Cortical Representations of Temporal Structure in Sound

Deborah A. Hall,1 Doug J. K. Barrett,1 Michael A. Akeroyd,2 and A. Quentin Summerfield1

1Medical Research Council Institute of Hearing Research, University Park, Nottingham; and 2Medical Research Council Institute of Hearing Research, Glasgow Royal Infirmary, Glasgow, United Kingdom

Submitted 14 March 2005; accepted in final form 5 July 2005

Harms and Melcher 2002; Liegeois-Chauvel et al. 2004) and human auditory cortex also responds to modulation rates >10 Hz, but the magnitude and shape of the response varies according to modulation rate. Steady-state evoked potentials can be recorded for stimuli that are modulated at rates between 1 and 200 Hz (Picton et al. 2003). Systematic effects of sinusoidal AM rate (10 to 98 Hz) on the steady-state MEG (magnetoencephalographic) response reveal a clearly detectable synchronized response at each rate, but as the rate increases to >40-Hz harmonics, the positive and negative deflections approximate more toward a sinusoid (Rob et al. 2000). Thus neural-coding mechanisms for temporal envelope and periodicity have been well explored.

Functional magnetic resonance imaging (fMRI) takes advantage of the association between neuronal activity and the local control of blood flow to localize activity in the brain. The fMRI signal from the human auditory cortex is able to determine activity within a few millimeters of space enabling response parameters to be mapped both within and across different auditory fields (e.g., frequency sensitivity; Talavage et al. 2004). Human fMRI data have identified degrees of functional segregation across regions of the auditory cortex both in terms of the range of best modulation rates (Giraud et al. 2000) and whether the response shape is sustained or phasic (Harms and Melcher 2003; Seifritz et al. 2002). The distribution of preferred responses is somewhat patchy and so no obvious organizational scheme has yet emerged from this research except to say that lateral Heschl’s gyrus (HG) and planum temporale (PT) prefer low modulation rates (Hart et al. 2003).

Address for reprint requests and other correspondence: D. A. Hall, MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, UK (E-mail: d.hall@ihr.mrc.ac.uk).

The costs of publication of this article were defrayed in part by the payment of page charges. The article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

INTRODUCTION

Temporal acoustic patterns are important determinants of the sound stimulus and can be broadly classified into three divisions: envelope (<50 Hz), periodicity (50–500 Hz), and fine structure (>500 Hz). For human listeners, temporal variations ≤50 Hz are important for accurate speech recognition (e.g., Druillem et al. 1994a,b; Shannon et al. 1995). When the frequency of the repeating pattern within the waveform exceeds 50 Hz the percept of pitch is created. Acoustic temporal structure can be present separately in each channel, but can also be created through binaural interaction across channels. The fine temporal relationships between the signals at the two ears generally determine spatial qualities such as the width and location of the sound source.

The auditory system must be able to represent temporal information across these different timescales. It is well established that the auditory cortex represents slow-rate envelope modulations explicitly using a temporal code that is phase locked to the stimulus. When amplitude-modulated (AM) stimuli are presented at modulation rates that span 1–500 Hz, most auditory cortical neurons have a discharge rate that is selectively tuned to low modulation rates; the best modulation frequencies range from about 3 to 30 Hz (e.g., Eggermont 1994; Schreiner and Urbas 1986, 1988; Wang et al. 2003). The observed selectivity is indeed a temporal, not a spectral, phenomenon because the same tuning pattern is found for frequency-modulated (FM) stimuli (Wang et al. 2003). It has been argued that fine-grained temporal patterns are more likely to be represented by the mean discharge rate of the population (Wang et al. 2003), although using moving ripple sounds it has been possible to demonstrate some preservation of precise spike timing (Elhilali et al. 2004). Results from human auditory neuroimaging have confirmed that sustained cortical responses occur predominantly for rates <10 Hz (Giraud et al. 2000; Harms and Melcher 2002; Liegeois-Chauvel et al. 2004) and equally for both AM and FM sounds (Hart et al. 2003). The human auditory cortex also responds to modulation rates >10 Hz, but the magnitude and shape of the response varies according to modulation rate. Steady-state evoked potentials can be recorded for stimuli that are modulated at rates between 1 and 200 Hz (Picton et al. 2003). Systematic effects of sinusoidal AM rate (10 to 98 Hz) on the steady-state MEG (magnetoencephalographic) response reveal a clearly detectable synchronized response at each rate, but as the rate increases to >40-Hz harmonics, the positive and negative deflections approximate more toward a sinusoid (Rob et al. 2000). Thus neural-coding mechanisms for temporal envelope and periodicity have been well explored.

Functional magnetic resonance imaging (fMRI) takes advantage of the association between neuronal activity and the local control of blood flow to localize activity in the brain. The fMRI signal from the human auditory cortex is able to determine activity within a few millimeters of space enabling response parameters to be mapped both within and across different auditory fields (e.g., frequency sensitivity; Talavage et al. 2004). Human fMRI data have identified degrees of functional segregation across regions of the auditory cortex both in terms of the range of best modulation rates (Giraud et al. 2000) and whether the response shape is sustained or phasic (Harms and Melcher 2003; Seifritz et al. 2002). The distribution of preferred responses is somewhat patchy and so no obvious organizational scheme has yet emerged from this research except to say that lateral Heschl’s gyrus (HG) and planum temporale (PT) prefer low modulation rates (Hart et al. 2003).

Address for reprint requests and other correspondence: D. A. Hall, MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, UK (E-mail: d.hall@ihr.mrc.ac.uk).

www.jn.org 0022-3077/05 $8.00 Copyright © 2005 The American Physiological Society 3181
Imaging studies have recently begun to map out the cortical distribution of responses to temporal acoustic patterns at rates >50 Hz and have begun to define those regions that are highly sensitive to periodicity (Griffiths et al. 1998, 2001; Krumbholz et al. 2003; Patterson et al. 2002; Penagos et al. 2004; Warren and Griffiths 2003) and fine structure (Budd et al. 2003; Krumbholz et al. 2005). Again HG and PT are strongly implicated in the analysis of these more rapid temporal variations in sound. Significant positive relationships have been revealed between activation in lateral HG and both the degree of monaural temporal regularity in a regular-interval noise (RIN) signal and the degree of correlation between a noise at the two ears (Budd et al. 2003; Griffiths et al. 1998; see also Krumbholz et al. 2003). Moreover, reanalysis of Budd’s data (computing the difference between correlated and uncorrelated noise conditions) confirmed that the sensitivity to correlated noise did not extend beyond lateral HG (see Fig. 1). This result has led to speculation that the analysis of temporal pitch and interaural correlation (IAC) might engage the same bilateral auditory region—lateral HG (Budd et al. 2003).

If the same region is implicated in the analysis of periodicity and fine temporal structure, is it the seat of the same neural process? Patterson et al. (2002) interpret their pitch result within the framework of a computational model of temporal pitch processing in which the role of lateral HG is to produce a summary histogram and locate its first peak. Although originally proposed to explain the fMRI data for pitch, this model is equally applicable to the IAC result of Budd et al. (2003). The computation is of a general form that combines temporal acoustic information across frequency channels to produce a summary correlogram for determining the perceptual salience of the sound feature. A mechanism of this sort is associated with pitch (Yost et al. 1996) and interaural time delays (e.g., Shackleton et al. 1992; Stern and Trahiotis 1995). In the early stages of the model algorithm, the individual frequency channels are first subjected either to a series of autocorrelations where the signal is correlated with successively delayed versions of itself (in the case of monaural temporal regularity) or to a series of cross-correlations where the signals at the two ears are correlated at a range of delay values (in the case of IAC). Both the auto- and cross-correlation outputs are then represented in a two-dimensional array with frequency on the ordinate and the correlation lag on the abscissa. If the sound has a salient percept associated with temporal pitch or an IAC of unity, the maximum correlation delays across frequencies will occur at the same correlation lag, thus generating a single high peak in the summary correlogram. In support of this association between the peak and the perceptual salience, pitch discrimination data clearly demonstrate that the pitch strength is an exponential function of the height of this peak (Yost 1996).

We predict that if the activation in lateral HG reflects a common computational step in the analysis of temporal structure in sound, then the effects of RIN and correlated noise will equally engage the same neural basis in the human auditory cortex. Previous results are inconclusive because periodicity and fine structure have been tested in separate groups of listeners. The present study reports a novel set of results that tests this hypothesis using a single set of listeners.

METHODS

Participants

Eighteen adults (ten male, eight female) aged between 18 and 42 yr participated in the experiment. None of the participants was a trained psychophysical listener. All participants were screened for normal hearing (<20 dB HL) for pure-tone octave frequencies between 500 and 8,000 Hz and none had a history of audiological or neurological impairment. All participants were familiarized with the scanning environment and listening task before giving written consent. The experimental procedures were approved by the local health service ethics committee.

Stimuli and task

The starting point for stimulus generation was a random noise burst that was band-passed from 500 to 1,500 Hz and sampled at a rate of 44,100 Hz. The high-pass cutoff was set to 500 Hz to exclude the region normally associated with the dominant, resolved harmonics for these low-pitched stimuli, whereas a low-pass cutoff of 1,500 Hz is close to the upper limit for the effectiveness of binaural time difference cues for sound location. Noise bursts were created digitally by taking both real and imaginary components within the pass-band from Gaussian distributions and setting those components outside the pass-band to zero. The real part of the inverse fast Fourier transform of this signal was stored for further processing. Monaural temporal regularity was generated by delaying a copy of the random noise and adding it back to the original. The pitch characteristics of the noise are directly related to the temporal patterning within the signal. The pitch frequency is equal to the reciprocal of the delay (i.e., the period of the repeating interval) and the salience of the pitch is exponentially related to the number of delay-and-add iterations (e.g., Yost 1996, 1998; Yost et al. 1996). RIN is a good candidate for investigating the cortical basis for pitch sensitivity because the signal can be high-pass filtered to minimize resolved spectral components in the excitation pattern and pitch salience can easily be varied parametrically (Griffiths et al. 1998, 2001; Patterson et al. 2002; Warren and Griffiths 2003). Three sound conditions were created that differed in the amount of regular temporal structure. The first RIN condition con-

![FIG. 1. Reanalysis of the data reported in Budd et al. (2003) demonstrating the pattern of significant differential activation by the sound with an interaural correlation value of unity compared with that with a value of zero (P < 0.001 uncorrected). Bilateral clusters are unmistakably located within the lateral portion of Heschl’s gyrus (HG), with peaks voxels at x = 54, y = −16, and z = 2 mm; and x = 62, y = −10, z = 0 mm. Reanalysis also confirmed that there was no significant sensitivity to uncorrelated noise. There is a 4-mm-slice separation between the 2 oblique brain images whose orientation is aligned with the surface of the superior temporal gyrus (middle image). Brain images are displayed in neurological convention (L = Left).](image-url)
tained 16 add-and-delay iterations, generating a clear periodic structure and thus a salient percept of temporal pitch. The second RIN condition contained only one add-and-delay iteration, thus generating a weak percept of temporal pitch. The third sound condition was random noise with no regular structure and thus no pitch.

These three levels of temporal regularity were crossed with two levels of IAC (correlations of zero and unity) to create six noise conditions. Correlated signals, with an IAC of unity, were created by copying the same noise waveform to left and right channels. This diotic signal is perceived as a compact source at the center of the head. Uncorrelated signals were created by generating two statistically independent waveforms so that no time delay would make the waveforms in the left and right channels match. Listeners often report an uncorrelated noise as two separate sound sources at the two ears (e.g., Blauert and Lindeman 1986), although some listeners report a spatially diffuse percept especially when they are naive psychophysical subjects. In all six noise conditions, the noise was spectrally equivalent because it always originated from the same samples of random noise. Additional supplementary information, including audio examples, is available.1

Each noise condition contained a sequence of 16 of the same type of noise bursts (450-ms duration with 10-ms raised cosine onset and offset ramps and 50-ms interburst intervals). The burst rate was 2 Hz and each sequence lasted 8 s 1 s. Thirty different noise sequences were generated for each of the six noise conditions. For the RIN conditions, the delays in the add-and-delay algorithm were chosen from a set of 16 steps ranging from 9.1 to 19.6 ms (i.e., 51–110 Hz). They were varied randomly from burst to burst to create a changing pitch sequence. Noise sequences plus 30, 8-s silent epochs were presented to listeners in a pseudorandomized order. Listeners were requested to press a button in response to each occurrence of the silent epoch. All noise sequences were presented at 80 dB SPL using high-fidelity headphones that had been specifically modified for use during fMRI (Palmer et al. 1998).

fMRI acquisition

Echo-planar brain images were acquired using a 3-T whole body scanner equipped with a head-gradient coil (Bowtell et al. 1994) and a custom-built radio-frequency receiver coil (Nova Medical, Tyngsboro, MA). Each image set constituted 22 slices that were selected to give the best coverage of the superior temporal gyrus. Coverage generally excluded the frontal and occipital poles and so before the functional experiment, an additional whole-brain set of 64 coronal slices was acquired to facilitate image postprocessing. Contiguous T2*-weighted functional images were acquired using the parameters TE = 35 ms and flip angle = 90°. Each set of images was acquired in the coronal orientation (64 × 64 matrix; 4-mm voxel resolution; 2.2-s clustered acquisition time). The time interval between each functional image set was 9.0 s. This sparse sampling method enabled the auditory stimulus to be delivered predominantly during the quiet period between image acquisitions to reduce acoustic masking and other contaminatory effects of background scanner noise (Edmister et al. 1999; Hall et al. 1999). A sustained hemodynamic response has been demonstrated for a burst rate of 2 Hz (Harms and Melcher 2002), although the additional nonlinear stepwise response to the presence of temporal pitch (variable 4) was a column of digits, one for each set of images. Our stimulus variables were 1) the presence of any noise condition (coded as 1 for presence and −1 for absence), 2) the values of IAC giving rise to different spatial percepts (coded as 1 for correlated and −1 for uncorrelated), and 3) the levels of temporal regularity giving different sensations of pitch salience (coded as 0, 3.0, and 12.3 with values being defined by the function “10 Log [number of add-and-delay iterations + 1]”; Griffiths et al. 1998). A fourth variable (4) was included to account for any additional nonlinearity in the response to temporal regularity. This variable took the form of a stepwise response that simply defined the presence of an RIN condition by a value of 1 and its absence by a value of −1.

To explore the question about a common neural basis for the sensitivity to temporal acoustic structure, we first considered the patterns of activation related to pitch and spatial width by defining four feature-related contrasts; the increasing response to temporal regularity (pitch salience) (variable 3), the additional nonlinear stepwise response to the presence of temporal pitch (variable 4), the response to correlated noise (compact source) (variable 2), and the response to uncorrelated noise (split, broad source) (defined by the negative sign of variable 2). Although these analyses were performed for each individual data set, we primarily consider the overall significance of the stimulus-related activation for the group of 18 listeners, determined using a second level of analysis that dealt with the between-subject variance. These group analyses were computed in SPM99 by performing one-sample t-test on the individual summary outputs from the four feature-related contrasts. Activation in the auditory cortex was reported only if it exceeded a voxel-level signif-

1 The Supplementary Material for this article (text and audio examples) is available online at http://jn.physiology.org/cgi/content/full/00271.2005/DC1.
Sensitivity related to pitch

First, we will discuss the spatial location of those voxels that displayed a growth in their response as a function of increasing temporal regularity (pitch strength) (Table 1). The result from the group analysis is presented (in blue) for five slices through the auditory cortex in Fig. 3A. There was widespread activation across the auditory cortex in both left and right hemispheres. Most of the peaks of activation were focused around the primary auditory cortex (medial HG), but activation extended out to the lateral convexity as well as anteriorly and posteriorly along the length of the superior temporal gyrus to include parts of the surrounding nonprimary auditory cortex on the PT and planum polare. Conversely, we inverted the linear growth contrast to identify candidate regions that decreased in their response as the pitch strength increased. No activation was found to reach significance.

The individual pitch-related activation maps showed much the same pattern as that of the group, with every one of the 18 listeners having distinct clusters of activation. Figure 3B (first row) shows the distribution of the individual activations for the rise in pitch-related activation. Individual activation maps have been summated to display the range of overlap between them. The color scale ranges from 1 (no overlap) to 16 (most maps overlap). We found no evidence for a difference in the size of the response across left and right hemispheres \( t(17) = 1.8, P = 0.09 \). The median size of the pitch-related sensitivity covered a 2.8-cm³-size region of the auditory cortex; equivalent to 354 voxels. The greatest overlap between listeners \( n = 16 \) occurred at the voxel location \( x = -50, y = -26, z = 4 \) mm, which lies on the posterior boundary of left HG, close to the peak voxel for the group. According to the probability maps given by Westbury et al. (1999) and Penhune et al. (1996), this point is more likely to be in PT than HG (49 vs. 1% probability). In summary, the sensitivity to pitch strength is widespread, is highly significant, and is replicable across listeners.

Next, we will discuss the spatial location of those voxels showing an additional stepwise response to the presence of temporal regularity (pitch). The result of the group analysis is presented (in cyan) in Fig. 3A. These response regions were primarily located in nonprimary auditory cortex (i.e., the lateral...
portion of HG) and extended below the surface of the superior temporal gyrus. In these voxels the pitch-related activity had both linear and nonlinear components. The additional nonlinear response was more limited than the area showing a purely linear response to pitch salience and some asymmetry was observed (right hemisphere activity greater than left). When the individual data were examined, the distribution of suprathreshold pitch-related activation was found to vary somewhat from listener to listener; in seven listeners it was purely right-sided, and in four it was absent.

Stepwise response to temporal regularity (pitch)

<table>
<thead>
<tr>
<th>Max Z Value</th>
<th>Side</th>
<th>Anatomical Region</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.87</td>
<td>L</td>
<td>Heschl’s sulcus/Superior temporal gyrus</td>
<td>41/22</td>
</tr>
<tr>
<td>5.18</td>
<td>L</td>
<td>Superior temporal gyrus</td>
<td>22</td>
</tr>
<tr>
<td>4.89</td>
<td>L</td>
<td>Medial Heschl’s gyrus</td>
<td>41</td>
</tr>
<tr>
<td>5.66</td>
<td>R</td>
<td>Medial Heschl’s gyrus</td>
<td>41</td>
</tr>
<tr>
<td>5.36</td>
<td>R</td>
<td>Superior temporal gyrus</td>
<td>22</td>
</tr>
<tr>
<td>5.27</td>
<td>R</td>
<td>Lateral Heschl’s gyrus</td>
<td>41</td>
</tr>
</tbody>
</table>

Correlated noise (compact sound source)

<table>
<thead>
<tr>
<th>Max Z Value</th>
<th>Side</th>
<th>Anatomical Region</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.30</td>
<td>L</td>
<td>Inferior parietal lobe</td>
<td>40</td>
</tr>
<tr>
<td>4.57</td>
<td>R</td>
<td>Inferior parietal lobe</td>
<td>40</td>
</tr>
<tr>
<td>3.93</td>
<td>R</td>
<td>Superior temporal gyrus</td>
<td>22</td>
</tr>
</tbody>
</table>

Uncorrelated noise (split, broad sound source)

<table>
<thead>
<tr>
<th>Max Z Value</th>
<th>Side</th>
<th>Anatomical Region</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.41</td>
<td>L</td>
<td>Heschl’s sulcus/Superior temporal gyrus</td>
<td>41/22</td>
</tr>
<tr>
<td>4.00</td>
<td>L</td>
<td>Middle temporal gyrus</td>
<td>21</td>
</tr>
<tr>
<td>4.12</td>
<td>R</td>
<td>Posterior insula</td>
<td>13</td>
</tr>
<tr>
<td>3.77</td>
<td>R</td>
<td>Heschl’s sulcus/Superior temporal gyrus</td>
<td>41/22</td>
</tr>
<tr>
<td>3.65</td>
<td>R</td>
<td>Heschl’s sulcus</td>
<td>41</td>
</tr>
</tbody>
</table>

Interaction between pitch salience and sound source location

<table>
<thead>
<tr>
<th>Max Z Value</th>
<th>Side</th>
<th>Anatomical Region</th>
<th>BA</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.13</td>
<td>L</td>
<td>Heschl’s sulcus/Superior temporal gyrus</td>
<td>41/22</td>
</tr>
<tr>
<td>4.57</td>
<td>R</td>
<td>Heschl’s sulcus/Superior temporal gyrus</td>
<td>41/22</td>
</tr>
</tbody>
</table>

Coordinates are reported in the standard MNI brain space. Z values are reported for the peak voxel within each cluster where the height threshold exceeds a level of $P < 0.001$. BA labels approximate to the corresponding Brodmann anatomical reference scheme.
FIG. 3. Results from the various group and individual analyses showing the patterns of sensitivity to pitch- and location-related acoustic attributes. A: group activation maps are derived from conventional $t$-test and the 4 stimulus effects are represented by solid colors. B: individual activations illustrate the variability of those effects across the separate listeners using probabilistic maps of the overlap between listeners. All maps are overlaid onto the same 5 horizontal brain images ($z = 16$ to $-16$ mm). This base image is an average of the 64-slice whole brain set that was acquired for all participants and is displayed in neurological convention (Left = Left).
voxx in the left and 42 voxels in the right hemisphere and which overlapped in two listeners. The centroid of these clusters was $x = -38$, $y = 2$, $z = 14$ mm and $x = 34$, $y = 6$, $z = 10$ mm, but the activity can also be seen on the overlap images at $z = 16$ and $8$ mm. The functional role of the anterior insula is unclear.

Sensitivity to the broad, split sound source revealed moderate clusters of activation in bilateral auditory cortex, which predominantly had a medial focus just behind and below HG (Fig. 3A, in yellow). The activation encompassed both primary and nonprimary regions of the auditory cortex on HG and PT, respectively. Although five of the individual listeners had no significant auditory activation in response to the broad, split sound source, the individual activation patterns for 13 listeners showed a distribution that was broadly similar to that of the group (Fig. 3B, bottom row). In six listeners it was bilateral, but in five it was present only on the left side and in two listeners it was present only on the right side. There was moderate consistency in the location of the response to the broad, split sound because eight of those listeners activated precisely the same left-sided medial location ($x = -50$, $y = -26$, $z = 4$ mm), immediately behind HG. This site was also close to the peak voxel for the group. The posterior insula activation that was obtained from the group analysis reached threshold in only one of the individual listeners.

In summary, several auditory cortical regions were sensitive to the spatial width of the sound source. However, these were many times smaller than that region, which was sensitive to pitch, as well as more variable across listeners. These results illustrate that the pattern of significant activation for the group not only reflects the suprathreshold activation in the individual listeners, but is also influenced by the individual’s subthreshold responses. It is interesting to note that the location of the most reliable response to the broad, split sound source was the same voxel in the PT that also showed the most statistically reliable response to pitch strength ($x = -50$, $y = -26$, $z = 4$ mm). In the following section, we quantify the degree to which regions of the auditory cortex are sensitive to both types of temporal acoustic structure.

Co-localization of responses to pitch and spatial width

By displaying the group results for the four feature-related contrasts on the same brain view (Fig. 3A), the overlap between the different activation maps can be appreciated. The schematic diagram in Fig. 4 quantifies the number of voxels that are coactivated by more than one of the contrasts: $\leq 90\% (213/230)$ of the voxels that responded to the presence of pitch were also sensitive to the rise in temporal regularity. Their location corresponds to the lateral portion of HG. In lateral HG, the stepwise response to the manipulation of temporal regularity lends further support to the idea of a perceptual center that plays a special role in representing both the presence of pitch and its salience.

A reasonable number of voxels responded to both pitch and spatial width cues. All voxels that responded to the compact sound source also responded to the rise in pitch salience. The median coordinate for the coactivated cluster was $x = 60$, $y = -2$, $z = -6$ mm, in a location that just undercut the lateral part of HG, most probably in nonprimary auditory cortex. A subset of this region also exhibited an additional stepwise sensitivity to the presence of pitch. A majority of voxels that responded to the broad, split sound source (68%, 405/596) were also sensitive to pitch salience. The median coordinates for the left- and right-sided coactivations both lay within the PT, immediately behind HG ($x = -48$, $y = -26$, $z = 2$ mm and $x = 52$, $y = -22$, $z = 8$ mm).

Interaction between the two sound features

The three-way interaction between pitch salience and spatial width reached significance in a bilateral auditory region located just behind HG, in Heschl’s sulcus and PT (see Fig. 5). The separate two-way interactions are informative about the form of the interaction and their results indicate that the region is particularly responsive to the effect of the uncorrelated noise when the sound also has a strong pitch compared with when it has no pitch. To confirm the shape of the interaction, the MR signal (adjusted for the constant term relating to the mean image value) was extracted for the left and right peak voxel coordinates for every listener and the group mean value for each of the six sound conditions was computed. As the graphs in Fig. 5 reveal, sensitivity to the broad, split sound source was greatest when the sound also had a strong temporal pitch. Indeed, as Table 1 reveals, the peak coordinates of the interaction lie very close to those voxels showing a significant sensitivity to the broad, split sound source. The shape of the interaction shown by the peak voxels was the same as that in the rest of the region where plots of the mean MR signal for the entire cluster followed a similar pattern.

Perceptual interactions between the two acoustic manipulations

The interaction result was unexpected and no clear explanation seemed immediately apparent. When creating the stimuli, we had made the assumption that because the manipulations of temporal iteration and IAC operated on different scales of temporal structure, their subsequent effects on perception and brain activation might be additive, but not interactive. Because our fMRI results revealed a superadditive response for uncorrelated nose and strong pitch, would there also be an interaction for perception? To test this prediction, we recruited a new set of 20 naïve psychophysical listeners because the original set of listeners were no longer available. Listeners were simply
asked to describe their experience of a set of sound stimuli. Our procedure followed that of Blauert and Lindeman (1986) in which listeners were asked to draw a map of each auditory event using schematic sections through the head’s vertical and horizontal planes. This method brings the listener’s attention to the position and extent of the auditory percept. For this test, we presented four examples of a binaural 8-s sound sequence, taken directly from each of the “no pitch” and “strong pitch” fMRI conditions. Thus the sequences were 1) random noise, IAC = 0, 2) random noise, IAC = 1, 3) RIN = 16 add-and-delay iterations, IAC = 0, and 4) RIN = 16 add-and-delay iterations, IAC = 1. Four monaural conditions were also presented to provide baseline measures. These were left- and right-ear versions of the “no pitch” and “strong pitch” conditions.

Figure 6 shows the distribution of the auditory images that were mapped for the monaural and binaural stimuli. Monaural noise bursts were reliably perceived to have a single sound source to the left or right, determined by the ear of presentation. Lateralization judgments were not influenced by whether the noise had a pitch. A majority of listeners (16/20) perceived both sequences of correlated noise bursts to have a single compact source around a central, midline location. Again judgments about lateralization were not strongly affected by any additional pitch information. In contrast, there was a good deal of variability in the perceived position of the binaurally uncorrelated stimuli. From our sample of 20 listeners, five listeners described the two sequences of uncorrelated noise as having two separate sources to the left and right, whereas three always described a single, more diffuse source. The remaining listeners described two lateralized sources on one occasion and a diffuse source on another. One striking observation was that seven of the listeners also stated that the source of the uncorrelated noise appeared to shift from burst to burst for the RIN sequence in which pitch varied from burst to burst. The perceived direction of the location shift was not systematic across listeners. It was either lateralized or crossed the midline.

FIG. 5. Interaction between pitch salience and sound source location showing significant bilateral effects in a medial part of the superior temporal gyrus. Again, activations are superimposed on the average of the 64-slice whole brain set that was acquired for all participants (Left = Left). Two graphs below plot the form of the interaction for the 2 peak voxels.

FIG. 6. Perceptual estimates of the intracranial localization of different sound sequences for a group of 20 naïve listeners. Each sound source is represented using vertical and horizontal sections through the listener’s head. Map from each listener contributes 5% of the gray scale, so that fully overlapping sources would be represented by the color black (20 × 5%). Judgments about monaural stimuli were equivalent for both random noise and regular-interval noise (RIN) = 16, “strong pitch” signals and so for illustration, only the maps for 2 monaural pitch conditions are shown here.
and either oscillating or circular. In contrast, the uncorrelated “no pitch” noise sequence was never perceived to be spatially varying. Note that the pitch sequences contained a random melody, yet changing the pitch is itself insufficient to induce a sensation of spatial shifts because listeners always described both the correlated noise with pitch and the monaural noise with pitch as being static. The perceptual data therefore also reveal an interaction between temporal regularity and interaural correlation that neatly mirrors the same interaction in the fMRI data. The response to the uncorrelated “strong pitch” noise appears to be most different from the other stimulus conditions, in terms of both the way it is perceived and its pattern of cortical activation.

DISCUSSION

A common neural process?

Overlap in the sensitivity to monaural temporal regularity and IAC occurred in several regions of the nonprimary auditory cortex (close to the lateral part of HG and in PT) providing the first evidence that these two sound features engage the same auditory regions. The strongest interpretation of this result would be to conclude that the coactivation reflects the common neural computation of temporal structure. However, two observations from the present study are incompatible with this conclusion. Most notably, our data clearly revealed two separate regions of IAC sensitivity determined by whether the signal was correlated or uncorrelated across the two ears. Across-frequency integration is a step that ought to be applied equally to both types of binaural signal because it is only the output of this process that differs according to the IAC value, not the process itself. Therefore this stage of the model would not predict correlated and uncorrelated noise to have separate neural substrates. Second, the group results demonstrate that the patterns of activation for pitch and spatial width were typically different to a greater degree than they were similar. Lateral HG was highly responsive to pitch, whereas its sensitivity to correlated noise reached significance only in the right hemisphere and even here the effect size (Z value) was reduced. This asymmetry suggests that any common neural computation for periodicity and fine structure is unlikely to be instantiated in lateral HG, but probably at an earlier stage in the auditory pathway. If cortical activation reflects a later operation then it would be performed on the output of the frequency integration process, possibly associated with the perceptual qualities of the sounds. We therefore infer that the cortical activation is a neural correlate of pitch and spatial width rather than monaural temporal regularity and IAC. Although fMRI cannot reveal what is the precise neural code in lateral HG, we expect the cortical code to be an abstract high-level one. For example, as a result of the general loss of rapid temporal synchrony, the cortex is proposed to operate on a transformed representation of the temporal information and not on the original rapid timing code (Eggermont 2001; Palmer 1995; Wang 2003). An abstract cortical code for pitch has some empirical support from human listeners because lateral HG is responsive to pitch irrespective of whether it is generated by the temporal pattern in noise (Griffiths et al. 1998; Patterson et al. 2002) or is a complex tone containing resolved or unresolved harmonics (Penagos et al. 2004).

The relevance of individual differences

We observed that the pitch-related activation was highly consistent across all the listeners, whereas that associated with different spatial widths was both less distinct and more highly variable across listeners. We view these individual differences to be informative because they support the poor association between the neural correlates of pitch and spatial width processing in human auditory cortex. Individual differences in spatial width activation may represent distinctions between the ease of perceiving that feature in the sound. Although somewhat anecdotal, listeners were more readily able to distinguish between the sound sequences along the dimension of pitch than along the dimension of spatial width. Un correlated noise is not commonly heard in the environment and so might be difficult to judge perceptually and to assign a verbal label. Given their relative unfamiliarity with these synthesized stimuli, different listeners might adopt different strategies for listening to and describing the spatial attributes of the sound. What we mean by the term strategic is intended to incorporate the idea of different listening modes. A well-known example is off-frequency listening, when listeners can use information in different frequency regions to improve their target discrimination performance in masking tasks (O’Loughlin and Moore 1981). An individual’s listening experience can also influence the listening mode for uncorrelated noise because trained and naïve listeners are reported to differ in the way they interpret the spatial characteristics of broadband signals when the value of IAC is <1 (Blauert 1978). Whereas practiced psychoacoustic listeners can distinguish separate auditory objects, naïve listeners are more likely to report a single broadened object (see also Blauert and Lindemann 1986). Indeed, in our survey of 20 naïve listeners, we recorded a wide range of perceptual experiences for the uncorrelated noise, including reports of either one diffuse or two separate lateralized auditory objects. Some of the naïve listeners differed even in their perception of the correlated noise. Although we were unable to collect perceptual data from those who participated in the scanning study, the qualitative data from the new listeners strongly indicate that the perceptual ambiguity in the spatial characteristics of the uncorrelated noise might be sufficient to underlie the observed patterns of results, particularly the intersubject variability obtained in the fMRI data. We conclude from the data presented here, a strong need to relate individual patterns of cortical activation with that of listener’s own perceptual experience, as well as with that of the acoustical attributes of the sound.

The interaction between pitch- and spatial-related processing

We discuss two possible functional interpretations of the interaction effect found in the PT: that PT reflects a site for either 1) processing sounds that vary in spatial location or 2) integrating object- and location-based sound information to segregate auditory objects. First, if we infer that the fMRI interaction is a correlate of the perceptual interaction, then PT might simply be implicated in the processing of auditory space.

2 T-test comparisons are particularly sensitive to the degree of response variance in either of the pairwise conditions and so variability in the response to uncorrelated noise conditions would be sufficient to reduce the effect size for correlated noise conditions.
The role of PT in spatial analysis is well established. It is responsive to sounds that move in space (Baumgart et al. 1999; Hart et al. 2004; Warren et al. 2002) or shift in their location from burst to burst (Krumhholz et al. 2005; Warren and Griffiths 2003). PT is also sensitive to multiple simultaneous sources that have a broad spatial distribution (Zatorre et al. 2002). In the present study, naïve listeners described the uncorrelated “pitch” noise as shifting in location from noise burst to noise burst. We speculate that this sensation could be induced because uncorrelated noise had a spatial ambiguity and because another (nonspatial) attribute of the sound was changing within the sequence (i.e., the frequency of the pitch). The key point here is the putative association between the reported sensation of shifts in spatial position and the greater PT activation by uncorrelated noise.

The second inference considers PT as a more general purpose system implicated in higher-order sound analysis including the representation and segregation of spectrotemporal sound patterns (Griffiths and Warren 2002). The interaction revealed by our fMRI data is consistent with a cortical process that integrates spectrotemporal features of a sound with spatially relevant information—an important process in sound segregation. Electrophysiological data from primates support the claim for a distinct region of the nonprimary auditory cortex that is responsible for the integration of object-related sound features with information about its location in space. A likely neural candidate for integration in the primate auditory cortex is the lateral auditory belt because here neurons are jointly sensitive to both the spatial position and the spectrotemporal features of a sound (Rauschecker and Tian 2000; Tian et al. 2001). For example, the response preference for a particular monkey call co-varies with that neuron’s spatial selectivity, especially in the posterior portion of the lateral belt (areas CL and CM). The human homologue of the lateral belt is most probably part of PT (see Appendix A in Hackett 2003). In human listeners, PT activation has been reported to co-vary with the spatial distribution of a set of acoustically distinctive sounds (Zatorre et al. 2002). Thus PT appears to be particularly sensitive to spatial position when object-related features are also present in the sound. Again, in the present experiment PT responded most strongly to the uncorrelated noise that had a salient pitch. Because this sound condition was the one judged to contain several auditory objects, the increased computational resource required to derive and integrate spatial and pitch-related outputs could account for the high level of PT activity. The observation (illustrated in Fig. 5) that decreasing spectrotemporal distinctiveness (i.e., pitch salience) was associated with a decrease in PT activity supports an integrative processing account in which the magnitude of the response is contingent on both the spectrotemporal and spatial attributes of the stimulus.

The results from this study clearly rule out the hypothesis that lateral HG plays an equivalent role in representing the temporal acoustic structure for temporal pitch and binaurally correlated stimuli. The interaction effect in PT was an unexpected result whose interpretation cannot be resolved by the present data. The precise functions of PT are not well defined because it is responsive to many different types of complex sound (Griffiths and Warren 2002) and, yet, not always jointly sensitive to pitch and location (Warren and Griffiths 2003). Important questions for future research should probe the role of PT in object formation by addressing the type of listening conditions that engage PT and exploring evidence for functional subdivisions.

Acknowledgments

The authors thank Prof. Stuart Gatehouse for insightful discussions. C. Shaw made a significant contribution to the artwork in Fig. 6. Scanning was conducted at the Sir Peter Mansfield Magnetic Resonance Centre with the help of K. Head.

Present address of Q. Summerfield: Department of Psychology, The University of York, Heslington, York, YO10 5DD, UK.

Grants

This research was fully funded by the Medical Research Council.

References


