tone. To make IRNo, a conventional IRN stimulus was generated as above. The IRN was sampled using a rectangular window with a duration equal to the IRN delay (10 ms). A fast Fourier transform (FFT) was used to generate the magnitude and phase spectra of the sample, and the phase of the components was randomized. An inverse FFT was then used to regenerate the time representation. The sampling window was advanced by half of the IRN delay (5 ms) and the process repeated. (Subsequent analysis suggests that the overlap was not necessary, i.e., the window could have been advanced by 10 ms with little effect on the stimulus characteristics.) The processed samples were overlapped and added (preserving the start times of the samples), adjusted to a spectrum level of 52 dB SPL and gated to produce a time waveform with a 580-ms steady state and 10-ms raised cosine ramps. The phase randomization process removes any correlation in the fine structure between samples, obliterating the harmonic structure and the pitch cue. Supplementary Figure 1 demonstrates that the height of the first peak in the autocorrelation function increases with increasing number of iterations for IRN and that removing the fine structure regularity eliminates the prominent peaks in the autocorrelation function for IRNo. The slowly varying broad spectral features are present in both stimulus types. These fluctuations are apparent when the spectrogram of IRN is smoothed in both time and frequency domains to remove any fine structure (Fig. 1). The process was repeated on all the IRN stimuli to produce 4 IRNo conditions. All stimuli included a noise masker, low-

746 A Human fMRI Study Using Iterated Ripple Noise: \( \cdot \) Barker et al.

Downloaded from https://academic.oup.com/cercor/article-abstract/22/4/745/416425
by guest on 14 May 2018
Cochlear Representations

To illustrate the representation of the stimuli in the peripheral auditory system, the stimuli were passed through a computational model (Plack et al. 2002). The model included a simulation of the middle ear and a nonlinear auditory filterbank that simulated the compressive frequency selective properties of the basilar membrane in the cochlea. The temporal response of the filterbank was smoothed by a sliding temporal integrator. The parameters of this version of the model were taken from Plack (2007). The spectrograms in Figure 1 show the output of the model as a function of time and filter center frequency for each condition, using different samples of IRN and IRNo for each repetition, and the mean of the SD of the level fluctuations, and 95% confidence intervals of this mean, were calculated. The results are shown in Figure 2. Fluctuation depth increases with number of iterations. The IRN and IRNo stimuli are quite closely matched, although the fluctuation depth for the IRN stimuli is a little greater than that for the IRNo stimuli at 16 and 64 iterations. The fluctuation depth for the processed noise control is slightly greater than that for the unprocessed noise control. Overall, these differences are small and so we did not expect them to markedly affect the fMRI results.

Psychophysical Testing

Prior to the scanning session, each participant performed a pitch discrimination task and a modulation discrimination task to measure the sensitivity to the pitch and modulation cues. Psychophysical testing was carried out in a sound-attenuating booth, and stimuli were delivered through Sennheiser HD 480 II headphones. Stimuli were presented through custom-made software that is supported by the Matlab platform (The MathWorks). Pitch discrimination thresholds output appears similar for IRN and IRNo stimuli, indicating that the processing used to generate the IRNo was successful in preserving the spectrotemporal features.

To provide a quantitative measure of these features, for each spectrogram, the standard deviation (SD) of the level fluctuations (in dB) was calculated across the whole sequence of center frequencies between 1 and 2 kHz. The calculation was performed 50 times for each condition, using different samples of IRN and IRNo for each repetition, and the mean of the SD of the level fluctuations, and 95% confidence intervals of this mean, were calculated. The results are shown in Figure 2. Fluctuation depth increases with number of iterations. The IRN and IRNo stimuli are quite closely matched, although the fluctuation depth for the IRN stimuli is a little greater than that for the IRNo stimuli at 16 and 64 iterations. The fluctuation depth for the processed noise control is slightly greater than that for the unprocessed noise control. Overall, these differences are small and so we did not expect them to markedly affect the fMRI results.

Figure 1. Simulated cochlear representations of IRN (top row) and IRNo (bottom row) in the form of spectrograms. The analysis smoothes the representation in both time and frequency domains to remove any fine structure. The bottom row shows processed version of the IRN stimuli in the top row.

Figure 2. The standard deviation (SD) of the cochlear representations of IRN and IRNo as a function of number of iterations, averaged over 50 replications. The values are measures of the fluctuation depth of the slowly varying modulations. The error bars show 95% confidence limits.
were measured for IRN using a 3 alternative forced-choice, two-down, one-up, adaptive procedure that targeted 70.7% performance (Levitt 1971). On the first trial, the f0 difference was 20% (20 Hz). The percent difference increased or decreased by a factor of 2 for the first 4 reversals and by a factor of 1.414 for the final 12 reversals. Discrimination threshold for each run was taken as the geometric mean of the f0 difference at the final 12 reversals. The percent difference was not allowed to increase above 200% (200 Hz).

Modulation discrimination performance was measured for IRN0 using a 3-alternative forced-choice “odd-one-out” paradigm in which participants were presented with 3 stimuli, 2 of which were different samples of the Gaussian noise control and one of which (chosen at random) was IRN0. The task was to select the interval that contained IRN0. Fifty trials were presented in each block, and the percentage of correct responses was taken. Responses were recorded and stored electronically. On each trial, feedback was given via a green (correct) or red (incorrect) light on the software interface. Participants completed 3 training runs for IRN and IRN0 with 16 iterations, and participants who did not perform above chance after the third run were excluded from further testing. There were 4 testing runs each for IRN and IRN0 with 2, 4, 16, and 64 iterations; pitch discrimination thresholds were taken as the geometric mean threshold of the last 4 runs.

fMRI Protocol

Scanning was performed on a Philips 3 T Intera Achieva using an 8-channel SENSE receiver head coil. A T1-weighted high-resolution (1 mm³) anatomical image (matrix size = 256 × 256, 160 sagittal slices, interscan interval = 7.8 ms, echo time = 3.7 ms) was collected for each subject. The anatomical scan was used to position the functional scan centrally on HG, and care was taken to include the entire superior temporal gyrus and to exclude the eyes. Functional scanning used a T2*-weighted echo-planar sequence with a voxel size of 3 mm³ (matrix size = 64 × 64, 32 oblique-axial slices, echo time = 36 ms). The scanning sequence was of a “sparse” type in which each set of 32 slices was clustered into an acquisition time of 1969 ms separated by an interscan interval of 7800 ms (Edmister et al. 1999; Hall et al. 1999). The fMRI response was measured at 2 fixed time points relative to each sound sequence; 7.3 s and 15.1 s after each stimulus onset. A SENSE factor of 2 was applied to reduce image distortions and a SoFone factor of 2 was used to reduce the acoustical scanner noise level by 9 dB. Functional data was acquired over 3 runs of 84 scans each and 1 run of 86 scans. Participants were requested to listen to the sounds but were not required to perform any task. A custom-built MR-compatible system delivered distortion-free sound using high-quality electrostatic head-phones (Sennheiser HE60 with high-voltage amplifier HEV70) that had been specifically modified for use during fMRI. An active noise control (ANC) device (Hall et al. 2009) was used for the first 7 sessions (#02, 05, 07, 09, 16, 17, 18), reducing the acoustical scanner noise by a further 35 dB at the main peak in the spectrum of the scanner noise (around 14 dB overall). For these listeners, 8 scans were appended to the beginning of the sequence in order to train the noise canceler. The ANC was not operative using subsequent sessions and so could not be used. We do not expect ANC to change the pattern of results in auditory cortex (see Blackman and Hall 2011), but effects of IRN and IRN0 were examined separately for the listeners who used the ANC and those who did not. Activation results for those experiencing ANC and those not experiencing ANC are reported in the Results section where appropriate.

Data Analysis

Images were analyzed separately for each of the 16 listeners using statistical parametric mapping (SPM5, http://www.fil.ion.ucl.ac.uk/spm). Preprocessing steps included realignment to correct for subject motion, normalization of individual scans to a standard image template, and smoothing with a Gaussian filter of 8 mm full-width at half-maximum (FWHM) for group analyses and 4-mm FWHM for incidence maps. Individual analyses were computed for the 4 runs, specifying the 2 stimuli types and the 4 iteration conditions and noise controls as separate regressors in the design.

First, the data for individual participants were analyzed using a first-level general linear model to assess the effects of interest with respect to the scan-to-scan variability. A high-pass filter removed any very low-frequency drifts in the time series data (up to 0.002 Hz). The resulting model estimated the fit of the design matrix (X) to the data (Y) within each voxel. Modeling yields parameter estimates (β), which represent the contribution of each effect of interest to the overall fMRI signal. For each participant, separate statistical contrasts were specified for individual sound conditions relative to the silent baseline (that was implicitly modeled in the design). A second-level (random effects) analysis contrasted the 2 control conditions (noise and processed noise) and confirmed that they elicited an equivalent brain response across auditory cortex. This fMRI result is consistent with the observation from Figure 2 that the 2 control signals were similar in terms of their fluctuation depth, and so the 2 conditions were combined for subsequent analyses to increase statistical power. The inputs for the second-level random effects analysis were therefore the contrast images for each IRN and IRN0 stimulus compared with the combined noise controls. A 2 × 4 repeated-measures analysis of variance (ANOVA) was created in SPM5, with stimulus type (IRN and IRN0) and number of iterations (2, 4, 16, and 64) as factors. Simple main effects and interactions were calculated using contrast weights (Friston et al. 2005). Typically, results are reported using a false discovery rate (FDR) threshold of P < 0.05 and small volume correction to control for type 1 errors (Genovese et al. 2002). The small volume defined the auditory cortex across the superior temporal gyrus (including HG, PT, and planum polare) and contained 4719 voxels in the left hemisphere and 5983 voxels in the right hemisphere. Estimates of peak localization within HG were made with reference to 3 cytoarchitectonic subdivisions; Te 1.2 (lateral HG), Te 1.0 (central HG), and Te 1.1 (medial HG) (Morosan et al. 2001; Eickhoff et al. 2005). Region of interest analysis used the same approach described by Hall and Plack (2009).

Results

Psychophysical Measures

Overall, results suggest that the perceptual salience of the spectrotemporal modulations increases as a function of the number of iterations, as does pitch salience for IRN. The results of the psychophysical measures are shown in Figure 3.

Figure 3. Top row: Pitch discrimination thresholds and values for the first peak in the autocorrelation function for IRN stimuli with increasing number of iterations. Bottom row: Modulation discrimination values for IRN0 with corresponding modulation depths, taken from the SD of the cochlear representations presented in Figure 2. Error bars represent 95% confidence limits.
**Pitch Discrimination Thresholds for IRN**

A one-way repeated-measures ANOVA confirmed that pitch discrimination thresholds for IRN were influenced by iteration ($F_{2,15,62.54} = 53.00$, $P < 0.001$). For this test, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($e = 0.72$). As expected (see Yost 1996), pitch discrimination thresholds decreased as a function of the number of iterations (and hence pitch strength). Polynomial contrasts indicated this linear relation to be significant ($F_{1,25} = 91.28$, $P < 0.001$). The results plotted in Figure 3 suggest that performance on pitch discrimination might plateau at around 16 iterations.

**Modulation Discrimination Performance for IRNo**

A further one-way repeated-measures ANOVA demonstrated performance for IRNo was affected by iteration ($F_{1,81.46,10} = 115.15$, $P < 0.001$). Degrees of freedom for this test were corrected using the same procedure ($e = 0.62$). Percentage correct increased as a function of number of iterations and again polynomial contrasts supported the significance of this linear relation ($F_{1,25} = 160.91$, $P < 0.001$).

**fMRI Results: Random-Effects Analysis**

To explore the distribution of stimulus-related activity, we first mapped the pattern of responses separately for IRN (all iterations combined) and for IRNo (all iterations combined) compared with the combined noise controls. We did this using planned comparisons within the 2 × 4 ANOVA. Both contrasts revealed significant feature-driven responses across the entire area of HG and PT, which survived correction ($P < 0.05$). As can be seen in Figure 4, there is considerable overlap of the activity related to IRN and IRNo, although there appears to be a slightly greater spread of activation for IRN than for IRNo. For IRN, the most significant peaks of activation fell close to the border between Te 1.2 (lateral HG) and Te 1.0 (central HG) in both hemispheres ($x = 50$, $y = 20$, $z = 4$ mm in the left and $x = 56$, $y = 14$, $z = 2$ mm in the right). This localization of IRN-related activity concurs with previous studies (Griffiths et al. 1998; Patterson et al. 2002; Krumbholz et al. 2003; Hall and Plack 2009). Peaks of IRNo-related activity fell close to those identified for IRN in the left hemisphere ($x = 50$, $y = 16$, $z = 2$ mm). On the right, the maximum statistical peak was in PT ($x = 68$, $y = 28$, $z = 10$ mm) but remained within the extent of IRN-related activity.

Second, we evaluated the effects of the stimulus type and the number of delay-and-add iterations on the pattern of auditory cortical activity. The main effect of stimulus type from the ANOVA indicated that no differential activity between IRN and IRNo survived correction (Table 1), although 3 small clusters were present at an uncorrected threshold ($P < 0.001$). Two of the clusters were in Te 1.0 bilaterally, with the third being in left PT. The direction of the trend was for a greater response to the IRN conditions with 16 and 64 iterations than to the other conditions. However, this difference was not shown to be reliable when the corresponding IRN and IRNo conditions (i.e., 16 and 64 iterations) were directly contrasted in a planned comparison ($P > 0.05$). Again for the main effect of iteration, no differential activity survived correction (Table 1). However, at the uncorrected threshold ($P < 0.001$), there was one small cluster in each hemisphere, on the left in Te 1.0 and on the right in PT. In summary, the nonsignificant effects suggest that the response patterns for IRN and IRNo were broadly equivalent.

Since there is no strong reason to expect the effects of iteration to be identical for the 2 types of stimulus, we explored the effects of iteration separately for IRN and IRNo. Planned comparisons within the 2 × 4 ANOVA indicated no significant effects at the corrected threshold ($P > 0.05$). We therefore have no evidence from this analysis that there is any differential response to the perceptual salience of the stimulus features for either IRN or IRNo.

**fMRI Results: ROI Analysis**

One of the theoretical perspectives outlined in the Introduction proposes a special role for lateral HG in pitch coding (Bendor and Wang 2005, 2006; see also Patterson et al. 2002). However, much of the evidence for lateral HG is based on studies that used IRN as the sole pitch-evoking stimulus. The results presented in Figure 4 show a response to the slowly varying spectrotemporal modulations of IRN within the vicinity of central and lateral HG. This novel finding suggests that the spectrotemporal modulations may have driven the response previously attributed to pitch and that the results of previous imaging studies using IRN may need to be reinterpreted.

Crucial to our claim that lateral HG is responsive to these slow-rate fluctuations is the ability to demonstrate that the...
influence of the number of delay-and-add iterations is the same for IRNo as it is for IRN signals because this manipulation increases the salience of these “nonpitch” features. To reliably conclude that the spectrottemporal features influence activity within the region corresponding to lateral HG, analysis should be performed at the larger spatial scale of the cortical region and not just individual voxels within it. The fMRI data to be analyzed in this way were therefore obtained using a region of interest (ROI) approach that computed the average magnitude of activity (b) from all voxels within lateral HG in response to each of the 8 stimulus conditions. The test of the within-subjects contrasts from the 2 × 4 ANOVA with stimulus type and number of delay-and-add iterations as factors tells us about the shape of the response as a function of the number of iterations, specifically by assessing the significance of the linear and quadratic trends in the data for all voxels within a given region. The interaction term tells us whether this relationship is different for the 2 classes of stimulus. The results (means and 95% confidence intervals) are represented in Figure 5, with error bars computed across listeners.

A 2 × 4 ANOVA with stimulus type and number of iterations as factors revealed that within lateral HG, there was an overall positive linear relationship between activity and the number of iterations (F1,15 = 25.96, P < 0.001), with no significant quadratic component (Fig. 5A). The iteration-by-stimulus type interaction term for the linear trend was not significant (F1,15 = 0.62, P > 0.05) and so there is no evidence that the number of iterations exerted different effects on the response to IRN and IRNo stimuli in lateral HG.

Since we had observed a widespread response to IRN and IRNo across auditory cortex (Fig. 4), we took this opportunity to examine the results for central HG (Te 1.0, Fig. 5B) and for PT (Fig. 5C) using the same procedures. The results were very much the same as for lateral HG. The tests of within-subjects contrasts again revealed a significant positive linear relationship between activity and the number of iterations (for central HG: F1,15 = 14.47, P < 0.01 and for PT: F1,15 = 9.38, P < 0.01), with no significant quadratic term. Similarly, the findings indicated a nonsignificant interaction for the linear trend (for central HG: F1,15 = 4.12, P > 0.05 and for PT: F1,15 = 3.16, P > 0.05). Hence, we have no evidence that the number of iterations exerts a differential effect on the response to IRN and IRNo stimuli in central HG and PT. The results from this ROI approach are consistent with the hypothesis that human auditory cortex is broadly responsive to the slowly varying spectrottemporal modulations in the signal.

To investigate the effects of ANC, a mixed-design ANOVA was performed separately for the 3 different ROIs, specifying ANC as a within-subjects factor. None of the regions indicated a significant effect of ANC (F1,14 = 0.967, P > 0.05 for Te 1.0, F1,14 = 0.002, P > 0.05 for Te 1.2 and F1,14 = 0.967, P > 0.05 for PT), with no interaction between ANC and stimulus or iteration.

fMRI Results: Incidence Maps

Given that the slowly varying modulations appear to contribute to the IRN-related response, we propose IRNo as a more appropriate noise control for examining the pitch evoked by

Figure 5. Plots showing the magnitude of activity for IRN and IRNo conditions with different numbers of iterations, taken from the 2 × 4 ANOVA for the 3 pitch-responsive regions, central HG (A), lateral HG (B), and PT (C). Error bars represent 95% confidence limits.

### Table 1

<table>
<thead>
<tr>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak coordinates</td>
<td>Z-score</td>
</tr>
<tr>
<td>IRN &gt; control noises</td>
<td>&amp; -50, -20, 4</td>
</tr>
<tr>
<td></td>
<td>&amp; -62, -28, 8</td>
</tr>
<tr>
<td></td>
<td>&amp; -52, -14, 2</td>
</tr>
<tr>
<td></td>
<td>&amp; -34, -26, 2</td>
</tr>
<tr>
<td></td>
<td>&amp; -50, -16, 2</td>
</tr>
<tr>
<td></td>
<td>&amp; -58, 24, 6</td>
</tr>
<tr>
<td></td>
<td>&amp; -52, -20, 4</td>
</tr>
<tr>
<td></td>
<td>&amp; -64, -28, 10</td>
</tr>
</tbody>
</table>

Note: Peak voxels are reported for the left and right hemispheres, respectively.

*FDR corrected.
IRN than the Gaussian noise used hitherto. An alternative demonstration of the impact of the choice of noise control is illustrated by the results of incidence maps created to display the distribution of IRN-related activity across individuals when either a Gaussian noise or IRNo is selected to be that noise control; for a description of the method, see Hall and Plack (2009). In that previous study, we reported that, compared with a Gaussian noise, IRN generated greater activity bilaterally around HG, especially just posterior to HG, close to the border with PT. The maximum consistency across the individual maps was 88% (14/16 individuals) sited anterior to HG (x –60, y –26, z 0 mm). In the right hemisphere, the maximum consistency was 88% (14/16 individuals) sited anterior to HG on the posterior edge of planum polare (x 60, y –4, z 0). In striking contrast are the results for the subtraction of IRNo from IRN (bottom row in Fig. 6). Although the distribution of activity was broadly similar, the degree of consistency across individuals was markedly reduced. On the left side, the maximum consistency across the individual maps of 38% (6/16 individuals) found at the border between the central portion of HG and PT (x –55, y –20, z 8 mm). In the right side, the maximum consistency was 44% (7/16 individuals) sited in PT at the anterior border with lateral HG (x 60, y –18, z 4). The incidence map could either reflect variability in whether or not individuals respond at all to the contrast IRN > IRNo or it could be that everyone in the group did respond significantly to IRN > IRNo but that this response occurred in different voxels across listeners. To determine which interpretation is correct, we reanalyzed the individual data sets. Results showed that 10 of 16 listeners had significant activity in auditory cortex for the IRN > noise contrast, whereas only 3 of 16 listeners showed a significant effect for the IRN > IRNo contrast (P < 0.05, corrected). In all cases, clusters sizes and effect sizes were smaller for the latter and so we conclude that the incidence maps are not obscuring a substantial pitch-specific response (IRN > IRNo) at varying locations across listeners.

These results show that when an appropriately matched control is used (i.e., when the contribution of slowly varying spectrotemporal fluctuations is ruled out), the magnitude of the response that can be attributed specifically to pitch is reduced. The distribution of the resulting pitch-related activity differs between listeners, as observed for other types of pitch-evoking stimuli (Hall and Plack 2009; García et al. 2010; Barker et al. 2011).

Discussion

Response to IRN May Result from Features Unrelated to Pitch

The present fMRI study introduced a novel type of auditory stimulus, IRNo (a “no-pitch” version of IRN), for use in a subtraction paradigm to investigate pitch-related activity using IRN. This stimulus was used to measure the potential contribution of incidental spectrotemporal modulations to the response previously attributed to pitch processing. Our original hypothesis was that the slowly varying modulations contribute to previously observed IRN responses (Hall and Plack 2009). The present study suggests that the IRN and IRNo response patterns within lateral HG, and across auditory cortex, are broadly similar, with little residual response that can be specifically attributed to pitch. When the effects of the modulations were controlled using an IRNo contrast, the residual response to IRN was much less consistent across individuals and more closely matched results from neuroimaging studies that used different types of pitch-evoking stimuli (e.g., Hall and Plack 2009). The presence of slowly varying spectrotemporal fluctuations in IRN mean that it is not possible to tell from comparisons with Gaussian noise whether observed IRN-related activity results from pitch, modulation, or a combination of the 2 features.

The contention is not regarding the use of IRN as a pitch-evoking stimulus rather it is the lack of a well-matched control in previous neuroimaging studies of pitch perception. The control stimuli used in previous IRN studies have not contained the slowly varying spectrotemporal characteristics of IRN and have therefore not provided a controlled comparison for the pitch stimuli. It is suggested that further studies seeking to use IRN as a pitch stimulus use a control that is well matched in terms of these features, such as IRNo.

It has been suggested that for a brain region to be classified as a pitch center, it should show an increase in activation with increasing pitch salience (and hence with increasing iterations for our IRN stimuli) (Bendor and Wang 2005; Hall and Plack 2009). The ROI analyses (Fig. 5) revealed a linear increase in

---

**Figure 6.** Distribution of IRN-related activation compared with Gaussian noise (top row) and with IRNo (bottom row). For the purpose of localization, outlines of the positions of lateral HG (yellow), middle HG (white), and PT (black) are overlaid onto the images. The incidence maps are overlaid onto 4 different axial slices through the group-averaged anatomical image. The color scale represents the percentage of IRN-related activation at every voxel and is calculated as a proportion of a possible maximum of 16. The left-hand side of the brain appears in the left-hand side of each axial slice.
Is There a Human Pitch Center?

In light of the current findings, it would be unwise to assign the title of pitch center to any area of auditory cortex based on the results of studies that have used IRN as their sole pitch-evoking stimulus (e.g., Griffiths et al. 1998, 2010; Patterson et al. 2002; Krumholz et al. 2003), as these studies have not used suitable controls that separate the pitch effect from the effects of slowly varying spectrotemporal modulations. Based on responses to resolved and unresolved harmonic complex tones, Penagos et al. (2004) argued for a salience-dependent pitch response in lateral HG. However, on inspection of their Figure 3, this response appears more posterior in most individual listeners than the group-averaged response in lateral HG that was reported. In addition, only 5 listeners were included in their analysis and the correction used (least significant difference) was much less stringent than the FDR correction used in the current study. Puschmann et al. (2010) also reported a pitch-related response in lateral HG for a tone-in-noise and 2 Huggins pitch stimuli. Again, however, their results indicate a large pitch-related response posterior to lateral HG, in PT (as observed in their Fig. 3). Warren et al. (2003) attribute the response in PT to their wideband harmonic complex tones specifically to pitch height (which provides a basis for sound segregation). Pitch chroma (which provides a basis for representing melodies) activated planum polare. Pitch chroma and pitch height were both found to activate lateral HG. However, wideband complex tones containing low-numbered harmonics have clear spectral features that are resolvable by the auditory system. It is therefore possible that the response to these stimuli was driven by the spectral features rather than by pitch per se. Finally, although Griffiths et al. (2010) used IRN as the pitch-evoking stimulus, the observed high gamma band oscillatory activity in medial and central HG (Te 1.1 and 1.0) occurred with an onset of about 70 ms after the onset of the IRN. It is unclear whether this particular response could have been related to the slow spectrotemporal features in IRN, since detection of these features depends on a long analysis window. Hence, although the functional significance of the high gamma band activity is not clearly understood, it may relate to some aspect of pitch processing in HG.

Considered together with our previous findings (Hall and Plack 2009; García et al. 2010; Barker et al. 2011), it would seem premature to describe lateral HG as a general pitch center until alternative explanations are ruled out. Based on our previous results and the incidence maps presented here, we suggest that anterior PT is a more likely candidate for general pitch processing.

Funding

PhD studentship awarded by the MRC Institute of Hearing Research and MR scanning was paid for through MRC infrastructure funding awarded to the same organization.

Supplementary Material

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

Notes

Two undergraduate students (Sparsh Chandan and Liam Hennessy) collected some of the psychophysical data reported and were supervised by the first author. The authors would like to thank Alain de Cheveigne for advice on the procedure for generating IRNo. We are grateful to Tim Griffiths and an anonymous reviewer for constructive criticisms of a previous version of the manuscript. Conflict of Interest

None declared.

References


Downloaded from https://academic.oup.com/cercor/article-abstract/22/4/745/416425
by guest on 14 May 2018