

Pitch Processing Sites in the Human Auditory Brain

Deborah A. Hall¹ and Christopher J. Plack²

¹MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK and ²Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK

Lateral Heschl's gyrus (HG), a subdivision of the human auditory cortex, is commonly believed to represent a general "pitch center," responding selectively to the pitch of sounds, irrespective of their spectral characteristics. However, most neuroimaging investigations have used only one specialized pitch-evoking stimulus: iterated-ripple noise (IRN). The present study used a novel experimental design in which a range of different pitch-evoking stimuli were presented to the same listeners. Pitch sites were identified by searching for voxels that responded well to the range of pitch-evoking stimuli. The first result suggested that parts of the planum temporale are more relevant for pitch processing than lateral HG. In some listeners, pitch responses occurred elsewhere, such as the temporo-parieto-occipital junction or prefrontal cortex. The second result demonstrated a different pattern of response to the IRN and raises the possibility that features of IRN unrelated to pitch might contribute to the earlier results. In conclusion, it seems premature to assign special status to lateral HG solely on the basis of neuroactivation patterns. Further work should consider the functional roles of these multiple pitch processing sites within the proposed network.

Keywords: lateral Heschl's gyrus, perceptual invariance, planum temporale

Introduction

Pitch is one of the primary auditory sensations. Pitch conveys prosodic information in English and semantic information in tonal languages such as Mandarin, is the most important perceptual dimension of Western music, and is one of the main cues that allow us to separate sounds arising from different sound sources (for example, 2 people speaking at once). Although pitch has been studied extensively using behavioral and neurophysiological techniques, we are only beginning to understand the neural mechanisms that underlie the sensation. A recent study in primates has reported pitch-selective neurons close to primary auditory cortex (Bendor and Wang 2005), consistent with human neuroimaging results, suggesting that a general code for pitch might emerge at, or below (Griffiths et al. 2001), this stage in the auditory pathway. However, this hypothesis has not been tested rigorously.

Pitch can be evoked using stimuli with very different physical characteristics. Most of the tones that we hear in everyday life, for example vowel sounds and the sounds from musical instruments, are complex tones, consisting of a series of harmonic sinusoidal components with frequencies equal to integer multiples of the repetition rate or fundamental frequency (F_0). The cochlea separates out the frequency components of sounds to a limited extent, so that the first 8 harmonics of a complex tone excite distinct places in the

cochlea and are said to be "resolved," whereas the higher harmonics are not separated, and are said to be "unresolved." The organization of the peripheral auditory system is tonotopic: different neurons respond to the different frequency components isolated by the cochlea. Each neuron has a characteristic frequency to which it is most sensitive, and the response of the neuron is largely determined by the spectrum and level of the incoming sound. However, we can associate the same pitch with sounds that have very different spectra, and which therefore activate different groups of neurons in the auditory nerve and in the auditory nuclei of the brainstem. For example, a complex tone with an F_0 of 200 Hz and harmonics 1-5 (200-1000 Hz) will evoke the same pitch as a complex tone with an F_0 of 200 Hz and harmonics 6-10 (1200-2000 Hz). In the latter case, the component with a frequency equal to its F_0 is absent, yet the pitch is unaffected. Although the F_0 component can be reintroduced by cochlear distortions (Pressnitzer and Patterson 2001), by masking the F_0 region with noise it has been demonstrated conclusively that a pitch corresponding to F_0 can be heard in the absence of spectral energy at F_0 (Licklider 1956). Pitches can also be evoked by specific patterns of correlation between the inputs to the 2 ears, without any monaural pitch information. One such binaural pitch is "Huggins pitch" (Cramer and Huggins 1958), produced by presenting the same broadband random noise to both ears, except for a narrow frequency range in which the input to the 2 ears is different (decorrelated). Listeners report hearing a pitch corresponding to the center of this frequency range, and this pitch can be used to produce musical melodies (Akeroyd et al. 2001), even though the inputs to each ear presented in isolation (monaurally) do not evoke a pitch. Stimuli that evoke Huggins pitch contain, at least at the level of each ear, no distinctive spectro-temporal features and so offer stringent experimental control for isolating the neural response to pitch.

Because humans can match sounds on the basis of pitch alone, at some stage in the auditory pathway we might hope to find neurons whose response is governed, not by the spectral or binaural characteristics of the stimulus, but by the pitch evoked by the stimulus. So we might find neurons tuned to a pitch corresponding to 200 Hz, irrespective of how that pitch is produced. A recent primate neurophysiological study (Bendor and Wang 2005) reported that some neurons within a region immediately anterolateral to primary auditory cortex were tuned to a characteristic F_0 , irrespective of harmonic content. These "pitch neurons" produced a selective response to a particular F_0 in the absence of spectral energy in the F_0 region. The results are consistent with human functional magnetic resonance imaging (fMRI) and positron emission

tomography (PET) studies indicating that lateral Heschl's gyrus (HG), the putative homologue of the region identified in the primate study, responds selectively to temporal regularity or periodicity (Griffiths et al. 1998; Patterson et al. 2002; Penagos et al. 2004; Hall et al. 2005, 2006; Barrett and Hall 2006). It has been suggested that lateral HG might function as a general "pitch center" (Bendor and Wang 2006).

Most of the fMRI and PET studies cited above have used one specialized type of pitch-evoking stimulus: iterated-ripple noise (IRN). IRN is produced by generating a random noise sample, delaying it, and adding it back to the original. IRN has a pitch corresponding to the inverse of the delay, and the pitch strength increases with the number of delay-and-add iterations. Although IRN is one stimulus that produces a clearly musical pitch, for a brain region to be confirmed as a *general* pitch center, it should respond to all pitch-evoking stimuli. Preliminary reports (Hertrich et al. 2005; Hall and Plack 2007) suggest somewhat different responses for IRN and Huggins pitch and so a more rigorous examination of the evidence for a pitch center is now required.

The response profile of a general pitch center should satisfy the following criteria: 1) *Pitch selectivity*. Responses to a pitch-evoking stimulus should be distinct from that to a control stimulus that does not evoke a pitch percept, but is matched as closely as possible with respect to its acoustic features. 2) *Elimination of peripheral phenomena*. When examining a region for pitch selectivity with missing F_0 stimuli, it must be possible to discount the contribution of peripheral effects, such as cochlear distortions (McAlpine 2004), to the pitch-evoked response; 3) *Pitch constancy*. Responses should occur for all pitch-evoking stimuli, whatever their spectral, temporal or binaural characteristics and irrespective of whether there is spectral energy at F_0 . The contrast observed between a pitch stimulus and its control might be due to a perceptual difference other than pitch, such as timbre or perceived spatial position. To rule out this possibility, the observation should be replicated with a range of pitch-producing stimuli. Our unique experimental design eliminates these confounds by using a wide range of pitch conditions and by searching for the center of activity that is present in all of the pitch contrasts; 4) *Covariation with salience*. Pitch-sensitive neurons in awake primate cortex are known to increase their discharge rate as a function of salience (where salience has been inferred from the temporal regularity of a click train or the number of iterations of IRN) (Bendor and Wang 2005). The capacity of neuroimaging techniques to detect changes in the response to pitch salience is not well proven. Early results from a PET study were driven by the difference between the pitch and the noise rather than by the increasing salience across pitch conditions (Griffiths et al. 1998), although more recent fMRI results have indicated an increase in activity as the number of iterations of IRN increases from 1 to 16 in several regions of the auditory cortex (Hall et al. 2005). Certainly, the electrophysiological findings would predict an increase in the magnitude of the pitch-related blood-oxygen-level-dependent (BOLD) activity with pitch salience.

The experiments reported here constitute the first attempt to identify pitch processing sites according to these criteria, using a range of pitch-evoking stimuli and combining psychophysical measures of pitch discriminability (as a marker of salience) with fMRI measures of the pitch response.

Materials and Methods

Stimuli

In Experiment 1, 5 different stimuli were created which each evoked a pitch corresponding to that of a 200-Hz pure tone (Supplementary Fig. 1):

- (1) T: A 200-Hz single-frequency tone;
- (2) WB: Wideband complex consisting of the harmonics of a 200-Hz F_0 added in cosine phase and low-pass filtered at 2 kHz;
- (3) Res: Resolved complex without an F_0 component consisting of the harmonics of a 200-Hz F_0 added in cosine phase and bandpass filtered between 1 and 2 kHz;
- (4) Unres: Unresolved complex without an F_0 component consisting of the harmonics of a 100-Hz F_0 added in alternating sine and cosine phase and bandpass filtered between 1 and 2 kHz to produce a pitch corresponding to 200 Hz;
- (5) Huggins: Huggins pitch stimulus consisting of a Gaussian noise low-pass filtered at 2 kHz and presented diotically, except for a frequency region from 190 to 210 Hz ($200 \text{ Hz} \pm 5\%$). This region was given a progressive phase shift, linear in frequency between 0 and 2π , in the left ear only. Huggins pitch stimuli contain no distinctive spectro-temporal features at either ear and so offer stringent experimental control to rule out the possibility that an F_0 component is introduced via peripheral nonlinearity (Pressnitzer and Patterson 2001; McAlpine 2004).

Signals were generated digitally with 16-bit resolution at a sampling rate of 48 kHz. A low-pass noise (filtered at 1 kHz) was added to the missing F_0 complexes to mask cochlear distortions. The single-frequency tone (1) included a bandpass noise (filtered between 500 Hz and 2 kHz) in order to match its gross spectral envelope to that of the other stimuli. A "nonpitch" control stimulus was also generated and, to match for the acoustic energy in each pitch stimulus, it consisted of a Gaussian noise low-pass filtered at 2 kHz. Low-pass noise has been the control stimulus of choice for most neuroimaging studies of pitch processing. All the stimuli were matched in terms of gross spectral envelope and overall level (83-dB SPL for the behavioral measurements and 90-dB SPL for the fMRI measurements, measured at the ear). For the behavioral measurements, the noise, when present, had a spectrum level (level in each 1-Hz wide band) of 50 dB (re. $2 \times 10^{-5} \text{ N/m}^2$), the single-frequency tone had a level of 77 dB SPL [$50 + 10 \log_{10}(500)$], the harmonics of the 200-Hz complexes had a level of 73 dB SPL [$50 + 10 \log_{10}(200)$], and the harmonics of the 100-Hz complex had a level of 70 dB SPL [$50 + 10 \log_{10}(100)$]. Hence the overall level of each stimulus was the same, and the gross spectral density (i.e., the average power per Hz) was constant from 0 to 2 kHz. With the exception of the Huggins stimulus, stimuli were presented diotically (i.e., the same stimulus to both ears). Stimuli had a total duration of 200 ms with 10-ms raised-cosine onset and offset ramps and were delivered via Sennheiser HD580 headphones. For the fMRI measurements the levels were increased by 7 dB and the stimulus duration was 500 ms, including 10-ms raised-cosine onset and offset ramps. Stimuli of one class were repeated in a 15.5-s sequence, with 50-ms gaps between each stimulus. The order of the stimulus conditions was fully counterbalanced. Listeners completed 2 h of psychophysical testing and a 50-min scanning session.

In Experiment 2, diotic IRN was generated by a delay-and-add process performed on a bandpass-filtered (1–2 kHz) Gaussian noise. A copy of the noise segment was added back onto the original after a delay of 10 ms had been imposed onto the copy. The delay-and-add process was repeated for 16 iterations to generate a salient pitch percept. Because many earlier neuroimaging studies have failed to adequately rule out the contribution of neural responses to low-frequency distortions for spectrally complex stimuli (but see Hall et al. 2006), here IRN was presented with and without a low-pass (0–1 kHz) Gaussian noise masker with the same spectrum level as the IRN to quantify the effects of cochlear distortion in temporal pitch coding (Supplementary Fig. 1). The low-pass noise masks distortion products at F_0 and its harmonics and so it is more conservative than narrowband maskers centered on the peak of the distortion product (Hall et al. 2006). For comparison with each IRN stimulus, a control Gaussian

noise with an equivalent bandwidth was also generated. Listeners completed a 30-min scanning session in which the stimulus duration and sound level were the same as in the fMRI session for Experiment 1.

Psychophysical Estimates of Pitch Salience

Pitch salience was estimated in a sound-proofed booth using a measure of individual pitch discrimination threshold. On each trial there were 2 observation intervals separated by 500 ms, containing a standard and a comparison tone, assigned at random. The frequency, F_0 , or (in the case of Huggins) center frequency of the phase-shifted region, of the standard was fixed to produce a nominal pitch corresponding to 200 Hz. The frequency of the comparison was greater than this. The discrimination task was pitch direction (“in which interval was the pitch higher?”). Discrimination thresholds were measured using a two-down, one-up, adaptive procedure that estimates the 71% correct point on the psychometric function (Levitt 1971); for every 2 consecutive correct responses, the frequency difference was decreased for the subsequent trial, and for every incorrect response the frequency difference was increased. The frequency difference between the standard and comparison intervals was varied using a geometric step size of 2 for the first 4 reversals (transitions between decreasing and increasing portions of the adaptive track), and 1.414 thereafter. In each block of trials, 16 reversals were measured and the threshold taken as the geometric mean frequency difference at the last 12. Five such estimates were made for each condition, and the final estimate was taken as the geometric mean of the last 4. Two of the subjects (#10 and #12) could not hear the Huggins pitch and had thresholds greater than 100%. The thresholds for these subjects were assumed to be 100% for the purpose of subsequent analysis.

fMRI Protocol

Scanning was performed on a Philips 3 T Intera using an 8-channel SENSE receiver head coil and a SENSE factor of 2 to reduce image distortions. For each listener, a 4.5-min T_1 -weighted image (1-mm³ resolution) was acquired first magnetization prepared rapid acquisition gradient echo (sequence; matrix = 256 × 256 × 160; time repetition [TR] = 8.2 ms; time echo [TE] = 3.7 ms; flip angle = 8°). This whole-head anatomical scan was used to position the subsequent functional scan centrally on HG. Functional scans consisted of 20 slices taken in an oblique-axial plane, with a voxel size of 3 mm³ (single shot fast field echo sequence; matrix = 64 × 64 × 20; TR = 8000 ms; TE = 36 ms; flip angle = 90°). We took care to include the superior temporal plane and superior temporal sulcus and to exclude the eyes. To eliminate the effect of the scanner noise on patterns of auditory cortical activation, functional scanning used a modification to the pulse sequence (SoftTone factor 2) to reduce the background scanner noise level (by 9 dB) and scans were collected at regular 8-s intervals, with the stimulus presented predominantly in the quiet periods between each scan. To equate the within-subject statistical power across the 2 experiments, each one comprised a total of 44 scans for each stimulus type and an additional 46 silent baseline scans, with the order of conditions randomized. Listeners were requested to attend to the sounds and to listen out for the pitch, but were not required to perform any task.

Analysis of the imaging data was conducted using SPM2 (www.fil.ion.ucl.ac.uk/spm) separately for each listener. Preprocessing steps included within-subject realignment and spatial normalization. For each subject, normalized images were up-sampled to a voxel resolution of 2 mm³ and smoothed by 4 mm full width at half maximum. This procedure meets the smoothness assumptions of the statistical model without compromising much of the original spatial resolution, so preserving the precise mapping between structure and function. Pitch-related brain activation was identified using the principal of the general linear model applied to the smoothed normalized images for each listener using standard procedures implemented in SPM2. The first-level individual analysis used a model that partitioned the observed response according to a sum of 6 weighted variables (the 5 pitch conditions and the noise control). Low-frequency artifacts in the time series, associated with physiological fluctuations, were handled by applying a high-pass filter with a cut-off of 0.002 Hz. After model estimation, statistical contrasts between each pitch condition and the noise control were specified by a linear combination of the corresponding variables and the

significance of each contrast was determined relative to the scan-to-scan residual variability. Individual contrasts were combined across the group using 2 approaches that each underpinned a different class of inference about the general pattern of pitch-related activation. A random-effects analysis expresses the typical characteristics of the population ($P < 0.05$, corrected for multiple comparisons) and it assesses the statistical significance of activity by comparing its mean value to its variability across subjects (Friston et al. 1999). However, when the between-subject variance is high and the mean activation signal is weak, this approach can prove rather unreliable and insensitive (Thirion et al. 2007). In such circumstances, an alternative and informative way to express the results is to plot an incidence (“probability”) map. This is a descriptive statistic that depicts the percentage of subjects that exhibit activity at a particular brain location and is generated by summing individual, thresholded statistical maps, typically thresholded between $P < 0.05$ (Keilholz et al. 2004; Moylan Governo et al. 2006) and $P < 0.001$ (Hall et al. 2005), uncorrected for multiple comparisons. In the present study, a probability threshold of $P < 0.01$ was chosen because it contributed information about the distribution of weak pitch-related activation for every listener (see also Hall and Plack 2007).

Listeners

Sixteen normally hearing listeners (≤ 25 dB hearing level between 250 Hz and 6 kHz) participated in Experiment 1. Their mean age was 24.5 years old, ranging from 18 to 40 years, and the group comprised 7 females and 9 males. A majority of listeners were musically trained; with only 2 listeners unable to read music or play an instrument (#10 and #14). All except one listener (#03) were right handed. Nine of these listeners volunteered to return and participate in Experiment 2. Recruitment of the same listeners reduces the effect of between-subject variability in functional neuroanatomy enabling more precise comparison of results across experiments. The study was approved by the University Medical School Ethics Committee and written informed consent was obtained from all participants.

Results

Experiment 1: Single-Frequency Tone, Complex Tones, and Huggins Pitch

Pitch discrimination thresholds (Fig. 1a) revealed that the single-frequency and WB complex tones were the most salient, followed by the resolved and unresolved complexes which elicited very similar discrimination thresholds, in contrast to previous reports of higher thresholds for unresolved than for resolved complexes (Shackleton and Carlyon 1994). As expected, although the Huggins pitch evoked a clear pitch percept (for all but #10 and #12) it was the least salient.

Pitch Selectivity is Broadly Posterior to Lateral HG

Pitch-related activity was determined by contrasting the response to each pitch-evoking stimulus with that to the noise control. This comparison examines pitch selectivity because significant effects indicate a greater response to a pitch-evoking stimulus than to an acoustically matched stimulus that does not evoke a pitch percept. A random-effects analysis that combined the 5 pitch contrasts using a repeated measures ANOVA revealed a small peak of activity in right planum temporale (PT) (x 64, y -18, z 4, 10 voxels), but this only reached significance at $P = 0.01$ and did not survive the correction for multiple comparisons (Fig. 2a). An incidence map showing the distribution of the pitch effects across the group was generated from the sum of all the individual “pitch versus noise” contrasts. The most reliable site of pitch-related activity (x 58, y -30, z 10) was close to the peak identified by

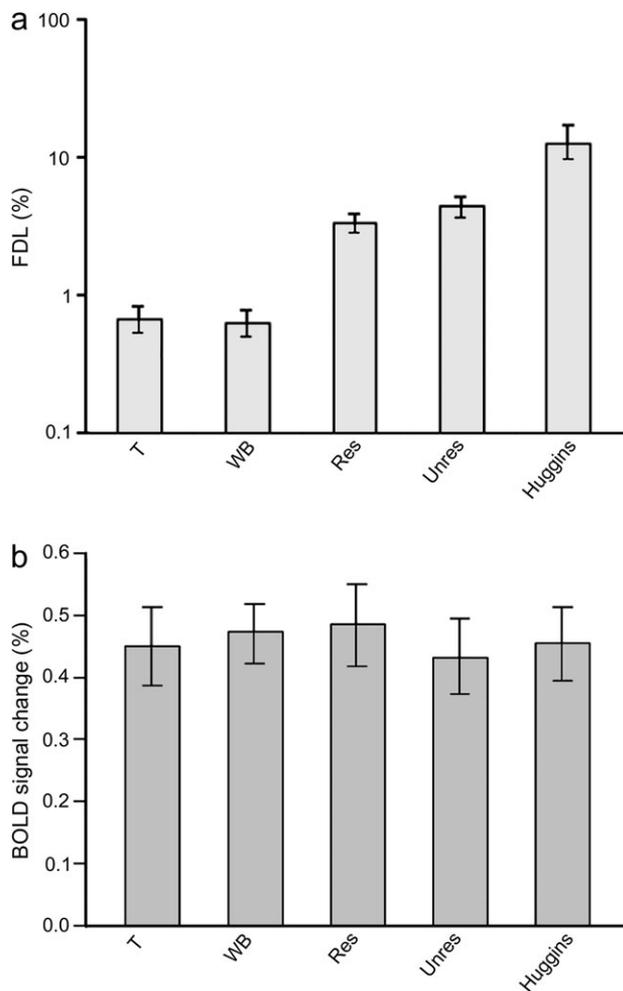


Figure 1. (a) Frequency difference limens (FDLs), derived from the discrimination task, provide a surrogate marker for pitch salience. (b) Pitch-related BOLD signal change measured in each of the individual pitch regions reported in Table 1. All error bars represent the standard error of the mean.

the more stringent random-effects analysis. Activity was predominantly posterior to HG, with the response mostly situated within PT (Fig. 2a). Moreover, this pattern was repeated across separate assessments of each pitch contrast (Fig. 2b); even for the harmonic complex stimuli with missing F_0 . Although this repeating pattern concurs with the criterion for pitch constancy, the consistency from listener to listener was surprisingly low. For example, even the WB complex produced a pitch-selective response in the same location in only 25% (4/16) of listeners (Fig. 2b). Therefore comparisons between pitch and noise were unable to identify a single pitch center that was common to all listeners.

Although lateral HG was highly responsive to all of the sound stimuli, when the average magnitude of the response within this region for each of the pitch conditions was compared with that of the spectrally matched noise control, t -statistics ($df = 15$) revealed no significant difference ($P > 0.05$) (Fig. 2c). This finding accounts for the general paucity of pitch-selective activation observed within this region in the present dataset. It is not the case that suprathreshold activity is completely absent from lateral HG. For example, there were at least 5 activated

voxels in 4 out of 16 listeners for the tone (T) contrast and in 7 listeners for the WB complex. However, although the typical group response profile of posterior regions in PT supports the claims for both pitch constancy and pitch selectivity, the profile of lateral HG does not. In contrast to previous fMRI findings (Griffiths et al. 1998; Patterson et al. 2002; Penagos et al. 2004; Hall et al. 2005, 2006; Barrett and Hall 2006), the present results demonstrate that lateral HG is much less responsive to other types of pitch-evoking stimuli than it is to IRN.

Pitch Constancy Occurs in Different Posterior Brain Sites Across Individuals

It is plausible that highly localized, pitch-selective responses are present within an individual brain, but are obscured by large variability between listeners and so another approach to assessing the evidence for pitch constancy is required. This section reports the results from the incidence maps generated for individual listeners using the sum of their 5 “pitch versus noise” contrasts. Eleven listeners produced consistent activation for at least 4 of the pitch contrasts, whereas a further 4 listeners produced consistent activation for at least 3 contrasts. This is a conservative approach which defines those pitch regions that strongly concur with the first 3 criteria. Again, the response was typically sited in the posterior auditory cortex (Table 1). In 9 listeners it was situated in PT and in 4 listeners it was around the temporo-parietal junction; although in 2 of these listeners activity occurred at both sites (#7, #15). As we suspected, the location of the pitch site varied markedly from listener to listener. In some cases, the site was not even in what is conventionally considered to be the auditory cortex (e.g., on the superior bank of the sylvian fissure in the prefrontal cortex). Nevertheless, the probability of this effect occurring by chance is extremely small ($P < 5 \times 10^{-8}$ for each listener) and so all of these sites provide good evidence for their involvement in generic pitch coding.

There was no support for hemispheric asymmetry in pitch coding (Patterson et al. 2002). In most listeners pitch sites were found bilaterally, but when only one hemisphere produced a significant response, it could occur either on the left or the right.

No Relationship Found between Response Magnitude and Pitch Salience

The final criterion for confirmation of a general pitch center is that of an association between the magnitude of the pitch response and that salience of the pitch. Response magnitude was quantified as the percentage BOLD signal change for each pitch condition relative to the noise control and was measured in each of the individual pitch regions reported in Table 1. We used psychophysical measures of the salience of the 5 pitch stimuli for the same set of listeners. Based on previous findings, one would expect to find a negative correlation in which a low pitch discrimination threshold is associated with a large pitch-related response. However, a partial correlation, controlling for the different pitch-evoking stimuli, failed to support this prediction ($r^2[112] = 0.14$, $P > 0.05$). A summary of the average pitch response magnitudes is presented in Figure 1b and this illustrates the disparity between pitch perception and the fMRI response across each of the 5 pitch conditions.

The lack of any significant covariation between response magnitude and salience does not strongly refute the claim for

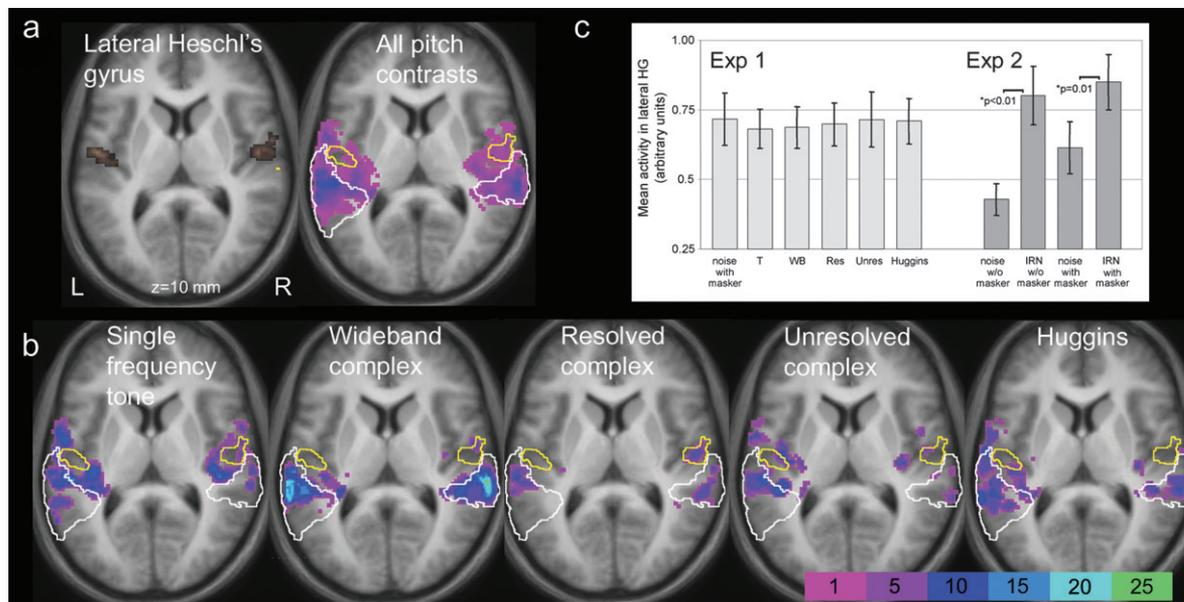


Figure 2. Distribution of pitch-selective activation across a horizontal section of the auditory cortex, shown as an incidence map of activation across the group. (a) Left: The spread of activation can be viewed relative to the position of lateral HG which, for our purposes, was defined by the anatomical area Te1.2 (Morosan et al. 2001) (containing 497 voxels). The yellow marker illustrates the x and y extent of the pitch-related activity identified by the stringent random-effects analysis. (a) Right: The total activity pattern generated by all 5 pitch contrasts in Experiment 1, summed across the group of 16 listeners. The yellow border denotes Te1.2 and the white border is PT (Westbury et al. 1999). Note the plane ($z = 10$ mm) contains the most probable site of pitch-related activity ($x = 58$, $y = -30$, $z = 10$ mm). (b) The distribution of activation separately for the different pitch contrasts. The color scale represents the percentage of pitch-selective activation at every voxel and is calculated as a proportion of a possible maximum of 80 in (a) (5 pitch effects \times 16 listeners) and of a possible maximum of 16 in (b). (c) The response magnitude to each sound condition relative to a “no sound” (silent) baseline, averaged across all voxels within the area Te1.2. Errors bars represent the standard error of the mean. For later comparison, the mean activity for those pitch stimuli in Experiment 2 is also shown.

Table 1

Individual cortical regions of pitch constancy (i.e., 4 or more pitch-selective responses occurring at the same co-ordinate, $P < 5 \times 10^{-8}$)

Left hemisphere										Right hemisphere								
Listener	Co-ordinate			Location	T	WB	Res	Unres	Huggins	Co-ordinate	Location	T	WB	Res	Unres	Huggins		
1	-64	-24	12	PT	X	X	X	—	X	66	-32	10	PT	X	X	X	—	X
2										64	-24	-2	STS	—	X	X	—	X
3										72	-38	2	PT	X	X	X	X	—
4										70	-10	0	PT	—	—	X	X	X
5										42	-44	24	PT	—	X	X	X	X
6	-52	-28	12	PT	X	X	—	X	X	38	-22	2	Medial HG	X	X	—	X	X
7	-56	-32	22	TPOJ	—	X	X	X	X	56	-36	30	PT	—	X	X	X	X
	-42	2	10	Insula	—	X	X	X	X									
8										62	-14	-8	STS	—	X	—	X	X
9	-60	10	4	Prefrontal	X	—	X	X	X									
10	-60	-32	12	PT	X	X	X	X	—									
11	-46	18	-2	Prefrontal	X	X	X	X	X									
12										62	-34	-8	MTG	X	X	—	—	X
13	-40	-32	16	PT	X	—	X	X	X	42	12	-20	PP	X	X	X	X	X
14	-54	-48	30	TPOJ	X	X	—	X	X									
15	-60	-39	18	TPOJ	X	X	X	X	-	48	-28	20	PT	X	X	X	X	X
16										62	-40	28	TPOJ	—	—	X	—	—

Note: Crosses identify which pitch stimuli contribute to the pitch constancy at each location. Key to abbreviations used: MTG = middle temporal gyrus, PP = planum polare; TPOJ = temporo-parieto-occipital junction.

a general pitch site because it may simply reflect a lack of sensitivity of the fMRI measurements or that response size does not reflect the accuracy of the representation of F_0 (our measure of salience).

Experiment 2: IRN

The results from Experiment 1 provide evidence to dispute the special role of lateral HG in human pitch coding. Given the important implications of these findings, it was essential to determine whether the discrepancy between our results and

earlier findings might be attributable to methodological differences. In particular, we were concerned that IRN stimuli used in previous studies might evoke responses due to the nonlinear distortions in the cochlea or to other specific features of IRN that are unrelated to pitch. Thus, Experiment 2 measured the response to IRN for 9 of the original listeners.

IRN Preferentially Engages Lateral HG

A random-effects analysis combined the 2 “IRN versus noise” contrasts (i.e., with and without the low-pass noise masker)

using a repeated measures ANOVA. Overall IRN-related activation encompassed HG and PT bilaterally ($x = -50, y = -14, z = 0$ mm, 259 voxels and $x = 58, y = -10, z = 4, 294$ voxels). Activation was highly reliable because it survived a statistical threshold of $P < 0.05$ (corrected for false discovery rate). In addition, for comparison with Experiment 1, 2 separate incidence maps were generated for each "IRN versus noise" contrast to depict the percentage of listeners exhibiting IRN-related activation at each point in the auditory cortex. Again, activity extended across the auditory cortex, including PT, the same region identified in the group analysis of Experiment 1. However, the distribution of IRN-related activation differed from the pitch-related activation observed in Experiment 1 in 3 key ways. First, unlike our preceding pitch contrasts, the effect of IRN was centered bilaterally on HG, particularly on its lateral portion (Fig. 3). The magnitude of the IRN activity within lateral HG is quantified in Figure 2c and t -statistics ($df = 8$) clearly reveal a significant response to IRN compared with its spectrally matched noise control ($P \leq 0.01$). The effect in lateral HG was not significant for any of the pitch stimuli used in Experiment 1. A second crucial difference was the consistency across listeners; indicative of the highly reproducible anatomical localization of the IRN-selective response. For example, the IRN with the noise masker produced an IRN-selective response with a maximum consistency across the individual maps of 55% (5/9 of listeners) in the left lateral HG ($x = -55, y = -12, z = 4$) and 78% (7/9 of listeners) in the right central HG ($x = 46, y = -18, z = 0$). None of the effects measured in Experiment 1 for pitch-evoking stimuli approached this degree of reliability (Fig. 2b). A third observation (not shown) was that although the IRN-selective activation was widespread, none of the pitch sites previously identified in individual listeners (Table 1) also responded to IRN.

The IRN effect was somewhat diminished when the low-pass noise masker was added to the stimuli to mask potential low-frequency cochlear distortions. The smaller effect is due to the increase in activity for the stimuli that included the low-pass noise, possibly as a consequence of the increased frequency bandwidth of the signal. However, it is important to note that both IRN stimuli generated reliable IRN-related activation in lateral HG. We conclude that the low-frequency masker did *not* abolish the IRN-selective response in HG and in PT indicating that the previously reported effects for IRN in lateral HG cannot simply be explained by a response to the distortion products.

It could be argued that lateral HG is better defined by the focus of some pitch-related activity than by canonical anatomical criteria, especially given known differences across brains (Morosan et al. 2001). Therefore an IRN-related region of interest was defined using the contrast for the IRN with the masker. It included reliable voxels that were present in at least 4 out of the 9 listeners (118 voxels). Despite positional differences between the functional and the anatomical regions of interest, the pattern of results remained broadly equivalent (Fig. 4). The only new result was a significant *deactivation* ($P < 0.05$) for the WB complex tone.

Discussion

No previous imaging study has searched for a generalized pitch response using such a wide range of stimuli. PT was reliably activated by many of the pitch stimuli suggesting that

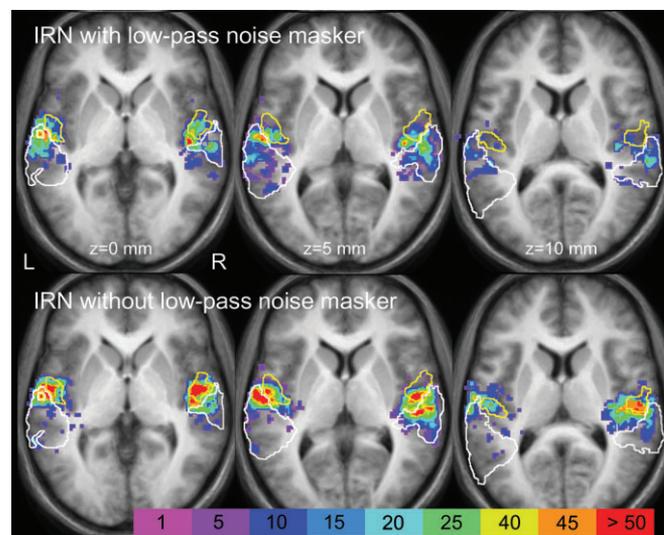


Figure 3. Distribution of IRN-selective activation across 3 horizontal sections of the auditory cortex, shown as an incidence map of activation across the group. Again, the yellow border denotes Te1.2 (Morosan et al. 2001) and the white border is PT (Westbury et al. 1999). IRN activity is denoted here as the summation of a binary map of activation ($P < 0.01$) that had been generated for IRN > noise contrasts for each of the 9 listeners in Experiment 2. The middle row shows the distribution for the IRN contrast with the low-pass noise masker and the bottom row shows that for the IRN contrast without the low-pass noise masker. The most probable peaks occur at $x = 46, y = -18, z = 0$ mm and $x = -62, y = -13, z = 5$ mm, respectively. For comparison within and between experiments, the effects of IRN are displayed for both of these axial planes, plus $z = 10$ mm. The color scale represents the percentage of IRN-selective activation at every voxel. The color range is directly comparable with Figure 2.

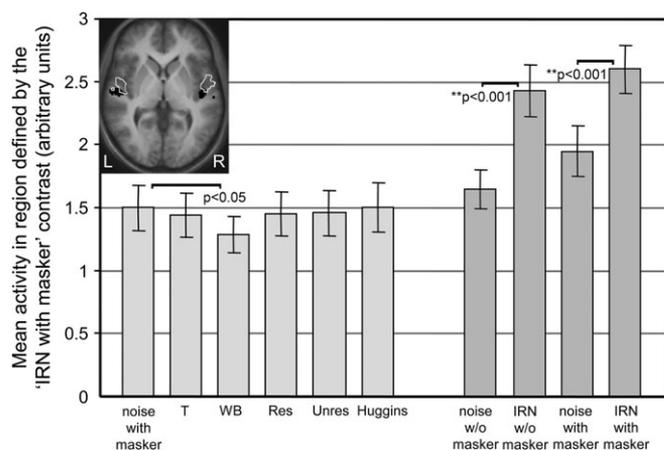


Figure 4. Applying the same procedures that were used to generate Figure 2c, pitch-related activity was recomputed for a different region of interest. This region was functionally defined by the focus of IRN-related activity obtained in Experiment 2 and its position and extent is shown in black in the inset ($z = 0$ mm). For comparison, the border of Te1.2 is also shown.

a generalized representation of pitch could be formed later in the auditory processing stream than previously considered. However, it would be unwise to assign special status to any particular brain region because our results demonstrate the involvement of multiple distinct sites in the human brain that might support different levels of pitch analysis. Future investigations that manipulate the pitch stimulus *and* the

pitch task would be required to tease apart the functional aspects of this network.

Differing Pattern of Activation for IRN May Result from Features not Related to Pitch

Although the IRN results replicate previous findings, none of the 5 other pitch-evoking stimuli generated reliable differential activation in lateral HG. These results raise the possibility that the IRN-related activity, believed until now to represent neural coding of pitch, does not reflect a response that is specific to the periodicity information conveyed in the signal (de Cheveigne 2007). This is a new claim because IRN is typically upheld as a well-controlled pitch stimulus (Griffiths et al. 1998, 2001; Patterson et al. 2002).

IRN is often preferred because pitch salience can be controlled easily by manipulating the number of delay-and-add iterations. If IRN is filtered into a high spectral region, the spectral peaks at a frequency spacing of $1/\text{delay}$ cannot be resolved by the auditory system, and hence such a stimulus is thought to evoke a pitch based on the fine-structure temporal regularity. However, in addition to fine-structure regularity, IRN also contains slowly varying spectro-temporal features which can be resolved by the ear. The strength of these features increases over the number of add-and-delay iterations. For illustrative purposes, samples of noise and IRN stimuli (0, 1, 4, and 16 iterations) were analyzed using a computational model of auditory processing (Plack et al. 2002) that includes a nonlinear filterbank to simulate frequency selectivity in the cochlea (Fig. 5). The response to IRN reveals broad spectral features that sweep in time and these features are not present in the response to noise. Moreover, the spectrograms clearly demonstrate that increasing IRN pitch salience is confounded by an increase in the strength of these broad spectral features. Thus, we suggest that these features of IRN, not the pitch-evoking features, may be responsible for generating the differential activation in lateral HG.

Electrophysiological recordings indicate that neurons in primary auditory cortex are sensitive to broad spectro-temporal features, including frequency modulations such as those seen in IRN (Depireux et al. 2001). Our proposal is further supported by a range of human fMRI studies using other types of complex sounds including modulated tones, dynamic spectral ripples and sine wave speech; whose modulation rates are typical of those observed in IRN. Notably, lateral HG and regions posterior to it generate the strongest and most sustained response to slow rates of sinusoidal modulation (4–8 Hz), whether this be slow fluctuations in frequency or in amplitude (Giraud et al. 2000; Hart et al. 2003). The same regions also respond best to the slowly changing large scale features of a dynamic ripple stimulus (i.e., those containing a combination of a slow rate of temporal modulation, for example, 2 Hz, and small number of spectral peaks in one octave) (Langers et al. 2003) and also greatly modify their response to sine wave words after training (Liebenthal et al. 2003). These latter results have been interpreted as a specialization toward the processing of phonetic qualities of speech including formant transitions.

Comparison with Other Studies

Surprisingly few neuroimaging investigations have mapped pitch selectivity using stimuli other than IRN. However, some

results appear to be consistent with our claim that pitch sites include PT. One study presented harmonic complex tones broadly similar to those used in Experiment 1 (Penagos et al. 2004). The authors did not map out the differential response to the pitches compared with the spectrally matched noise, and so it is difficult to make a definitive comparison with our own data, but instead they did compare the response to harmonic complex tones containing either resolved or unresolved components with the assumed salience of the tones (strong and weak respectively) in a small number of listeners ($N = 5$). Notwithstanding the alternative view that the effect of salience could be due to the detection of gross spectral features present for the resolved harmonics, we can reconsider the location of the putative salience-related effects by looking carefully at the individual activation patterns reported. Salience-related activity occurred in multiple, small, scattered patches whose precise location varied from listener to listener. Although the study focused on the response in lateral HG, in 4 out of the 5 listeners, some of these patches are clearly located in PT bilaterally. This result concurs with our findings. A second study that is relevant to this discussion used harmonic complex tones containing both resolved and unresolved components to investigate the response to orthogonal pitch dimensions (height and chroma) (Warren et al. 2003). Conveniently, the authors also contrasted the pitch-evoking stimuli with a broadband Gaussian noise and demonstrated bilateral activation posteriorly in PT and anteriorly in planum polare, as well as in lateral HG. However, it is important to note that these pitch stimuli comprised tone sequences with changing height and/or changing chroma, whereas our pitch sequences contained fixed pitch attributes. As Patterson et al. (2002) have shown, the dynamic changes are sufficient to account for the more extensive pattern of auditory cortical activity and these are more closely related to processing of melody or other slowly varying sound characteristics than to pitch processing per se. A melody would produce clear spectro-temporal fluctuations similar to those that we attribute to IRN.

For well-motivated reasons, 4 of the 5 pitch stimuli used here were presented in the context of noise (in the case of the 2 missing F_0 stimuli, to exclude distortion products), but this implies that the pitch-evoking stimulus was presented as a figure against a noise-like ground from which the listener had to segregate it. Griffiths and Warren (2002) predicted a crucial involvement of PT in tasks of identifying and localizing a single sound object in space; particularly in the presence of a competing sound source. However, this proposal has received little empirical support. The presence of a noise masker has been shown to engage frontal and parietal cortices, suggesting that selective attention plays a critical role in sound segregation (Scott et al. 2004). Where auditory regions have been shown to play a role in figure-ground separation, it has been lateral HG, not PT (Scheich et al. 1998). Several observations from the present experiments are also inconsistent with Griffiths and Warren's prediction. In Experiment 1, the pitch stimulus with no background noise (WB) actually produced the most consistent activity in PT (Fig. 2*b*). In Experiment 2, adding the low-frequency noise masker to IRN decreased PT activity (Fig. 3). Consequently, it is unlikely that the pitch-related activity observed at this higher processing stage can be ascribed to figure-ground separation.

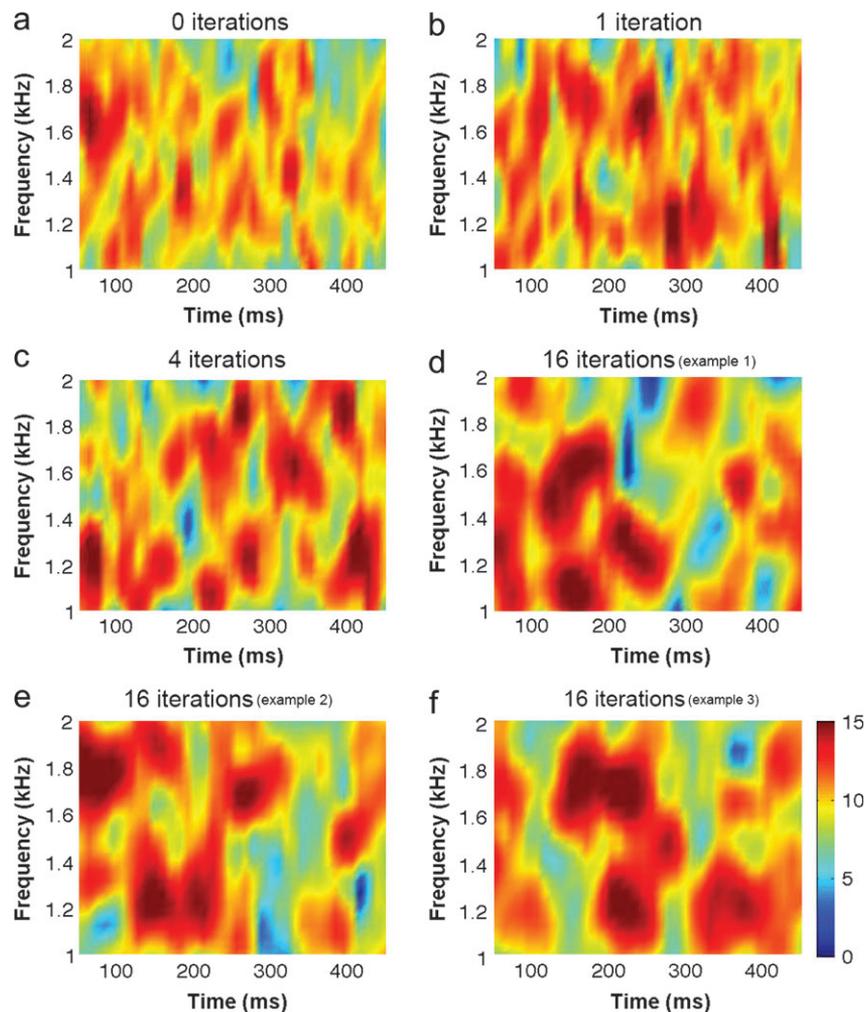


Figure 5. Simulated output of the cochlea in response to the Gaussian noise (0 iterations) (*a*) and to IRN stimuli at a range of add-and-delay iterations (*b–f*). The model output in dB is plotted as a function of time and of the center frequency of each auditory frequency channel (or each place in the cochlea) across a bandwidth of 1–2 kHz. These smoothed spectrograms clearly illustrate the slowly varying spectro-temporal features increasingly present in IRN. Three different examples of the 16-iteration IRN (*d–f*) demonstrate the consistent pattern across IRNs generated using different noise carriers. The relative amplitude of these features in the physical stimulus is greater than depicted here due to the basilar-membrane compression implemented in the model.

The Cortical Pitch Hierarchy, from Pitch Extraction to Melody

Evidence supports the popular view that music is preferably coded in the right hemisphere and speech in the left (Ivry and Robertson 1998; Zatorre et al. 2002; Poeppel 2003). Hemispheric specialization provides one solution to the computational problem posed by the need for optimized coding of low-level acoustic cues. For example, cochlear implant research has shown that music appreciation is special in the sense that it requires fine spectral detail to allow the extraction of complex harmonic pitch in signals where the spectral regions contain fully or partially resolved harmonics (Shannon 2005). One prevailing interpretation is that the right hemisphere has better spectral resolving power—necessary for detecting changes in timbre and phrasing—whereas the left hemisphere is selectively better at coding temporal differences of the order of tens of milliseconds—necessary for perceiving phonetic categories (Zatorre et al. 2002). Supporting evidence comes from neuroimaging studies carried out in normal listeners. For example,

increased responses to the complexity of spectral modulations were found in right-lateralized parts of HG and the antero-lateral border of PT (Schönwiesner et al. 2005). Even the perception of melodies wholly created using temporal pitch (IRN) stimuli produced significant activation in the right-lateralized auditory cortex, including the planum polare and the superior temporal sulcus (Patterson et al. 2002). These authors proposed a hierarchy of pitch processing in the auditory cortex, in which pitch is extracted bilaterally in lateral HG (part of the primary auditory region), whereas long-term variations in pitch are subsequently processed in the superior temporal gyrus and/or planum polare on the right side (part of the nonprimary auditory region). Variations in pitch can also engage anterior parts of right PT (Schönwiesner et al. 2005). Although our critique of pitch coding does not apply to the findings regarding melody processing, our finding that the pitch site occurred more frequently in PT suggests that pitch extraction continues to occur at higher stage of auditory processing than suggested by these authors. It remains unclear what are the precise neural correlates of the functional

hierarchy for pitch processing because our pitch centers seemed to differ between individuals and sometimes involved regions outside the auditory parabelt (Hackett 2003). This variability may be due to the rather unconstrained nature of the pitch listening task used in the present experiment. However, we tentatively suggest a new hierarchy in which a generalized representation of pitch is not fully complete until at least PT and is then conveyed to neighboring regions for the analysis of pitch melody.

Supplementary Material

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

Funding

Medical Research Council (UK); and a Knowledge Transfer Grant from the The Royal National Institute for Deaf People.

Notes

Both authors contributed equally to design, data acquisition, analysis, and the writing of the manuscript. *Conflict of Interest*: None declared.

Address correspondence to Deborah A. Hall, BSc, PhD, MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK. Email: d.hall@ihr.mrc.ac.uk.

References

Akeroyd MA, Moore BCJ, Moore GA. 2001. Melody recognition using three types of dichotic-pitch stimulus. *J Acoust Soc Am*. 110: 1498-1504.

Barrett DJK, Hall DA. 2006. Response preferences for 'what' and 'where' in human non-primary auditory cortex. *Neuroimage*. 32: 968-977.

Bendor D, Wang X. 2005. The neuronal representation of pitch in primate auditory cortex. *Nature*. 436:1161-1165.

Bendor D, Wang X. 2006. Cortical representations of pitch in monkeys and humans. *Curr Opin Neurobiol*. 16:1-9.

Cramer EM, Huggins WH. 1958. Creation of pitch through binaural interaction. *J Acoust Soc Am*. 30:413-417.

de Cheveigne A. 2007. Comment on 'Searching for a pitch centre in human auditory cortex' by Hall DA and Plack CJ. In: Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S, Verhey J, editors. *Hearing—From sensory processing to perception*. Heidelberg: Springer. p. 90-91.

Depireux DA, Simon JZ, Klein DJ, Shamma SA. 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol*. 85:1220-1234.

Friston KJ, Holmes AP, Worsley KJ. 1999. How many subjects constitute a study? *Neuroimage*. 10:1-5.

Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A. 2000. Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol*. 84: 1588-1598.

Griffiths TD, Büchel C, Frackowiak RSJ, Patterson RD. 1998. Analysis of temporal structure in sound by the human brain. *Nat Neurosci*. 1:422-427.

Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD. 2001. Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci*. 4:633-637.

Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. *Trends Neurosci*. 25:348-353.

Hackett TA. 2003. The comparative anatomy of the primate auditory cortex. In: Ghazanfar AA, editor. *Primate audition: ethology and neurobiology*. Boca Raton (FL): CRC Press. p. 199-225.

Hall DA, Barrett DJK, Akeroyd MA, Summerfield AQ. 2005. Cortical representations of temporal structure in sound. *J Neurophysiol*. 94:3181-3191.

Hall DA, Edmondson-Jones AM, Fridriksson J. 2006. Frequency and periodicity coding in human auditory cortex. *Eur J Neurosci*. 24:3601-3610.

Hall DA, Plack CJ. 2007. The human 'pitch center' responds differently to iterated noise and Huggins pitch. *Neuroreport*. 18: 323-327.

Hart HC, Palmer AR, Hall DA. 2003. Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex*. 13:773-781.

Hertrich I, Mathiak K, Menning H, Lutzenberger W, Ackermann H. 2005. MEG responses to rippled noise and Huggins pitch reveal similar cortical representations. *Neuroreport*. 16:193-196.

Ivry RB, Robertson LC. 1998. *The two sides of perception*. Cambridge (MA): MIT Press.

Keilholz SD, Silva AC, Raman M, Merkle H, Koretsky. 2004. Functional MRI of the rodent somatosensory pathway using multislice echo planar imaging. *Magn Reson Med*. 52:89-99.

Langers DRM, Backes WH, van Dijk P. 2003. Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples. *Neuroimage*. 20:265-275.

Levitt H. 1971. Transformed up-down methods in psychoacoustics. *J Acoust Soc Am*. 49:467-477.

Licklider JCR. 1956. Auditory frequency analysis. In: Cherry C, editor. *Information theory*. New York: Academic. p. 253-268.

Liebenthal E, Binder JR, Piorkowski RL, Remez RE. 2003. Short-term reorganization of auditory analysis induced by phonetic experience. *J Cogn Neurosci*. 15:549-558.

McAlpine D. 2004. Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. *J Neurophysiol*. 92:1295-1311.

Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T, Zilles K. 2001. Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage*. 13:684-701.

Moylan Governo RJ, Morris PG, Prior MJW, Marsden CA, Chapman V. 2006. Capsaicin-evoked brain activation and central sensitization in anaesthetized rats: a functional magnetic resonance imaging study. *Pain*. 129:35-45.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. 2002. The processing of temporal pitch and melody information in auditory cortex. *Neuron*. 36:767-776.

Penagos H, Melcher JR, Oxenham AJ. 2004. A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci*. 24:6810-6815.

Plack CJ, Oxenham AJ, Drga V. 2002. Linear and nonlinear processes in temporal masking. *Acustica*. 88:348-358.

Poeppl D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun*. 41:245-255.

Pressnitzer D, Patterson RD. 2001. Distortion products and the perceived pitch of harmonic complex tones. In: Breebaart DJ, Houtsma AJM, Kohlrausch A, Prijs VF, Schoonhoven R, editors. *Physiological and psychophysical bases of auditory function*. Maastricht: Shaker Publishing BV. p. 97-104.

Scheich H, Baumgart F, Gaschler-Markefski B, Tegeler C, Tempelmann C, Heinze HJ, Schindler F, Stiller D. 1998. Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background decomposition. *Eur J Neurosci*. 10:803-809.

Schönwiesner M, Rübsamen R, von Cramon DY. 2005. Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci*. 22: 1521-1528.

Scott SK, Rosen S, Wickham L, Wise RJS. 2004. A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. *J Acoust Soc Am*. 115: 813-821.

- Shackleton TM, Carlyon RP. 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J Acoust Soc Am.* 95:3529-3540.
- Shannon RV. 2005. Speech and music have different requirements for spectral resolution. *Int Rev Neurobiol.* 70:121-134.
- Thirion B, Pinel P, Meriaux S, Roche A, Dehaene S, Poline JB. 2007. Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *Neuroimage.* 35:105-120.
- Warren JD, Uppenkamp S, Patterson RD, Griffiths TD. 2003. Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci USA.* 100:10038-10042.
- Westbury CF, Zatorre RJ, Evans AC. 1999. Quantifying variability in the planum temporale: a probability map. *Cereb Cortex.* 9:392-405.
- Zatorre RJ, Belin P, Penhune VB. 2002. Structure and function of auditory cortex: music and speech. *Trends Cogn Sci.* 6:37-46.