METHODOLOGICAL CONSIDERATIONS FOR fMRI STUDIES
OF PITCH PROCESSING

A thesis submitted to The University of Manchester for the degree of
Doctor of Philosophy

Faculty of Medical and Human Sciences

2010

DAPHNE ANNE GARCÍA

SCHOOL OF PSYCHOLOGICAL SCIENCES
# Table of Contents

List of Figures .................................................................................................................. 4  
Abbreviations .................................................................................................................... 9  
Abstract ............................................................................................................................11  
Declaration .........................................................................................................................12  
Copyright Statement .........................................................................................................13  
Acknowledgments ..............................................................................................................15  
Chapter 1. Introduction ....................................................................................................16  
Chapter 2. Coding of basic acoustical and perceptual components of sound in human auditory cortex .................................................................................................................25  
Chapter 3. General methodology .....................................................................................77  
  3.1 Participants ....................................................................................................................77  
  3.2 Equipment ...................................................................................................................78  
    3.2.1 Psychophysical Testing ..........................................................................................78  
    3.2.2 fMRI Scanning ......................................................................................................78  
  3.3 Stimuli .........................................................................................................................79  
    3.3.1 Psychophysical Stimuli .........................................................................................79  
    3.3.2 Scanning Stimuli ...................................................................................................80  
  3.4 Psychophysical Procedure .........................................................................................81  
  3.5 fMRI Protocol ............................................................................................................82  
    3.5.1 Anatomical Scanning ............................................................................................83  
    3.5.2 Functional Scanning .............................................................................................84  
    3.5.3 Data Analysis .......................................................................................................84  
Chapter 4. The effect of stimulus context on pitch representations in the human auditory cortex .........................................................................................................................91
Chapter 5. The cortical pitch response to jittered pulse trains and harmonic complex tones reveals no increase with increasing pitch salience ......................................................... 92

Chapter 6. Reexamining the evidence for a pitch sensitive region based on a human fMRI study using iterated ripple noise ............................................................................. 117

Chapter 7. Non-pitch attributes of sound influence the auditory cortical responses to pitch ....................................................................................................................... 148

Chapter 8. General discussion and conclusions ............................................................................. 176

References .................................................................................................................................. 183
List of Tables

Chapter 4

Table 1. fMRI results for Huggins pitch and unresolved harmonic complex contrasts.

Chapter 6

Table 1. Significant clusters of activity for main effects of IRN and IRNo, and for the subtraction of IRNo from IRN activity.

Chapter 7

Table 1. Table of previously identified pitch-related responses

Table 2. fMRI results for the maximal effects of pitch and modulation and for the conjunction between pitch and modulation in auditory cortex.

List of Figures

Chapter 2

Figure 7.1. Schematic representation of human and macaque parcellation schemes in auditory cortex.

Figure 7.2. Schematic representation of frequency coding in the central and peripheral auditory system.
Figure 7.3. Spatial distribution of responses to a single-frequency tone and a harmonic-complex tone across the supratemporal plane.

Figure 7.4. Schematic drawings of temporal envelope and distribution of fMRI response shapes for modulated tones in Heschl’s gyrus and superior temporal gyrus.

Figure 7.5. Spatial distribution of the responses to steady-state and frequency-modulated tones across the supratemporal plane.

Figure 7.6. Examples of systematic changes in auditory cortical activity as a function of sound level in ipsi- and contralateral auditory cortices.

Figure 7.7. Simulated cochlear outputs in response to random noise and iterated ripple noise.

Figure 7.8. Incidence map showing auditory cortical increases in activity as a function of pitch salience.

Figure 7.9. Incidence map showing inter-listener consistency for pitch-related activation for five pitch stimuli.

Chapter 4

Figure 1. Graphical representation of the linear and non-linear response models depicting the coupling between neural activity and fMRI activation.
Figure 2. Schematic diagram of the stimulus components for the continuous stimulation paradigm.

Figure 3. Functional imaging results showing the locations of pitch- and context-related activity, as well as regions that are sensitive to conjunction and interaction between the main effects.

Figure 4. Peak locations of pitch-related activity in the current study and their counterparts reported by a previous pitch study (Puschmann et al. 2010), overlaid onto a mean normalized anatomical scan and a corresponding Talairach brain slice (using data reported by Arnott et al. 2004), including coordinates for nonspatial auditory activity.

Figure 5. Graphical representation of the effects of context, pitch and their interactions, plotted separately for the two pitch stimuli in three different regions of auditory cortex.

Chapter 5

Figure 1. 100 ms samples of waveforms for all of the stimuli used in the experiment.

Figure 2. Pitch discrimination thresholds for pulse train and unresolved harmonic complex stimuli with increasing pitch salience.
Figure 3. Distribution of pitch-related activation across horizontal (axial) and vertical (saggital) sections of auditory cortex, shown as an incidence map of activation across the 15 listeners.

Chapter 6

Figure 1. Simulated cochlear representations of original and processed iterated ripple noise in the form of spectrograms.

Figure 2. Graphical representation of the fluctuation depth of the slowly varying modulations for original and processed iterated ripple noise with 0, 4, 16 and 64 iterations.

Figure 3. Psychophysical and corresponding blood oxygenation level dependent responses for the original and processed iterated ripple noise stimuli used in the experiment.

Figure 4. Pattern of fMRI responses for original and processed iterated ripple noise, compared to matched noise controls, showing areas of overlap between the two responses.

Figure 5. Activation values for original and processed iterated ripple noise within the peak voxel for the effect of iteration.
**Figure 6.** Profile plots of activity for original and processed iterated ripple noise with different numbers of iterations for the three pitch-responsive regions within auditory cortex.

**Figure 7.** Incidence maps showing the distribution of activation elicited by iterated ripple noise compared to Gaussian noise and compared to processed iterated ripple noise.

**Chapter 7**

**Figure 1.** Simulated cochlear representations of original and processed iterated ripple noise in the form of spectrograms.

**Figure 2.** Schematic representations of four possible models for an interaction between modulation and pitch.

**Figure 3.** Schematic representation of the stimuli entered into the 2x2 factorial analysis.

**Figure 4.** Activation map showing locations for the main effects of modulation, pitch and a conjunction for the two features across auditory cortex.

**Figure 5.** Activation values for the four stimuli in the 2x2 factorial analysis within lateral Heschl’s gyrus.
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Auditory 1</td>
</tr>
<tr>
<td>AA</td>
<td>Anterior area</td>
</tr>
<tr>
<td>ALA</td>
<td>Anterolateral area</td>
</tr>
<tr>
<td>BOLD</td>
<td>Blood oxygenation level dependent</td>
</tr>
<tr>
<td>cHP</td>
<td>Complex Huggins pitch</td>
</tr>
<tr>
<td>dB</td>
<td>Decibels</td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion tensor imaging</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalography</td>
</tr>
<tr>
<td>f0</td>
<td>Fundamental frequency</td>
</tr>
<tr>
<td>fMRI</td>
<td>Functional magnetic resonance imaging</td>
</tr>
<tr>
<td>HG</td>
<td>Heschl’s gyrus</td>
</tr>
<tr>
<td>IRN</td>
<td>Iterated ripple noise</td>
</tr>
<tr>
<td>LA</td>
<td>Lateral area</td>
</tr>
<tr>
<td>MA</td>
<td>Medial area</td>
</tr>
<tr>
<td>MEG</td>
<td>Magnetoencephalography</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic resonance</td>
</tr>
<tr>
<td>PA</td>
<td>Posterior area</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PP</td>
<td>Planum polare</td>
</tr>
<tr>
<td>PT</td>
<td>Planum temporale</td>
</tr>
<tr>
<td>R</td>
<td>Rostral</td>
</tr>
<tr>
<td>rCBF</td>
<td>Regional cerebral blood flow</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>RT</td>
<td>Rostro-temporal</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>SENSE</td>
<td>Sensitivity encoding</td>
</tr>
<tr>
<td>SPL</td>
<td>Sound pressure level</td>
</tr>
<tr>
<td>STA</td>
<td>Superior temporal area</td>
</tr>
<tr>
<td>T</td>
<td>Tesla</td>
</tr>
<tr>
<td>UNRES</td>
<td>Unresolved harmonic complex tone</td>
</tr>
</tbody>
</table>
Four functional magnetic resonance imaging (fMRI) studies of pitch processing in auditory cortex were designed to reduce the impact of a number of methodological issues that have hitherto limited previous research findings.

Due to adaptation effects, it is necessary to repeatedly present short stimulus bursts rather than long-duration stimuli. Thus, conventionally, in neuroimaging studies of pitch perception, a number of short bursts of the pitch stimulus, separated by silent intervals, are compared to a Gaussian noise presented in the same way. The results of the first experiment indicate that replacing the silent intervals with an energetically matched noise context increases the pitch-specific response by removing the ‘energy-onset response’ that saturates the overall response if silent intervals are used. In the second experiment, a particular pitch-evoking stimulus, iterated ripple noise (IRN), which is commonly used in neuroimaging studies of pitch perception, was examined. Hall and Plack (Cerebral Cortex 2009;19:576-585) showed that IRN contains slowly varying spectro-temporal features unrelated to pitch, and suggested that these features could account for at least some of the cortical activation produced by IRN. The results support this hypothesis, but also suggest that there is an additional pitch-dependent effect in the same region of auditory cortex.

The third experiment assessed the effect of using a different control stimulus to the usual Gaussian noise. The new matched controls were a pulse train with randomly jittered inter-pulse intervals and a random-phase unresolved harmonic complex tone. These low-pitch-salience controls were compared to a regular interval pulse train, which is identical to a cosine-phase unresolved harmonic complex tone. The third experiment did not provide evidence for sensitivity to pitch-salience in pitch-responsive regions of auditory cortex. The fourth and final experiment was a factorial design seeking to answer two main questions: 1) Is the pitch-sensitive region of auditory cortex responsive to the salience of other sound features (e.g. modulation)? 2) Are the responses to pitch and to modulation within this region co-located? Two different pitch-evoking stimuli with different levels of pitch salience were used, presented in a noise context. Results indicate that the pitch-sensitive region contains representations for both pitch and modulation. Furthermore, there was no evidence for an interaction between pitch and modulation, suggesting that the two responses are independent.

Overall, the results suggest that careful stimulus design, and appropriate experimental control, is necessary to obtain reliable information on the cortical response to pitch. In addition, the results have shed further light on the likely neural substrates of pitch processing in the cortex.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright Statement

• The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the “Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

• Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

• The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the “Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

• Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the
University Library, The University Library’s regulations (see http://www.campus.manchester.ac.uk/library/aboutus/regulations) and in The University’s policy on presentation of Theses.
Acknowledgments

I would like to dedicate my thesis to the three most important people in my life: my betrothed, Rob, and my two sisters Nicole and Brittany.

I would like to express my sincere gratitude to my supervisors Prof. Deb Hall and Prof. Chris Plack for the continuous support of my PhD study and research, for their patience, motivation, enthusiasm and infinite wisdom. Their guidance was unfathomable and I could not have achieved anywhere near what I have without their help.

I would also like to thank staff at the Medical Research Council’s Institute of Hearing Research at Nottingham University and at the Sir Peter Mansfield Magnetic Resonance Centre in Nottingham for their help with scanning, testing and analyzing, the participants who gave up their time and the wonderful people at The University of Manchester.

Thanks to the reviewers for their helpful comments, and also to those who provided valuable feedback from various conferences and meetings, specifically Alain de Cheveigné for his advice on the procedure for generating IRNo, and my internal and external examiners Tim Griffiths and Kevin Munro.

All the students who have worked with me on various projects: Simon Müller, Sparsh Chandan, Liam Hennessy and Natasha Bashir – thank you.

Last but not least, my family (including my in-laws-to-be) and the family that I chose; my friends who have kept me sane by supplying encouragement and wine – I love you all!
Chapter 1. Introduction

Pitch is one of the primary auditory sensations. It is the feature of sound from which musical melodies are formed and can be ordered on a scale extending from low to high, or from A to G in musical terms, making it one of the most important features of Western music. As well as its role in music, pitch is vital for the comprehension of tonal languages used in South-East Asia and sub-Saharan Africa, where pitch carries semantic meaning. Even in non-tonal languages such as English, pitch carries a large amount of grammatical information used in comprehension of sentence structure. Pitch is also one of the most important cues we use to separate sounds from different sources.

Understanding how pitch is processed in the brain by normally hearing individuals is imperative in understanding the impact of impairment on pitch perception. An understanding of the mechanisms of pitch perception will provide a basis for further research into the improvement of pitch perception for hearing-impaired individuals, which is currently poor for people receiving certain types of clinical interventions such as cochlear implants (Moore 2003; Moore and Carlyon 2005), brainstem implants (Kuchta et al. 2004; Colletti et al. 2005; Otto et al. 2008) and auditory midbrain implants (Lim et al. 2007, 2008). An improvement of pitch perception for these individuals would have a beneficial effect on the enjoyment of music, overall hearing, and speech comprehension both in quiet and noise, leading to an increased quality of life (Drennan and Rubinstein 2008). Although none of the experiments presented here directly relate to clinical aspects of hearing, they will provide important information that can be taken forward in translational studies that can be used for the development of clinical research. Chapter 2 of this thesis provides an in-depth
introduction to the different types of pitch-evoking stimuli, the theories of pitch perception and the methods used for studying pitch in humans and non-human species, as well as describing parcellation schemes for different areas of human and non-human auditory cortex.

Pitch is a subjective feature of sound. The physical correlates of pitch are frequency (e.g. in the case of single-frequency tones) and periodicity, the repeating pattern of the waveform of the sound. There are innumerable ways in which sound waves can combine to produce a repeating waveform, whose repetition rate determines the value of the pitch (i.e. its fundamental frequency, or f0) (Plack 2005). Sounds with the same f0 evoke the same pitch percept, even if other sound qualities vary greatly. Thus, many auditory neuroscientists have postulated the existence of a region in the brain that is responsible for coding the pitch percept, regardless of its physical features: a human ‘pitch center’\(^1\). It has been suggested that in order for any region of human auditory cortex to be considered a pitch center, it must satisfy four criteria (Hall and Plack 2009). First, it must be selectively responsive to pitch, meaning that it must respond to pitch, but not to a closely matched acoustic stimulus that does not evoke a pitch percept. Second, the contribution of peripheral phenomena such as cochlear distortions (e.g. McAlpine 2004) to the signal must be eliminated. The third criterion is that it must respond to all pitch-evoking stimuli, regardless of the physical generators of the pitch percept. Finally, the magnitude of the pitch response must increase along with the pitch strength, or salience.

\(^1\) The American spelling of the word ‘centre’ is used throughout the thesis because the papers were submitted to American journals. The same spelling was also used in the chapters that were not submitted to journals in the interests of consistency.
This thesis comprises a series of four experiments that each address one or more of the requirements for a pitch center, as well as approaching different methodological issues known to influence the outcome of neuroimaging auditory research. All four of the experiments presented here use a neuroimaging method known as blood oxygenation level dependent (BOLD) functional magnetic resonance imaging (fMRI). This works on the premise that any part of the brain that is responding to an external stimulus uses more oxygen than non-active parts of the brain. This results in a higher ratio of oxy- to deoxyhaemoglobin in active areas, which changes the magnetic properties of the local draining venules and veins that provide oxygen to the active area (Ogawa et al. 1993). Active regions require more oxygen than inactive regions, and so the increased blood supply to active areas results in an overshoot of oxyhaemoglobin. Oxyhaemoglobin is diamagnetic, which means that it repels magnetic charge, whereas deoxyhaemoglobin is paramagnetic, meaning that it attracts magnetic charge. The signal that is detected by the scanner is not only influenced by the oxygenation properties of the local tissue, but by many other factors including heart beat, respiration, temperature etc. These cause large individual differences in magnetic properties of cerebral tissue, so these properties must be measured at rest for each individual before test stimuli are presented. When an area of the brain is active (i.e. responding to external stimulation), the elevated ratio of oxy- to deoxyhaemoglobin in the active region changes the small local distortions in the magnetic field, and these are picked up by the scanner (Matthews 2001). The resting values provide a baseline from which deviations in magnetic properties due to a higher oxy- to deoxyhaemoglobin ratio in active areas can be measured. The scanner contains three electromagnets whose currents interact to alter the strength and direction of the main magnetic field within the scanner (Glover 2001). The interaction
of the electromagnetic forces with the main field creates air pressure waves that generate a significant amount of scanner noise (typically 114 dB SPL in a 3 Tesla scanner). Although the interaction of the electromagnetic fields generates the loudest noise associated with the acquisition of the magnetic resonance (MR) signal, there are additional processes involved in MR scanning that contribute to the acoustic noise. For example, the magnet also requires additional coolant pumps and air ventilation, both of which increase the acoustic noise within the scanner (Ravicz et al. 2000).

The hostile acoustic environment of the MR scanner is particularly problematic for studies of auditory perception. Not only does the scanner noise evoke acoustic cortical activity, it also reduces the ability of the listener to attend to specific auditory stimuli in auditory scanner paradigms. There are various methods that can be used to reduce the impact of scanner noise. These include the use of earmuffs or earplugs (Ravicz and Melcher 2001), clustering scans and increasing time between scans and thus reducing the overall acoustic energy associated with image acquisition (Hall et al 1999) and using an active noise cancellation device (Hall et al. 2009). Although all of these methods are somewhat effective at reducing the scanner noise, none of them provide enough of a reduction to eliminate the effects of scanner noise on attention and activation levels. It is therefore of great importance to find ways in which the sensitivity to specific feature-related responses to test stimuli can be increased.

The fourth chapter of this thesis addresses a method of increasing sensitivity to specific sound features in neuroimaging studies. It is well known that the brain’s response to a continuous auditory stimulus decreases a short time after the stimulus
onset. This is a phenomenon known as adaptation (Robson et al. 1998) and is biologically useful to enable the detection of novel stimuli. However, it is counter-productive in the fMRI environment which relies on the detection of very small signals, so any further reduction in signal reduces the detectability of an already small signal. To get around this problem, auditory fMRI researchers have traditionally pulsed stimuli to evoke multiple onset responses that are integrated to represent an overall feature-specific response. Each time a sound is presented from silence, a significant amount of acoustic energy is produced. This is known as the ‘energy-onset response’, and is a nuisance variable that has, until recently, been difficult to overcome. Traditionally in neuroimagining studies of pitch processing, a pitch-evoking stimulus is pulsed on and off in one presentation block, and a noise (non-pitch) stimulus is pulsed on and off in another ‘control’ presentation block. The activation associated with the noise stimulus is subtracted from that associated with the pitch stimulus with the assumption that the residual response is a pure representation of pitch. Such an assumption works on the premise that the BOLD response is linear, and that the addition of stimulus features does not result in any interactions between the features. However, the brain is known to be susceptible to non-linearities (Sidtis et al. 1999; Friston et al. 2000; Devor et al. 2003), so it is unlikely that interaction effects would not be present between multiple stimulus features such as energy and pitch. A magnetoencephalography (MEG) study developed a method of presenting sound pitch stimuli in a way that minimizes the energy-onset response. This method was termed the ‘continuous stimulation’ paradigm (Krumbholz et al. 2003). Continuous stimulation essentially involves filling the inter-pulse intervals with noise such that there is no dip in acoustic energy between the pulses and hence no energy onset at the beginning of each pitch pulse. Each transition in the stimulus from noise to pitch
adds a single stimulus feature (pitch) rather than two in the traditional paradigm. The continuous stimulation paradigm has been used in a number of MEG studies, all of which report an increase in sensitivity to pitch-related activation compared to the traditional paradigm (Krumholz et al. 2003; Chait et al. 2006; Seither-Priesler et al. 2004, 2006). There are large differences between MEG and fMRI. MEG is a good method to study the timecourse of neural responses, but it does not have the spatial resolution of other techniques, such as fMRI. fMRI, on the other hand, has poor temporal resolution, but can locate neural responses in the order of mm (Matthews 2001). Chapter 4 of this thesis constitutes one of the first fMRI studies to employ the continuous stimulation paradigm, providing a spatial location for the generators of the compound pitch onset response. This study also reveals regions of auditory cortex that show an interaction between the responses to pitch and to sound energy. In order to increase the specific BOLD sensitivity to pitch, the continuous stimulation paradigm is used in three of the four experiments presented here (see Chapter 4, 5 and 7). Due to the specific hypothesis under test, the third experiment (Chapter 6) employed the traditional stimulation paradigm. The reason is that this chapter explicitly addresses controversies surrounding a particular type of pitch-evoking auditory stimulus known as iterated ripple noise (IRN), which is widely used in studies looking for a pitch center. In order to ensure compatibility of the results of the third paper with results from previous studies, it was necessary to use the same stimulation paradigm as was previously used (i.e. ‘pulsed’ stimulation).

Many previous pitch-perception studies have suggested the existence and location of a pitch center based on the use of stimuli that do not satisfy all four criteria set out by Hall and Plack (2009). The first results to suggest a pitch center were based on
animal findings that did appear to satisfy all four criteria (Bendor and Wang 2005), but due to the invasive nature of procedures used in animal models that cannot be used in human subjects, anatomical and methodological differences cannot unequivocably be directly compared. None of the human studies on pitch perception have satisfied all four criteria for a pitch center using a neuroimaging method with high spatial resolution. Many previous studies have studied pitch using MEG (e.g. Krumbholz et al. 2003; Seither-Priesler et al. 2004; Gutschalk et al. 2004; Ritter et al. 2005; Chait et al. 2006). These studies are informative for the timescale of the pitch response, such as onsets and offsets, but are not able to locate the generators of these responses with a high degree of precision. As mentioned previously, fMRI studies have high spatial resolution, but most previous fMRI studies have utilized a single type of pitch-evoking stimulus. As such it is not possible for these studies to satisfy the ‘pitch constancy’ criterion for a pitch center. Positron emission tomography (PET) studies are somewhat similar to fMRI, but are slightly more invasive as they involve ingestion of a neuromagnetic dye to track responses. They have similar spatial resolution to fMRI. PET and fMRI studies have provided spatial support for the pitch response found in MEG (e.g. Griffiths et al. 1998; Patterson et al. 2002; Penagos et al. 2004; Hall et al. 2006), but as previously mentioned, these studies each used a single type of pitch-evoking stimulus. Finally, many neuroimaging studies that have implicated an area of auditory cortex claimed to be involved in pitch processing have failed to control for non-pitch features in their pitch stimulus, which means that these studies do not satisfy the ‘pitch specificity’ criterion for a pitch center (e.g. Griffiths et al. 1998; Patterson et al. 2002; Krumbholz et al. 2003).
This thesis considers and attempts to overcome criticisms of previous neuroimaging studies of pitch perception by presenting a series of experiments that aim to improve the sensitivity to, and specificity of, the human pitch response, and ultimately addressing all four criteria for a pitch center set out by Hall and Plack (2009). Individual hypotheses are addressed in each paper, but the specific aims of this thesis are as follows:

1. To investigate the effects of stimulus presentation context on pitch-related responses
2. To determine to what extent non-pitch features within a pitch-evoking stimulus affect pitch-related responses
3. To determine whether or not there is a region of human auditory cortex that satisfies all four criteria of a pitch center and the location of any such region

The second chapter of this thesis is a book chapter that was submitted for the Springer Handbook of Auditory Research series, and constitutes an in-depth review of the methods, stimuli and terms used for the following papers. This is followed by a general methodology employed in all four studies, in Chapter 3. Chapters 4 through 7 inclusive are experiments that are thought to be of importance to the field of pitch perception and thus it was decided that all four should be submitted to peer-reviewed journals as soon as possible. It was for this reason that a request for submission by ‘alternative format’ was sought and obtained from the Graduate Office in the Faculty of Medical and Human Sciences at the University of Manchester. As each set of results was analyzed, a paper was written for submission to a peer-reviewed journal. This thesis includes a chapter that has been written and submitted for publication in a
book (Chapter 2), one published paper (Chapter 4, presented as a reprint from the journal), and three papers that are in preparation to be submitted for publication (Chapters 5, 6 and 7, presented in the format in which they will be submitted).

The 'alternative format' in which this thesis is presented necessitates some repetition of material (e.g. introduction, method and reference sections may overlap) but this is allowed under the university regulations. The two PhD supervisors: Prof. Christopher J Plack and Prof. Deborah A Hall are also co-authors on all four publications, and Prof. Deborah A Hall is a co-author on the book chapter. The experimental design, data collection, analysis and writing for all four manuscripts presented in this thesis were undertaken by PhD student Daphne García.
Chapter 2. Coding of basic acoustical and perceptual components of sound in human auditory cortex

Deborah Hall¹ and Daphne García²,³

¹Department of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK
²MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK
³Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK

This chapter has been submitted as a book chapter for inclusion in the Springer Handbook of Auditory Research: Human Auditory Cortex, edited by Richard R. Fay and Arthur N. Popper. It is presented in the format in which it was submitted. Although the candidate is the second author for the chapter, she contributed around 50% of the work for the chapter.

Publication 1
Section B  The Principal Computational Challenges
Chapter 7  Coding of basic acoustical and perceptual components of sound in human auditory cortex

Short title  Basic constituents of sound
Authors  Deborah Hall ¹ (corresponding author)
         Daphne Garcia ²

¹ Division of Psychology, School of Social Sciences, Nottingham Trent University,
Burton Street, Nottingham, UK. NG1 4BU
Email: deborah.hall@ntu.ac.uk
Office phone: +44 (0) 115 848 2377, Office fax: +44 (0) 115 848 6829

² MRC Institute of Hearing Research, University Park, Nottingham, UK. NG7 2RD
Email: daphne.garcia@ihr.mrc.ac.uk
Office phone: +44 (0) 115 922 3431, Office fax: +44 (0) 115 951 8503
1 Introduction

Neuroimaging studies are important for developing an understanding of the functional organization of the human auditory cortex. This chapter summarizes the contributions from human neuroimaging studies that have examined cortical responses to different types of sound stimuli. While being somewhat simpler than natural sounds, laboratory-generated sounds represent fundamental elements that are nonetheless interesting because they enable tight experimental control over other, potentially confounding, acoustical variables. Such synthesized sounds include single-frequency tones, broadband signals, sound level, sinusoidal spectrotemporal modulation and pitch. Examples are mostly presented from the field of functional magnetic resonance imaging (fMRI), but other neuroimaging modalities are also discussed. Central auditory neuroscience represents a bridge between many other disciplines, and hence recent progress in computer neuroscience, engineering, and physics has made a significant contribution to rapid developments in this field too. With this in mind, the chapter concludes with some examples of how novel approaches to experimental design and analysis are beginning to reveal how auditory stimulus attributes have spatially overlapping organizations.

1.1 A Scheme for Parcellating Human Auditory Cortex

Most neuroimaging work on the human brain has focused on the functional architecture of macroscopic brain areas. This focus has been largely influenced by available methodology. For example, most experimental designs use time-integrated averaging procedures and usually analyse the data by means of subtracting one stimulus condition from another. fMRI acquisition protocols often use a 3 mm$^3$ resolution and the data from neighboring volume elements are
averaged (via spatial smoothing) to reduce noise (Talavage and Johnsrude, Chapter 6).

To interpret the areas of feature-sensitive activation with reference to the underlying neuroanatomy, auditory neuroscientists have made widespread use of supplementary information obtained using anatomical mapping techniques and functional recording methods in animals and in humans. In the case of non-invasive recordings of human central auditory function using neuroimaging methods such as fMRI, there is no definitive approach for parcellating living human auditory cortex into its major microanatomical divisions. A traditional strategy in the neurosciences has been to link specific auditory processes to their gyral and sulcal locations in the human brain because it has been understood that these macroscopic anatomical landmarks had an important physiological relevance. However, the advent of more sophisticated methods for studying the microanatomy has shown this to be a rather simplistic view of structure-function relations. Nowadays, the neuroimaging field relies heavily on the results of electrophysiological and anatomical studies in animals and on post-mortem studies of human anatomy to interpret and to localize human functional data. Here in Chapter 7, both macroanatomical and microanatomical approaches are used. This Introduction therefore explains the schemes that are adopted for labelling different subdivisions of the human auditory cortex and introduces the terminology.
Figure 7.1. A: Surface of human left hemisphere with a cut through the Sylvian fissure to reveal the macroanatomical structure of the auditory cortex on the inner surface, including Heschl’s gyrus, planum polare and planum temporale. In this panel, the position of Heschl’s gyrus (the core region) is shown by the dotted grey region. A suggestion for how belt and parabelt regions might be organized is shown by the dark (belt) and light grey (parabelt) shading. B: Summary diagram of the microanatomical structure of the human supratemporal plane (left hemisphere) based on modifications of Figure 10 in Rivier and Clarke (1997) and Figure 6 in Wallace et al. (2002). C: A closer look at Heschl’s gyrus illustrates the microanatomical structure adopted in Chapter 7 (c.f. Morosan et al. 2001). D: Summary diagram of the microanatomical structure of the auditory cortex in macaque monkey (Kaas and Hackett 2000). In both Panels B and D, regions corresponding to the auditory core are dotted and regions possibly corresponding to the auditory belt are hatched. See text for an explanation of the abbreviations.
Most of the human auditory cortex lies within a deep fold that forms the boundary between the temporal and frontal lobes. The auditory cortex itself occupies a region called the supratemporal plane, the upper surface of the superior temporal gyrus. In the human brain, there is considerable anatomical variability in the position, size, and shape of cortical structures; this being particularly true for the supratemporal plane. Nevertheless, three key macroscopic features are consistently present (Heschl’s gyrus, planum temporale and planum polare, Fig. 7.1A). Heschl’s gyrus is a distinctive visible landmark. It cuts obliquely across the supratemporal plane in an antero-lateral to postero-medial direction. The gyrus has anterior and posterior borders that are clearly visible in an MR scan, while the insula forms its medial boundary. Planum temporale refers to the large undulating cortical surface extending behind Heschl’s gyrus, while planum polare describes the cortical surface in front of Heschl’s gyrus. Sound-related activity usually covers parts of these three regions, and this is especially true for hearing acoustically complex sounds and for tasks that involve active listening.

The cortex displays a high degree of tissue differentiation and connectivity which, when stained appropriately, can be viewed under the microscope and quantified. Animal studies demonstrate that microscopic anatomical landmarks, such as patterns of cell staining across the cortical layers, are more tightly coupled with functional specificity and neural processing than are macroscopic features (Morel et al. 1993; Kosaki et al. 1997). Staining profiles can be used as criteria for differentiating the auditory cortex into subdivisions that have anatomical and (hopefully) functional significance. Further details of these methods are discussed in Clarke, Chapter 2. The following is simply a summary to help place the subsequent localization of basic auditory feature coding in the context of a specific architectonic scheme.
In primates, anatomical and functional studies have provided a rich array of evidence for an auditory ‘core’ surrounded by ‘belt’ and ‘parabelt’ regions (e.g. Pandya and Sanides 1973; Morel et al. 1993; Rauschecker et al. 1995; Kaas and Hackett 2000; Rauschecker et al. 2002). The core forms the primary auditory cortex and receives thalamic inputs from the ventral medial geniculate body. Belt regions receive projections primarily from the core and more sparsely from dorsal and medial divisions of the medial geniculate body. Parabelt regions receive inputs from the belt and dorsal division of the medial geniculate body, with minimal connections with the core and ventral division of the medial geniculate body. These underlying concepts provide the basis for the organization of auditory cortex across numerous primate species, including humans (see Hackett 2003 for a review, see also Fig. 7.1A). A dominant model is that of a hierarchically organized auditory cortex in which the superior temporal gyrus contains specialized areas among which the neural processing of a sound proceeds from the analysis of its low-level physical constituents (in core regions) to higher perceptual dimensions (in belt and parabelt regions and even in prefrontal cortex, Romanski and Goldman-Rakic 2002). Although the precise number and location of all the core and belt fields has not yet been determined with any certainty there has been some broad consensus across studies. In humans, the core is typically centred on the medial two-thirds of Heschl’s gyrus and so Heschl’s gyrus provides a convenient macroanatomical landmark for defining primary auditory cortex (Fig. 7.1B and C). In both primates and humans, cell staining has enabled the core to be further subdivided into two fields. In primates, these fields commonly take the labels A1 (auditory 1) and R (rostral) (see Hackett 2003 for a review, see also Fig. 7.1D). The pattern of frequency tuning for each of these core fields shows a clear tonotopy, with the two gradients being mirror reversed at their shared low-
frequency border (e.g. Morel et al. 1993). The potential human homologue of A1 and R have been denoted Te 1.1 and Te 1.0 by Morosan et al. (2001) on the medial and central portions of Heschl’s gyrus, respectively (Fig. 7.1C). Adjacent to the core lies a field whose cellular characteristics are described as ‘transitional’ between core and belt regions (von Economo and Koskinas 1925; Morosan et al. 2001). In primates, this field is known as RT (rostro-temporal). Very little is known about RT, but it is narrowly tuned to tone frequencies and the direction of this tonotopic gradient appears to be reversed relative to R and shares a high-frequency border with it (Kaas and Hackett 1998). On the basis of its microanatomical profile, it has been suggested that a possible human homologue is area Te 1.2, sited on the lateral third of Heschl’s gyrus (Morosan et al. 2001, see Fig. 7.1C). The same region has been referred to as anterolateral area (ALA, see Fig. 7.1B) (Wallace et al. 2002).

A number of distinctive nonprimary fields have been identified in primates and in humans. One primate scheme subdivides belt regions into seven distinct fields and parabelt regions into two subdivisions (e.g. Kaas and Hackett 1998, 2000). One human scheme subdivides the planum polare and planum temporale into at least five fields (Rivier and Clarke 1997; Wallace et al. 2002, see Fig. 7.1B). Immediately behind Heschl’s sulcus are three fields laid out adjacent to one another along a medial-to-lateral axis. Rivier and Clarke (1997) refer to these as the posterior area (PA), lateral area (LA) and superior temporal area (STA). In the same study, two small fields were also identified in front of Heschl’s gyrus, one known as the anterior area (AA) and another known as the medial area (MA).

One of the goals that still motivates many human neuroimaging studies concerns the relationship between the localization of functional activity and the underlying
Microanatomy. Where it is possible to do so, the cortical representation of basic acoustic constituents are interpreted in terms of both macroanatomical and microanatomical definitions. The schemes of Morosan et al. (2001) and Rivier and Clarke (1997) are popular for speculating on the underlying microanatomical landscape of the observed feature-related auditory activity. Perhaps one of the main reasons for their favor is attributable to the authors’ efforts to present their schemes in formats that are compatible with human functional images, most notably in terms of their transformation into a brain space that has standardized 3-dimensional co-ordinates. Section 2.1 draws heavily on the delineation of Heschl’s gyrus into Te 1.0, 1.1 and 1.2 when describing the pattern of frequency-dependent responses that characterize the tonotopic organization of the human primary auditory cortex.

2 Single-Frequency Tones

Single-frequency tones (sinusoids) are the simplest type of acoustic signal since they form the building blocks from which all natural sounds can be expressed. Indeed, such form of frequency segregation is naturally performed by the cochlea for frequencies ranging from 20 Hz to 20 kHz. When a sinusoidal sound pressure wave is transmitted to the inner ear, it maximally vibrates a single place along the basilar membrane that is frequency specific (see Fig. 7.2A). Hair cells at the place of maximum vibration serve to transduce the mechanical energy into neural impulses. Hence, taken along its entire length, the basilar membrane can be thought of as behaving like a series of frequency channels transmitting frequency information to the auditory nerve (see Fig. 7.2B). In reality, the amount of excitation along the basilar membrane is not discrete but rather it decreases with successive shifts away from the best frequency. The resultant neural tuning curve reflects the degree of frequency selectivity (or width of each frequency channel).
Using psychophysical methods, the width of a frequency channel has been estimated to be about 12 per cent of the centre frequency, for frequencies between 750 Hz to 5 kHz (Moore 2004).

The gradient of frequency-specific coding along the cochlea is known as cochleotopy, although this orderly representation is maintained throughout the ascending auditory system and is found in all major auditory nuclei prior to the auditory cortex. Within central auditory structures, the same gradient of frequency-specific coding is known as tonotopy. Numerous electrophysiological studies have recorded tonotopic responses in the mammalian auditory system. The best frequency of a neuron corresponds to the frequency at which the neuron is most responsive at low sound levels.

In primates, frequency selectivity has been shown to be greatest in primary auditory cortex with neurons becoming increasingly more broadly tuned in nonprimary regions of the belt and parabelt cortex (Morel et al. 1993). A prediction therefore is that the most convincing demonstration of human tonotopy should occur for primary auditory cortex rather than for nonprimary regions. Moreover, while single-frequency tones might be sufficient to stimulate primary auditory cortex, more complex sounds such as narrow-band noise bursts are preferable for investigating the response properties of surrounding areas.

2.1 Frequency Coding in Primary Auditory Cortex.

At the advent of human neuroimaging, non-invasive measurements of electrical and magnetic field potentials were instrumental in documenting the tonotopic array in human auditory cortex (e.g. Romani et al. 1982; Pantev et al. 1988, 1989). The temporal acuity of these methods has been harnessed to accurately measure both
transient (e.g. at sound onset and offset) and sustained (e.g. throughout the stimulus epoch) frequency-sensitive responses. From this early work, there is evidence that the latency of particular transient responses reflects the underlying tonotopy. The source of the frequency-sensitive activity has been estimated using statistical methods to identify the location and orientation of the most likely dipole source. Dipole modelling of the transient evoked response has been applied to middle latency (10-50 ms) and longer latency (~100 ms) responses to single frequency tones (e.g. Pantev et al. 1988, 1989, 1995; Verkindt et al. 1995), again with a high level of intra- and inter-individual consistency. Within human auditory cortex, these results have suggested either a single tonotopic gradient (Pantev et al. 1988, 1989; Verkindt et al. 1995) or two mirror-image tonotopic gradients as depicted in Figure 7.2C (Pantev et al. 1995). In the case of a single frequency-sensitive gradient, the most commonly reported orientation is that of a high (medial) to low (lateral) axis, probably centered around Heschl’s gyrus. Dipole modelling of the sustained response also supports the same interpretation (Pantev et al. 1996). More recently the focus of investigation has moved towards that of fMRI because it makes fewer assumptions about the underlying activity, rendering it more suitable for examining the spatial organization of fine-grained feature-specific coding in human auditory cortex (see Talavage and Johnsrude, Chapter 6). It is important to note that in fMRI the responses to an individual tone frequency cannot be measured directly. Instead, the response to a stimulus condition is compared to the response to a different stimulus condition. For example, to highlight regions most responsive to low frequencies, a low-frequency tone condition would typically be contrasted with a high-frequency tone condition. In terms of tonotopic mapping, it is important to clarify that this type of statistical contrast would not identify regions of low-frequency specificity, but would instead highlight regions with a preference for low-frequency sounds instead of high-
frequency sounds. Nevertheless, this method is adequate for mapping out any loose tonotopic organization of the sort expected in human auditory cortex.

Figure 7.2 A: A highly schematic illustration of the basilar membrane in the cochlea as it might appear if it were unwound with the narrow, basal end being sensitive to high frequencies and the wide, apical end being sensitive to low frequencies. B: A popular model of the cochlea in which the frequency selectivity of the basilar membrane is represented as an array of overlapping frequency channels. C: A diagram showing the spatial organization of frequency coding in primary auditory cortex (fields Te1.0 and Te 1.1 on Heschl’s gyrus). Within each field there is a systematic progression of isofrequency bands. The dark shading indicates high frequencies and the light shading represents low frequencies.

Some of the earliest fMRI studies to investigate tonotopy in human auditory cortex did not necessarily capitalize on the best spatial resolution achievable (Wessinger et al. 1997; Bilecen et al. 1998) and contrasted responses to only one low-frequency tone (55 and 500 Hz, respectively) and one high-frequency tone (880 and 8000 Hz, respectively). More recent fMRI studies on tonotopy have addressed both of these issues. For example, Talavage and colleagues presented four pairs of narrow-band stimuli restricted to low (below 660 Hz) and high (above 2490 Hz) frequencies (Talavage et al. 2000). Theirs was the first known fMRI study to have provided evidence for not one, but two frequency-dependent regions across Heschl’s gyrus, and these shared a low-frequency border as in the primate
core region. The locations of these frequency-dependent regions appear to correspond to areas Te 1.0 and Te 1.1 on the middle two-thirds of Heschl’s gyrus (see Fig. 7.2C). All twelve hemispheres studied demonstrated these ‘mirror-image’ tonotopic regions, with high frequencies being represented at the postero-medial and antero-lateral endpoints and low frequencies at the common border in between.

Subsequently using a 3 Tesla scanner and four frequency-modulated tones each with different centre frequencies (250 Hz to 8 kHz), Schönwiesner and colleagues (2002) cast some doubt on the ability to convincingly demonstrate tonotopy using fMRI. Although the results obtained from this study showed very similar low- and high-frequency dependent activation foci to those found by Talavage and his co-workers, the authors were uncertain about attributing them to two tonotopic maps because no systematic frequency-response gradients were observed and also because the foci lay on or near possible boundaries of other auditory fields.

Since the initial research by Talavage et al. (2000), at least three further human fMRI studies have identified two, mirror-image tonotopic maps across Heschl’s gyrus (Formisano et al. 2003; Talavage et al. 2004; Upadhaya et al. 2007). The study by Formisano and colleagues used an ultra-high-field (7 Tesla) scanner to measure responses to six tone frequencies (300 Hz to 3 kHz). In the medial portion of Heschl’s gyrus, their results documented a high (postero-medial) to low (antero-lateral) frequency gradient that was reasonably consistent across the six listeners who participated in the study. The low-frequency response region shared a border with a second frequency gradient in the central portion of Heschl’s gyrus which further extended towards the antero-lateral tip of the gyrus. In terms of the correspondence between these tonotopic maps and predictions about the
underlying microanatomy, the medial gradient is consistent with the Te 1.1 and the central gradient is consistent with Te 1.0 (see Fig. 7.1C). Demonstrating tonotopy still remains a challenge and not all recent fMRI studies have confirmed two mirror-image tonotopic maps (e.g. Langers et al. 2007a). This study found firm support only for a single gradient in Heschl’s gyrus with a low-frequency response at the postero-medial end and a high-frequency response at the antero-lateral end.

As a complementary approach to fMRI, the mapping of neuronal fiber projections provides another technique for examining the functional role of different auditory cortical regions. Diffusion tensor imaging (DTI) is a non-invasive MR method for identifying white matter fiber tracks and so is a useful way to investigate cortico-cortical connectivity. Upadhaya et al. (2007) used both imaging methods in a 3 Tesla scanner to re-examine tonotopy across Heschl’s gyrus. The fMRI data confirmed the mirror-image fields on Heschl’s gyrus. The DTI data revealed significant (isofrequency) projections between the two foci of high-frequency sensitivity and between the focus of low-frequency sensitivity and (non-isofrequency) projections between the high-frequency foci and their shared low-frequency border. Again, these projections are consistent with two core tonotopic fields.

2.2 Frequency Coding in Nonprimary Auditory Cortex

In contrast with the general consensus of two mirror-image frequency-gradients across Heschl’s gyrus, the spatial arrangement of frequency sensitivity across nonprimary regions is less well defined. Talavage and colleagues (2000) postulated the existence of up to five nonprimary auditory fields, marked by four high-frequency and four low-frequency endpoints. Attributing these fields to cytoarchitectonic areas is somewhat dependent on the way in which the endpoints
are ‘joined’ up to form putative gradients and also on the parcellation scheme adopted. For example, in reference to the scheme shown in Figure 7.1B, one of these gradients could be located in area PA, another in AA and a third at the border of STA and LA (and so could be attributed to both or either field). Of course, without further evidence of a linear progression between the endpoints the interpretation of these data remains rather speculative and so the authors conducted a further study that used a technique of phase mapping to measure responses across a more complete range of frequencies (Talavage et al. 2004). Specifically, the stimulus in this experiment was a narrow bandwidth, amplitude-modulated noise with a center-frequency that was swept back and forth between 125 Hz and 8 kHz. The results confirmed tonotopicity in four of the five nonprimary areas defined previously. The fifth region showed a broader-tuned response that was not sufficiently frequency selective to yield consistent results.

More recently, an fMRI study by Langers et al. (2007a) failed to provide reliable evidence of any tonotopically arranged fields outside primary auditory cortex, finding only small-scale variations in the optimal stimulus frequency in planum temporale. These authors concluded that frequency as an organizing principle was no longer obvious because at this stage in the auditory hierarchy, the sound signals were perhaps recoded to represent auditory scene analysis and auditory objects (see also Griffiths, Micheyl, and Overath, Chapter 8).

3 Broadband Signals

Another acoustic dimension associated with single frequency tones is that of signal bandwidth. Single frequencies form one endpoint of this dimension, while broadband noise forms the other. Bandwidth is therefore one of the most basic variables with which to characterize central auditory function. Broadband signals
are generally more effective than single-frequency tones in evoking a neuronal response. This may be especially true in regions of nonprimary auditory cortex where single neurons respond more strongly to broadband stimuli than to single-frequency tones (Rauschecker et al. 1995). Several fMRI studies have demonstrated the large scale consequences of this in terms of a relative increase in BOLD (blood oxygen level dependent) activity across human auditory cortex for broadband signals (e.g. Wessinger et al. 2001; Hall et al. 2002). For example, Hall et al. (2002) compared activity for a single-frequency tone at 500 Hz and a harmonic-complex tone ($F_0 = 186$ Hz, harmonics 1-5) that spanned 2.6 octaves. They reported significantly more activity to the latter stimulus in Heschl’s gyrus and in the lateral part of the supratemporal plane (Fig. 7.3). Comparing the peaks of activity with the architectonic scheme suggested that the increased activity by spectral cues might involve the fields LA and STA, as well as Te 1.2. These effects were significant at the group level and also showed good consistency across participants (i.e. for 5 out of 6). The effect of bandwidth has also been quantified parametrically by varying the bandwidth of a continuous noise stimulus across a third, one, or two octaves each with a fixed centre frequency of 1 kHz (Hawley et al. 2005). In this study, only the brainstem and midbrain nuclei (cochlear nucleus, superior olivary complex and inferior colliculus) were examined but in all three structures, a significant monotonic increase in the amplitude of the BOLD signal was observed.
Figure 7.3. A linear cut across the right and left supratemporal plane showing the spatial distribution of the response to the single-frequency tone (upper panel) and harmonic-complex tone (lower panel). The orientation of the long axis of Heschl’s gyrus is plotted as a red line and the approximate central locations of the surrounding cytoarchitectonic fields are also shown. A version of this figure was presented at the 24th Association for Research in Otolaryngology MidWinter Meeting, 2001, Florida, USA, and the data were reported in Hall et al. (2002).

There are three possible functional interpretations for the observed growth in activity as a function of bandwidth. First, it is possible that the increase directly reflects the recruitment of neurons that perform spectral integration and thus have receptive fields that span large bandwidths. Conversely, it is also possible that the increase could be attributed to populations of neurons that each have a single best frequency and an excitatory response to sound, since this would lead to a spread of activity within tonotopic fields. These two explanations are rather difficult to
separate using fMRI alone. The third explanation draws attention to sound level because it is an important acoustical feature that may contribute to the observed differences. Moreover, effects of both sound level and bandwidth have been found in overlapping regions of auditory cortex (Hall et al. 2001). Where details are reported, fMRI studies that manipulate bandwidth have sought to control for sound level by equating overall sound energy (e.g., Wessinger et al. 1997; Hawley et al. 2005), or spectrum level (Hawley et al. 2005). It is likely that perceptual bases for matching, such as via a loudness model (e.g., Moore et al. 1997) would have a greater physiological validity at the cortical level, but this is unlikely to markedly change the current state of understanding about the effect of bandwidth on the pattern of auditory cortical activity.

4 Modulation

Natural sounds rarely contain acoustic features that are constant over time. Rather, they contain some kind of modulation over time either in frequency (FM) or in amplitude (AM). Typically, slow-rate modulations (< 50Hz) are important for perceiving speech and recognizing melodies, while fast-rate modulations convey other types of sensations such as pitch and roughness. Common modulations in speech include frequency changes. Formant transitions are a good example. These are complex sounds that contain multiple spectral peaks that sweep upwards or downwards in frequency over time, and also possess phonemic qualities. Further details about speech and music coding can be read in Chapter 9 (Giraud and Poeppel) and Chapter 10 (Zatorre and TBD), respectively. To simplify their experimental investigation, many investigators have chosen to present synthesized signals containing a single modulation component (e.g. sinusoidal amplitude modulation or a repeated train of noise bursts). It is those studies that are reviewed here.
In the auditory nerve, temporal modulation is represented faithfully in temporal discharge patterns (Joris and Yin 1992). However, as one ascends the auditory system, neurons have an increasingly limited capacity to represent time-varying signals and so the temporal attributes of the signal become more indirectly represented by the neural code. This successive degradation in temporal precision is partly due to the temporal integration of inputs that occurs from one processing stage to the next and partly due to the biophysical properties of neurons along the ascending pathway (e.g. Wang and Sachs 1995). A good example of the cortical response to modulated signals is an electrophysiological study in marmoset monkeys (Lu et al. 2001). Results showed that cortical neurons in primary auditory cortex encode temporal modulation in terms of the temporal firing pattern and the mean firing rate, depending on the rate of modulation. Specifically, at slow modulation rates of up to 16 Hz, approximately 20-55% of neurons coded the signal in an explicit manner, as a temporal discharge code. Whereas when the modulation rate exceeded 20 Hz, this proportion shifted to 20-40% of neurons coding the signal in an implicit manner, using a discharge rate code. For the first time, this study highlighted the importance of the rate code for temporal information in the awake animal and it extended the range of the neural code to more closely match the wide perceptual sensitivities to low and high modulation rates. The rate code is highly relevant for fMRI since this method is more sensitive to changes in overall sustained discharge rate than to changes in neural synchrony (Logothetis 2008).
4.1 Sustained and Transient Responses to Modulated Signals

fMRI studies have also shown that slow and fast modulation rates evoke different patterns of cortical activity particularly in terms of its sustained and transient components. One of the early experiments to investigate this issue measured the response within a number of auditory structures to amplitude-modulated noise.
presented at rates of 4 to 256 Hz (Giraud et al. 2000). In auditory cortex, the preferred stimulus had a modulation rate of 4-8 Hz. This evoked the largest response and activity was sustained at a high level across the entire 30-s stimulus duration. In midbrain structures, such as inferior colliculus, a different pattern was observed. Here, the greatest response was to the noise modulated at 256 Hz and activity was restricted to the period immediately following stimulus onset (i.e. it was transient). The auditory cortical response to modulation has been more fully explored by Harms and Melcher (2002) and Harms et al. (2005). In these fMRI studies, stimuli were trains of noise bursts presented at rates of 1 to 35 Hz. There was a non-monotonic relationship between rate and overall activity, with activity increasing from 1-2 Hz and then decreasing from 10-35 Hz. This can again be explained by the temporal envelope of the BOLD response over the 30-s stimulus duration. Activity was sustained for the slowest rates of modulation and then became more transient above 10 Hz (Fig. 7.4A). The authors suggested that the change to the shape of the BOLD response from sustained to transient with increasing modulation rate reflected the perceptual shift from individually resolved bursts (i.e. 1 and 2 Hz) to fused bursts (i.e. 10 and 35 Hz) forming a single ‘continuous’ perceptual event. Activity was characterized separately for Heschl’s gyrus and the superior temporal gyrus, but appeared to be very comparable. The later study in 2005 demonstrated that the transient response tended to be larger on the superior temporal gyrus than on Heschl’s gyrus (Harms et al. 2005, see Fig. 7.4B), but the exact reason for this is unclear. It is possible that the larger amplitude of the transient response reflects the greater role of that region in segregating the auditory scene into distinct meaningful events (Griffiths, Micheyl and Overath, Chapter 8).
4.2 Sensitivity to Slow-rate Modulation within Subdivisions of the Auditory Brain

A number of fMRI studies have sought to identify which regions of human auditory cortex are most sensitive to slow-rate modulations (Hall et al. 2002; Hart et al. 2003a, 2004). In all of these studies, the signal was modulated at a rate of 5 Hz and the stimulus for baseline comparison was a steady-state sound, matched in all other acoustic features. Hall et al. (2002) reported that the response to frequency-modulated tones occurred in Heschl’s gyrus and in lateral parts of the supratemporal plane (possibly corresponding to regions LA and STA) (Fig. 7.5). A
particularly large response was seen just behind the lateral part of Heschl’s gyrus in a region that might correspond to Te 1.2. The 2002 finding has since been replicated several times (e.g. Hart et al. 2003a; 2004). Of final note is an independent fMRI study that reported a disproportionately large response to upward and downward linear frequency sweeps in a large region posterior and lateral to Heschl’s gyrus (termed T3) (Brechmann et al. 2002). The previous modulation-related activity that was ascribed to Te 1.2 is broadly encompassed within area T3, although the borders of the different anatomical subdivisions differ.

It is interesting to note that Brechmann et al. (2002) showed the modulation-related activity in this cortical region to be level independent. This finding suggests that the neural code for modulation in this nonprimary auditory cortical region perhaps reflects an abstract representation of the perceptual attribute of the stimulus. However, it has also been noted that this region appears to respond to other acoustic cues such as bandwidth (Hall et al. 2002) indicating no clear systematic segregation of response preference.

4.3 A Common Representation of Modulation Rate?

While amplitude and frequency modulated sounds differ significantly in their spectral contents, they share the same modulation waveform that gives rise to their perceived time-varying properties. Until recently, it has been unclear whether cortical neurons might apply a common temporal processing mechanism to such a variety of time-varying signals. One way to answer this question is to systematically measure cortical responses to sinusoidally amplitude- and frequency-modulated signals since these are two examples that are easy to manipulate and are representative of natural sounds. For instance, amplitude and frequency modulations are important components of communication sounds of
animals and are found in a wide range of species-specific vocalizations including human speech. One relevant study reporting data recorded from single neurons in primary auditory cortex of awake marmosets was that by Liang et al. (2002). Electrophysiological recordings were made for both types of sinusoidally modulated stimuli presented at rates of 1-512 Hz, increasing in a base-2 logarithmic scale. Results showed a high degree of similarity between cortical responses to both classes of stimuli. It was possible to identify a particular modulation frequency for which a neuron was selective, either by assessing its temporal firing pattern or its mean firing rate. Critically, this selectivity was shown to be similar regardless of whether the temporal modulation was created in the amplitude or frequency domain.

A comparable study in human auditory cortex has been conducted using fMRI to measure sustained cortical responses to signals that were modulated at a rate of 5 Hz in the time domain and separately in the frequency domain (Hart et al. 2003a). In this study, two carrier signals were used to provide some internal validation of the effects; a single-frequency tone and a harmonic-complex tone, both with \( f_0 = 300 \) Hz. When compared with their matched steady-state carriers, both types of modulation evoked significantly greater activity in the lateral portion of Heschl’s gyrus (possibly Te 1.2) and in adjacent parts of the planum temporale (possibly LA and STA), replicating the previous findings. The most important finding was that the two activation patterns were largely overlapping supporting the view of a common neural code. In summary, these results indicate that cortical neurons extract the temporal profiles of modulated tones by the same mechanism, regardless of the spectral content of the sounds. Results from this human fMRI study suggest that this function is not restricted to the primary auditory cortex (namely Te 1.0 and 1.1).
5 Sound Level

Like frequency, level is one of the most basic attributes of sound and is coded at the first stage of cochlear transduction. At the auditory periphery, sound level is represented by the firing rates of neurons at the centre of the excitation pattern (e.g., Liberman 1978), by the spread of the excitation pattern (e.g. Chatterjee and Zwislocki 1998) and by temporal synchrony in the pattern of neural firing (e.g. Brosch and Schreiner 1999). The dynamic range of human hearing is extremely broad and yet is exquisitely sensitive to discriminating very small changes in pressure variations in the air across this range (Viemeister and Bacon 1988). At 1 kHz, the lowest detectable sound pressure level is about $10^{-12}$ watts/m$^2$. This corresponds to 0 dB SPL (decibels sound pressure level). Arguably, the highest sound level that can be tolerated without causing intense pain and cochlear damage is about $10^{13}$ watts/m$^2$ (120 dB SPL). Although the dynamic range of hearing exceeds 100 dB, individual auditory neurons are sensitive to a much narrower range of levels (generally 20-30 dB). Sensitivity to sound level is improved because different neurons adjust their input–output functions according to the prevailing distribution of levels (Dean et al. 2008).

Mapping sound level representations in auditory cortex is made difficult because there is no unitary code for sound level and there appears to be no spatially discrete region that is specialized for coding sound level alone. Neurophysiological studies in animals indicate that sound level may be represented by neurons which are distributed within populations that subserve other functions (e.g., Taniguchi and Nasu 1993; Heil et al. 1994), including the sharpness of frequency tuning to pure tones (Recanzone et al. 1999). Certainly, individual neural firing patterns have been shown to be influenced by both the level and the frequency of a sound.
At low sound levels, activated neurons show sharp frequency tuning close to the stimulating frequency, but at higher intensities of the same tone frequency there is a spread of excitation to neurons with characteristic frequencies both higher and lower than the stimulating frequency (Phillips et al. 1994). The spread of excitation is determined by the frequency of the stimulus. For low-frequency tones, animals studies in which cochlear action potentials have been recorded indicate activity across almost the whole auditory nerve at quite modest sound levels (Kim and Molnar 1979), while for high-frequency tones, the spread of activity across the auditory nerve fibers is more restricted (Palmer and Evans 1995). This result can be explained by considering the shape of the frequency response profiles. For low frequencies, the low- and high-frequency borders of the response area are relatively sharp whereas, for high frequencies, the low-frequency tails of the response areas are relatively shallow. Hence, for low-frequency tones, there is a rapid recruitment of fibers tuned to high frequencies when the sound level is sufficient to encroach on the low-frequency tail of their response areas.

Within auditory cortex, the response of the neural population to sound level becomes highly complex. Temporal coding has largely disappeared and rate coding is a mixture of both monotonic and non-monotonic neuronal responses to increasing sound level (e.g., Heil et al. 1994). Monotonic units are those showing a progressive increase in discharge rate as a function of sound level. In such units, a maximum firing rate is reached above which further increases in sound level have no effect. In contrast, non-monotonic units are those for which further increases in sound level result in a progressive decrease in activity from the maximum value. In other words, non-monotonic units are tuned to particular best SPLs (Pfingst and O’Connor 1981). Monotonic rate-level functions appear to be in the substantial majority throughout the central auditory system, at least for broadband noise
stimuli (Phillips et al. 1985). Thus, perhaps one might predict that the neuroimaging response to broadband noise should also show monotonic dependencies on sound level, since these techniques provide an indication of the summed activity of a neural population. For single-frequency tones, the predictions become less clear because there is a high proportion of non-monotonic rate-level functions in auditory cortex (Phillips et al. 1985, 1994; Heil et al. 1994). For single-frequency tones, neurons showing monotonic and non-monotonic behavior will contribute substantially to the level dependence of cortical activity. Human neuroimaging studies have therefore taken an exploratory approach to characterizing the predominant relationship between sound level and amount of sound-related activity using different stimuli and different measures of sound-related activity.

Figure 7.6. An example of the systematic changes in auditory cortical activity as a function of sound level, in response to a 300-Hz tone. To be classed as ‘activated’, voxels had to reach a significance threshold of p<0.001. The number of activated voxels was calculated separately for each sound level contrast (i.e. tone – silent condition) for each of 10 normal-hearing subjects. A version of this figure was presented at the 24th Association for Research in Otolaryngology MidWinter Meeting, 2001, Florida, USA. The group means are published in Hart et al. (2002).
5.1 Monotonic Level-Dependent Functions in Human Auditory Cortex

EEG/MEG (electroencephalography/magnetoencephalography) studies have reported an effect of increasing sound level on various parameters of the human auditory evoked response including an increase in the N100(m) amplitude, a reduction in the N100(m) latency and an increase in the N1-P2 peak-to-peak amplitude (Stufflebeam et al. 1998; Mulert et al. 2005). fMRI and PET (positron emission tomography) have also been used to measure sound-related activity and results have similarly indicated a growth in activity with increasing sound level across human auditory cortex (e.g., Jäncke et al. 1998; Lockwood et al. 1999; Hart et al. 2002; 2003b; Langers et al. 2007b). Not all studies have the sensitivity to determine the shape of the level-dependent function. Some have been somewhat limited by their narrow sampling of the full dynamic range and their choice of large step sizes (e.g. Jäncke et al. 1998; Lasota et al. 2003; Mulert et al. 2005). In those studies that have used a more optimal parametric design, the extent of activation and response magnitude both tend to increase monotonically (e.g. Hall et al. 2001; Hart et al. 2002; 2003b; Sigalovsky and Melcher 2006; Langers et al. 2007b). One exception is the PET study reported by Lockwood et al. (1999) in which rCBF (regional cerebral blood flow) for a 500-Hz tone showed a somewhat U-shaped function. As a more representative example, Figure 7.6 illustrates data reported by Hart et al. (2002) for a 300-Hz tone. Analysis confirmed that the number of activated voxels in auditory cortex was significantly determined by sound level across the 42-96 dB SPL range \[F(9,81)= 17.51, \ p<0.001\]. Such a pattern was observed in both hemispheres, but was strongest in the hemisphere contralateral to the monaural stimulus. Moreover, on this contralateral side, the growth was particularly sharp at the highest sound levels \[\text{significant quadratic component: } F(1,9)= 8.52, \ p<0.05\]. Typically, the level-dependent function continues its upward trajectory even at intense sound levels. The response seems to show no evidence
of non-monotonicity nor of reaching a plateau. Similar results have been reported for a range of different sound stimuli including a 300-Hz tone presented up to 96 dB SPL (Hart et al. 2002), two frequency-modulated tones spanning the spectral range 0.5–1.0 kHz and 4-8 kHz presented up to 80 dB sensation level (Langers et al. 2007b); a 4.75-kHz tone presented up to 96 dB SPL (Hart et al. 2003b), a 4-kHz tone presented up to 90 dB SPL (Lockwood et al. 1999), and a continuous broadband noise presented up to 99 dB SPL (Sigalovsky and Melcher 2006). The rate of growth as a function of sound level does not appear to be the same across all frequencies. In a study that directly compared the effect of two tone frequencies, Hart et al. (2003) demonstrated that, within Heschl’s gyrus, the response to a low-frequency tone was flat between 42 and 66 dB SPL and then showed a rapid growth that continued up to the highest level studied (96 dB SPL). In contrast, the response to a high-frequency tone increased steadily across the same range of levels. These results concur with physiological evidence suggesting that recruitment of primary auditory cortical neurons may be different at high and low frequencies (Phillips et al. 1994).

Systematic increases in both extent and magnitude of the response do not always co-occur in the same dataset. For example, for syllables and pure tones presented at levels of 75, 85 and 95 dB SPL, Jäncke et al. (1998) found a significant increase in the extent of auditory cortical activity, but no significant effect on response magnitude. Likewise, for monosyllabic words presented at levels from 65 to 110 dB (measured on a C-weighted scale), Mohr et al. (1999) found a reliable increase in response magnitude, but not extent. Comparable outcomes for extent and magnitude might be expected because, at a simplistic level of interpretation, growth with sound level is physiologically consistent with a regional increase in the general activity of the underlying neuronal population. A dissociation between the
shape of the level-dependent function for extent and magnitude might simply reflect lack of sensitivity in the (BOLD or rCBF) neuroimaging measure. Indeed, it has been suggested that extent is perhaps a less reliable measure of activation than magnitude (Hall et al. 2001; Mohr et al. 1999), especially in experiments with many stimulus conditions. An alternative explanation, especially in those studies utilizing fine spatial resolution, is that a dissociation between the extent and magnitude measures might represent either neural recruitment or a local increase in neural activity, respectively. The preceding discussion has hopefully emphasized the point that comparisons between animal and human data on level sensitivity are unlikely to be straightforward. Although it is reasonable to anticipate neural recruitment for high sound levels (see Hart et al. 2002), increases in BOLD/rCBF responses are not necessarily indicative of increases in neural firing rate, especially given the contribution of non-monotonic units to sound level coding. At the cortical level, there are profuse local inhibitory influences (Manunta and Edeline 1998; Logothetis 2008), although a direct local contribution to the observed non-monotonicity of rate-level functions has yet to be demonstrated. Nevertheless, if non-monotonic responses are mediated by summation of excitatory and inhibitory inputs to cortical neurons, an increase in subthreshold activity at high sound levels would occur despite the reduction in the output from such units. The greater metabolic demand caused by such a rise in synaptic activity would most likely be responsible for an increase in the BOLD/rCBF response (Logothetis 2008).

5.2 Sensitivity to Sound Level within Subdivisions of the Auditory Brain

At every major stage of the ascending auditory pathway, significant rate-level functions have been demonstrated in humans. To our knowledge, only one fMRI study has so far quantified level-dependence of activation within subcortical
auditory structures (Sigalovsky and Melcher 2006). Using a broadband continuous noise stimulus presented binaurally at 30, 50 and 70 dB sensation levels (equivalent to 50–99 dB SPL), the main trend was again one of a monotonic increase in activity. This pattern was observed in the cochlear nucleus, superior olivary complex, inferior colliculus and medial geniculate body (and auditory cortex).

A small number of neuroimaging studies have distinguished level-dependent functions in different anatomically and functionally distinct subdivisions of human auditory cortex. One of the first fMRI studies to investigate this issue was conducted by Hart et al. (2002). These authors quantified the response to sound level within three anatomically defined regions of human auditory cortex; (i) Heschl’s gyrus (probably incorporating the primary fields Te 1.0 and Te 1.1), (ii) the small region immediately lateral to Heschl’s gyrus (representing Te 1.2) and (iii) planum temporale (possibly including LA, STA and PA). Within these three regions, Hart and colleagues plotted the proportion of suprathreshold (p<0.001) voxels and the mean scaled per cent signal change as a function of sound level. In this study, the range of sound levels spanned 42-96 dB SPL in 6-dB steps and the stimulus was a 300-Hz tone. Of the three anatomically defined regions, the response centred on Heschl’s gyrus was the most sensitive to increasing sound level for both magnitude and extent measures of activity. Consistent with this finding was a subsequent fMRI study demonstrating a monotonic increase in the percentage of voxels within Heschl’s gyrus that reached the chosen threshold of p<0.0001 (Lasota et al. 2003). This study used a 1-kHz tone presented at a range of sound levels (0-50 dB hearing level). Langers et al. (2007b) also commented that Heschl’s gyrus was the dominant source for their sound-level dependencies.
Although not specifically commenting on putative differences between cortical regions in their sensitivity to level, Sigalovsky and Melcher (2006) examined four regions of interest that defined broad subdivisions of auditory cortex. i) The postero-medial two-thirds of Heschl’s gyrus was intended to approximate Te1.0 and Te 1.1, ii) the remaining antero-lateral third of Heschl’s gyrus was probably equivalent to Te 1.2 (as shown in Fig. 7.1), iii) the entire planum temporale was assumed to incorporate lateral belt regions (LA, PA and STA), and iv) an antero-medial region, located in front of Heschl’s gyrus up to the circular sulcus, was possibly the human homologue of medial belt regions (MA and AA). The authors applied a number of independent measures of sound-related activity. The primary ‘magnitude’ analyses first identified voxels reaching significance at p<0.01 and then across subjects and hemispheres calculated the average maximum percent change at the onset of the noise stimulus (relative to a silent baseline) and the average maximum percent change at the offset of the noise across each sound level condition. A supplementary ‘extent’ analysis counted numbers of voxels within the region of interest that exceeded a probability of activation of p=0.01. Comparing the 30 and 70 dB conditions, there was an increase (p<0.05) in both the onset and offset percent change in all of the subdivisions except the anterior medial non-primary auditory cortex where the same trend did not reach significance. However, this region was generally less responsive to sound stimulation than the other cortical regions. Again, the most significant level-dependent change occurred in primary auditory cortex; albeit for the magnitude of the offset response, not the onset response.

5.3 Searching for a Topographic Representation of Sound Level

In the mammalian primary auditory cortex, an orderly spatial organization of a number of parameters related to the encoding of sound level has been
demonstrated. Organizing principles include minimum threshold, dynamic range, best SPL and non-monotonicity of intensity functions (e.g., Heil et al. 1994). The analysis of several neuroimaging datasets has explored the evidence for a systematic relationship between sound level and the location of auditory activity (ampliotopy). On balance the results are somewhat negative (see Hart et al. 2002; Sigalovsky and Melcher 2006). For one study that did report a positive effect (Lockwood et al. 1999), on closer inspection the data do not appear very convincing. To support their conclusion, the authors drew attention to the 8 mm shift (inferior to superior) in the peak location of ipsilateral response as sound level increased. Given that the width of the smoothing kernel applied to the image data during spatial pre-processing was 10 mm, the spatial sensitivity to shifts smaller than this value is rather limited. In summary, human neuroimaging studies have so far failed to demonstrate ampliotopy. This does not necessarily rule out the possibility that ampliotopy does exist. It may simply remain obscured by current measurement techniques.

5.4 A Physical or Perceptual Representation of Sound Level?
A range of scales are available for measuring sound level. A common objective measure of sound level (‘intensity’) is the decibel (dB) scale which relates to the power of the sound energy. Decibels represent the ratio of a given intensity ($10^x$ watts/m$^2$) to the standard threshold of hearing, so that the threshold of hearing corresponds to 0 dB. However, listeners do not describe sounds in terms of dB, but instead use language such as ‘soft’ or ‘loud’. Intensity and loudness are measures of different sound level characteristics. Two different 60-dB sounds will rarely have the same loudness because the judgement of loudness takes into consideration the ear’s sensitivity to the component frequencies of the sound. A common ‘loudness’ scale is that measured in phons. The basis for the phon scale
references each sound to the equivalent dB level for a 1-kHz tone. So, if a given sound is judged to be as loud as a 1-kHz tone at 60 dB, then it is said to have a loudness of 60 phons. For broadband signals, the loudness is determined by the auditory excitation pattern, integrated across frequency (Moore et al. 1997).

Hall et al. (2001) considered the issue of control over sound level in the context of comparing auditory cortical activity for single-frequency tones and broadband signals. If intensity is fixed while signal bandwidth is increased, then loudness nevertheless increases because the signal spans a greater number of frequency channels. The question therefore arises, “should one match stimuli for intensity or loudness?” To address this, Hall and colleagues presented a range of single-frequency tones and harmonic-complex tones that were matched either in dB or phons. When the fMRI data were collapsed across stimulus class, neither activation extent nor magnitude significantly correlated with the dB scale ($r=0.04$, $p=0.59$ and $r=0.06$, $p=0.48$, respectively). In contrast, both extent and magnitude correlated significantly with the phons scale ($r=0.36$, $p<0.001$ and $r=0.35$, $p<0.001$). On the basis of these results, the authors speculated that loudness may be an important aspect of the auditory cortical representation of sound.

More recently, Langers et al. (2007b) considered auditory cortical responses as a function of intensity and loudness using low- and high-frequency stimuli presented across a 70 dB range, in steps of 10 dB. To address whether intensity or loudness was the main characteristic driving the pattern of level-dependent activation, the authors compared two groups of listeners; one with normal hearing and one with age-related sensorineural hearing loss. This type of impairment reduces high-frequency hearing sensitivity and is accompanied by loudness recruitment at high frequencies (a disproportionate rise in loudness ratings as a function of intensity).
If loudness were the driving factor, then a dissociation would be predicted between dB and equivalent loudness curves across the two groups of participants at high frequencies. Typically, the fMRI results revealed monotonic increases in the magnitude of activation across intensity and loudness. At low frequencies, the steepness of the intensity- and loudness-dependent functions did not differ across the hearing impaired and normal hearing groups. This was also true at high frequencies for the loudness-dependent function. However, at high frequencies the intensity-dependent function was significantly steeper in the hearing impaired group than in the group with normal hearing (mean slope was 37 and 21 $10^{-3}\% / dB$, respectively). These results therefore support the conclusion that loudness relates more strongly to cortical activation than does intensity. This interpretation is also consistent with the general view that cortical activation reflects the correlate of the subjective strength of the stimulus percept.

5.5 The Role of the Auditory Cortex in Level Discrimination

A region in the posterior temporal lobe of the right hemisphere has been identified during an intensity discrimination task performed in the PET scanner (Belin et al. 1998). This region is perhaps located more posterior to the non-primary auditory fields that have been discussed so far with respect to level coding per se. It is more likely that this higher auditory brain centre plays a role in computing sound-intensity differences since the magnitude of activation was not influenced by task performance ($d' = 4.5, 3.5, 2.5, \text{ and } 1.5$). Although decreasing discriminability did not increase activation in the posterior temporal region, it did so in a number of right-sided frontoparietal regions; namely inferior frontal gyrus, precentral sulcus and inferior parietal lobe. It is possible that these regions therefore may play a more general role in allocating attentional resources to perform the discrimination task.
6 Pitch

Pitch is one of the most fundamental auditory percepts. It can be defined in musical terms by any sound that can be used to produce a melody, and can be ordered on a scale from low to high. Pitch plays an important role in music perception and in language (conveying prosody and, in some languages, semantic information). Pitch is a perceptual attribute of sound, but it is determined by physical characteristics of the acoustic signal including its frequency (e.g. in the case of single-frequency tones) or its temporal periodicity (e.g. in the case of complex sounds). These two physical cues form the basis of two mechanisms for the neural coding of pitch: a rate-place code and a time code. Harmonic-complex tones are an interesting example because depending on whether their frequency components are ‘resolved’ or ‘unresolved’, the pitch can be conveyed by either, or both, neural codes. Defining each harmonic as ‘resolved’ or ‘unresolved’ depends on its neural activation pattern within the peripheral auditory system. The low-numbered (resolved) harmonic components tend to fall within individual frequency channels producing a characteristic excitation pattern across the membrane in which there is a one-to-one mapping between the spectral peaks in the acoustic signal and the peaks of excitation. The sensation of pitch could therefore arise from a detection of the harmonically related, resolved peaks of neural activity. This is the rate-place code. Although it is still debated at what point the harmonics cease to be resolved along the basilar membrane, it is generally accepted that harmonics below the seventh are resolved and those above the thirteenth are unresolved (Houtsma and Smurzynski 1990). The unresolved harmonics are not individually represented on the membrane, but instead multiple harmonics fall within a single frequency channel and the resulting excitation pattern contains no distinct spectral peaks. The pitch of these stimuli can be determined instead from the output of a single channel containing many interacting harmonics, whose
repetition rate corresponds to the f0 (i.e. the pitch) of the complex tone (Houtsma and Smurzynski 1990; Carlyon et al. 1992; Micheyl and Oxenham 2004). This is the time code.

Although pitch processing mechanisms most probably exploit both spectral and temporal information (Carlyon et al. 1992; Shamma and Klein 2000), many neuroimaging investigations have sought to eliminate the spectral cues for pitch in order to isolate the neural representation of the time code. Stimuli for which the dominant cue for pitch is temporal rather than spectral include unresolved harmonic-complex tones, amplitude-modulated tones, regular interval sounds and dichotic pitches (Fig. 7.7). For these stimuli, pitch cues are not carried in the spectral (i.e. tonotopic) pattern of neural activity and pitch coding may therefore engage additional regions of the auditory cortex that are not so sharply tuned to frequency. One popular type of regular interval sound is iterated ripple noise (IRN). IRN is created by generating a sample of random noise, delaying it, and adding or subtracting the duplicate to or from the original (Yost 1996). The pitch of an IRN is equivalent to the reciprocal of the delay imposed. The pitch strength (salience) can be increased by increasing the number of delay-and-add iterations (Yost et al. 1996). Both pitch value and strength can be manipulated in a systematic manner, with little effect on the spectral content of the stimulus, as long as a suitable high-pass filter is used so that only unresolved harmonics are present (and thereby eliminating distortion products produced by low-numbered harmonics).
6.1 Pitch Sensitivity within Subdivisions of the Auditory Brain

One way to identify pitch-sensitive activity is to compare the response to IRN with that to a random noise signal that has the same spectral content. When Patterson and colleagues (2002) contrasted a sequence of IRN bursts with a fixed pitch and a sequence of random noise bursts, they found activation in lateral Heschl’s gyrus (although there was also more medial activation in central Heschl’s gyrus, see Griffiths et al. 2010). This result was consistent in eight of the nine listeners. The putative anatomical field corresponding to this region is Te 1.2 (see Fig. 7.1C). A number of other PET and fMRI studies provide convergent evidence that lateral Heschl’s gyrus is maximally responsive to IRN (e.g., Griffiths et al. 1998; Hall et al. 2005; Hall and Plack 2009). Moreover, two of these studies have demonstrated a systematic increase in the response within lateral Heschl’s gyrus as a function of increasing pitch strength (Griffiths et al. 1998; Hall et al. 2005), as shown in Figure 7.8. This relationship was examined using IRN signals in which the number of delay-and-add iterations ranged from 0 to 16.
Figure 7.8. An incidence map showing auditory cortical increases in activity as a function of pitch salience (an increase in activity for IRN with 0, 1 and 16 add-and-delay iterations). The color code illustrates the variability of the effect across 16 listeners. All maps are overlaid onto the same 5 horizontal brain images (z = +16 to -16 mm) in neurological convention (i.e. left = left). The original version of this figure was published in Hall et al. 2005 J. Neurophysiol. 94:3181-3191.

If this region is to be called a ‘pitch center’ then it should represent subjective pitch regardless of the spectral, temporal, or binaural characteristics of the stimulus. One fMRI study filtered harmonic-complex tones into low and high spectral regions to produce resolved complex tones evoking a strong sense of pitch and an unresolved complex tone evoking a weak sense of pitch (Penagos et al. 2004). Contrasting these two stimulus conditions again revealed patches of activity around lateral Heschl’s gyrus. The amplitude of the BOLD response was significantly smaller for the weak pitch condition than the strong pitch condition.

Figure 7.9. Incidence maps showing the consistency of pitch-related activation for five pitch stimuli presented to six listeners. Activity was calculated separately for
each pitch contrast (i.e. pitch – noise condition) using a significance threshold of \( p < 0.01 \). For each listener, the activity maps were combined and the resulting color coding indicates how many of the pitch stimuli evoked activity at a particular voxel (blue = 1, cyan = 2, green = 3, yellow = 4, red = 5). All maps are overlaid onto the individual anatomical brain image in neurological convention (i.e. left = left). A version of this figure was presented at the 12th Annual Meeting of the Organization for Human Brain Mapping, 2006, Florence, Italy. Group mean data are published in Hall and Plack (2009).

In a recent fMRI study, Hall and Plack (2009) measured cortical responses to seven different pitch-evoking stimuli, each with different spectral and temporal characteristics (pure tone, resolved and unresolved harmonic complex tones, a wideband harmonic-complex tone, a binaural pitch stimulus (Huggins pitch) and two types of IRN). The results for the IRN stimulus showed good agreement with previous studies. However, a different pattern of activation was reported for the other five pitch-evoking stimuli. Instead of lateral Heschl’s gyrus, planum temporale was most consistently activated across listeners. However, even in this region there was a high degree of individual variability (illustrated in Fig. 7.9). From this subset of six listeners, three showed planum temporale activity for many of the pitch stimuli presented but for three other listeners activity was located elsewhere. This finding would indicate that it is rather premature to assign special status to lateral Heschl’s gyrus solely on the basis of activation patterns. A recent fMRI study used a novel form of group analysis to explore the cortical representations of pitch and sound objects (Staeren et al. 2009). Stimuli were chosen from four different sound categories (complex tones, singers, cats and guitars) and each contained examples at three different pitch values (250, 500 and 1000 Hz). Responses that discriminated between the pitch values were distributed across patches of postero-lateral Heschl’s gyrus and planum temporale, in accordance with previous measures of pitch-related activity. At the time of writing, the search for a generalized human pitch centre is ongoing.
6.2 Pitch Onset

Neuroimaging investigations of pitch processing have typically presented sequences of bursts of pitch-evoking stimuli separated by intervals of silence. Neural responses to the control condition (e.g. a sequence of random noise bursts) are subtracted from the pitch condition, with the residual activation identified as the ‘pitch-specific’ response. It is well known that many auditory cortical neurons are highly responsive at stimulus onset (e.g. Lu et al. 2001; Liang et al. 2002) and so one might therefore expect a large transient energy response at each sound onset for these stimulus sequences. It is possible that neuroimaging measures have confounded pitch onset and energy onset responses. However, careful design of the stimulation paradigm is able to separate out the transient response to the pitch onset from that to energy onset (e.g. Krumbholz et al. 2003; Chait et al. 2006). In the continuous stimulation paradigm, bursts of pitch-evoking stimuli are introduced into an ongoing noise signal thus removing the changes in energy at the transition from baseline to pitch. Furthermore, the temporal resolution of EEG and MEG is ideally suited to isolating the transient onset responses. Using this paradigm in the context of an MEG study, Krumbholz et al. (2003) found a positive deflection with a latency of about 150 ms at the transition from random noise to IRN. Such a deflection was not seen for the transition from IRN to random noise and so it was termed the ‘pitch onset response’. In addition, the amplitude of the pitch onset response increased with increasing pitch strength and the latency of the pitch onset response decreased as f0 increased. Crucially, the pitch onset response appears to be consistent across different types of pitch-evoking stimuli because a similar pattern of results has been obtained for both a tone-in-noise and a binaural (Huggins) pitch (Chait et al. 2006).
The neural generators of the pitch onset response have been estimated using dipole source modeling (Krumbholz et al. 2003; Seither-Preisler et al. 2004; Gutschalk et al. 2004; Ritter et al. 2005; Chait et al. 2006). According to these results, the source is typically located close to Heschl’s gyrus but is unlikely to be sited within primary auditory cortex. However, the spatial resolution of these methods does not allow for precise localization (Chait et al. 2006). Depth-electrode recordings in patients who are candidates for epilepsy surgery do allow for more accurate localization of the stimulus-evoked electrical signals. A recent study presented IRN in the context of the continuous stimulation paradigm to a single patient undergoing surgery (Schönwiesner and Zatorre 2008). A depth electrode was directed within the lower bank of the Sylvian fissure about 5 mm behind Heschl’s gyrus running parallel to it, so that five of the nine electrode contacts recorded electrical activity from this gyrus. Contacts 2 and 3 (close to the medial two-thirds of Heschl’s gyrus) responded strongly to the energy onset response, while contact 5 (on the supratemporal plane close to lateral Heschl’s gyrus) responded best to the pitch onset. Although the spatial accuracy is much improved, the signal-to-noise ratio of the data was rather poor owing to the low number of repetitions afforded by the method. The findings from this study would seem to concur with those of surface magnetoelectrical activity (Krumbholz et al. 2003; Seither-Preisler et al. 2004; Gutschalk et al. 2004; Ritter et al. 2005; Chait et al. 2006). In general conclusion, a continuous stimulation paradigm would appear to improve specificity of pitch-related activity by eliminating activation related to energy onset.

6.3 Listening to Melodies

When different pitches are presented in a temporal sequence, they form a melody. Melody plays a critical role in music perception and in the recognition of familiar
tunes. In terms of the stages of sound processing, melody perception can be construed as one of the highest levels. Functional neuroimaging methods have revealed areas in nonprimary auditory cortex (in belt and parabelt regions) to be responsible for melody processing (Patterson et al. 2002; Brown and Martinez 2007; Zatorre et al. 1994). In their fMRI study of melody processing, Patterson et al. (2002) presented two different types of melody, one in which 32 sequential IRN bursts produced a novel diatonic melody and one in which the IRN bursts produced a random note melody. Contrasting these two conditions with one in which there was a sequence of IRN bursts with a fixed pitch revealed activity within planum polare and superior temporal gyrus. Moreover this activity was greater in the right hemisphere. The asymmetry emerged only for the effect of melody and was not present for the simple effect of pitch (defined by contrasting the fixed pitch sequence with a random noise condition). This finding is consistent with the hemispheric specialization hypothesis which claims that the right hemisphere plays a dominant role in coding small and precise changes in frequency (pitch) over relatively long temporal durations (see Zatorre et al. 2002 for a review).

The concept of a spatially segregated hierarchy of pitch coding has been proposed to explain the results presented (Patterson et al. 2002; Zatorre et al. 2002). At the first stage (possibly subcortical) temporal regularity is extracted from separate frequency channels of the incoming signal, while at the second stage (possibly lateral Heschl’s gyrus) this temporal pattern information is integrated across frequency channels to code pitch. Higher-level processes such as pitch tracking and melody extraction occur at the third stage especially in distributed regions of the right superior temporal gyrus and prefrontal cortex (Zatorre et al. 1994).
7 Summary

One broad framework for central auditory processing that has been around for some time proposes that the coding of information relating to the sound object and information relating to its spatial location remain independent up to and beyond the auditory cortex. The dual route model of modularity was originally proposed for the visual system (Ungerleider and Mishkin 1982). Corresponding evidence for the auditory system originated from research in primate anatomy and function. Anatomically, two major cortico-cortical projections were identified, each from lateral belt and parabelt regions to discrete regions of the prefrontal cortex (Romanski et al. 1999; Romanski and Goldman-Rakic 2002). These two routes are illustrated in Figure 7.1A. Functionally, neurons in the anterior lateral belt are primarily responsive to the spectrotemporal features of a sound that code object identity and are consistent with a ‘what’ pathway for object recognition; while neurons in the posterior lateral belt are more sensitive to the spatial properties of a sound, consistent with a ‘where’ stream for object localization (Rauschecker et al. 1995, 2002; Rauschecker and Tian 2000).

Like the model for pitch and melody processing described above, this model views the coding of higher-level sound properties as a process that is spatially segregated and hierarchical. In other words, sound recognition proceeds through several anatomically discrete and functionally specialized cortical areas culminating in higher centers where perceptual discriminations and other behaviourally relevant judgements are performed. The neuroimaging results presented in Chapter 7 show that a wide range of sounds from pure tones, through harmonic complex tones, modulated signals and pitches stimulate primary and nonprimary regions of human auditory cortex. These data do not provide any clear sense in which key functional roles can be ascribed to the different anatomical
regions illustrated in Figure 1 and are thus rather difficult to reconcile with the modular framework. A potential conclusion might be simply that the auditory cortex is highly sensitive to dynamic complex sounds without any distinguishable topographic organization. An alternative conclusion is that sound representations are topographically organized, but are spatially distributed across the surface of the auditory cortex.

In the visual system, a body of evidence is beginning to demonstrate how macroanatomical regions previously ascribed with a single function might actually perform several different functions and how cortical representations that were previously absent in the data might in reality be present (Grill-Spector et al. 2006; Logothetis 2008). Clever experimental methodology is the first key to revealing organizations that might previously have been obscured. High-resolution imaging and fMRI adaptation designs are two examples that have been applied in the auditory domain. For example, Formisano et al. (2003) used a combination of ultra-high field (7 Tesla) and surface coil fMRI to achieve a fine-grained spatial resolution (1.20 x 1.48 x 2.00 mm). High-resolution fMRI detected activity on a much finer spatial scale than had been reported hitherto, enabling mirror-symmetric frequency gradients on Heschl’s gyrus to be measured systematically in each individual listener. fMRI adaptation designs are particularly recommended for investigating the functional properties of a brain region that has spatially overlapping or close neural populations that encode different stimulus categories (Grill-Spector et al. 2006). It is sensitive to differential fMRI responses within a region. The method takes advantage of the observation that the BOLD response decreases with repeated presentation of the same stimuli. In the auditory domain, fMRI adaptation studies have so far concerned the representation of perceptual categories (such as phonemes, Ahveninen et al. 2006 and animal vocalizations,
Altmann et al. (2007) instead of basic sound features. For example, Altmann et al. (2007) reported that the response amplitude across the left superior temporal gyrus was significantly weaker for trials in which the same animal vocalization was repeated compared to trials in which the two animal vocalizations were different, thus indicating a selective representation of this sound category in left nonprimary auditory cortex.

The second key to discovering new principles of organization is to use clever analysis in order to maximize the potential afforded by clever design. Phase-encoded stimulus mapping and multivoxel pattern analysis are two examples that have been applied in the auditory domain. Unlike conventional pairwise contrast analysis, phase-encoded mapping compares the responses to a set of stimuli and estimates the most effective stimulus. For example, Talavage and colleagues (2004) were able to identify multiple tonotopic gradients systematically in individual listeners by mapping areas of auditory cortex that showed a progressive linear change in the frequency of maximal sensitivity. Another approach is to take into account the full spatial pattern of brain activity by applying a classification algorithm to decode what patterns are present across the cortical surface. Compared with univariate analysis, the particular strength of multivoxel pattern analysis is in revealing the representation of different perceptual categories within a single region of activity, often using discriminative responses that are weak but consistent across different sound examples. For example, using this method it has been shown that four sound categories evoke distinctive patterns of activity across the superior temporal gyrus (Staeren et al. 2009). A distributed cortical coding of sound properties could explain why several auditory regions have been implicated in the processing of many different auditory attributes. It is even possible that auditory cortical regions encoding relatively basic attributes of sounds (such as
pitch) and higher level properties (such as category) are not mutually exclusive. Much more is known about basic sound processing in the human auditory cortex than a decade or so ago. With recent interest in the application of novel approaches to fMRI design and analysis, there is every reason to be optimistic for the future.

8 Acknowledgements

The second author received support for her work from the Medical Research Council through a PhD studentship that was awarded from the MRC Institute of Hearing Research.

9 References


Chapter 3. General methodology

All psychophysical testing was performed at the Medical Research Council’s Institute of Hearing Research, Nottingham University section, Nottingham, UK. Scanning sessions were undertaken at the Sir Peter Mansfield Magnetic Resonance Centre, Nottingham University.

3.1 Participants

Altogether there were 26 listeners who participated in psychophysical experiments and fMRI scanning (10 males 16 females, age range 20 to 47 years). Five of these listeners participated in all four experiments (s02, s03, s05, s07 and s09). Additionally, 30 listeners were recruited by undergraduate project students at Nottingham University to take part in psychophysical experiments whose results were included in chapters 6 and 7. These students were supervised on a day-to-day basis by the principal investigator. All additional listeners were psychology students at the University of Nottingham who participated for course credits. No further information is available for any of these listeners. None of the listeners recruited by project students participated in any of the scanning sessions. Absolute hearing thresholds were measured following the British Society of Audiology recommended procedure (British Society of Audiology, 2004). The initial descending familiarization step size was 10 dB. Once the participant stopped responding, thresholds were determined using a 5 dB ascending and 10 dB descending procedure. Thresholds for all participants were below 20 dB HL between 0.5 and 8 kHz. The study received ethics approval from the Medical School Research Ethics Committee, University of Nottingham (ethics code A/1/2005/3.5 for chapter 4, A/1/2005/3.7 for chapters 5 and 7, A/1/2005/3.6 for chapter 6). All listeners provided written informed consent. Medical questionnaires
filled out by every listener indicated that none had a history of neurological or hearing impairment.

3.2 Equipment

3.2.1 Psychophysical Testing

For the psychophysical testing, participants sat in a double-walled sound-attenuated booth with a monitor that displayed the visual components of the test stimuli, a mouse to click the icon that initialized testing, and a three-button custom-made response box. The experimenter was in a separate control room adjacent to the booth. The experimenter and participant were able to communicate via an intercom system. A PC computer system with Microsoft Windows 2000 operating system and high-fidelity soundcard were used to create stimuli and record responses. Stimuli were delivered through Sennheiser HD 480 II headphones.

3.2.2 fMRI Scanning

The standard unit of measurement for magnetic field is the Tesla (T). All scanning for this thesis was performed on a 3 T Philips Intera Acheiva scanner (for reference, the Earth’s magnetic field is ~5 x 10^{-5} T). Sensitivity encoding (SENSE) is an intervention developed by Philips to reduce total scan time and to reduce image distortions arising at longer times to echo. All experiments in this thesis used an 8-channel SENSE receiver head coil for improved sensitivity relative to the standard single channel quadrature head coil. A SENSE factor of 2 was applied in all four experiments. A SofTone factor of 2 was applied to slow down the ramps on the gradient switching in order to further reduce acoustic noise by 9 dB. A custom-built MR compatible system delivered distortion-free sound using high-
quality electrostatic headphones (Sennheiser HE60 with high-voltage amplifier HEV70) that had been specifically modelled with no ferromagnetic components to be safe for use in fMRI.

3.3 Stimuli

3.3.1 Psychophysical Stimuli

All pitch stimuli evoked a pitch corresponding to either a 100-Hz (Chapters 5 and 6) or a 200-Hz (Chapter 4) tone. The study described in Chapter 7 did not contain a psychophysical paradigm. Temporal and spectral characteristics and presentation level varied between experiments, but some features were common to all pitch stimuli. All psychophysical pitch stimuli had a total duration of 200 ms and an inter-stimulus interval of 500 ms. With the exception of the complex Huggins pitch (cHP) used in Chapter 4, all stimuli were band-pass filtered to include only harmonics that cannot be resolved by the auditory system. Resolved harmonics were excluded because they provide tonotopic features that could provide non-pitch cues that elicit differential activation to noise, thus providing a nuisance variable that affects the response to pitch stimuli. The cHP included both low and high numbered harmonics due to the fact that the dichotic nature of the stimulus makes peripheral resolution of harmonics impossible. All stimuli except cHP contained a low-pass noise masker to mask cochlear distortion products such as combination tones (Yost 2000). As the name suggests, combination tones are created by the combination of harmonics present in the stimulus, and can introduce components at frequencies that represent peripherally resolvable harmonics. Any such distortions were masked by the addition of a low-pass noise. Chapter 6 included a non-pitch stimulus that contained slowly-varying spectro-temporal modulation (referred to as IRNo). This stimulus was matched in level,
bandwidth and masker, but the modulation was not very salient when the stimulus was 200 ms in duration. To increase the salience of this feature, the duration of IRNo stimuli was increased to 600 ms. All stimuli included 10-ms linear-intensity onset and offset ramps to avoid artifacts such as clicks that can arise from abrupt onsets and offsets.

Calibration for all psychophysical stimuli was performed to ensure that presentation levels were correct. Stimuli were presented to a KEMAR manikin (Burkhard and Sachs 1975) fitted with a Bruel and Kjaer half-inch microphone type 4134 (serial no. 906663), Zwischenlocki occluded ear simulator (Knowles model no. DB-100) and Bruel and Kjaer measureing amplifier type 2636 (Serial no. 1324093).

3.3.2 Scanning Stimuli

The stimuli that were presented in the MR scanner were similar to those used for the psychophysical experiments, but the level and duration of the stimuli were adapted for optimum response during fMRI. For some of the stimuli in Chapter 4, and all of the stimuli in Chapters 5 and 7, we adopted a continuous stimulation paradigm. This involved interleaving the experimental stimuli with a Gaussian noise, matched in bandwidth and overall level to the experimental stimuli to produce a stable envelope for the stimulus and thus reducing the contribution of multiple energy onsets (see Figure 1 from Chapter 4 for a diagram). This paradigm was employed to increase sensitivity to pitch and avoid any nonlinear effects that may affect activation patterns (see Chapter 4).

For the other half of the stimuli in Chapter 4, and all the stimuli in Chapter 6, bursts of experimental stimuli were separated by silence. A continuous stimulation
paradigm was not used in Chapter 6 because it had not been used for fMRI studies prior to the experiment described in Chapter 4. As the experiment in Chapter 6 was specifically designed to examine stimulus properties used in previous studies, it was important to match the procedure as closely as possible to those studies. Therefore, for Chapter 6, the experimental presentation paradigm was matched to that used in previous studies (traditional, or ‘classical’ presentation paradigm). A Gaussian noise stimulus matched in level and bandwidth was included in all experiments as a control for pitch and modulation stimuli.

3.4 Psychophysical Procedure

For Chapters 4, 5 and 6, scanning sessions were preceded by psychophysical testing sessions to assess accuracy in distinguishing experimental features in the sound stimuli and to expose listeners to the sound features of interest. Stimuli were presented through custom made software that is supported by the MatLab platform (The MathWorks, Natick, MA). Pitch discrimination thresholds were measured using a three-alternative forced-choice, two-down, one-up adaptive procedure that targeted 70.7% performance. Two observation intervals contained the standard tone as described in the Psychophysical Stimuli section. The remaining interval (chosen at random) contained a comparison tone with a higher f0. The monitor screen in the sound-attenuated booth contained six boxes; three empty boxes on the top row and boxes labelled 1, 2 and 3 on the bottom row. As each interval was presented, the box in the top row corresponding to that interval flashed white. Once all three intervals had been presented, the listener was required to select the button on the response box corresponding to the interval that contained the higher f0. On each trial, feedback was given via a green (correct) or
red (incorrect) light in the top-row box corresponding to the chosen interval. On the first trial, the f0 difference between standard and comparison tones was 20%. The percent difference increased or decreased by a factor of two for the first four reversals, and by a factor of 1.414 for the final 12 reversals. Discrimination threshold was taken as the geometric mean of the f0 difference at the final 12 reversals and the responses were recorded and stored electronically. The adaptive track was limited at 200%. No listeners performed below chance. There were five runs for each of the different stimuli; the first was considered as a practice and the pitch-discrimination threshold was taken as the average of the last four runs.

The psychophysical testing paradigm for modulation stimuli was as described above, with a few changes as an adaptive paradigm could not be used for IRNo. Instead of using an adaptive paradigm, the modulation testing used a three-alternative forced choice ‘odd-one-out’ paradigm where two of the intervals contained a Gaussian noise, and the other (chosen at random) was an IRNo stimulus. The task was to select the interval that contained the IRNo. Each run consisted of 50 trials and the percentage of correct responses was taken.

3.5 fMRI Protocol

Magnetic resonance imaging (MRI) makes use of the inherent magnetic properties of hydrogen nuclei which are abundant in the human body. The MR scanner transmits radiofrequency pulses that cause the hydrogen atoms to align at 90 degrees with the main field of the scanner, and the emission of energy on their return to equilibrium state is measured and recorded as the MR signal. The transition between low (equilibrium) and high (aligned) energy states has multiple
components whose representations can be differentially weighted to produce images with different contrasts. The differential concentration of water (and thus hydrogen) in different types of tissue (i.e. white and grey matter), determine the rate at which the hydrogen atoms switch between energy states, and show up as either dark or light areas depending on the weighting chosen. T1-weighted images are generally used for high-contrast anatomical images because of the sharpness of the image they produce. T2*-weighting is more typical for functional images, as it captures contributions from surrounding tissue to the MR signal, and hence provides greater sensitivity to local differences in blood oxygenation (a marker for active, or responsive, brain regions). The two main parameters that influence the weighting of the different response components are the time between radio-frequency pulses (TR, time to repeat) and the time between pulses and the rephased signal or ‘echo’ (TE, time to echo).

3.5.1 Anatomical Scanning

In order to provide individualised information on which to overlay activation maps, a high-resolution anatomical image was collected for each listener. The anatomical scan was a T1-weighted image (matrix size = 256 x 256, 160 saggital slices, TR = 8.2 ms, TE = 3.7 ms) with 1 mm$^3$ resolution. The anatomical scan was used to position the functional scan centrally on HG, and care was taken to include the entire superior temporal gyrus and to exclude the eyes. It was important to exclude the eyes because they contain a high proportion of water, which can cause artifacts known as nyquist ghosts. These arise from phase differences between gradients and manifest as reproductions of the eyes that have been shifted by half the field of view. Essentially, this would mean that a reflection of the eyes would appear around the middle of the brain, in the superior temporal gyrus.
3.5.2 Functional Scanning

Functional scanning used a T2*-weighted echo-planar sequence with 3 mm$^3$ resolution (matrix size = 64 x 64, 32 oblique-axial slices, TE = 36 ms). For the functional runs, scans were clustered into a 1969 ms period with a TR of 8000 ms. This is known as ‘sparse’ imaging and reduces the contribution of the auditory cortical responses to the background acoustic noise to the response to the sound of interest (Edmister et al. 1999; Hall et al. 1999). Functional data were acquired over two sequential scanning runs in Chapter 4 and over 4 sequential scanning runs in Chapters 5, 6 and 7. For all experiments, stimuli were presented in a quasi-random order, with stimulus conditions divided evenly across runs so that each run contained the same number of presentations for each stimulus type. Where this was not possible (e.g. where there were 15 presentations of each stimulus and four runs), the stimuli would be divided evenly and any residual stimuli would be randomly added to any of the runs. An additional rule was that the same stimulus condition was not presented twice in succession. Listeners were requested to listen to the sounds presented to them in the scanner but were not required to perform any task during scanning sessions.

3.5.3 Data Analysis

Four different types of analysis were used for the experiments in this thesis. Not all of the different analyses were utilized for each experiment, but this section provides an overall description of each of the different analysis methods. Information on experiment-specific analyses can be found in Chapters 4 – 7.
Pre-processing

Analysis of the functional imaging data was conducted using statistical parametric mapping, SPM5 (www.fil.ion.ucl.ac.uk/spm/software/spm5) separately for each listener. Due to large individual differences in brain anatomy and morphology, a number of pre-processing steps had to be followed before images could be analyzed. Pre-processing involved realigning and resizing brains so that they matched (as closely as possible) a template image. This process is crucial for comparison of stimulus-specific activation across individuals, and the pre-processing steps were common to all studies. The first pre-processing step was reorientation of the individual anatomical and functional images to a template. The template image for the anatomical pre-processing was a T1-weighted group template created from 152 individual brains from the Montreal Neurological Institute (MNI), and the functional template was a T2-weighted average group template created from the same brains. Both templates were aligned to MNI305 reference space (Eickhoff et al. 2005). The second step was realignment for inter-scan subject motion to reduce the movement-related signal and reduce variance in subsequent analyses. The output of the realignment process provided a graphical representation of inter-scan movement in three translation parameters (x, y and z in mm) and the three rotation parameters (pitch, roll and yaw in radians). The realignment process also created a mean functional image for the following step. Movement did not exceed a translation of 3 mm or a rotation of 3 radians. The next step was coregistration of the anatomical image to the mean image, to enable functional images to be overlaid onto the high-resolution anatomical image. Following coregistration, the anatomical image was segmented into white matter, grey matter and cerebrospinal fluid (Ashburner and Friston 2005). Normalization involves adjustment of the anatomical and functional images onto a template image. The anatomical image was normalized to a template in MNI (ICBM) space
and the functional images were normalized to the corresponding normalized anatomical scan. The final stage in pre-processing was smoothing. In this process, data to be included in individual analyses were smoothed by 4 mm full-width at half maximum (FWHM) of the Gaussian smoothing kernel, and data for inclusion in group analyses were smoothed by 8 mm FWHM. This procedure meets the smoothness assumptions of SPM without compromising much of the original spatial resolution, so preserving the precise mapping between structure and function (Turner et al. 1998). Individual data for group analysis were pre-processed as described above, but were smoothed to 8 mm to allow for activation at the same place in different brains to be detected (due to the high variability between individual brain anatomy) (Mikl et al. 2008). When deciding on the amount of smoothing to apply to brain images, it is important to consider the trade-off between facilitating activation between different subjects and the corresponding decrease in spatial resolution (Brett et al. 2002).

Just as computer images are made up of a large number of pixels, brain images are made of a large number of voxels (3-D pixels, or volume-pixels). SPM computes activation maps (SPM images) in MNI brain-space by carrying out a T- or F-test for each voxel in the normalized brain scan and tracking the activity of each voxel across scans. SPM creates activation maps by using the general linear model, which is explained by the equation $Y = X\beta + \varepsilon$. In this equation, $X$ is the design matrix (explained in more detail later), $\beta$ is the contribution of that regressor to the overall MR signal (calculated by SPM) and $\varepsilon$ is an error term. Regressors entered into the general linear model for each experiment are described in the corresponding Chapter for that experiment. Before starting the analysis, a design template was created for each of the experimental runs for each participant. The design template was a matrix of 1s (stimulus present in scan) and 0s (stimulus
absent in scan) in which each regressor (stimulus condition) made up one column, with six additional columns tracking translational and rotational head movement on the x, y and z axes (in order to remove head movement as a source of error). Each row represented a scan.

Image analysis: Individual data

The first step of creating the SPM images was to perform a 1<sup>st</sup>-level specification. This step models the scan-to-scan variability within each participant. In all experiments, scans were determined as the units for the design with an interscan interval (TR) of 8.2 s. Each listener’s 1<sup>st</sup>-level specification consisted the number of experimental runs in that study (2 for Chapter 4 and 4 for Chapters 5 – 7). For the first session, all normalized functional scans for run 1 were selected with multiple regressors input as a text file with the design template from run 1 and a high-pass filter cut-off of 420 s (in each study this was ample to ensure long enough cut-off), rounded up to the nearest 10) or 1/420 Hz. This process was repeated for each subsequent experimental run. The 1<sup>st</sup>-level analysis output a design matrix with one column for each regressor (split between the sessions) plus one column per session at the end, which modelled the average activity for each session over the total number of regressors. There was one row per scan. On completion of 1<sup>st</sup>-level specification, the design matrix was reviewed to ensure it appeared as expected, and the model was estimated. This step fits the model design (X) to the data (Y) in each voxel to provide a β-value for each regressor, explaining the contribution of that regressor to the overall MR signal (assuming that the process is linear (Turner et al. 1998)). The process also creates a matrix of normally distributed error terms (e). The null hypothesis for SPM images is that all βs are zero (i.e. that none of the regressors have an effect on the MR signal in the area being scanned).
Once calculated, SPM used β-values to compare regressors using t- or F-tests and create activation images. The experimental hypotheses in each experiment necessitated the use of different statistical thresholds; see individual Chapters for information on the statistical threshold used for each experiment. Activation patterns for individual data were mapped onto each listener’s own normalized anatomical image. An all-inclusive t-test comparing the pooled activation of all sound conditions against silence was performed for each listener to ensure data quality and the correct input of information into the general linear model. A visual inspection was carried out for this t-test in each listener before any subsequent analysis, to ensure data had been input correctly. Where region-of-interest analyses were performed across pre-defined auditory regions specified by cytoarchitectonic data, individual β-values were mapped onto mask images of each auditory region of interest. Masks of auditory areas Te 1.0, Te 1.1 and Te 1.2 in Heschl’s gyrus (HG) were based on cytoarchitectonic probability maps created by Morosan et al (2001). The mask of planum temporale (PT) was based on morphological details provided by a previous study quantifying the variability in PT (Westbury et al. 1999). The planum polare (PP) mask was also based on morphology, and was constructed in-house by staff at the Institute of Hearing Research, by tracing the outline of the anterior portion of the superior temporal gyrus using a group-averaged normalized anatomical image. All masks were mutually exclusive, with overlapping voxels being attributed to the area with which it had the highest probability of membership. A MatLab script was applied to extract the region-averaged β-values for each stimulus type for each participant, and for each auditory area. The number of β-values for each stimulus type varied between experiments, with a value corresponding to each experimental run in left and right hemispheres, respectively. The outputs of this process were
standardized values of response size that were used to produce tables and graphs to compare activation in response to the different stimuli in different auditory areas. It was necessary to obtain this information because, although the SPM comparisons are useful for visual inspection of the data, SPM looks at voxel significance over time, not over regions. Although individual analysis provides valuable information about fixed-effects within participants, group data are necessary to address a number of more generalizable hypotheses.

Image analysis: Group data

To provide input contrast images for the 2\textsuperscript{nd}-level group specification, a number of comparisons were performed on each 1\textsuperscript{st}-level individual analysis. The 2\textsuperscript{nd}-level specification modeled inter-subject variability, mapping voxels that were activated consistently across listeners (known as a random effects (RFX) analysis). This analysis shows the invariant behaviour of the population from which the particular sample is drawn. The 2\textsuperscript{nd}-level analysis output a design matrix comprising a column for each regressor and a row for each individual contrast scan per column (see individual Chapters for experiment-specific details). As in the 1\textsuperscript{st}-level specification, the 2\textsuperscript{nd}-level design matrix was reviewed to ensure it appeared as expected, and the model was estimated. Individual t- and F-contrasts were then performed on the data. Each contrast computation provided an output table of all supra-threshold voxels for each contrast, which included corrected and uncorrected p-values, z-values and co-ordinates at voxel-level, and the same information plus cluster size (no. of activated voxels) at the cluster level. The location of each auditory cluster could be identified for each contrast using an SPM toolbox that combines probabilistic cytoarchitectonic maps with functional imaging data (Eickhoff et al. 2005). Voxel of interest (VOI) analysis could be carried out on supra-threshold voxels using masks to evaluate the statistical
significance of contrast-related activity in that voxel. Activation patterns for group data were mapped onto a group-averaged anatomical image that was created from each of the normalized individual brains used in each experiment. After mapping the activation onto MNI brain space, tables could be created in SPM containing information about the location and size of clusters of activation present in specified contrasts. The MNI coordinates provided in these tables could then be explored in a toolbox that converts MNI coordinates to the standard brain atlas defined by Talairach and Tournoux (Talairach and Tournoux 1988). The anatomy toolbox in SPM contains a Talairach and Tournoux stereotaxic atlas with auditory areas labeled according to cytoarchitectonic subdivisions determined in post-mortem human brain studies (Morosan et al. 2001). The toolbox gives probability maps stating the likelihood of each cluster of activation being located in a labeled area or areas. There are, however, disadvantages of using a template brain – namely that the high inter-subject variability increases the likelihood of mapping activity to anatomical areas that do not correspond to the functional activity observed (Brett et al. 2002). Additionally, the toolbox does not yet contain an area corresponding to PT. Although we can map activity onto the region described in the ROI analysis as PT, the fact that this region is not defined in the toolbox means that it is not possible to obtain corresponding probability values for voxels in this region.
Chapter 4. The effect of stimulus context on pitch representations in the human auditory cortex

Daphne García¹,², Deborah A. Hall¹,³ and Christopher J. Plack²

¹MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK
²Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK
³Department of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK

This chapter was published as a paper and is presented in the format in which it was published.

Publication 2

The effect of stimulus context on pitch representations in the human auditory cortex

Daphne Garcia a,b,⁎, Deborah A. Hall a,c, Christopher J. Plack b

a MRC Institute of Hearing Research, University Park, Nottingham NG7 2BD, UK
b Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK
c Division of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK

A R T I C L E   I N F O

Article history:
Received 15 December 2009
Revised 6 February 2010
Accepted 26 February 2010
Available online 6 March 2010

A B S T R A C T

Neuroimaging studies of pitch coding seek to identify pitch-related responses separate from responses to other properties of the stimulus, such as its energy onset, and other general aspects of the listening context. The current study reports the first attempt to evaluate these modulatory influences using functional magnetic resonance imaging (fMRI) measures of cortical pitch representations. Stimulus context was manipulated using a ‘classical stimulation paradigm’ (whereby successive pitch stimuli were separated by gaps of silence) and a ‘continuous stimulation paradigm’ (whereby successive pitch stimuli were interspersed with noise to maintain a stable envelope). Pitch responses were measured for two types of pitch-evoking stimuli; a harmonic-complex tone and a complex Huggins pitch. Results for a group of 15 normally hearing listeners revealed that context effects were mostly observed in primary auditory regions, while the most significant pitch responses were localized to posterior nonprimary auditory cortex, specifically planum temporale. Sensitivity to pitch was greater for the continuous stimulation conditions perhaps because they better controlled for concurrent responses to the noise energy onset and reduced the potential problem of a non-linear fMRI response becoming saturated. These results provide support for hierarchical processing within human auditory cortex, with some parts of primary auditory cortex engaged by general auditory energy, some parts of planum temporale specifically responsible for representing pitch information and adjacent regions that are responsible for complex higher-level auditory processing such as representing pitch information as a function of listening context.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Pitch is an important feature of auditory perception. It is arguably the most important perceptual feature of music and is a key component of tonal languages used in many parts of the world, such as sub-Saharan Africa and East Asia. In non-tonal languages such as English, we use pitch to recognize the gender and identity of different speakers as well as using intonation to discriminate between different types of sentence (e.g. a question or a statement, Chatterjee and Peng, 2008) and as a cue to stress. Pitch is also one of the main cues used by the auditory system to segregate sounds from different sources (Singh, 1987). Most pitch stimuli, whether natural or laboratory-made, are ‘complex’ tones made up of a number of harmonic sinusoidal components with frequencies that are integer multiples of the repetition rate or fundamental frequency (f0). In an early psychophysical study on the frequency analytical power of the human ear, Plomp (1964) discovered that the human ear is capable of ‘hearing out’ the first five to eight harmonics of a complex harmonic tone. These are the harmonics that each excite a different place on the basilar membrane, and are said to be ‘resolved’. The basilar membrane can be modeled as a bank of bandpass filters, with a width corresponding to about 12% of the center frequency, for frequencies between 750 and 5000 Hz (Moore, 2003). Resolved harmonics fall within individual filters so that pitch may be determined by the distinctive pattern of spectral peaks in the neural excitation pattern. In contrast, for unresolved harmonics, multiple harmonics excite the same filter. For these stimuli, the pitch can be determined from the waveform produced by the interaction of the harmonics, whose repetition rate corresponds to the f0 of the complex tone (Houtsma and Smurzynski, 1990; Carlyon et al., 1992; Michiel and Oxenham, 2004).

Although most pitch-evoking stimuli encountered in the environment are harmonic-complex tones, a pitch sensation can be evoked by manipulating noise signals, for example amplitude or frequency modulation (Mahaffey, 1967; Darwin et al., 1994) and spectral rippling (Yost and Hill, 1979). Cramer and Huggins (1958) found that pitch can even be conveyed through binaural interaction, with signals that contain no spectral or temporal pitch information when played individually to each ear. They presented the same wideband noise to both ears, except for a narrow frequency band, which was out of phase between the ears. A pitch was heard corresponding to the
center frequency of the band. This ‘Huggins pitch’ (HP) is one of a number of binaural pitches that have now been identified (Plack and Oxenham, 2005).

Over recent years, neuroimaging methods such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG), magnetoencephalography (MEG) and positron emission tomography (PET) have been used to search for the neural substrates of pitch processing in human listeners. These studies do not necessarily claim that pitch is first extracted in the auditory cortex, instead they simply seek to demonstrate that pitch is one of the organizing principles of sound coding at the level of the auditory cortex. Nevertheless, there are a number of discrepancies in the neuroimaging literature, especially in terms of neural mechanisms for pitch coding and the localization of those neural representations. Some authors have suggested that the same cortical neurons that represent pitch information are also involved in coding other aspects of sound, such as energy onset (Nääätänen and Picton, 1987), while others claim that these properties are processed separately (Schönwiesner and Zatorre, 2008). Some authors have proposed that different pitch percept can be elicited by sounds that possess very different spectral, temporal, and/or binaural characteristics, there should be a unified representation of pitch (Hall and Plack, 2009), while others argue that the physiological support for this claim is rather weak (Nelken et al., 2008).

A number of human neuroimaging studies have identified a pitch-sensitive region in the auditory cortex and have localized it to Heschl’s gyrius (HG, see Fig. 3) (Griffiths et al., 1998; Gutschalk et al., 2002; Patterson et al., 2002; Hall et al., 2006; Puschmann et al., 2010). A number of these studies have favored the use of a single type of pitch-evoking stimulus known as iterated ripple noise (IRN). IRN is created by generating a sample of noise and imposing a delay before adding (or subtracting) the noise back to (or from) the original. The pitch sensation of the resulting sound is related to the reciprocal of the delay, and its salience is determined by the number of delay-and-add (or subtract) iterations and the gain applied to the delayed sample (Yost, 1996). One of the earliest studies to localize pitch representa- tion was a PET study that identified areas of human auditory cortex that were sensitive to pitch salience (Griffiths et al., 1998). Here, salience was manipulated by systematically increasing the temporal regularity of IRN signals; with 0, 1, 2, 4, 8 and 16 iterations, respectively. Around HG, bilateral pitch-related activity was found to increase in magnitude with increasing pitch salience. Despite the data smoothing applied to the PET images, the focus of activity in the right hemisphere appeared to be close to the central portion of HG, while the focus in the left hemisphere appeared to be centered on lateral HG. Consistent with this finding, some MEG studies have implicated lateral HG in pitch processing by using click trains (Gutschalk et al., 2002, 2004, 2007). Furthermore, these findings are consistent with a non-human primate model of pitch coding that localizes pitch-selective neurons to a discrete cortical region near the anterolateral border of the primary auditory cortex (Bendor and Wang, 2005). This low-frequency region is proposed to correspond to lateral HG in humans.

Other neuroimaging studies have benefited from the greater spatial specificity of fMRI. For example, Hall et al. (2006) confirmed that pitch-related activity was present in lateral HG and tended to overlap with a primary-like region that was sensitive to low-frequency tones, irrespective of the spectral content of the (IRN) pitch-evoking stimuli. Results from a number of studies agree that the pitch-sensitive response is not confined to lateral HG, but spreads into adjacent posterior or anterior regions of the superior temporal gyrus (Patterson et al., 2002; Barrett and Hall, 2006; Penagos et al., 2004).

Hall and Plack (2009) have called into question the assumption that lateral HG operates as the main center for the cortical representation for pitch. Hall and Plack argued that evidence from one type of pitch-evoking stimulus alone does not constitute reliable evidence for a ‘pitch centre’. The motivation for their study was therefore to examine whether pitch-related responses in lateral HG were consistently present for a range of different pitch-evoking stimuli, each with different physical characteristics. Pitch-evoking stimuli included IRN, single-frequency tones, wideband complex tones, missing f0 complex tones containing resolved or unresolved harmonics, and an HP stimulus. While IRN generated a pitch-sensitive response in lateral HG, the other pitch-evoking stimuli were more likely to produce activity in planum temporale (PT) than in lateral HG. The authors concluded that there was insufficient consistency across pitch effects to label any one region a ‘pitch center’. However, these conclusions have recently been questioned by findings from an fMRI study that obtained significant responses in lateral HG and PT for two different HP stimuli and for a single-frequency tone-in-noise signal (Puschmann et al., 2010). Typically, fMRI studies of pitch processing favor the presentation of a sequence of pitch-evoking sounds, each separated by silent intervals. This ‘classical stimulation paradigm’ is preferred because a slow repetition rate of stimulus bursts is known to evoke a robust and sustained fMRI response in auditory cortex (Harms and Melcher, 2002). Pitch-related activation is computed by subtracting from this condition the response to a baseline condition containing a matched sequence of noise bursts (Friston et al., 1996). Such subtraction methods rely on the assumption that the context of the stimulus presentation has no effect on the magnitude of the pitch-related response. In other subject areas, such as language processing (Price et al., 1997), this assumption has been shown to be untrue.

With regard to pitch, there is some evidence that auditory evoked responses are sensitive to the abrupt onset of sound energy (the energy-onset response) as well as to the abrupt onset of pitch (the pitch-onset response) (e.g. Krumbholz et al., 2003; Chait et al., 2006; Seither-Prieler et al., 2004). Thus, it is plausible that the pattern of energy onsets in the stimulus sequence might modulate the pitch-related response in a context-dependent manner.

The relationship between pitch- and energy-onset auditory evoked responses has been examined in detail in human listeners using MEG. The benefit of using MEG is that, unlike fMRI, it has millisecond temporal resolution that allows for the reliable detection of individual transient deflections (for a review see König et al., 2007). Krumbholz et al. (2003) separated the evoked response to the energy onset from that to the pitch onset by using a continuous stimulation paradigm. Here, the stimulus has a fixed spectral energy, but the perceptual features alternate between noise and pitch. The authors observed a transient deflection at about 150 ms after the transition from noise to pitch (the pitch-onset response), but not from pitch to noise, nor from one sample of noise to another. Moreover, the amplitude of the pitch-onset response increased as a function of pitch salience (number of iterations) and the latency of the pitch-onset response decreased as a function of pitch value (IRN delay). These results confirmed to the authors that the observed response was not simply related to detecting a perceptual change in the stimulus, but was indicative of pitch-specific coding. The study by Krumbholz et al. (2003) exclusively measured IRN, but similar properties of the pitch-onset response have been reported for a tone-in-noise stimulus and for HP (Chait et al., 2006).

Results from a recent depth-electrode study by Schönwiesner and Zatorre (2008) extend these findings from surface recordings of electromagnetic activity. The patient’s pattern of brain activity revealed a double dissociation between the pitch-onset response and the energy-onset response. The former stimulated electrodes placed across lateral portions of HG while the latter stimulated elec- trodes placed across medial portions of HG. These results refute the idea that the same, or overlapping, populations of auditory cortical neurons respond to energy and pitch onsets. Although the relatively poor temporal resolution of fMRI does not allow for the identification of individual evoked responses, we suggest that fMRI activation represents the accumulated activity resulting from a sequence of transient responses. Thus, the experiment described here constitutes the first attempt to investigate the differential
consequences of energy and pitch responses on fMRI measures of brain activity.

A difference between EEG and MEG measures of pitch coding and those of fMRI concerns the degree to which the response that is measured saturates at the upper limits of the response function. The fMRI response is known to be highly susceptible to non-linearities (Sidtis et al., 1999; Friston et al., 2000; Devor et al., 2003). Hence, the response to a sound stimulus that contains a combination of response-evoking features (e.g. energy and pitch onsets) will be most likely to exhibit saturation. If the fMRI response to a noise stimulus is brought close to saturation by the repeated onset of acoustical energy, any additional response (i.e. the addition of a pitch) will be limited by the saturation of the fMRI signal. This is illustrated in Fig. 1a. If the non-linear response model is correct, then the pitch-related activity (pitch condition minus noise condition) might be expected to be greater in the noise context than in the silent context because the former comparison is less affected by the saturating upper limit. This model has been invoked to explain previous auditory fMRI results (Melcher et al., 2000). The alternative model that proposes a linear system (e.g. Dale and Buckner, 1997) would predict an additive rather than a sub-additive response and this would be reflected in an equivalent pattern of pitch-related activation, irrespective of the stimulus context. The predictions of the linear model are illustrated in Fig. 1b.

In the present study, the energy-onset response was manipulated by presenting a sequence of pitch-evoking signals either within a silent context (akin to a ‘classical stimulation’ paradigm, see Hall and Plack, 2009) or a noise context (akin to the ‘continuous stimulation’ paradigm, see Krumbholz et al., 2003). For the silent context, we assume that the onset of each pitch will evoke both energy- and pitch-onset responses. For the noise context, we assume that the onset of each pitch will evoke only a pitch-onset response. We assume that sustained responses to pitch (see Gutschalk et al., 2004, 2007) are not markedly affected by the stimulus context and so do not contribute to any observed differences in pitch-related activity. To ensure the findings were not specific to a particular stimulus, the hypothesis was examined using two different pitch-evoking stimuli; an unresolved harmonic-complex tone (UNRES) and a complex HP (cHP).

Materials and methods

Listeners

Fifteen listeners (8 male, 7 female; age range 23–48 years) with normal hearing (≤20 dB hearing level between 250 Hz and 8 kHz) took part in this study. All but one listener (#01) was right-handed (laterality index = 50, Oldfield, 1971). Seven listeners were musically trained between grade 3 and diploma level (# 01, 02, 07, 08, 10, 12 and 15) while five others reported informal musical experience (self-taught/ungraded, # 04, 05, 09, 13 and 16). One listener (#11) completed the psychophysical testing but was not able to return for the fMRI session. None had a history of any neurological or hearing impairment. Listeners gave written informed consent and the study was approved by the Medical School Research Ethics Committee, University of Nottingham.

Stimuli

All stimuli evoked a pitch corresponding to a 200-Hz tone. One stimulus was a (diotic) unresolved harmonic-complex tone with harmonics 10–20 (henceforth referred to as UNRES). The level of each harmonic was 23 dB greater than the spectrum level of the control noise so that the gross spectral density of all the stimuli was the same. The UNRES stimulus was filtered between 2 and 4 kHz with a noise masker (49 dB SPL spectrum level) from 0 to 2 kHz (to mask cochlear distortion products). The other stimulus was a (dichotic) complex HP (henceforth referred to as cHP) in which the pitch cue was only available via integration of the signals from each ear (dichotic). The cHP stimulus was created from a diotic Gaussian noise (49 dB SPL spectrum level) filtered between 0 and 4 kHz. In one ear, a phase shift was introduced in eight 30-Hz wide frequency bands, centered on the first eight harmonics. The noise control stimulus was a Gaussian noise (49 dB SPL spectrum level), again low-pass filtered at 4 kHz. The three signals (UNRES, cHP and noise) were matched in bandwidth (0–4 kHz) and spectral density (and hence overall energy). It is probably impossible to generate stimuli that differ in pitch strength but are perfectly matched for every other perceptual feature. While the Gaussian noise is a good control for cHP (the only other perceptual difference between the two is the spatiality of the decorrelated band), it is perhaps less so for UNRES because the signals differ in other respects, such as their envelope structure. The logic of ‘common activity’ has been applied to reduce the risk of attributing these potential differences to pitch (Hall and Plack, 2009; Puschmann et al., 2010) and can also be applied to the analysis of the current experiment.

For the psychophysical testing, each pitch stimulus was 350 ms in duration (including 10-ms linear-intensity onset and offset ramps) and the inter-stimulus interval was 500 ms. Reference stimuli had an f0 of 200 Hz. The stimuli were presented at an overall level of 85 dB SPL, calibrated using a KEMAR manikin (Burkhard and Sachs, 1975) fitted with Brüel and Kjaer half-inch microphone type 4134 (serial no. 906663), Zwislocki occluded ear simulator (Knowles model no. DB-100) and Brüel and Kjær measuring amplifier type 2636 (serial no. 1324093), scaled from 22.4 Hz to 22.4 kHz using fast time constant (125 ms) on maximum hold. Due to the metallic components in the KEMAR system, calibration inside the scanner was not possible.

In the scanner, stimulus conditions each comprised a 15.41-s alternating sequence of 450-ms experimental sounds each separated by 230 ms. In the ‘pitch-in-noise-context’ conditions, the separation contained a Gaussian noise as the context. In the ‘pitch-in-silent-context’ conditions, the pitch signals were separated by 230 ms silence. The first and last components of each sequence were the context. Each pitch and noise signal was generated using 10 ms linear-intensity onset and offset ramps, which were overlapped at the 3 dB points to produce a stable envelope for the stimulus (see Fig. 2).
Eighteen sample sequences were created for each condition. The control noise conditions were created in the same way.

Psychophysical testing

Prior to the scanning session, each participant performed a pitch-discrimination test to assess accuracy in distinguishing the pitch cues. Psychophysical testing was carried out in a sound-attenuating booth and stimuli were delivered through Sennheiser HD 480 II headphones. Stimuli were presented through custom-made software that is supported by the Matlab platform (The MathWorks, Natick, MA). Pitch-discrimination thresholds were measured for cHP and UNRES using a three alternative forced-choice, two-down, one-up, adaptive procedure that targeted 70.7% performance (Levitt, 1971). Two observation intervals contained the standard tone (UNRES or cHP) with an f0 of 200 Hz. The remaining interval (chosen at random) contained a comparison tone with a higher f0 which the listener was required to select as the 'odd one out'. On the first trial, the f0 difference between standard and comparison was 20% (40 Hz). The percent difference increased or decreased by a factor of two for the first four reversals, and by a factor of 1.414 for the final 12 reversals. Discrimination threshold was taken as the geometric mean of the f0 difference at the final 12 reversals. The adaptive track was not allowed to increase above 200% (600 Hz). Responses were recorded and stored electronically. On each trial, feedback was given via a green (correct) or red (incorrect) light on the software interface. There were five runs each for cHP and UNRES; the first was considered as practice and so the pitch-discrimination threshold was taken as the average of the last four runs.

fMRI protocol

Scanning was performed on a Philips 3 T Intera Achieva using an 8-channel SENSE receiver head coil. A T1-weighted high-resolution (1 mm³) anatomical image (matrix size = 256 × 256, 160 sagittal slices, TR = 8.2 ms, TE = 3.7 ms) was collected for each subject. The anatomical scan was used to position the functional scan centrally on the head. Functional scanning used a T2*-weighted echo-planar sequence with a voxel size of 3 mm³ (matrix size = 64 × 64, 32 oblique-axial slices, TE = 36 ms). Sparse imaging with a TR of 8000 ms and a clustered acquisition time of 1969 ms was used (Edmister et al., 1999; Hall et al., 1999). A SENSE factor of 2 was applied to reduce image distortions and a SoftTone factor of 2 was used to reduce the background scanner noise level by 9 dB. Functional data were acquired over two runs of 128 scans each, with the sounds presented in a quasi-random order, and with the rule that the same stimulus condition was not presented twice in succession. Listeners were requested to listen to the sounds, but were not required to perform any task. A custom-built MR compatible system delivered distortion-free sound using high-quality electrostatic headphones (Sennheiser HJE60 with high-voltage amplifier HEV70) that had been specifically modified for use during fMRI.

Data analysis

Images were analyzed separately for each listener using statistical parametric mapping (SPM5, http://www.fil.ion.ucl.ac.uk/spm). Pre-processing steps included realignment to correct for subject motion, normalization of individual scans to a standard image template, and smoothing with a Gaussian filter of 8 mm full width at half maximum. Individual analyses were computed for the two runs (256 scans), specifying the two pitch and the two noise conditions as separate regressors in the design. In the individual analysis, we specified separate statistical contrasts for each sound condition relative to the silent baseline that was implicitly modeled in the design. A high-pass filter cutoff of 420 s was used.

First, the data for individual participants was analyzed using a first-level general linear model to assess the effects of interest with respect to the scan-to-scan variability. The resulting model estimated the fit of the design matrix (X) to the data (Y) in each voxel in order to provide t values (the contribution of a single regressor to the overall fMRI signal). In order to obtain activation maps for individual analysis, SPM was used to fit the GLM to each individual voxel in the functional image, and to compute individual t statistics. The effect of each stimulus condition was identified and the resulting (unthresholded) contrast images were entered into a group-level random effects analysis in order to assess the effects of interest with respect to the inter-subject variability. At this group level, 2 × 2 repeated measures ANOVAs were created, with signal (pitch present and pitch absent, i.e. noise) and context (noise and silent contexts) as factors. Separate ANOVAs were computed for UNRES and cHP conditions and within each ANOVA, simple main effects and interactions were calculated using t statistics (Friston et al., 2005). Although initial SPM t contrasts were defined using an uncorrected threshold of p < 0.001, all results are reported after small volume correction (SVC) to control for type 1 errors using a false discovery rate (FDR) threshold of p < 0.05 (Genovese et al., 2002). The small volume defined the auditory cortex across the superior temporal gyrus (including HG, PT, and planum polare) and contained 4719 voxels in the left hemisphere and 5983 voxels in the right hemisphere. Activations were localized using an SPM toolbox that overlaid an SPM thresholded map onto a set of probabilistic maps of the three cytoarchitectonic subdivisions of HG (Te 1.0, Te 1.1 and Te 1.2, Morosan et al., 2001; Eichhoff et al., 2006).

Results

Behavioral results

The mean geometric discrimination threshold across the listeners for cHP was 2.93 Hz and for UNRES was 3.54 Hz. A paired t-test showed that thresholds for the two stimuli did not differ significantly t(11) = 0.053, p = 0.821. For comparison, previous research suggests that the threshold for a 200-Hz pure tone is ~ 1 Hz (Wier et al., 1977), and that for a 200-Hz unresolved harmonic-complex tone is ~ 5 Hz (Houtsma and Smurzynski, 1990).
fMRI results: effect of stimulus context

For the main effect of context, bilateral clusters of activation ($p < 0.05$ SVC) were revealed for both pitch types (UNRES and cHP, Fig. 3). These areas in which activation was greater for the silent context than for the noise context. Both UNRES and cHP showed the same pattern of context-related activation. In the left hemisphere, the most significant activation was in the medial portion of HG (Te1.1, Morosan et al., 2001) ($x = −38, y = −26, z = 6$ mm for both pitch types, Table 1). In the right hemisphere, the most significant activation was in PT (cHP: $x = 64, y = −30, z = 12$ mm; UNRES: $x = 66, y = −22, z = 10$ mm). However, for both pitch types there was a substantial spread of context-related activation across bilateral HG (areas Te1.0, central HG; Te1.1, medial HG, and Te1.2, lateral HG) and PT (Fig. 3). There were no voxels that showed a greater response for the noise context than the silent context, which suggests that the human auditory cortex is more responsive to successive energy onsets than it is to the overall energy in the stimulus.

Table 1

<table>
<thead>
<tr>
<th>Left hemisphere</th>
<th>Right hemisphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak coordinates</td>
<td>Z-score</td>
</tr>
<tr>
<td>cHP contrasts</td>
<td></td>
</tr>
<tr>
<td>Context: silence &gt; noise</td>
<td>$−38 − 26 6$</td>
</tr>
<tr>
<td></td>
<td>$−66 − 42 20$</td>
</tr>
<tr>
<td></td>
<td>$−42 − 14 20$</td>
</tr>
<tr>
<td></td>
<td>$−44 − 20 14$</td>
</tr>
<tr>
<td>Pitch: cHP &gt; noise</td>
<td>$−58 − 24 8$</td>
</tr>
<tr>
<td></td>
<td>$−48 − 6 − 8$</td>
</tr>
<tr>
<td>Interaction</td>
<td>No suprathreshold voxels</td>
</tr>
<tr>
<td>UNRES contrasts</td>
<td></td>
</tr>
<tr>
<td>Context: silence &gt; noise</td>
<td>$−38 − 26 6$</td>
</tr>
<tr>
<td></td>
<td>$−60 − 38 14$</td>
</tr>
<tr>
<td></td>
<td>$−36 − 18 16$</td>
</tr>
<tr>
<td>Pitch: UNRES &gt; noise</td>
<td>$−62 − 24 8$</td>
</tr>
<tr>
<td></td>
<td>$−46 − 8 − 6$</td>
</tr>
<tr>
<td></td>
<td>$−60 − 12 4$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$−46 − 40 20$</td>
</tr>
<tr>
<td></td>
<td>$−46 − 18 − 2$</td>
</tr>
<tr>
<td></td>
<td>$−44 − 38 20$</td>
</tr>
</tbody>
</table>

$^a$ FDR-corrected.
fMRI results: effect of pitch

The effect of signal (pitch present versus pitch absent) was also computed from the $2 \times 2$ full factorial ANOVA to identify auditory cortical regions in which activation was greater for the pitch condition than for the spectrally matched noise control. Both cHP and UNRES contrasts revealed large bilateral clusters of pitch-related activation (Table 1). For both pitch types, the greatest response was located bilaterally in PT. Peak voxels were within 4 mm of each other ($x = -58$, $y = -24$, $z = 8$ mm and $x = -62$, $y = -24$, $z = 8$ mm in the left hemisphere for cHP and UNRES respectively; and again $x = 64$, $y = -16$, $z = 6$ mm and $x = 66$, $y = -18$, $z = 6$ mm in the right). While it is true that the pitch-related activity for UNRES appeared somewhat more widespread than that for cHP, there was an extremely high agreement between the most significant peaks (Table 1). This finding increases our confidence that this activation focus represents a response to the pitch quality of the stimuli, and not to some other feature that was not perfectly matched between conditions. There were no voxels that showed a significantly greater response to the control noise than to the pitch stimulus.

According to a number of slices displayed in Fig. 3, some of the pitch-related activity appears to span Heschl’s sulcus (the posterior border of HG) and so we explored the data further to establish where the central focus of activity was located in relation to this landmark, as well as in relation to the lateral HG response reported by Puschmann et al. (2010), in Table 2 of their paper. The spatial coordinate of the most significant response to cHP was extracted for each individual listener and these data were used to compute a mean coordinate and its 95% confidence intervals in each dimension. The cHP contrast was chosen as it was most comparable to the HP stimulus reported by Puschmann et al. Panel a (Fig. 4) shows that the peaks were separated by 11 and 8 mm in the left and right hemispheres, respectively, with our focus being posterior to that of Puschmann et al. (2010) on the posterior side of Heschl’s sulcus. It is perhaps also worth noting here that our result is not entirely contradictory with the neuroimaging literature. Indeed, even Puschmann et al. (2010) observed some significant bilateral pitch-related activity in PT. The anterolateral portion of PT has been widely associated with the representation of nonspatial auditory features, as shown by a meta-analysis (Arnott et al., 2004). This is illustrated in Fig. 4b.

Regional differences in the response to context and pitch

The activation maps generated from the SPM analysis indicated a preference for stimulus context in HG and a preference for pitch in PT and lack of any clear preference for either feature in lateral HG. To quantify these putative differences between regions, we conducted a number of region-of-interest analyses enabling direct statistical comparison between regions. Three spatially discrete regions were defined using the probabilistic values for areas Te 1.0 and Te 1.2 (Morosan et al., 2001), and for PT (Westbury et al., 1999). Using all voxels within each region-of-interest, the mean estimate of the size of the fMRI response to each stimulus was computed separately for each listener. For UNRES and cHP, repeated measures ANOVAs were specified to examine differential responses to the stimulus context across Te 1.0 and PT. For both types of pitch, there was a significant interaction between context and region ($F(1,26) = 17.53$, $p < 0.001$ for UNRES and $F(1,26) = 18.43$, $p < 0.001$ for cHP). While Te 1.0 was highly sensitive to stimulus context (Fig. 5a), PT was significantly less so (Fig. 5c).

Fig. 5c also demonstrates how PT is sensitive to UNRES and cHP. For both contrasts, the main effect of pitch within PT reached significance ($F(1,26) = 38.76$, $p < 0.001$ for UNRES and $F(1,26) = 10.03$, $p = 0.004$ for cHP). In Te 1.2, there was support for a preference for UNRES compared to the control noise ($F(1,26) = 9.72$, $p = 0.004$), but this was not true for cHP ($p = 0.146$) (see Fig. 5b). To examine whether these apparent regional differences were significant, we performed a direct comparison between Te 1.2 and PT, again using ANOVA statistics. The overall effect of pitch was significant ($F(1,26) = 18.82$, $p < 0.001$ for UNRES and $F(1,26) = 4.58$, $p = 0.042$ for cHP) but there were no significant interactions between pitch and region for either stimulus type. In other words, the effect of pitch was not significantly greater in PT than in Te 1.2. It should be noted that the estimates of pitch-related activity in PT are rather conservative because the estimates of response magnitude were averaged over a large number of voxels, and many of those in posterior PT were unresponsive to any of the stimulus features. Thus, taken together, these analyses indicate a clear pitch-related response in PT, but they do not rule out the possibility of a similar pitch response in Te 1.2 (i.e., lateral HG).
Concurring with previous Discussion, the former comparison was less affected by the saturating upper limit. In the noise context than in the silent context would be expected if the observed activity, and (iii) the significant interaction between pitch and context in which the sum of those responses has a maximum magnitude of the pitch response was determined by some higher-order property of the stimulus, which was determined by a particular combination of features (i.e., pitch and context). A reasonable conjecture is that the pattern of results within these subdivisions of PT is consistent with the (non-linear response) model; in which the same neural population is responsive to both energy onsets and pitch onsets and in which the sum of those responses has a maximum saturating limit. Non-linearity is perhaps not restricted to the fMRI methodology since neuromagnetic studies of the pitch-onset response have also reported a greater sensitivity to pitch when the energy-onset response has been eliminated by presenting the pitch signals in a noise context rather than in a silent context (Krumhholz et al., 2003; Seither-Priesler et al., 2004; Chait et al., 2006). These MEG studies have convincingly demonstrated that these pitch-onset effects cannot be attributed to a general response to stimulus change. Furthermore, if the context effects seen here were simply a non-specific response to stimulus alternation, one would expect that the subtraction of the continuous noise condition from the pitch-in-noise conditions would elicit the equivalent pattern of activity as the subtraction of the continuous noise from the noise-in-silence condition. This was not the case.

Although the response to energy onsets was greatest within primary auditory cortex (including Te 1.0, Fig. 5a), this effect was by no means restricted to primary regions. Given the widespread effects of stimulus context, it is cautionary to note that previous fMRI reports of pitch-related activity could have perhaps underestimated the cortical representations of pitch whenever those experiments utilized the ‘classical stimulation’ paradigm. For example, we speculate that this explanation might account for the rather low consistency of pitch-related activity reported for the 16 listeners in Hall and Plack (2009).

**Interaction showing modulatory effect of context on pitch-related activity**

Fig. 3 illustrates some overlap between the effects of context and pitch. The interaction term from the factorial ANOVA was examined to determine the pattern of co-activation because a significant interaction would demonstrate that the pitch-related activity was modulated by the stimulus context. Cortical regions showing a significant interaction between pitch and context are shown in yellow in Fig. 3 and are reported in Table 1. For the UNRES conditions, a number of small foci of bilateral activity were located in PT and one in left central HG (Te 1.0). No significant interaction was observed for the cHP conditions, although there was some evidence for a similar distribution of activity at the uncorrected threshold (p < 0.001, not shown).

To understand the shape of the interaction, a post-hoc region-of-interest analysis was computed again using the estimates of size of the response in PT. As Fig. 5c shows, although the responses in the noise context were marginally smaller than in the silent context, the difference between the pitch and noise conditions was much more marked in the noise context. Post-hoc testing demonstrated that this difference reached significance (p < 0.05) in the UNRES condition.

Three observations are consistent with the non-linear model; (i) the region is sensitive to detecting changes in energy and in pitch, but the two effects are not additive, (ii) the larger responses in the silent context than in the noise context are slight but at least consistent with the interpretation that the accumulated response to successive energy onsets contributed to the overall magnitude of the observed activity, and (iii) the significantly greater pitch-related activity in the noise context than in the silent context would be expected if the former comparison was less affected by the saturating upper limit.

**Discussion**

The present fMRI study reports a novel attempt to measure the effects of stimulus context on the cortical representation of pitch. Concurring with previous findings from the same authors (Hall and Plack, 2009), the most significant pitch-related activity was centered on posterior auditory cortex, in lateral PT. The results are consistent with the view that posterior auditory cortex is engaged in nonspatial, as well as in spatial, auditory analysis (Arnott et al., 2004). Some pitch-related activity was identified in lateral HG (Te 1.2). However, there was no convincing evidence for a general sensitivity to pitch in this region because the effect was only significant for UNRES. The UNRES contrast is not ideally controlled for non-pitch features such as temporal modulations in the UNRES signal that are not present in the control noise.

Focal subdivisions of PT revealed a modulatory effect such that the magnitude of the pitch response was determined by some higher-order property of the stimulus, which was determined by a particular combination of features (i.e., pitch and context). A reasonable conjecture is that the pattern of results within these subdivisions of PT is consistent with the (non-linear response) model; in which the same neural population is responsive to both energy onsets and pitch onsets and in which the sum of those responses has a maximum saturating limit. Non-linearity is perhaps not restricted to the fMRI methodology since neuromagnetic studies of the pitch-onset response have also reported a greater sensitivity to pitch when the energy-onset response has been eliminated by presenting the pitch signals in a noise context rather than in a silent context (Krumhholz et al., 2003; Seither-Priesler et al., 2004; Chait et al., 2006). These MEG studies have convincingly demonstrated that these pitch-onset effects cannot be attributed to a general response to stimulus change. Furthermore, if the context effects seen here were simply a non-specific response to stimulus alternation, one would expect that the subtraction of the continuous noise condition from the pitch-in-noise conditions would elicit the equivalent pattern of activity as the subtraction of the continuous noise from the noise-in-silence condition. This was not the case.

Although the response to energy onsets was greatest within primary auditory cortex (including Te 1.0, Fig. 5a), this effect was by no means restricted to primary regions. Given the widespread effects of stimulus context, it is cautionary to note that previous fMRI reports of pitch-related activity could have perhaps underestimated the cortical representations of pitch whenever those experiments utilized the ‘classical stimulation’ paradigm. For example, we speculate that this explanation might account for the rather low consistency of pitch-related activity reported for the 16 listeners in Hall and Plack (2009).

**fMRI evidence for a hierarchy of auditory processing across primary and nonprimary regions**

The results from the present fMRI study indicate a hierarchy of auditory processing where physical properties related to the temporal
structure of the acoustic energy engage HG (primary auditory cortex) and perceptual features such as pitch dominate the response in PT (nonprimary auditory cortex). In addition, subdivisions of PT appear sensitive to particular combinations of features (i.e. context and pitch). It is possible to speculate that the observed effects of context and pitch represent a modulation of stimulus processing by the changing stimulus features involuntarily capturing the focus of selective attention, and do not isolate stimulus-driven processing. Two reasons lead us to believe this not to be the case. First, a carefully designed fMRI study seeking to measure the effects of pitch-related attention in the auditory cortex by manipulating the focus of attention towards or away from the pitch of an IRN stimulus, failed to find any such effects (Krumbholz et al., 2007, see also Altmann et al., 2008; Paltoglou et al., 2009). Second, it is unclear why attentional capture by changes in pitch or changes in stimulus energy might engage different parts of the auditory cortex, as observed in the current set of results.

With respect to the present localization of the energy-onset response in human primary auditory cortex, it is perhaps worthwhile returning to the findings from the depth-electrode study reported by Schönwiesner and Zatorre (2008). Recall, they reported energy-onset responses recorded at an electrode placed on medial HG and pitch-onset responses at an electrode placed on lateral HG, suggesting spatially separate neural populations responsive to the two sound attributes. Their results are somewhat consistent with the hierarchical model that we propose to account for our fMRI data. Certainly, we would agree that medial HG is strongly responsive to energy onsets because this was the main activation site where the fMRI response was significantly diminished when energy onsets were eliminated by the noise context. In the present study, there was partial evidence for the engagement of lateral HG in the response to pitch. However, the present fMRI data would predict that if one is to place an electrode array intracranially across the surface of PT instead of across HG, then an even greater pitch response might be recorded.

It is generally accepted that ‘higher’ cortical regions encode perceptual qualities, such as spatial location or speech sound identity (Davis and Johnsrude, 2003; Nelken, 2008; Rauschecker et al., 1995; Rauschecker, 1998). The human nonprimary region PT would fit into this conceptualization. However, we do not claim that the role of primary auditory cortex is merely to encode simple (i.e. physical) attributes of a stimulus. In fact, it would perhaps be rather naïve to do so. For example, electrophysiological recordings have established that primary auditory cortical neurons have complex response properties, showing sensitivity to both low-level and high-level features of sounds (Nelken, 2008; Kelly and Sally., 1988). Nevertheless, the present study demonstrates a reasonably clear division between HG and PT in terms of representations of context and pitch, respectively. Moreover, only subdivisions of PT (the ‘higher’ cortical region) were sensitive to response interactions between stimulus context and pitch representations.

**Pitch-related activity in planum temporale (PT), for both UNRES and cHP**

The location of pitch-related activity found in the present factorial fMRI experiment is more posterior than previously suggested by fMRI studies that have used IRN as the pitch-evoking stimulus (e.g. Patterson et al., 2002; Hall et al., 2005), have applied dipole source modeling to MEG data in order to localize the neural generators of the pitch response (Krumbholz et al., 2003; Gutschalk et al., 2002, 2004, 2007) or have applied a region-of-interest mask defining lateral HG (Puschmann et al., 2010). Our data demonstrate that the most reliable location of pitch-related activity appears to be immediately behind lateral HG, in PT.

To ensure the findings were not specific to a particular stimulus, the effect of context on pitch representations was examined using two different pitch-evoking stimuli. The finding that the location of responses to UNRES and cHP are broadly comparable suggests that the pitch responsive region observed in the present study is not stimulus specific. However, UNRES elicited activation in a greater number of voxels. Although inter-interpreter consistency was low in the Hall and Plack (2009) study, an incidence map of their 16 individual listeners also demonstrates similar patterns for their unresolved harmonic complex and simple HP. Information on the number of activated voxels was not available for that study, but they did not find a significant difference in percentage signal change for the two conditions. In the present study, most of the pitch-related activity was centered on PT in accordance with Hall and Plack’s (2009) findings. Future investigations using a greater number of different pitch-evoking stimuli would be required to determine whether the results reported here are general to all pitch stimuli or are specific to certain types of pitch stimuli.

**Acknowledgments**

This study was supported by a PhD studentship awarded by the MRC Institute of Hearing Research and MR scanning was paid for through MRC infrastructure funding awarded to the same organization. The authors would also like to thank Deafness Research UK who provided a travel grant to present this study at the 2009 ARO MidWinter Meeting in Baltimore, MA. The authors thank the Editor and two anonymous reviewers for constructive comments on an earlier version of the manuscript.

**References**


Chapter 5. The cortical pitch response to jittered pulse trains and harmonic complex tones reveals no increase with increasing pitch salience

Daphne Barker"1,2, Christopher J. Plack2 and Deborah A. Hall3

1MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK
2Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK
3Department of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK

This chapter was prepared for publication and is presented in the format in which it was prepared.

Publication 3

* Daphne García will be married in early 2011 and the paper will be submitted in her married name, Daphne Barker.
The cortical pitch response to jittered pulse trains and harmonic complex tones reveals no increase with increasing pitch salience

Daphne Barker \textsuperscript{a,b}
Christopher J Plack \textsuperscript{b}
Deborah A Hall \textsuperscript{a,c}

\textsuperscript{a} MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD
\textsuperscript{b} Division of Human Communication and Deafness, The University of Manchester, Manchester, M13 9PL
\textsuperscript{c} Division of Psychology, Nottingham Trent University, Burton Street, Nottingham, NG1 4BU

Corresponding author: Daphne Barker
Email: d.garcia@ihr.mrc.ac.uk
Tel: +44 115 922 3431
Fax: +44 115 951 8503
Abstract

Neuroimaging evidence for a pitch processing region in auditory cortex tends to support one of two different conclusions: either that there is a pitch center in the lateral portion of Heschl’s gyrus, or that the pitch response is centered more posterior in planum temporale. It has been proposed that for a region to be considered pitch-specific, its response must also covary with the perceived strength (salience) of the pitch. Thus far, there is no compelling neuroimaging evidence for an effect of salience. The current experiment attempted to determine which region in human auditory cortex is most likely to be sensitive to pitch and whether or not this region also responds to pitch salience. Jittered pulse trains and unresolved complex tones, with harmonics in variable phase relations, were created with three levels of pitch salience. Results support a pitch-processing region in planum temporale, with no pitch response found in lateral Heschl’s gyrus. We failed to identify any regions that indicated an increased pitch response with increasing pitch salience; the implications of which are discussed.
Introduction

Pitch is a fundamental sound feature. It is one of the main cues by which we group sounds or segregate sounds from different sources. It is also an important feature of speech, as intonation is used in tonal languages to convey semantic information, and in non-tonal languages to convey prosodic information. Additionally, pitch is one of the primary features of Western music - it has been defined as the sensation whose variation is associated with musical melodies (Plack 2005). Pitch is a perceptual feature of sound and is related either to the harmonic spectral distribution of a tone, its temporal regularity or a mixture of the two.

Complex tones are made up of a number of spectral components called harmonics. The frequency of each harmonic is an integer multiple of the fundamental frequency (f0). Low-numbered harmonics, up to around the tenth (Plack, 2005) excite distinct places along the basilar membrane and are said to be ‘resolved’. Higher-numbered harmonics are not individually represented on the basilar membrane, with several harmonics falling within a single frequency channel, and these are said to be ‘unresolved’.

The pitch of a stimulus containing resolved harmonics can be represented by a ‘rate-place’ code, in which pitch is determined by individual peaks of neural activity along the tonotopic array, or by a temporal code, in which harmonic frequencies are represented by synchronized (phase-locked) neural activity (Pierce 1990; Langner 1997; Cedolin and Delgutte 2005). In either case, pitch can be derived by a combination of the activity in different channels. The excitation pattern of an unresolved harmonic complex tone does not contain any spectral peaks, but it still
evokes a pitch percept. This cannot be represented by the rate-place code, but can be explained by the temporal code. Each place on the basilar membrane responds to two or more interacting harmonics, and the resultant waveform repeats at regular intervals, the rate of which determines the pitch.

The physical features responsible for the creation of a pitch percept determine the pitch strength, or salience. Salience is related to the harmonic content of the stimulus. Stimuli in which low-numbered (resolved) harmonics are present produce the strongest salience. Tones containing only higher number harmonics create a weak pitch salience (Houtsma and Smurzynski 1990). Furthermore, for unresolved harmonics, the relative phases of the harmonics determine salience. Stimuli in which all the harmonics have the same phase (e.g. sine or cosine) contain ‘peaky’ envelopes (Figure 1). Stimuli in which the phases of the spectral components are entered in random phases contain less-peaky envelopes, which result in a weak pitch salience. “Schroeder phase” complexes also have flat envelopes, but the fine temporal structure within the envelope varies systematically rather than randomly (Figure 1). The salience of Schroeder-phase complexes is almost identical to sine-phase complexes (and thus would be the similar for cosine-phase complexes) when the lowest harmonic present is 10 or lower (Houtsma and Smurzynski 1990). Pitch discrimination thresholds between sine- and Schroeder-phase complexes begin to diverge when the lowest present harmonic is above 13. Schroeder-phase complexes with a lowest harmonic number above 13 produce a pitch percept that is less salient than equivalent constant-phase complexes, but more salient than equivalent random-phase complexes.
Regular pulse trains are harmonic complex tones with all harmonics in cosine phase to produce a maximally peaky waveform. However, it is possible to alter the regularity of the pulses, so that the inter-pulse interval varies between pulses. For example, each individual pulse of a 100-Hz pulse train with 50% random jitter could occur anywhere up to 2.5 ms either side of its isochronous location (Figure 1). The greater the amount of jitter, the lower the regularity of the pulse train, and the lower the pitch salience (Gutschalk et al. 2007).

![Waveform Diagram](image)

**Figure 1.** 100 ms samples of waveforms for all of the stimuli used in the experiment. The low signal-to-noise ratio in the random phase unresolved harmonic complex (top
panel) is apparent in the signal by the high degree of spectral splatter around each harmonic. The frequency sweeps in the Schroeder phase unresolved harmonic complex can be seen in the second panel. The third and fourth panels show the irregularity introduced by jittering the pulse trains. The bottom panel indicates that the high salience condition is the same for the cosine-phase unresolved harmonic complex and the regular pulse train. The stimuli are shown without a noise masker for clarity, but a low-pass noise masker was applied to all pitch stimuli in the experiment.

The pitch value of two sounds with the same f0 is the same regardless of how the sound is created, leading many researchers to postulate on the existence of a general ‘pitch center’ in the auditory cortex (Patterson et al. 2002; Krumbholz et al. 2003; Hall and Plack 2009; Puschmann et al. 2010). For a region to be labelled as such, it should ideally satisfy four criteria (Bendor and Wang 2005, 2006; Hall and Plack 2009). The first is that it would be selectively responsive to pitch, and not to any other sound feature (such as timbre). Second, its response would have to be determined by central mechanisms, meaning that its activation could not be attributable to peripheral effects such as cochlear distortions (McAlpine 2004). Third, it should be responsive to all pitch-evoking stimuli, regardless of their spectral, temporal or binaural characteristics. Finally, the activity in such a region should covary with pitch salience, such that a stronger pitch salience activates the region to a greater extent than does a stimulus with a weaker pitch salience.

Results from previous physiological and neuroimaging studies have been taken as evidence that a discrete region of auditory cortex could be the pitch center. In a study using single-unit extra-cellular recordings in marmoset monkeys, Bendor and Wang (2005) found a region of auditory cortex that satisfied all four criteria for a pitch center. This region was located in a low-frequency responsive field near the anterolateral border of primary auditory cortex. However, even without considering the anatomical differences between monkeys and humans, there is no way to know whether invasive
animal methods are identifying the same processes of neural computation as non-invasive methods that can be used safely and ethically in humans.

Patterson et al. (2002) performed a functional magnetic resonance (fMRI) study, whose results led them to suggest that an area directly antero-lateral to primary auditory cortex acts as a pitch center in humans. This study used a type of regular-interval pitch stimulus known as iterated ripple noise (IRN). IRN is created by generating a sample of Gaussian noise, delaying it, and adding or subtracting the delayed sample back to or from the original. The more times this iterative process is repeated, the more salient the pitch of the IRN becomes. The f0 of IRN is equal to the inverse of the delay imposed. When contrasted with noise, IRN elicited a response in lateral Heschl’s gyrus (lateral HG) that was consistent in eight of their nine listeners. It has been suggested that lateral HG is the human anatomical and physiological homologue of the pitch-sensitive region reported by Bendor and Wang (2005) in non-human primates (Hackett 2003; Bendor and Wang 2006).

Using a variety of methods to measure human auditory cortical responses to IRN, a number of human neuroimaging studies have come to the same conclusion; depth electrodes (Schönwiesner and Zatorre 2008), positron emission tomography (Griffiths et al. 1998) and magnetoencephalography (MEG) (Krumbholz et al. 2003; Hertrich et al. 2005; Seither-Preisler et al. 2006). It has recently been suggested, however, that the response elicited by IRN could be driven, at least in part, by slowly-varying spectro-temporal modulations inherent in the stimulus that are not related to pitch (de Cheveigné 2007; Hall and Plack 2009; García et al. 2010; Barker et al. in preparation). Few studies have looked for a pitch center using pitch-evoking stimuli other than IRN, and the results of those studies are inconsistent. Results from MEG
research using click trains (Gutschalk et al. 2002, 2004, 2007) and from fMRI research using harmonic complex tones (Penagos et al. 2004) concur with IRN studies in finding pitch activity consistent with a location in lateral HG. Other fMRI studies using a wide range of pitch-evoking stimuli, including spectral, temporal and dichotic pitches, have found only weak evidence for a pitch center, with the most consistent activity across listeners spanning a number of different primary and non-primary auditory regions (Hall et al. 2006; Hall and Plack 2007, 2009; García et al. 2010). These regions include central HG, lateral HG, and (most consistently) anterolateral planum temporale (PT).

The evidence for an effect of pitch salience in auditory cortex is even more sparse. A PET study that used IRN found that brain activity in lateral HG increased with number of iterations (and hence stronger pitch salience) (Griffiths et al. 1998). However, the inclusion of Gaussian noise (0 iterations) in the correlation undoubtedly influenced the result. It would seem from Figure 3 in the paper that if noise were excluded from the analysis, there would be very little effect of varying iterations. A recent paper shed some light on this issue by using a novel stimulus with the same slowly-varying modulations as IRN, but without the fine structure responsible for its pitch. Barker et al. (in preparation) compared IRN and this novel stimulus with 2, 4, 16 and 64 iterations to determine whether the slowly-varying modulations contribute to the pitch response seen previously for IRN and whether activation magnitude increases as a function of number of iterations. Results indicate that slowly-varying modulations contribute to the pitch-related response seen for IRN. Nevertheless a residual pitch-related response remained even when these modulations were controlled for. Additionally, Barker et al. reported an increase in auditory activity within central and
lateral HG, and within PT, with increasing number of iterations. However, crucially, this effect was seen for both stimulus types. Therefore, in this study, the increase in activity with number of iterations was attributed to the increasing depth of the slowly-varying fluctuations rather than to pitch salience.

An fMRI study used harmonic-complex tones with different f0s filtered into low and high spectral regions to create tones with resolved and unresolved harmonics that elicit strong and weak pitch percepts (Penagos et al. 2004). This acoustic manipulation avoids the covariation in slowly-varying spectro-temporal features as a function of pitch salience seen in IRN. The paper also reported an effect of salience in lateral auditory regions, but the conclusion was based on the results of only five listeners, and the analysis used a very lenient correction (least significant difference) that is susceptible to type I errors (Bullmore et al. 1996).

Some MEG studies have found that the amplitude of auditory evoked field components thought to be involved in pitch perception – the N100m and the pitch onset response (POR) – increase as a function of pitch salience (Seither-Preisler et al. 2003; Krumbholz et al. 2003). Amplitude is only one measure of an evoked potential, the latency of the response peak is also an informative response measure. For example, in the visual system, more salient stimuli require less integration time, which manifests as shorter latencies (Sorrentino et al. 2006). No changes in latency were observed for pitches with differing salience in any of the studies, and in one study, there was no effect of salience on latency or amplitude (Seither-Preisler et al. 2006b). A recent fMRI study that used a number of pitch stimuli varying in physical characteristics and salience found no association between pitch salience and magnitude of activation (Hall and Plack 2009).
In the present experiment we investigated the response to pitch and to pitch salience using fMRI. We measured responses for two different pitch-evoking stimuli with three levels of salience and we avoided the previous limitation of including the Gaussian noise in the salience analysis. In summary, our research questions were:

I. Is the pitch response (pitch vs. noise) in auditory cortex located in lateral HG, PT, or another region?

II. Is the auditory cortex responsive to pitch salience?

III. Are the responses to pitch and pitch salience co-located?

Materials and Methods

Listeners
Sixteen listeners (11 male, 5 female; age range 20 - 47 years) with normal hearing ($\leq$20 dB hearing level between 250 Hz and 8 kHz) took part in both the psychophysical and fMRI testing. All listeners were right-handed (laterality index = 50, Oldfield 1971). Seven listeners were musically trained between grade 2 and grade 7 (# 02, 07, 18, 19, 22, 23 and 25) while five others reported informal musical experience (self-taught/ungraded, # 05, 09, 16, 17, 21). None had a history of any neurological or hearing impairment. Listeners gave written informed consent and the study was approved by the Medical School Research Ethics Committee, University of Nottingham. One of the listeners (#16) was excluded because their scan-to-scan movement during the fMRI experiment was greater than 3 mm and 3°.
Conditions
The current study employed two types of pitch-evoking stimuli (pulse trains and unresolved harmonic complex tones). Each pitch stimulus had three levels of salience: low, medium and high. The high-salience condition was the same stimulus for the two pitch types. The non-pitch control condition was a Gaussian noise and a silent condition was included as a baseline.

Stimuli
All stimuli evoked a pitch corresponding to a 100-Hz tone. One stimulus type was a pulse train. The other stimulus type was an unresolved harmonic complex tone. The level of each harmonic was 20 dB greater than the spectrum level of the control noise so that the gross spectral density of all the stimuli was the same. The individual pulses in the pulse train were either regular at 10-ms intervals (regular pulse train, high salience) randomly jittered (uniform distribution) within 2.5-ms either side of the 10-ms interval (50% jittered pulse train, medium salience) or randomly jittered within 5-ms either side of the regular interval (100% jittered pulse train, low salience). Our limit of 5-ms is the maximum range of jitter without successive pulses overlapping. Jittered pulse trains provide a well-controlled way of varying salience because all of the stimuli have the same energy and average pulse rate as the regular pulse train. The high-salience (cosine phase) unresolved harmonic complex was the same condition as regular pulse train. For the medium salience unresolved harmonic complex, each harmonic was given a phase based on the following equation:

\[ \theta_n = \pi n(n+1)/N \]
Where $\theta_n$ is the phase of the $n$th harmonic, $n$ is the $n$th harmonic and $N$ is the total number of harmonics in the complex (Schroeder-phase unresolved harmonic complex) (Drennan et al. 2008). The phase of each component in the low salience (random-phase) unresolved harmonic complex was random for each presentation. Both stimuli were filtered between 1.5 and 2.5 kHz to remove low-numbered harmonics that can be resolved by the peripheral auditory system and thereby provide tonotopic features that could drive a differential response compared to noise that is unrelated to pitch. A bandwidth of 1.5-2.5 kHz was chosen so that the cosine-phase harmonic complex was substantially more salient than the Schroeder-phase harmonic complex (Houtsma and Smurzynski 1990). All stimuli included a noise masker, low-pass filtered at 1.5 kHz with a spectrum level of 51 dB SPL in the scanner and 45 dB SPL for psychophysical testing to mask cochlear distortion products. The control stimulus was a Gaussian noise (51 dB SPL spectrum level), low-pass filtered at 2.5 kHz. The three signals (pulse trains, unresolved harmonic complexes and Gaussian noise) were matched in bandwidth (0-2.5 kHz) and spectral density, and hence in overall power (85 dB SPL in the scanner and 79 dB SPL for psychophysical testing).

For the psychophysical testing, each pitch stimulus was 200 ms in duration (including 10-ms raised-cosine onset and offset ramps) and the inter-stimulus interval was 500 ms. Reference stimuli had a nominal f0 of 100 Hz.

The energy onset response is an effect that dominates sound responses in the auditory cortex, so that sensitivity to pitch is compromised (Krumbholz et al. 2003; Seither-Preisler et al. 2004; García et al. 2010). ‘Continuous stimulation’ is a valuable
paradigm in which pitch stimuli are presented in the context of noise in order to reduce the energy onset response and to increase pitch sensitivity (García et al. 2010). In the MR scanner, stimulus conditions each comprised a 14.75-s alternating sequence of 450-ms experimental sounds (including 10-ms linear-intensity onset and offset ramps) each separated by 250 ms Gaussian noise. The first and last components of each sequence were Gaussian noise. Each pitch and noise signal was generated using 10 ms linear-intensity onset and offset ramps, which were overlapped at the 3 dB points to produce a stable envelope for the stimulus. Sixteen sample sequences were created for each condition. A different set of stimuli was generated for each participant.

**Psychophysical Testing**

Prior to the scanning session, each participant performed a pitch-discrimination task to obtain estimates of pitch salience. Psychophysical testing was carried out in a sound-attenuating booth. Stimuli were delivered through Sennheiser HD 480 II headphones at a level of 79 dB driven by custom-made software supported by the Matlab platform (The MathWorks, Natick, MA). Pitch discrimination thresholds were measured using a three alternative forced-choice, two-down, one-up, adaptive procedure that targeted 70.7% performance (Levitt 1971). On the first trial, the f0 difference was 20% (20 Hz). The percent difference increased or decreased by a factor of two for the first four reversals, and by a factor of 1.414 for the final 12 reversals. Discrimination threshold was taken as the geometric mean of the f0 difference at the final 12 reversals. The adaptive track was not allowed to increase above 200% (300 Hz). Responses were recorded and stored electronically. On each trial, feedback was given via a green (correct) or red (incorrect) light on the software
interface. There were five testing runs for each of the different stimuli; pitch
discrimination thresholds were taken as the geometric mean threshold of each of the
last four runs.

**fMRI Protocol**

Scanning was performed on a Philips 3 T Intera Achieva using an 8-channel SENSE
receiver head coil. A $T1$-weighted high-resolution (1mm$^3$) anatomical image ($\text{matrix size} = 256 \times 256$, 160 saggital slices, $TR = 7.8 \text{ ms}$, $TE = 3.7 \text{ ms}$) was collected for each
subject. The anatomical scan was used to position the functional scan centrally on
HG, and care was taken to include the entire superior temporal gyrus and to exclude
the eyes. Functional scanning used a $T2^*$-weighted echo-planar sequence with a
voxel size of 3mm$^3$ ($\text{matrix size} = 64 \times 64$, 32 oblique-axial slices, $TE = 36 \text{ ms}$). Sparse
imaging with a $TR$ of 7800 ms and a clustered acquisition time of 1969 ms was used
(Edmister et al. 1999; Hall et al. 1999). A SENSE factor of 2 was applied to reduce
image distortions and a SofTone factor of 2 was used to reduce the background
scanner noise level by 9 dB. Functional data was acquired over four runs of 58 scans
each. Each sound condition had a total of 32 scans, with 40 scans for the silent
baseline. Listeners were requested to listen to the sounds, but were not required to
perform any task. A custom-built MR compatible system delivered distortion-free
sound using high-quality electrostatic headphones (Sennheiser HE60 with high-
voltage amplifier HEV70) that had been specifically modified for use during fMRI.
Stimuli were presented at an overall level of 85 dB SPL, calibrated using a KEMAR
manikin (Burkhard and Sachs 1975) fitted with Brueel and Kjaer half-inch microphone
type 4134 (serial no. 906663), Zwislocki occluded ear simulator (Knowles model no.
DB-100) and Brueel and Kjaer measuring amplifier type 2636 (serial no. 1324093),
scaled from 22.4-Hz to 22.4 kHz using fast time constant (125 ms) on maximum hold. Due to the metallic nature of components in the KEMAR system, calibration inside the MR scanner was not possible.

**Data Analysis**

Images were analyzed separately for each listener using statistical parametric mapping (SPM5, [http://www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)). Preprocessing steps included realignment to correct for subject motion, normalization of individual scans to a standard image template, and smoothing with a Gaussian filter of 8 mm full width at half maximum (FWHM) of the Gaussian smoothing kernel. The realignment process generated estimates of the scan-to-scan movement for three translations (x, y and z planes) and three rotations (roll, pitch and yaw). These were included as variables in the individual design specification in addition to the stimulus conditions and the scanning runs. The silent baseline was implicitly modeled in the design. The first-level general linear model assessed the variables of interest with respect to the scan-to-scan variability. A high-pass filter cutoff of 420 s was used to remove low frequency confounds. The resulting model estimated the fit of the design matrix (X) to the data (Y) in each voxel in order to provide β values (the contribution of a single regressor to the overall fMRI signal). Separate statistical contrasts for each sound condition were specified relative to the silent baseline. To investigate the differential responses across conditions, a one-way ANOVA was specified at the second level using the preceding contrast images for each individual. Details of the specific contrasts performed and statistical thresholds applied are reported in the Results section below.
Estimates of peak localization within HG were made with reference to three cytoarchitectonic subdivisions; Te 1.2 (lateral HG), Te 1.0 (central HG) and Te 1.1 (medial HG) (Morosan et al. 2001; Eickhoff et al. 2005). Data extraction for the region-of-interest analysis used the approach described by Hall and Plack (2009).

Results

Behavioural Group Results: Degree of Regularity and Harmonic Phase Affect Pitch Discrimination Thresholds

Repeated-measures ANOVA were performed separately for each stimulus type (pulse trains and unresolved harmonic complexes) with presumed salience (low, medium and high) and run (1, 2, 3 and 4) as factors. There was a significant effect of presumed salience for both pulse trains and unresolved harmonic complexes [F(2,28) = 67.38, p < 0.001 and F(2, 28) = 3.801, p < 0.05, respectively], with no significant effect of run. Both stimulus types exhibited a positive linear relationship between presumed pitch salience and discrimination threshold [F(1,14) = 87.82, p < 0.05 and F(1, 14) = 6.38, p < 0.05 for pulse trains and unresolved harmonic complexes, respectively] (Figure 2).

Figure 2. Pitch discrimination thresholds for pulse train (grey line) and unresolved harmonic complex (black line) stimuli with increasing pitch salience. Error bars represent 95% confidence limits.
fMRI Group Results: Pitch-Related Activity

All five pitch conditions were combined for comparison with the Gaussian noise. From the one-way ANOVA in SPM5, a t-test to compare responses evoked by pitch and those evoked by Gaussian noise revealed bilateral clusters of pitch-related activity in PT. The right hemisphere cluster contained 38 voxels with a peak at x 64, y -22, z 4 mm \( [t(1,84) = 4.00, p < 0.05, \text{uncorrected}] \) and the left hemisphere cluster contained 14 voxels with a peak at x -62, y -28, z 8 mm \( [t(1,84) = 3.74, p < 0.05] \). There was also a cluster of 3 voxels in right planum polare (PP) with a peak at x 48 y -10 z -4 mm \( [t(1,84) = 3.36, p < 0.001, \text{uncorrected}] \). Voxels were significant at an uncorrected level of \( p < 0.001 \) but none survived false discovery rate (FDR) correction for multiple comparisons within the volume of the superior temporal gyrus \( (p > 0.05) \).

This finding prompted us to explore the distribution of pitch-related activity across the group and to determine the maximum percentage of listeners showing pitch-related activity in the same voxel. To do this, we generated an incidence map by summing the individual statistical maps \( (p < 0.01) \), each with an extent threshold of 50 voxels to rule out contributions from individual voxels that simply reflect residual noise. The resulting summed activation map was overlaid onto the group-averaged normalized anatomical scan (Figure 2). The incidence map revealed extensive activation along the superior temporal gyrus, predominantly located posterior to HG with a second smaller focal region in anterior auditory cortex (PP). The main peaks identified in the statistical analysis were encompassed within the incidence map. On the right, the maximum incidence was eight listeners \( (53\%) \) in anterolateral PT \( (x 64, y -22, z 4 \text{ mm}) \) and eight listeners in PP \( (x 55, y -2, z -4 \text{ mm}) \). On the left, the maximum
incidence was seven listeners (47%) in anteromedial PT (x -46, y -34, z 10 mm) and eight listeners in PP (x -54, y -6, z -4 mm).

Figure 3. Distribution of pitch-related activation across horizontal (axial) and vertical (saggital) sections of auditory cortex, shown as an incidence map of activation across the 15 listeners. The colour scale represents the percentage of pitch-related activation at every voxel and is calculated as a possible maximum of 15. The activation is overlaid onto the average anatomical image from all listeners. The left hemisphere is on the left-hand side of the image.

To seek supportive evidence for a response to pitch within a specific region of auditory cortex, we performed a region-of-interest (ROI) analysis. This has the benefit of considering the overall response within a region, rather than at the level of an individual voxel. The regions considered for further analysis were the three subdivisions of HG (Te 1.0, Te 1.1 and Te 1.2), PT and PP. In each region, a paired t-test was conducted on the β values to compare the pitch and noise responses. There was a significant effect of pitch in Te 1.0 [t(14) = -2.44, p < 0.05] and in PT [t(14) = -3.43, p < 0.05]. In both cases, the direction was for a greater response to the pitch conditions than to the noise.

fMRI Group Results: Salience-Related Activity

From the one-way ANOVA in SPM5, separate F-tests were performed for the pulse trains and unresolved harmonic complexes, on the three levels of salience. Neither of
these F-tests revealed a significant effect of salience within auditory cortex (p > 0.05, uncorrected).

ROI analyses were also conducted for the five regions investigated for pitch-related activity. For each region, separate one-way ANOVA with three levels of salience (low, medium, and high) were conducted for the two stimulus types (unresolved harmonic complexes and pulse trains). Of the 10 ANOVA, an effect of salience was found in two regions. For the unresolved harmonic complex stimuli, there was a main effect of salience in Te1.2 [F(2,26) = 3.91, p < 0.05] with a significant linear trend [F(1,13) = 7.47, p < 0.05]. For the pulse train stimuli, there was a main effect of salience in Te1.1 [F(2,26) = 5.80, p < 0.05], with a significant linear trend [F(1,13) = 7.45, p < 0.05]. Both these contrasts survived Bonferroni correction. However, in both cases the linear trend was for a decrease in the magnitude of the response as salience increased. Further support for a negative relationship in Te1.1 between pitch salience and activation values was provided from a partial correlation analysis. For this analysis, the β values for each condition were averaged across run and hemisphere to provide one value for each of the five pitch conditions per listener in each auditory region. The βs from each region were compared to pitch discrimination thresholds separately, and for each comparison the effects of listener were partialled out. The only significant result was a positive correlation in Te1.1 [r(72) = 0.33, p < 0.05]. Although the correlation is significant, the 95% confidence intervals are extremely broad (the population correlation is estimated to be between -0.03 and 0.71) indicating a large variability between individuals.
Discussion

The present study attempted to determine first, whether there is a response to pitch in lateral HG, in PT, or in any other region of auditory cortex. We also aimed to find regions of auditory cortex that were sensitive to pitch salience, and then to confirm whether or not the pitch and pitch salience effects were co-located. Although we identified regions that responded to pitch, we failed to identify any region of auditory cortex that showed increased activation with increasing pitch salience.

The Most Likely Site for a Pitch Center is Planum Temporale

The pitch response that was revealed when all pitch conditions for each stimulus type were combined and contrasted with noise provides additional evidence that the most likely site for a pitch-specific region in human auditory cortex is in PT (Hall and Plack 2009; García et al. 2010), with no such response evident in lateral HG. Although incidence maps indicated a focal region of pitch-related activity in PP, activity in this region did not appear in the group-averaged result, nor was it significant in the ROI analysis. PT was the only region that revealed a reliable sensitivity to pitch across all three methods of analysis. This result concurs with previous findings from the same authors (Hall and Plack 2009; García et al. 2010), but the interpretation of the pitch-responsive region differs from other studies (Patterson et al. 2002; Krumholz et al. 2003; Hertrich et al. 2005; Seither-Preisler et al. 2006; Schönwiesner and Zatorre 2008; Puschmann et al. 2010).

Pitch-Responsive Regions Are Not Sensitive to Pitch Salience

Although we found PT to be the most likely candidate for a pitch center, it did not satisfy the criterion of a systematic increase in activation with increasing pitch.
salience. We propose a number of possible reasons why we did not find this effect. The first of these is that the pitch response simply reflects a mechanism for detecting pitch and is not sensitive to salience. Within their pitch region, Bendor and Wang (2005) found 51 neurons that responded significantly to increases in pitch salience measured using IRN with increasing iterations, click trains with decreasing jitter and harmonic complex tones with a decreasing number of the lowest harmonic present. There are, however, challenges in using neurophysiological findings to predict what we should find using fMRI. There are millions of neurons within each voxel in the brain, whose collective impact on the blood oxygenation level determines the fMRI response. Considering that our statistical threshold for defining significant activity required a contiguous set of 50 voxels within individual maps, a sparse neural code for pitch salience would remain undetectable by our methods. The second plausible reason is that the response that is attributed to pitch is actually sensitive to factors related to the segregation of the pitch sound from the concurrent background noise (i.e. the noise masker), and not to pitch per se. The suggestion that PT plays a role in sound segregation is not a new one. Evidence for the role of PT in the segregation of auditory objects has previously been suggested (Griffiths and Warren 2002). The third possibility is that, within pitch responsive regions, the effects of salience are smaller than the overall effect of pitch and that the present study was just not sufficiently sensitive.

Conclusion

While our results confirm pitch-selectivity in the anterior part of PT, we failed to find any evidence that pitch-responsive regions show increased activation with increasing pitch salience. One interpretation of this pattern is that pitch-sensitive regions are
responsive to the presence or absence of pitch and not to other attributes of pitch.
This result is inconsistent with the fourth criterion for a pitch center suggested by Hall and Plack (2009).

Acknowledgements
This study was supported by a PhD studentship awarded by the MRC Institute of Hearing Research and MR scanning was paid for through MRC infrastructure funding awarded to the same organization. The authors would like to thank Simon Müller for his contribution to the analysis of fMRI data.

References


Chapter 6. Reexamining the evidence for a pitch sensitive region based on a human fMRI study using iterated ripple noise

Daphne Barker*1,2, Christopher J. Plack2 and Deborah A. Hall3

1MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK
2Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK
3Department of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK

This chapter was prepared for publication and is presented in the format in which it was prepared.

Publication 4

* Daphne García will be married in early 2011 and the paper will be submitted in her married name, Daphne Barker.
Reexamining the evidence for a pitch-sensitive region based on a human fMRI study using iterated ripple noise

Daphne Barker a,b
Christopher J Plack b
Deborah A Hall a,c

a MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD
b Division of Human Communication and Deafness, The University of Manchester, Manchester, M13 9PL
c Division of Psychology, Nottingham Trent University, Burton Street, Nottingham, NG1 4BU

Corresponding author: Daphne Barker
Email: d.garcia@ihr.mrc.ac.uk
Tel: +44 115 922 3431
Fax: +44 115 951 8503
Abstract

Human neuroimaging studies have identified a region of auditory cortex, lateral Heschl’s gyrus, that shows a greater response to iterated ripple noise (IRN) than to a Gaussian noise control. Based in part on results using this pitch-evoking stimulus, it has been argued that lateral Heschl’s gyrus is a general ‘pitch center’. However, IRN contains slowly varying spectro-temporal modulations, unrelated to pitch, that are not found in the control stimulus. Hence it is possible that the cortical response to IRN is driven in part by these modulations. The current study reports the first attempt to control for these modulations. This was achieved using a novel type of stimulus that was generated by processing IRN to remove the temporal regularities (and thus the pitch), but leave the slowly varying modulations. This ‘no-pitch IRN’ stimulus is referred to as IRNo. Results showed a widespread response to the spectro-temporal modulations across auditory cortex. When IRN was contrasted with IRNo rather than with Gaussian noise, the effect was no longer statistically significant. Our findings raise the possibility that the response to sound features unrelated to pitch could previously have been erroneously attributed to pitch coding.
Introduction

Pitch is one of the primary auditory percepts and so has been of interest to psychoacousticians, neuroscientists, psychologists and linguists. Despite a large body of research examining the neural correlates of pitch perception, debate continues as to whether there exists an area of auditory cortex that is responsible specifically for representing the percept of pitch rather than the physical attributes responsible for its creation. The fact that the same pitch can be elicited by sounds with different spectral and temporal characteristics has led many researchers to postulate the existence of neurons selectively responsive to pitch. It has been suggested that for a region to be considered a pitch center, it must satisfy four criteria; 1) It must respond selectively to pitch compared to an appropriately matched noise. 2) It must still be present after the elimination of peripheral effects such as cochlear distortions. 3) It must respond to all pitch-evoking stimuli, regardless of physical attributes. 4) It must show an increase in activation with increasing pitch salience (Hall and Plack 2009). A landmark primate study used single-unit extra-cellular recordings in the vicinity of primary auditory cortex to find such a region (Bendor and Wang 2005). This study identified a cluster of neurons in the antero-lateral border of the primary auditory cortex that met all four criteria for a pitch center. However, there are a number of problems involved in translating such results into the domain of human cognitive neuroscience. For example, Bendor and Wang (2005) were recording spiking activity from a population of 131 individual units, 51 of which exhibited a significant pitch response. In contrast, fMRI detects changes in blood oxygenation levels that occur as an indirect consequence of population neural activity, not all of which needs to be suprathreshold. Furthermore, BOLD activation is thought to represent local field potentials rather than spiking activity measured in
single- and multi-unit recording (Logothetis et al. 2001). Hence it is unclear that the same effects should necessarily be observed using the two different methods even if both species possess pitch-sensitivity at the neuronal level.

A popular method for isolating pitch-specific responses is to contrast pitch-evoking stimuli with control stimuli that are matched in terms of spectral content but do not evoke a pitch percept. One contrast of this type is that of an iterated ripple noise (IRN) pitch stimulus and a Gaussian noise control (Griffiths et al. 1998; Patterson et al. 2002; Krumbholz et al. 2003). IRN is created by generating a sample of Gaussian noise, imposing a delay to the noise, and adding (or subtracting) the delayed version back to (or from) the original (Yost 1996). An IRN signal can be high-pass or band-pass filtered so that it contains no perceptually resolvable spectral peaks at harmonic frequencies. Instead, the pitch percept is determined by fast rate temporal regularities in the stimulus. The pitch sensation of IRN is related to the reciprocal of the imposed delay. One appeal of IRN is that its salience can be easily manipulated by changing the number of delay-and-add (or subtract) iterations. Increasing the number of iterations increases the salience of the pitch (Yost 1996). Pitch salience has been expressed in terms of the height of the first peak in the autocorrelation function, which increases with increasing iterations, and correlates well with the perceived pitch strength (Yost 1996).

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

At present, the evidence for lateral HG as a pitch center is somewhat mixed (Penagos et al. 2004; Chait et al. 2006; Hall and Plack 2007, 2009; García et al. 2010; Puschmann et al. 2010). When results with a wide range of pitch-evoking stimuli are more closely scrutinized, most appear to indicate that pitch-related activity also engages regions of auditory cortex surrounding lateral HG. For example, Hall and Plack (2009) found that planum temporale (PT) was typically responsive to many different pitch-evoking stimuli, including tone-in-noise, wideband harmonic complex and Huggins pitch. In contrast, lateral HG was found to respond no differently to these stimuli than to the spectrally matched noise control. However, consistent with the earlier results, lateral HG did respond significantly to two types of IRN stimulus compared to the corresponding spectrally matched noise control. Although the IRN-related response was highly consistent across listeners (> 50%), the activation in PT produced by the other pitch-evoking stimuli was less so (< 25%). The lack of
consistency led the authors to conclude that no one region could reliably be assigned the label of ‘pitch center’.

To explain the discrepancy in the spatial distribution and consistency of activity for the different pitch contrasts, Hall and Plack (2009) demonstrated that IRN contains acoustic features unrelated to pitch that are not present in the other pitch-evoking stimuli nor in the noise control. The iterative delay-and-add process introduces unpredictable spectro-temporal variations that occur over a longer time scale (hundreds of milliseconds) than the temporal regularity responsible for pitch (tens of milliseconds). Increasing the number of delay-and-add iterations in the IRN not only increases pitch salience, but also increases the depth of the modulations across time and frequency, hence increasing the perceptual salience of those modulations. Earlier, de Cheveigné (2007) had argued that the ‘spectral ripple’ in IRN could set it apart from other pitch-evoking stimuli, and that these additional features could explain the disparity in results from studies using IRN and those using different pitch-evoking stimuli. Previous research provides support for Hall and Plack’s (2009) suggestion that the modulations, rather than the pitch of IRN, might be responsible for the robust activation obtained (Schönwiesner and Zatorre 2009). The fMRI study indicated strong selectivity to specific properties of dynamic spectral ripples on HG and around Heschl’s sulcus.

To quantify the contribution of the spectro-temporal fluctuations to activity within auditory cortex and to the IRN-related response, the current study uses a novel type of stimulus that preserves the slowly-varying spectro-temporal modulations of IRN, but removes the fine temporal structure responsible for the pitch percept. We have
called this new stimulus ‘no-pitch IRN’ (IRNo). If the response to IRNo in lateral HG is
the same as it is to IRN, this would suggest that the region is responsive to the
longer-term spectral fluctuations, not to the pitch, forcing a reinterpretation of previous
results. This provided the main motivation for the current experiment. This hypothesis
was examined by manipulating the number of delay-and-add iterations (2, 4, 16 and
64) of the IRN and the IRNo signals. By comparing each stimulus type to a spectrally
matched control noise we were able to identify auditory regions that responded
differently to the two classes of stimulus. By assessing the linear and quadratic
components of the response function across a number of delay-and-add iterations,
we were able to evaluate the degree to which the response is driven by the slowly
varying spectro-temporal modulations alone or in combination with the pitch features.

Materials and Methods

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

**Stimuli**

Diotic IRN stimuli were generated by a delay-and-add process performed on a Gaussian noise. The noise was bandpass-filtered (1–2 kHz) to remove low-numbered harmonics that are resolved (i.e. separated out) by the peripheral auditory system. A delay of 10 ms was imposed before adding the delayed noise back to the original sample. This process was repeated 2, 4, 16 or 64 times to make all four IRN stimulus conditions, each with a pitch corresponding to a 100-Hz tone. To make IRNo, a conventional IRN stimulus was generated as above. The IRN was sampled using a rectangular window with a 10-ms duration. A fast Fourier transform (FFT) was used to generate the magnitude and phase spectra of the sample, and the phase of the components was randomized. An inverse FFT was then used to regenerate the time representation. The sampling window was advanced by half of the IRN delay (5 ms) and the process repeated. The processed samples were overlapped and added (preserving the start-times of the samples), adjusted to a spectrum level of 52 dB SPL and gated in time with onset and offset ramps as described below. The phase randomization process removes any correlation in the fine structure between samples, obliterating the harmonic structure and the pitch cue. However, the slowly varying broad spectral features are preserved. These fluctuations are apparent when the spectrogram of IRN is smoothed in both time and frequency domains to remove any fine structure (Figure 1). The process was repeated on all the IRN stimuli to produce four IRNo conditions. All stimuli included a noise masker, low-pass filtered at 1 kHz and with a spectrum level of 52 dB SPL, to mask cochlear distortion products.
Figure 1. Simulated cochlear representations of IRN (top row) and IRNo (bottom row) in the form of spectrograms. The analysis smooths the representation in both time and frequency domains to remove any fine structure. The bottom row shows processed version of the IRN stimuli in the top row. The color bar shows model output in dB.

There were two noise controls for this study. The first was a Gaussian noise, low-pass filtered at 2 kHz. The second was identical to the first, but the 1-2 kHz region was processed in the same way as for the IRNo stimuli. All sounds (IRN, IRNo, noise and processed noise) were matched in bandwidth (0-2 kHz) and spectral density (and hence overall energy).

For testing the pitch discrimination thresholds for IRN, each stimulus was 200 ms in duration (including 10-ms raised-cosine onset and offset ramps) and the inter-stimulus interval was 500 ms. Reference stimuli had a fundamental frequency (f0) of 100 Hz. For testing modulation discrimination performance for IRNo, each stimulus was 600 ms in duration (including 10-ms linear-intensity onset and offset ramps) and the inter-stimulus interval was 500 ms. Stimuli were presented at an overall level of 85 dB SPL, calibrated using a KEMAR manikin (Burkhard and Sachs 1975) fitted with Bruel and Kjaer half-inch microphone type 4134 (serial no. 906663), Zwislocki occluded ear simulator (Knowles model no. DB-100) and Bruel and Kjaer measuring
amplifier type 2636 (serial no. 1324093), scaled from 22.4-Hz to 22.4 kHz using fast
time constant (125 ms) on maximum hold. Due to the metallic nature of components
in the KEMAR system, calibration inside the scanner was not possible.

In the scanner, stimulus conditions each comprised a 14.25-s alternating sequence of
600-ms experimental sounds (including 10-ms linear-intensity onset and offset ramps)
each separated by 50 ms silence. Sixteen sample sequences were created for each
condition and a different set of stimuli was generated for each participant.

**Cochlear Representations**

To illustrate the representation of the stimuli in the peripheral auditory system, the
stimuli were passed through a computational model (Plack et al. 2002). The model
included a simulation of the middle ear and a non-linear auditory filterbank that
simulated the compressive frequency selective properties of the basilar membrane in
the cochlea. The temporal response of the filterbank was smoothed by a sliding
temporal integrator. The parameters of this version of the model were taken from
Plack (2007). The spectrograms in Figure 1 show the output of the model as a
function of time and filter center frequency for examples of the IRN and IRNo stimuli
used in the experiment. For the purpose of illustration, the IRNo stimuli shown in the
bottom row are processed versions of the IRN stimuli shown in the top row. Because
the bandwidth of the auditory filters is greater than the spacing between the
harmonics in the IRN, the harmonic frequencies do not appear as horizontal lines in
the plots (in other words, the harmonics are unresolved by the cochlea). Instead, the
model reveals the broad spectro-temporal fluctuations that increase in depth as the
number of iterations is increased. For the same number of iterations, the model
output appears similar for IRN and IRNo stimuli, indicating that the processing used to
generate the IRNo was successful in preserving the spectro-temporal features.

![Figure 2](image_url)

**Figure 2.** The standard deviation of the cochlear representations of IRN and IRNo as
a function of number of iterations, averaged over 50 replications. The values are
measures of the fluctuation depth of the slowly varying modulations. The error bars
show 95% confidence limits.

To provide a quantitative measure of the modulation, for each spectrogram the
standard deviation of the level fluctuations (in dB) was calculated across the whole
response pattern for center frequencies between 1 and 2 kHz. The calculation was
performed 50 times for each condition, using different samples of IRN and IRNo for
each repetition, and the mean of the standard deviation and 95% confidence intervals
calculated. The results are shown in Figure 2. Fluctuation depth increases with
number of iterations. The IRN and IRNo stimuli are quite closely matched. The
fluctuation depth for the IRN stimuli is slightly greater than that for the IRNo stimuli at
16 and 64 iterations. The fluctuation depth for the processed noise control is slightly
greater than that for the unprocessed noise control. These slight differences in
modulation between IRN and IRNo were not expected to influence the fMRI results
greatly.
Psychophysical Testing

Prior to the scanning session, each participant performed a pitch-discrimination task and a modulation-discrimination task to measure the perceptual salience of the pitch and modulation cues. Psychophysical testing was carried out in a sound-attenuating booth and stimuli were delivered through Sennheiser HD 480 II headphones. Stimuli were presented through custom-made software that is supported by the Matlab platform (The MathWorks, Natick, MA). Pitch discrimination thresholds were measured for IRN using a three alternative forced-choice, two-down, one-up, adaptive procedure that targeted 70.7% performance (Levitt 1971). On the first trial, the f0 difference was 20% (20 Hz). The percent difference increased or decreased by a factor of two for the first four reversals, and by a factor of 1.414 for the final 12 reversals. Discrimination threshold for each run was taken as the geometric mean of the f0 difference at the final 12 reversals. The percent difference was not allowed to increase above 200% (200 Hz).

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

**fMRI Protocol**

Scanning was performed on a Philips 3 T Intera Acheiva using an 8-channel SENSE receiver head coil. A $T1$-weighted high-resolution (1mm$^3$) anatomical image (matrix size = 256x256, 160 sagittal slices, $TR = 7.8$ ms, $TE = 3.7$ ms) was collected for each subject. The anatomical scan was used to position the functional scan centrally on HG, and care was taken to include the entire superior temporal gyrus and to exclude the eyes. Functional scanning used a $T2^*$-weighted echo-planar sequence with a voxel size of 3mm$^3$ (matrix size = 64x64, 32 oblique-axial slices, $TE = 36$ ms). Sparse imaging with a $TR$ of 7800 ms and a clustered acquisition time of 1969 ms was used (Edmister et al. 1999; Hall et al. 1999). A SENSE factor of 2 was applied to reduce image distortions and a SofTone factor of 2 was used to reduce the background scanner noise level by 9 dB. Functional data was acquired over three runs of 84 scans each and one run of 86 scans. Listeners were requested to listen to the sounds, but were not required to perform any task. A custom-built MR compatible system delivered distortion-free sound using high-quality electrostatic headphones (Sennheiser HE60 with high-voltage amplifier HEV70) that had been specifically modified for use during fMRI. An active noise control (ANC) device (Hall et al. 2009) was used for the first seven sessions (#02, 05, 07, 09, 16, 17, 18), reducing the acoustical scanner noise by a further 35 dB at the main peak in the spectrum of the scanner noise (around 14 dB overall). For these listeners, eight scans were appended to the beginning of the sequence in order to train the noise canceller. The ANC was not operative using subsequent sessions and so could not be used. We do not expect
ANC to change the pattern of results, but effects of IRN and IRNo were examined separately for the listeners who used the ANC, and those who did not. Activation results for those experiencing ANC and those not experiencing ANC are reported in the Results section where appropriate.

Data Analysis

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

First, the data for individual participants was analyzed using a first-level general linear model to assess the effects of interest with respect to the scan-to-scan variability. The resulting model estimated the fit of the design matrix (X) to the data (Y) in each voxel in order to provide parameter estimates (β) which represent the contribution of a single regressor to the overall fMRI signal. In order to obtain contrast images at the first level, the analysis fit the general linear model to each individual voxel in the functional image, and computed individual t statistics. At the first level, we specified separate statistical contrasts for each sound condition relative to the silent baseline that was implicitly modeled in the design. A high-pass filter cutoff of 420 s was used to remove low frequency confounds. A second level (random effects) analysis with
contrast images for the two control conditions with zero delay-and-add iterations (i.e. the noise and processed noise) confirmed that they elicited an equivalent brain response across auditory cortex. This fMRI result is also consistent with the observation from Figure 2 that the two control signals were similar in terms of their fluctuation depth, and so the two conditions were combined for subsequent analyses to increase statistical power. The inputs for the second level random effects analysis were therefore the contrast images for each IRN and IRNo stimulus compared to the combined noise controls. A 2x4 repeated measures ANOVA was created in SPM5, with stimulus type (IRN and IRNo) and number of iterations (2, 4, 16 and 64) as factors. Simple main effects and interactions were calculated using contrast weights (Friston et al. 2005). Typically, results are reported after small volume correction (SVC) to control for type I errors using a false discovery rate (FDR) threshold of p < 0.05 (Genovese et al. 2002). The small volume defined the auditory cortex across the superior temporal gyrus (including HG, PT and planum polare) and contained 4719 voxels in the left hemisphere and 5983 voxels in the right hemisphere. Estimates of peak localization within HG were made with reference to three cytoarchitectonic subdivisions; Te 1.2 (lateral HG), Te 1.0 (central HG) and Te 1.1 (medial HG) (Morosan et al. 2001; Eickhoff et al. 2005). Region of interest analysis used the same approach described by Hall and Plack (2009).

Results

Psychophysical Measures

The results of the psychophysical measures are shown in Figure 3. To evaluate the effect of the number of delay-and-add iterations on discrimination performance, a one-way repeated-measures ANOVA was conducted. Pitch discrimination thresholds
for IRN were significantly affected by iteration ($F(2.15, 62.34) = 53.00, p < 0.001$). For this test, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = 0.72$). Modulation discrimination performance for IRNo was also significantly affected by iteration ($F(1.84, 46.10) = 115.15, p < 0.001$). Degrees of freedom for this test were also corrected using the same procedure ($\epsilon = 0.62$). Polynomial contrasts indicated a linear trend for both IRN and IRNo ($F(1, 29) = 91.28, p < 0.001$ and $F(1, 25) = 160.91, p < 0.001$, respectively). For IRN, pitch discrimination thresholds decreased as a function of the number of iterations, whereas for IRNo the percentage correct increased as a function of number of iterations. The polynomial contrasts also indicated an additional significant quadratic trend for IRN ($F(1, 29) = 12.32, p < 0.01$), but not for IRNo. Considering Figure 3, this trend is seen as a plateau in pitch discrimination performance beyond 16 iterations. These results suggest that the perceptual salience of the spectro-temporal modulations increases as a function of the number of iterations as does pitch salience for IRN (Yost 1996).

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

**fMRI Results: Random-Effects Analysis**

First, we explored the pattern of responses separately for IRN (all iterations combined) and for IRNo compared to the combined noise controls, using planned comparisons within the 2x4 ANOVA. Both contrasts revealed significant feature-driven responses across the entire area of HG and PT, which survived correction (p < 0.05) (Figure 4). As can be seen in Figure 4, there is considerable overlap of the activity related to IRN and IRNo, although there appears to be a slightly greater spread of activation for IRN than for IRNo. For IRN, the most significant peaks of activation fell close to the border between Te 1.2 (lateral HG) and Te 1.0 (central HG) in both hemispheres (x-54 y-16 z 4 mm in the left and x58 y-2 z-2 mm in the right). This localization of IRN-related activity concurs with previous results (Hall and Plack 2009). The peaks of IRNo-related activity fell within 2 mm of those identified for IRN (x-52 y-16 z 4 and x60 y-4 z 0 mm).
Second, we evaluated the effects of the stimulus and the number of delay-and-add iterations on the pattern of auditory cortical activity. The main effect of stimulus indicates whether there are any significant differences between IRN and IRNo contrasts. No differential activity survived correction, although two small clusters were present in the left hemisphere at an uncorrected threshold (p < 0.001) (Table 1). Hence, the responses to IRN and IRNo were broadly equivalent.
For the main effect of iteration, there was one small cluster in the left hemisphere close to the border between Te 1.0 (central HG) and Te 1.2 (lateral HG) (p < 0.001, uncorrected) (Table 1). According to the probability map for these subdivisions of HG (Eickoff et al. 2005), the peak was most probably located in Te 1.0 (30-60% probability), which is also where the largest proportion of the voxels in the cluster were located (67%). Although this result did not survive more stringent statistical correction, we consider it informative to explore the way in which iteration might affect the magnitude of activity since this peak voxel is the most likely candidate for a location that is responsive to the perceptual salience of the stimulus. To explore the contributions of pitch salience and the salience of the slowly varying spectro-temporal modulations, we plotted the parameter estimates (β) for this voxel, separately for each delay-and-add iteration for IRN and IRNo. The results are shown in Figure 5. From this figure, it is evident that there is a considerable increase in activity between 4 and 16 iterations for IRN, which flattens after 16 iterations. For IRNo, the positive increase is more gradual across all iteration values. The random effects analysis revealed no significant interaction between stimulus and number of iterations. We therefore conclude that, for this voxel at least, a response to perceptual salience of the stimulus features appears for both IRN and IRNo signals, hence the slowly-varying modulations covary with the iterations.
Table 1. Significant clusters of activity for main effects of IRN and IRNo, and for the subtraction of IRNo from IRN activity. The peak voxels are reported for the left and right hemispheres, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Left hemisphere</th>
<th></th>
<th>Right hemisphere</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peak coordinates</td>
<td>Z-score</td>
<td>Voxel-level p-value</td>
<td>Cluster size</td>
</tr>
<tr>
<td><strong>Main effect of IRN</strong></td>
<td>-54 -16 4</td>
<td>Inf &lt;0.001</td>
<td>161</td>
<td>58 -2 -2</td>
</tr>
<tr>
<td></td>
<td>-48 -10 6</td>
<td>4.37</td>
<td>&lt;0.001</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>-14 -14 2</td>
<td>3.02</td>
<td>0.002</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>-50 8 -8</td>
<td>1.78</td>
<td>0.048</td>
<td>3</td>
</tr>
<tr>
<td><strong>Main effect of IRNo</strong></td>
<td>-50 -16 2</td>
<td>6.12</td>
<td>&lt;0.001</td>
<td>1496</td>
</tr>
<tr>
<td></td>
<td>-50 -54 38</td>
<td>3.04</td>
<td>0.01</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>-60 -46 44</td>
<td>2.33</td>
<td>0.048</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>IRN&gt;IRNo</strong></td>
<td>-54 -16 4</td>
<td>3.33</td>
<td>0.074</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>-58 -14 4</td>
<td>3.26</td>
<td>0.074</td>
<td>1</td>
</tr>
</tbody>
</table>

*FDR Corrected

Figure 5. Activation values for IRN (dark grey) and IRNo (light grey) within the peak voxel (x-50 y-14 z2 mm) for the effect of iteration. Error bars represent 95% confidence limits.

**fMRI Results: Region-of-Interest (ROI) Analysis**

One of the theoretical perspectives outlined in the Introduction proposes a special role for lateral HG in pitch coding (Bendor and Wang 2005, 2006; see also Patterson et al. 2002). The results presented in Figure 4, and (tentatively) for the peak voxel in Figure 5, suggest a response to the slowly varying spectro-temporal modulations within the vicinity of central and lateral HG. This novel finding potentially weakens
previous interpretations of neuroimaging data based on evidence using IRN as a sole pitch-evoking stimulus.

Crucial to this claim is the ability to demonstrate that the influence of the number of delay-and-add iterations is the same for IRNo as it is for IRN signals. Moreover any further analyses should ideally consider the whole of the regional subdivision, not just a single voxel, in order to be certain that this confound might influence patterns of activity across regions of the auditory cortex. To test this hypothesis statistically, we interrogated the profile plots from the 2x4 ANOVA using tests of the within-subjects contrasts. These tests tell us about the shape of the response as a function of the number of delay-and-add iterations, specifically by assessing the significance of the linear and quadratic trends in the data. The interaction term tells us whether this relationship is different for the two classes of stimulus. For example, perhaps a quadratic relationship might emerge for IRN given the psychophysical measurements of pitch discrimination thresholds. The fMRI data to be analysed in this way were obtained using a region of interest (ROI) approach that computed the average magnitude of activity (β) from all voxels within lateral HG in response to each of the eight stimulus conditions. The profile plots (means and 95% confidence intervals) are represented in Figure 6, with error bars computed across listeners.
Figure 6. Profile plots of activity for IRN and IRNo with different numbers of iterations, taken from the 2x4 ANOVA for the three pitch-responsive regions, central HG (top row), lateral HG (middle row) and PT (bottom row). Error bars represent 95% confidence limits.

Within lateral HG, there was a positive linear relationship between activity and the number of iterations ($F(1, 15) = 25.96, p < 0.001$), with no significant quadratic component (Figure 6A). The interaction term for the linear trend was not significant ($F(1, 15) = 0.62, p > 0.05$) and so there is no evidence that the number of iterations exerted different effects on the response to IRN and IRNo stimuli in lateral HG.

Since we had observed a widespread response to IRN and IRNo across auditory cortex (Figure 4), we took this opportunity to examine the profile plots for central HG (Figure 6B) and for PT (Figure 6C) using the same procedures. The results were very much the same as for lateral HG. The tests of within-subjects contrasts revealed a significant positive linear relationship between activity and the number of iterations.
(for central HG: F(1, 15) = 14.47, p < 0.01 and for PT: F(1, 15) = 9.38, p < 0.01), with no significant quadratic term. Similarly, the findings indicated a non-significant interaction for the linear trend (for central HG: F(1, 15) = 4.12, p > 0.05 and for PT: F(1, 15) = 3.16, p > 0.05). Hence we draw the same conclusion that the number of iterations exerts an equivalent effect on the response to IRN and IRNo stimuli in central HG and PT. The results from this ROI approach provide stringent statistical evidence that human auditory cortex is somewhat broadly responsive to the slowly varying spectro-temporal modulations in the signal.

To investigate the effects of ANC, a mixed design ANOVA was performed separately for the three different ROIs, specifying ANC as a between-subject factor. None of the regions indicated a significant effect of ANC (F(1, 14) = 0.967, p > 0.05 for Te 1.0, F(1, 14) = 0.002, p > 0.05 for Te 1.2 and F(1, 14) = 0.967, p > 0.05 for PT), with no interaction between ANC and stimulus or iteration.

fMRI Results: Incidence Maps

Given that the slowly varying modulations contribute to the IRN-related response, we propose IRNo as a more appropriate noise control for examining the pitch evoked by IRN than the Gaussian random noise used hitherto. An alternative demonstration showing the impact of the choice of noise control is illustrated by the results of incidence maps created to display the distribution of IRN-related activity across individuals when either a Gaussian noise or IRNo is selected to be that noise control (see Hall and Plack, 2009, for a description of the method). In a previous study, we reported that compared to a Gaussian random noise, IRN generated greater activity bilaterally around HG, especially just posterior to HG, close to the border with PT. The
maximum consistency across the individual maps was 55% (5/9 listeners) in the left lateral HG (x=-55 y=12 z=4 mm) and 78% (7/9 listeners) in right central HG (x=46 y=-18 z=0 mm).

Figure 7. Distribution of IRN-related activation compared to Gaussian noise (top row) and to IRNo (bottom row). For the purpose of localization, outlines of the positions of lateral HG (yellow), middle HG (white) and PT (black) are overlaid onto the images. The incidence maps are overlaid onto four different axial slices through the group-averaged anatomical image. The colour scale represents the percentage of IRN-related activation at every voxel and is calculated as a proportion of a possible maximum of 16.

For the present study, the same statistical contrast generated activity centered around HG spreading posteriorly and anteriorly across auditory cortex. The top row in
Figure 7 illustrates this result. On the left side, the maximum consistency across the individual maps was 75% (12/16 individuals), centered in antero-lateral PT, close to the border with lateral and central portions of HG ($x$-60 $y$-26 $z$8 mm). In the right side, the maximum consistency was 88% (14/16 individuals) sited anterior to HG on the posterior edge of planum polare ($x$60 $y$-4 $z$0). In striking contrast are the results for the comparison between IRN and IRNo conditions (bottom row in Figure 7). Although the distribution of activity was broadly similar, the degree of consistency across individuals was markedly reduced. On the left side, the maximum consistency across the individual maps of 38% (6/16 individuals) found at the border between the central portion of HG and PT ($x$-55,$y$-20 $z$8 mm). In the right side, the maximum consistency was 44% (7/16 individuals) sited in PT at the anterior border with lateral HG ($x$ 60, $y$ -18, $z$ 4).

These results demonstrate that the introduction of an appropriately matched control greatly reduces the magnitude of the response that can be attributed specifically to pitch when the contribution from any slowly varying spectro-temporal fluctuations is controlled. The resulting pitch-related activity for IRN is less consistent between listeners and is more similar to that found for other types of pitch-evoking stimuli (Hall and Plack 2009; García et al. 2010).

**Discussion**

*Response to IRN May Result from Features Unrelated to Pitch*

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

The contention here is not in the use of IRN as a pitch-evoking stimulus, rather it is the lack of a well-matched control in previous studies of pitch perception using IRN. These studies have not controlled for the slowly-varying spectro-temporal characteristics of IRN that contribute to the response, but are not related to pitch. It is suggested that further studies seeking to use IRN as a pitch stimulus use a control that is well matched in terms of these features, such as IRNo.

It has been suggested that for a brain region to be classified as a pitch center it should show an increase in activation with increasing pitch salience (and hence with increasing iterations for our IRN stimuli) (Hall and Plack 2009; Bendor and Wang 2005). The ROI analyses (Figure 6) revealed a linear increase in activity with increasing iterations in lateral HG, central HG, and PT. However, there was no
evidence for a differential effect for IRN and IRNo. Hence, we infer that the linear trends in auditory cortical activity were more strongly driven by the response to the depth of the spectro-temporal fluctuations than by the response to pitch salience.

*Is there a human ‘Pitch Center’?*

In light of the current findings, it would be unwise to assign the title of ‘pitch center’ to any area of auditory cortex based on the results of studies that have used IRN as their sole pitch-evoking stimulus (e.g. Griffiths et al. 1998; Patterson et al. 2002; Krumbholz et al. 2003), as these studies have not used suitable controls that separate the pitch effect from the effects of slowly varying spectro-temporal modulations. Based on responses to resolved and unresolved harmonic complex tones, as described above, Penagos et al. (2004) argued for a salience-dependent pitch response in lateral HG. However, on inspection of their Figure 3, the salience-dependent pitch response appears more posterior in most listeners than the group-averaged lateral HG response reported. In addition, only five listeners were included in their analysis and the correction used (least significant difference) was much less stringent than the FDR correction used in the current paper (and prone to type I errors). Puschmann et al. (2010) also reported a pitch-related response in lateral HG for a tone-in-noise and two Huggins pitch stimuli. Again, however, their results indicate a large pitch-related response posterior to lateral HG, in PT (as observed in their Figure 3). Warren et al. (2003) attribute the response in PT to their wideband harmonic complex tones specifically to pitch height (which provides a basis for sound segregation). Pitch chroma (which provides a basis for representing melodies) activated planum polare. Pitch chroma and pitch height were both found to activate lateral HG.
In summary, neuroimaging studies to date do not provide strong evidence for a single pitch center that responds to all pitch-evoking stimuli, and that is also responsive to changes in pitch salience. However, it has been shown recently that the pitch-related response can be enhanced by the use of a ‘noise context’ between the bursts of pitch stimuli (García et al. 2010). This reduces the effects of energy onsets which tend to saturate the response. It is possible that the use of a noise context could increase the chance of observing a pitch-specific salience-dependent response.

Acknowledgments

This study was supported by a PhD studentship awarded by the MRC Institute of Hearing Research and MR scanning was paid for through MRC infrastructure funding awarded to the same organization. Two undergraduate students (Sparsh Chandan and Liam Hennessy) collected some of the psychophysical data reported and were supervised by the first author. The authors would like to thank Alain de Cheveigné for advice on the procedure for generating IRNo.

References


Chapter 7. Non-pitch attributes of sound influence the auditory cortical responses to pitch

Daphne Barker\textsuperscript{1,2}, Christopher J. Plack\textsuperscript{2} and Deborah A. Hall\textsuperscript{3}

\textsuperscript{1}MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, UK
\textsuperscript{2}Division of Human Communication and Deafness, The University of Manchester, Manchester M13 9PL, UK
\textsuperscript{3}Department of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK

This chapter was prepared for publication and is presented in the format in which it was prepared.

Publication 5

\* Daphne García will be married in early 2011 and the paper will be submitted in her married name, Daphne Barker.
Non-pitch attributes of sound influence the auditory cortical responses to pitch

Daphne Barker\textsuperscript{a,b}

Christopher J Plack\textsuperscript{b}

Deborah A Hall\textsuperscript{a,c}

\textsuperscript{a} MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD

\textsuperscript{b} Division of Human Communication and Deafness, The University of Manchester, Manchester, M13 9PL

\textsuperscript{c} Division of Psychology, Nottingham Trent University, Burton Street, Nottingham, NG1 4BU

Corresponding author: Daphne Barker

Email: daphne@ihr.mrc.ac.uk

Tel: +44 115 922 3431

Fax: +44 115 951 8503
Abstract

Iterated ripple noise (IRN) is a type of pitch-evoking stimulus that is commonly used in neuroimaging studies of pitch processing. It is known to produce a consistent response in a region of auditory cortex located antero-lateral to primary auditory cortex (lateral HG). The IRN-related response has often been attributed to pitch, although recent evidence suggests that it is more likely driven by slowly varying spectro-temporal modulations not related to pitch. The factorial design used in this functional magnetic resonance imaging (fMRI) study confirmed that both pitch and modulations elicited a significantly greater response than a baseline Gaussian noise in a region of non-primary auditory cortex (planum temporale). Moreover, this response appears sensitive to pitch salience, but not to the salience of modulation depth. The results reported in this paper suggest that the most likely site for a human ‘pitch center’ is planum temporale. This region appears to contain separate representations of both pitch and modulation, since there is no evidence for an interaction between these two features.
Introduction

Pitch is one of the primary auditory percepts. It is arguably the most important feature of music and is one of the main cues in language. The importance of pitch in hearing makes it a feature of great interest for auditory scientists. There are many different physical features that can elicit the same pitch percept. For example, although a middle C played on the piano sounds very different to a middle C played on the guitar or sung, it is still recognized as the same note. It is this phenomenon that has led auditory scientists to postulate on the existence of a ‘pitch center’ – a region of auditory cortex responsible for coding pitch, regardless of the physical attributes from which it arises. It has been assumed that such a region would elicit a greater response to pitch stimuli with stronger pitch salience (the strength of the pitch percept) than it would to stimuli with weaker pitch salience (Griffiths et al. 1998, 2001; Krumbholz et al. 2003; Penagos et al. 2004; Hall and Plack 2009).

Iterated ripple noise (IRN) is a type of pitch stimulus that is created by generating a sample of noise, imposing a delay, and adding or subtracting the delayed version to or from the original (Yost 1996). The delay-and-add process introduces temporal regularity, which evokes a pitch percept that is equal to the reciprocal of the delay. The more times this delay-and-add process is repeated, the more salient the pitch becomes (Yost 1996). The fact that pitch salience can be increased by repeating the iterative process without changing any other fundamental features of the stimulus has made IRN a popular choice of pitch stimulus for use in neuroimaging studies searching for a pitch center. These studies worked on the subtractive assumption that deducting the activation produced by Gaussian noise from that produced by IRN leaves a representation of the pitch response. The IRN response that has been
attributed to pitch is highly consistent across individual listeners and is also reproducible between studies (Patterson et al. 2002; Krumbholz et al. 2003; Hertrich et al. 2005; Seither-Preisler et al. 2004, 2006; Schönwiesner and Zatorre 2008; Hall and Plack 2009 (experiment 2)). Most of these studies have revealed an IRN-related response in an auditory region located antero-lateral to primary auditory cortex, in the lateral portion of Heschl’s gyrus (HG). When pitch stimuli other than IRN are used, however, the inter-listener consistency decreases and the pitch response appears more posterior than the IRN response (Hall and Plack 2007; 2009 (experiment 1); García et al. 2010; Barker et al. in preparation a). A logical conclusion for this difference is that IRN contains an additional acoustic feature, not present in other pitch-evoking stimuli, that elicits a greater response than other pitch stimuli.

The delay-and-add process involved in the generation of IRN stimuli produces slowly varying spectro-temporal modulations that are not related to the pitch percept. IRN is made from a sample of Gaussian noise, which has a random waveform. However, the iterative delay-and-add process introduces broad spectro-temporal features into the noise (Hall and Plack 2009; Barker et al. in preparation b) (Figure 1). Most previous pitch studies using IRN as their sole pitch-evoking stimulus have not been designed to separate the pitch response from the response to slowly-varying spectro-temporal fluctuations. Hence, it is not precisely clear which feature is responsible for the IRN response. In order to determine whether it is the pitch, the slowly-varying modulations or an interaction between the two that drives the IRN-related response, Barker et al. (in preparation b) created a new type of stimulus. This novel stimulus consists of IRN that has been processed in a way that removes the temporal fine structure responsible for the pitch percept, whilst leaving the slowly-varying spectro-temporal
features intact. IRN that is processed in this way is denoted as ‘no-pitch IRN’ (IRNo).
It is interesting to note that results from psychophysical testing indicate that the perceptual discriminability of IRNo modulations improves with increasing number of iterations, in the same way that pitch discrimination thresholds reduce for IRN (Barker et al. in preparation b).

Figure 1. Simulated cochlear representations of IRN (top row) and IRNo (bottom row) in the form of spectrograms. The analysis smoothes the representation in both time and frequency domains to remove any fine structure. All stimuli were created from the same original sample of Gaussian noise, and the IRNo stimuli on the bottom row are processed versions of the stimuli on the top row (IRN). The color bar shows model output in dB. See Barker et al. (in preparation b) for details of the model.

Since these two features appear to covary, studies that have examined the neural response to pitch salience, using IRN as the sole pitch-evoking stimulus (Griffiths et al. 1998; Patterson et al. 2002; Krumbholz et al. 2003; Seither-Preisler et al. 2004, 2006; Schönwiesner and Zatorre 2008; Hall and Plack 2009 (experiment 2)), are also confounded by the response to the depth of the slowly varying spectro-temporal modulations. In a previous fMRI study we have demonstrated that the response to these modulations contributes to the cortical auditory response that previous IRN studies have interpreted as pitch specificity (Barker et al. in preparation b).
There are four potential explanations for a combined effect of modulation and IRN on the auditory cortical response. First is the linear model where both modulation and IRN features evoke a response that is additive (Figure 2A). There is some support for this model in the visual system; a roughly linear increase in haemodynamic response was reported for increasing number of trials presented within a block for brief full-field visual stimuli (Dale and Buckner 1997). Second is the saturation model which suggests that both modulation and IRN features evoke a response, but the magnitude of that response is restricted to some maximum value (saturation). Hence, the presence of IRN-related pitch has little additional effect when the signal already contains slow-rate modulation (Figure 2B). García et al. (2010) provided support for the saturation model when they demonstrated that the sensitivity to pitch-related activation in auditory cortex can be significantly increased by removing the effects of multiple energy onsets during stimulus presentation. Third is the enhancement model in which the dynamic nature of the spectro-temporal ripples actually enhances the pitch response (Figure 2C). The effect of enhancement in human auditory cortex has been demonstrated in a study that examined the effects of auditory selective attention on task-related processes (Paltoglou et al. 2009). Finally, the suppression model suggests the response to modulation suppresses the response to pitch (Figure 2D). Paltoglou et al (2009) provided partial support for suppression in some feature-specific auditory regions when attention is directed toward different auditory features.
**Figure 2.** Schematic representations of the four models for an interaction between modulation and pitch. For each model, the stimuli containing modulation (IRN and IRNo) are represented by solid lines and the stimuli that do not contain modulation (unresolved harmonic complex and noise) are represented by dashed lines. The stimuli that do not contain pitch (IRNo and noise) are on the left-hand side of each graph, and the stimuli that contain pitch (IRN and unresolved harmonic complex) are on the right. A) The linear model is not an interaction model; it suggests an additive effect of stimulus features. B) The saturation model suggests a maximal response capacity in pitch-responsive regions. C) The enhancement model suggests that the presence of modulation increases the pitch-related response. D) The suppression model suggests that the presence of modulation in a signal decreases the response to pitch.

Current evidence cannot distinguish between these different models and so the primary motivation for the current study was to quantify the relationship between cortical responses to pitch (in general) and to slow rate spectro-temporal modulations using stimulus conditions that were created to provide a factorial investigation of these features. In order to determine which of these models is correct, the interaction between pitch and modulation must be analyzed to determine whether the presence of modulation impacts on the size of the pitch-related response. The factorial analysis addressed the size of the contribution of each of these features to the overall fMRI response and this question was examined within a spherical region-of-interest centered anatomically on an *a priori* estimate of the location of the pitch center. The previous study (Barker et al. in preparation b) could not have revealed any
interactions between the pitch and modulation features of IRN because the design did not include a condition with pitch but without any slow-rate modulation and thus the modulatory effect of one feature on the other could not be established.

The location of pitch- and modulation-related activity within auditory cortex was also of interest. We sought to determine, out of the two regions that are in contest for the title of ‘pitch center’, whether lateral HG or PT is the most likely candidate. We did this by considering the probability of the anatomical localization of the peaks of the pitch-related response with respect to estimates of the underlying cytoarchitecture, within the spherical region mentioned previously. The final hypothesis addressed by the current study concerned the effect of pitch salience on the fMRI response in the pitch-responsive region. Pitch salience was manipulated using IRN with different numbers of iterations and an unresolved harmonic complex with and without a noise masker. Additionally, IRNo stimuli (with a corresponding number of iterations) were used to determine whether activation increases with increasing modulation depth. The goal of the present experiment was therefore to provide definitive answers to four questions by using a factorial design that enables the evaluation of each feature individually and of interactions between the features. In summary, the research questions addressed here are:

I. Is there an interaction between the responses to modulation and to pitch?

II. Is IRN-related activity driven by slowly varying spectro-temporal modulations, or by pitch?

III. Is lateral HG or PT the most likely site for a pitch center?
IV. Are the generators of the pitch and modulation responses sensitive to differing levels of salience for these features?

Materials and Methods

Listeners

Fourteen listeners (7 male, 7 female; age range 22 - 48 years) with normal hearing (≤20 dB hearing level between 250 Hz and 8 kHz) took part in fMRI testing. All listeners were right-handed (laterality index = 50, Oldfield 1971). Seven listeners were musically trained between grade 2 and grade 7 (# 02, 07, 12, 19, 22, 23 and 25) while three others reported informal musical experience (self-taught/ungraded, # 05, 09 and 13). None had a history of any neurological or hearing impairment. Listeners gave written informed consent and the study was approved by the Medical School Research Ethics Committee, University of Nottingham. The scanning session for one of the listeners (# 25) had to be terminated due to a significant region of unilateral local MR signal decay around auditory cortex, possibly due to a shimming artifact which could not be rectified. Another subject (# 19) had to be excluded from the analysis because she failed the subjective quality control on two counts. First, there was a significant amount of head motion and second there was an absence of reliable sound-related activity.

Conditions

The experimental design comprised 10 sound conditions which part crossed the factors pitch, spectro-temporal modulation and salience. Two types of pitch-evoking stimuli were employed; IRN and unresolved harmonic complex tones (unres). IRN stimuli comprised three levels of pitch salience (4, 16 and 64 iterations – denoted
IRN4, IRN16 and IRN64, respectively), while the unres had two levels of pitch salience (masked and unmasked unres). Another stimulus contained slowly-varying spectro-temporal fluctuations, but did not evoke a pitch percept (IRNo). This stimulus had three levels of fluctuation salience (4, 16 and 64 iterations – denoted IRNo4, IRNo16 and IRNo64, respectively). The design also included two control conditions. The first was a Gaussian noise (noise) and the second was a Gaussian noise that had been processed in the same way as the IRNo stimuli (processed noise).

**Stimuli**

All IRN and unres stimuli evoked a pitch corresponding to a 100-Hz tone. For the unmasked unres condition, the level of each harmonic was 20 dB greater than the spectrum level of the control noise so that the gross spectral density of all the stimuli was the same. All components of the unresolved harmonic complexes were added in cosine phase, and the stimuli were bandpass-filtered between 1 and 2 kHz to remove low-numbered harmonics that are resolved (i.e. separated out) by the peripheral auditory system. To make the low-pitch-salience (masked) unres, a bandpass-filtered (1 – 2 kHz) Gaussian noise masker was added to the unmasked unres so that the level of the complex tone equaled the level of the masking noise (0 dB signal-to-noise ratio). The addition of a noise masker in the same spectral region of the unmasked unres reduces the perceptual salience of the pitch. A pilot psychophysical study using nine listeners revealed that f0 discrimination thresholds for masked unres were on average 11% higher than for unmasked unres. The level of each harmonic was 17 dB greater than the spectrum level of the control noise, and the spectrum level of the added noise masker was 3 dB below the spectrum level of the control noise. IRN stimuli were generated by a delay-and-add process performed on a Gaussian noise.
The noise was again bandpass filtered (1 – 2 kHz) to remove the resolved harmonics. A delay of 10 ms was imposed before adding the delayed noise back to the original sample. The delay-and-add process was repeated 4, 16 or 64 times to generate the three IRN conditions. To create IRNo, a conventional IRN stimulus was generated as above. The IRN was sampled using a rectangular window with a 10-ms duration. A fast Fourier transform (FFT) was used to generate the magnitude and phase spectra of the sample, and the phase of the components was randomized. An inverse FFT was then used to regenerate the time representation. The sampling window was advanced by half of the IRN delay (5 ms) and the process repeated. The processed samples were overlapped and added (preserving the start-times of the samples), adjusted to a spectrum level of 52 dB SPL and gated to produce a time waveform with a 580-ms steady state and 10-ms linear-intensity ramps. The phase randomization process removes any correlation in the fine structure between samples, obliterating the harmonic structure and the pitch cue. However, the slowly varying broad spectral features are preserved. These fluctuations are visible in the spectrogram representation of IRN when it is smoothed in both time and frequency domains to remove any fine structure (Figure 1). The process was repeated using the IRN₄, IRN₁₆ and IRN₆₄ conditions to generate the three IRNo conditions. All IRN and IRNo stimuli included a noise masker, low-pass filtered at 1 kHz and with a spectrum level of 52 dB SPL, to mask cochlear distortion products. The parameters of the noise control were 52 dB SPL spectrum level and low-pass filtered at 2 kHz. The processed noise control was generated in the same way as the IRNo, but was otherwise identical to the noise control. All sounds (unres, IRN, IRNo, noise and processed noise) were matched in bandwidth (0 - 2 kHz) and spectral density, and hence overall energy (85 dB SPL).
The energy onset response is an effect that dominates responses in the auditory cortex to repeated bursts of sounds, so that sensitivity to pitch is reduced (Krumbholz et al. 2003; Seither-Preisler et al. 2004; García et al. 2010). To improve sensitivity to the features of interest, we therefore employed a ‘continuous stimulation’ paradigm in which experimental sounds were interspersed by short bursts of noise. In the MR scanner, stimulus conditions each comprised a 15.19-s alternating sequence of 600-ms experimental sounds (including 10-ms linear-intensity onset and offset ramps) each separated by 250 ms Gaussian noise, with the same overall sound level as the experimental sounds. The first and last components of each sequence were Gaussian noise. Each pitch and noise signal was generated using 10-ms linear-intensity onset and offset ramps, which were overlapped at the 3 dB points to produce a stable envelope for the stimulus. Sixteen sample sequences were created for each condition and a different set of stimuli was generated for each participant.

fMRI Protocol

Scanning was performed on a Philips 3 T Intera Acheiva using an 8-channel SENSE receiver head coil. A T1-weighted high-resolution (1mm³) anatomical image (matrix size = 256x256, 160 sagittal slices, TR = 8.2 s, TE = 3.6 ms) was collected for each subject. The anatomical scan was used to position the functional scan centrally on HG, and care was taken to include the entire superior temporal gyrus and to exclude the eyes. Functional scanning used a T2*-weighted echo-planar sequence with a voxel size of 3mm³ (matrix size = 64x64, 32 oblique-axial slices, TE = 36 ms). Sparse imaging with a TR of 8188 ms and a clustered acquisition time of 1990 ms was used (Edmister et al. 1999; Hall et al. 1999). A SENSE factor of 2 was applied to reduce image distortions and a SofTone factor of 2 was used to reduce the background
scanner noise level by 9 dB. Functional data was acquired over four runs of 98 scans each. Each sound condition had a total of 32 scans, with 34 scans for the silent baseline. Listeners were requested to listen attentively to the sounds, but were not required to perform any task. A custom-built MR compatible system delivered distortion-free sound using high-quality electrostatic headphones (Sennheiser HE60 with high-voltage amplifier HEV70) with passive noise attenuation. An active noise control (ANC) device (Hall et al. 2009) was used to reduce the overall acoustical scanner noise by a further 14 dB. Eight scans were appended to the beginning of the run in order to initialize the noise cancelling device. These scans were excluded from the analysis.

Data Analysis
Images were analyzed separately for each listener using statistical parametric mapping (SPM5, http://www.fil.ion.ucl.ac.uk/spm). Preprocessing steps included realignment to correct for subject motion, normalization of individual scans to a standard image template, and smoothing with a Gaussian filter of 8 mm full width at half maximum (FWHM). The realignment process generated estimates of the scan-to-scan movement for three translations (x, y and z planes) and three rotations (roll, pitch and yaw). These were included as variables in the individual design specification in addition to the 10 sound conditions and the 4 scanning runs. The silent baseline was implicitly modeled in the design. The first-level general linear model assessed the variables of interest with respect to the scan-to-scan variability. A high-pass filter cutoff of 420 s was used to remove low frequency confounds. The resulting model estimated the fit of the design matrix (X) to the data (Y) in each voxel in order to provide β values (the contribution of a single regressor to the overall fMRI
signal). Separate statistical contrasts for each sound condition were specified relative to the silent baseline. To investigate the differential responses across conditions, a one-way ANOVA was specified at the second level using the preceding contrast images for each individual. We defined the model in this way because it provides maximum flexibility for assessing the different effects of interest and embedded within the model a subset of the data can be considered as a 2x2 factorial ANOVA with pitch (pitch/noise) and spectro-temporal modulation (modulations/no modulations) as factors. For this analysis, two conditions contributed to each cell. The design is represented schematically in Figure 3.

<table>
<thead>
<tr>
<th>Modulation</th>
<th>Pitch</th>
<th>Noise</th>
<th>Unmasked Unres</th>
<th>Processed noise</th>
<th>Masked Unres</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Unres</td>
<td>No</td>
<td>Unres</td>
</tr>
<tr>
<td>Yes</td>
<td>IRN016</td>
<td>IRN064</td>
<td>IRN16</td>
<td>IRN64</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. Schematic representation of the stimuli entered into the 2x2 factorial analysis. Each cell except for the ‘no pitch, no modulation’ cell contains two levels of salience. There is no difference between the noise and the processed noise and so the design is not fully factorial.

Estimates of peak localization within lateral HG were made with reference to the cytoarchitectonic subdivision Te 1.2 (Morosan et al. 2001; Eickhoff et al. 2005). The toolbox used to estimate the activity in Te 1.2 does not define PT, but the area of the sphere that is not defined as Te 1.0 or Te 1.2 falls entirely within our definition of PT. Therefore, unless a peak falls within Te 1.0 or Te 1.2, it is assigned to PT. Estimates of peak localization were required to exceed 50% in Te 1.0 or Te 1.2 in order to be confident that the activity could be assigned to either of those areas.
Although 14 listeners were scanned, only 12 were included in the analyses (the reasons for excluding subjects 19 and 25 are mentioned in the *Listeners* section above). To improve external validity, our interpretation of the pitch- and modulation-related activity was informed by a spherical ROI (radius 10 mm) that was centered on the average peak co-ordinates that had been derived from six previous pitch studies (Table 1). This spherical ROI encompassed parts of central and lateral HG and PT, and pitch-related activation within this region was interpreted to represent a highly consistent pitch response across studies.

**Table 1.** Location (MNI coordinates) of the most significant pitch-related responses identified by previous fMRI studies using various pitch-evoking stimuli.

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x y z Probability</td>
<td>x y z Probability</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patterson et al. 2002</td>
<td>-55 -13 2 10% 40%</td>
<td>57 -9 -2 20% 10%</td>
</tr>
<tr>
<td>Hall and Plack 2009</td>
<td>No left hemisphere clusters</td>
<td>64 -18 4 -</td>
</tr>
<tr>
<td>Puschmann et al. 2010</td>
<td>-50 -20 5 50% 30%</td>
<td>58 -12 7 60% -</td>
</tr>
<tr>
<td>García et al. 2010 (cHP)</td>
<td>-58 -24 8 20% -</td>
<td>64 -16 6 - -</td>
</tr>
<tr>
<td>García et al. 2010 (unres)</td>
<td>-62 -24 8 10% -</td>
<td>66 -18 6 - -</td>
</tr>
<tr>
<td>Barker et al. in prep a</td>
<td>-62 -28 8 - -</td>
<td>64 -22 4 - -</td>
</tr>
<tr>
<td>Barker et al. in prep b</td>
<td>-54 -16 4 10% 30%</td>
<td>58 -2 -2 - 60%</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>-57 -21 6 - -</td>
<td>62 -14 3 - -</td>
</tr>
</tbody>
</table>

Effects of modulation and an interaction between modulation and pitch were also limited to the spherical region described above to determine the effects of modulation and any interactions specifically within the pitch-responsive region. Unless otherwise specified, results are reported after volume correction based on the spherical ROI to control for type I errors using a false discovery rate (FDR) threshold of \( p < 0.05 \) (Genovese et al. 2002).
Results

Figure 4. Activation map from the 2x2 factorial ANOVA showing locations for the main effects of modulation (blue) and pitch (red), and a conjunction for the two features (pink). The yellow border denotes Te 1.2 (lateral portion) (Morosan et al. 2001) and the black border outlines PT (Westbury et al. 1999). Activation is overlaid onto an average anatomical image made from the 12 individual listeners. The left hemisphere is on the left-hand side of each anatomical image. These images used an uncorrected threshold p < 0.05. This figure demonstrates the patterns of activation across the entire cortex, although the analyses were restricted to a 10-mm sphere centered on the white spots in the center panel, which represents the mean center point of the pitch response from the six previous studies listed in Table 1.

Sensitivity to Pitch and to Modulation

The main effect of pitch was determined by the comparison between the four most salient pitch conditions (masked unres, unmasked unres, IRN$_{16}$, IRN$_{64}$) and four matched no-pitch conditions (noise, processed noise, IRNo$_{16}$ and IRNo$_{64}$) (Figure 4). Within the spherical ROI, this contrast highlighted bilateral peaks of pitch-related activity with maxima in PT ($x$=-64 $y$=-28 $z$=6 in the left hemisphere and $x$=66 $y$=-22 $z$=8 in the right, Table 2). The cluster in the left hemisphere contained two further maxima in central HG, one of which encompassed part of lateral HG. The right hemisphere cluster also contained two further maxima; both were located in PT, and one encompassed part of lateral HG (Table 2), although this maximum did not survive
correction (FDR p > 0.05). The probability of the maxima encompassing lateral HG actually being located within that region was low (20% in the left hemisphere and 30% in the right).

Table 2. Location (MNI coordinates) of the maximal effects of pitch and modulation, and of the conjunction between pitch and modulation, including probabilities of individual peaks being located in Te 1.0 and/or Te 1.2. Unless the probability within either of these regions is greater than 50%, the activation is assigned to PT. Voxels significant at p < 0.05 FDR corrected within the spherical ROI. n indicates the number of voxels within each cluster.

<table>
<thead>
<tr>
<th>Peak</th>
<th>Left</th>
<th>Probability</th>
<th>Right</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
<td>z</td>
<td>Te 1.0</td>
</tr>
<tr>
<td>Main Effect of Pitch</td>
<td>-64</td>
<td>-28</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-64</td>
<td>-28</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-54</td>
<td>80</td>
<td>30%</td>
<td>20%</td>
</tr>
<tr>
<td>3</td>
<td>-50</td>
<td>-20</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>Main Effect of Modulation</td>
<td>-58</td>
<td>-14</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-58</td>
<td>-14</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-52</td>
<td>-18</td>
<td>0</td>
<td>30%</td>
</tr>
<tr>
<td>3</td>
<td>-64</td>
<td>-26</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-64</td>
<td>-26</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-64</td>
<td>-26</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>Conjunction</td>
<td>-56</td>
<td>-20</td>
<td>8</td>
<td>20%</td>
</tr>
<tr>
<td>1</td>
<td>-56</td>
<td>-20</td>
<td>8</td>
<td>20%</td>
</tr>
<tr>
<td>2</td>
<td>-50</td>
<td>-20</td>
<td>2</td>
<td>40%</td>
</tr>
<tr>
<td>3</td>
<td>-64</td>
<td>-26</td>
<td>10</td>
<td>-</td>
</tr>
</tbody>
</table>

Due to the low probability (< 50%) of peaks being located in lateral HG, this region appears to be a poor candidate for a pitch center. Given that the area of the spherical ROI that is not located in Te 1.2 or Te 1.0 falls entirely within our definition of PT, it is reasonable to assume that activity within our sphere that is not assigned to Te 1.0 or Te 1.2 can be localized to PT. We therefore conclude that PT is a more likely candidate for a human pitch center than lateral (or central) HG.

The main effect of modulation was determined by contrasting IRNo$_{16}$, IRNo$_{64}$, IRN$_{16}$, and IRN$_{64}$ with noise, processed noise, masked and unmasked unres (Figure 4). This contrast generated bilateral clusters of activity within our spherical ROI (x: -58 y: -14 z: 4 in the left hemisphere and x: 64 y: -10 z: 2 in the right). The most significant peak within
each of these clusters was located in PT, although there were further peaks that encompassed parts of central and lateral HG (Table 2). In the left hemisphere there were three maxima, one which encompassed part of lateral HG (30% probability), and in the right there were five, three of which encompassed part of lateral HG (30%, 30% and 70% probability). The maxima in the right hemisphere all remained significant when corrected (FDR p < 0.05), but none of the maxima in the left survived correction. A conjunction analysis demonstrated regions that were sensitive to both pitch and spectro-temporal modulation (Figure 4). This analysis revealed bilateral clusters in PT that spread into central and lateral HG on the left, and into lateral HG on the right. Neither of the clusters survived correction (FDR, p > 0.05). An interaction between pitch and modulation also revealed bilateral clusters in PT (with no spread into other regions) that did not survive correction (FDR, p > 0.05). Since an interaction is not reliably supported by the evidence, the linear model seems to best explain these results. Although we reserve the possibility that the small sample size does not provide enough statistical power for the effect to survive correction for multiple comparisons.

In order to determine the overall response to each feature within the whole of Te 1.2, a region-based analysis was conducted for each listener and combined across listeners. Data extraction for the region-based analysis used the approach described by Hall and Plack (2009). To determine whether the IRN-related activation in lateral HG reported in previous studies (Patterson et al. 2002; Krumbholz et al. 2003; Hertrich et al. 2005; Seither-Preisler et al. 2004, 2006; Schönwiesner and Zatorre 2008; Hall and Plack 2009) was driven by pitch or by modulation, the region-based analysis of the response magnitude (i.e. mean $\beta$ values) was calculated for each of
the different sound conditions within this region. In this analysis, values for the two conditions in each cell of the 2x2 ANOVA depicted in Figure 3 were averaged (Figure 5). The region-based analysis revealed no effect of pitch, but there was a significant effect of modulation \( F(1, 11) = 19.34, p < 0.05 \), with no significant interaction.

**Figure 5.** Activation values for the four stimuli in the 2x2 factorial analysis within Te 1.2. ‘Noise’ includes both the Gaussian noise and the processed noise, ‘IRN’ includes IRN16 and IRN64, ‘IRNo’ includes IRNo16 and IRNo64 and ‘Unres’ includes masked and unmasked unres. Error bars represent one standard deviation.

In addition, a contrast of the pooled responses for the IRN conditions (IRN4, IRN16 and IRN64) with those for the IRNo conditions (IRNo4, IRNo16 and IRNo64) failed to reveal any significant clusters in auditory cortex. In light of these findings, it seems that lateral HG is responsive to the slowly varying spectro-temporal modulations in IRN, and not to the pitch. The significant response to pitch in PT when the spherical ROI was applied suggests that PT is a more likely site for a pitch center than lateral HG, although a response to modulations was also found within this region.

---

* The area defined as Te 1.2 in our ROI used different boundary criteria than the probability maps so there may be a slight discrepancy between our Te 1.2 mask and the area defined as Te 1.2 by the SPM Anatomy toolbox. This discrepancy arises due to the uncertainty in defining specific borders for group-based data.
Salience-related activity

A pilot exploration using eight listeners demonstrated that the pitch discrimination thresholds for the IRN conditions were considerably higher than for any of the unres stimuli. This finding implies that IRN stimuli were much less salient than unres stimuli and so these pitch comparisons were analyzed separately. Since the research question relates to an effect of salience within a pitch-responsive region, the spherical ROI described previously was applied. For unres stimuli, the subtraction (unmasked unres – masked unres) examined the effect of pitch salience. Within the spherical ROI, this contrast highlighted bilateral clusters in auditory cortex, with peaks located in PT (x-58 y-30 z8 in the left hemisphere and x60 y-22 z6 in the right). The left cluster contained 4 maxima, of which one spread into lateral HG (x-56 y-18 z10, 20% probability). The cluster in the right hemisphere contained 3 maxima including one that incorporated part of lateral HG (x62 y-6 z4, 30% probability). To investigate the effect of pitch salience for the IRN stimuli, the subtraction (IRN_{64} - IRN_{4}) was performed. Again, within the spherical ROI, this contrast revealed bilateral clusters with maxima located in PT (x-64 y-30 z6 in the left hemisphere and x62 y-14 z0 in the right), although the increasing depth of spectro-temporal modulations with increasing iterations may have driven this effect (Barker et al. in preparation b). One of the two maxima in the left hemisphere (x-56 y-18 z8) encompassed part of lateral HG (20% probability), but none of the three maxima in the right hemisphere encroached into lateral HG. None of the maxima for salience-related activity for IRN remained significant when corrected for multiple comparisons (FDR, p > 0.05). Some activity appeared generally sensitive to pitch salience, irrespective of the type of pitch-evoking sound since a conjunction analysis for the above contrasts revealed bilateral supra-threshold clusters. These clusters were centered bilaterally in PT (x-64 y-30 z6
in the left hemisphere and \( x_{62} \ y_{-14} \ z_{0} \) in the right). One of the two maxima in the left hemisphere included Te 1.2 (\( x_{-56} \ y_{-18} \ z_{8} \), 20% probability), but none of the three maxima in the right hemisphere spread into lateral HG. The evidence for a general sensitivity to pitch salience is weak because none of the activation revealed in the conjunction analysis survived correction for multiple comparisons (FDR, \( p > 0.05 \)). To investigate the effect of modulation salience, the subtraction (IRNo_{64} - IRNo_{4}) was performed. This contrast did not reveal any supra-threshold clusters. These results suggest that pitch responsive regions are sensitive to pitch salience, but we have no evidence to suggest that they are sensitive to the depth of slowly-varying modulations. No interaction in either direction was observed between stimulus (IRN and IRNo) and iteration (4 and 64), suggesting that although our results do not provide evidence for a sensitivity to modulation salience, it may still contribute to the salience effect for IRN.

**Discussion**

The current study used a factorial design to examine the effects of pitch and slowly-varying modulations in the human auditory cortex (Figure 1). This is the first study that has allowed for the responses to pitch and to slowly-varying spectro-temporal modulations in IRN to be separated and for any interactions to be revealed. The use of IRN and IRNo with different numbers of iterations and of masked and unmasked unresolved harmonic complex stimuli revealed regions of auditory cortex that were sensitive to pitch and to modulation salience.

*No Evidence for an Interaction Between Pitch and Modulation Responses*
The first question addressed in this experiment was whether there was an interaction between the response to pitch and the response to modulation. A previous study (Barker et al. in preparation b) revealed that the slowly-varying spectro-temporal modulations created by the delay-and-add iterative process influence the IRN response, but the results could not determine the precise nature of this contribution. The results from the current factorial design suggest that the pitch-sensitive region is sensitive both to pitch and to modulation, with no significant interaction between the two features. Hence, we cannot reject the linear response model (Figure 2A). This model implies that the two features of IRN are additive. Auditory cortex responds to individual sound features such as the pitch in the unresolved harmonic complex or the slow-rate modulation in IRNo, and the addition of other sound features further increases the magnitude of the response. If the linear response model were correct, however, one would expect a greater response to IRN than to IRNo, but this was not the case. One reason for this could be that although the pitch and modulation effects show an additive response, the individual variability is large and thus the means and 95% confidence intervals overlap.

The IRN Response is Driven by Modulations with an Additional Effect of Pitch

The lack of any significant difference between IRN and IRNo within the spherical ROI in this study, along with the lack of any interaction between the two effects, suggests that the IRN response is driven by slowly varying spectro-temporal modulations, with no significant additional effect of pitch. This result concurs with Barker et al (in preparation b), who found broadly similar response patterns for IRN and for IRNo within central and lateral HG and within PT. However, both studies indicated a small additional effect of the pitch in IRN over and above the modulation response elicited
by IRNo. In the current study, there were significant clusters of activation for the high versus low salience IRN contrast at an uncorrected level but not for the equivalent IRNo contrast. In the previous study, there was a significant linear trend for number of iterations for IRN, but not for IRNo. Furthermore, the contrast (IRN – IRNo) in Barker et al. (in preparation b)’s study revealed a bilateral pitch-related response for IRN that was co-located for up to seven of their 16 listeners. Therefore, although modulation accounts for the majority of the IRN response magnitude, there is some evidence that pitch does contribute in a small way.

**Planum Temporale is the Most Likely Site For a Pitch Center**

This is perhaps the most highly contested and widely researched of the four hypotheses. Early neuroimaging studies postulated the existence of a human pitch center in lateral HG (Patterson et al. 2002; Krumbholz et al. 2003; Hertrich et al. 2005; Seither-Preisler et al. 2004, 2006; Schönwiesner et al. 2008; Hall and Plack 2009 (experiment 2)). However, most of these studies exclusively used IRN as their pitch-evoking stimulus. Research using a wider range of pitch-evoking stimuli, such as click trains (Gutschalk et al. 2004), harmonic complex tones (Penagos et al. 2004; Barker et al. in preparation a, b) and Huggins pitch (Hall and Plack 2009; García et al 2010) has demonstrated an effect of pitch focused in, or extending into PT. The current study examined the location of the pitch response reported in six previous studies and found that the average location was in anterior PT. It is possible that pitch-responsive neurons are located on both sides of the border between lateral HG and PT (across the sulcus lying behind Heschl’s gyrus), and thus both of these regions exhibit a response to pitch. The variability in brain structure and morphometry across individuals could also explain the slight discrepancy in the observed pitch
response. Based on the fact that most of the data indicating a response in lateral HG used IRN as a pitch-evoking stimulus, and most other pitch studies employing different pitch stimuli found a pitch response in PT, the most likely reason for the discrepancy is the effect of modulation on the pitch response.

Evidence for Sensitivity to Pitch Salience but not to Modulation Salience

Most previous research has suggested a sensitivity to pitch salience in auditory cortex (Griffiths et al. 1998; Penagos et al. 2004; Bendor and Wang 2005; Gutschalk et al. 2007), although this finding is not universal. For example, using pulse trains with different amounts of jitter and unresolved harmonic complexes with different relative phases, Barker et al. (in preparation a) actually found a decrease in activation with increasing pitch salience. Results from the current experiment provide some evidence for a sensitivity to pitch salience within pitch-sensitive regions, although the effect for IRN did not survive correction. A psychophysical pilot experiment revealed that IRN stimuli were much less salient than unres stimuli, even when the unres stimuli were masked to reduce the signal-to-noise ratio. With that in mind, it is possible that the high-salience IRN condition was not sufficiently salient to produce an increase in the magnitude of the fMRI response that was large enough to survive correction.

We did not find any evidence for a sensitivity to modulation in the pitch-responsive region. However, if modulation salience is not related to BOLD response and pitch salience is, one would expect an interaction between IRN and IRNo in terms of the effect of iteration, whereby the two IRNo conditions are equivalent and the high-salience IRN response is greater than the low-salience IRN response. No such interaction was observed, and so a contribution from modulation can not be ruled out.
of the uncorrected salience response for IRN (Figure 5). Considering a highly significant effect of salience for unres, a weak effect of salience for IRN, and no effect of salience for IRNo, it is possible that modulation has an inhibitory or suppressive effect on pitch salience for IRN (such as in Figure 2d). Results from Barker et al. (in preparation b), indicating a significant linear increase in activity with increasing number of iterations with no significant interaction between IRN and IRNo suggest that this is unlikely. A possible explanation for the disparity between the current results and those from the previous study could be the number of listeners that were tested. The previous study recruited 16 listeners, but only 12 listeners contributed to the current analysis and thus the current study had less statistical power than the previous.

To summarise, the results of the salience analyses suggest that the cortical representation of pitch is sensitive to differing levels of salience, and although it is not sensitive to differing levels of modulation alone, modulation may affect the salience response for IRN.

Summary
The results reported in the current paper support the suggestion made by Barker et al. (in preparation b) that the slowly-varying spectro-temporal modulations in IRN affect the pitch response. This main finding implies that future studies using IRN as a pitch-evoking stimulus should employ a baseline that controls for these modulations (such as IRNo) and that interpretations from results of previous studies using IRN as their sole pitch-evoking stimulus should be carefully considered. Furthermore, our results suggest that if there is a human 'pitch center', it is more likely to be located in
PT than in lateral HG. Finally, the finding of a sensitivity to pitch salience in the auditory cortex suggests that future neuroimaging pitch studies should employ a variety of different pitch-evoking stimuli, each with differing levels of pitch salience. Greater confidence in a ‘pitch center’ would be gained by observing an area of auditory cortex that responds to all different kinds of pitch-evoking stimuli, and whose activation increases with increasing pitch salience.

Acknowledgements

This study was supported by a PhD studentship awarded by the MRC Institute of Hearing Research and MR scanning was paid for through MRC infrastructure funding awarded to the same organization. The authors would like to thank Simon Müller for his contribution to the analysis of fMRI data.

References


Chapter 8. General discussion and conclusions

This chapter will discuss the main findings of each experimental paper (Chapters 4 – 7) in relation to the primary aims of the PhD as set out in the Introduction, and to the implications for further research. The aims were as follows:

1. To investigate the effects of stimulus presentation context on pitch-related responses
2. To determine to what extent non-pitch features within a pitch-evoking stimulus affect pitch-related responses
3. To determine whether or not there is a region of human auditory cortex that satisfies all four criteria of a pitch center (pitch specificity, elimination of peripheral phenomena, pitch constancy and covariation with salience) and the location of any such region.

The first experiment (Chapter 4) is the first fMRI study to employ a continuous stimulation paradigm, as previously used in MEG studies (Krumbholz et al. 2003; Seither-Priesler et al. 2004, 2006; Chait et al. 2006). These studies suggest that the responses to energy onset and to pitch onset can be separated, and that the sensitivity to pitch can be increased by increasing the duration between energy onset and pitch onset. A human depth-electrode study suggested that the two responses arise from neural generators in separate regions of auditory cortex, with a distinction between the energy-onset response located in primary auditory cortex and the pitch-onset response located more lateral in a region of non-primary auditory cortex (Schönwiesner and Zattore 2008). Chapter 4 was the first study to employ the novel continuous stimulation paradigm in fMRI, which enabled the effects of context on the
pitch response to be determined. In this experiment, we found evidence for differential patterns for a compound energy-onset response and compound pitch-onset response, but did not provide support for a clear spatial distinction between the responses. Instead, our results suggest that in the region where the two responses overlap, the energy-related response modulates the pitch-related response. Specifically, these results suggest that when the traditional (classical) stimulation paradigm is used the compound energy-onset response dominates the pitch-related response so that there is little residual response capacity for pitch before a saturation level is reached. The implication of these results on future neuroimaging studies of pitch perception is that the use of a noise context (i.e. using a continuous stimulation paradigm) is advised in order to increase the pitch-related response, and reduce the likelihood of type II errors.

The second experiment of this PhD (Chapter 5) concerned the effect of pitch salience on the size of the pitch-related response. It was hypothesized that the pitch-related response in auditory cortex would increase in magnitude with increasing pitch salience. Previous evidence for a specific region of auditory cortex that is responsive to the effect of salience of the pitch percept has been unconvincing (Penagos et al. 2004, Griffiths et al. 2010). Chapter 5 considered the effect of pitch salience on fMRI representations for two different pitch stimuli. The results from this experiment identified a region of human auditory cortex that fulfilled two of the four criteria for a pitch center (aim 3 of this PhD). This region was located in PT, slightly more posterior to the pitch center suggested by previous neuroimaging studies (Patterson et al. 2002; Krumbholz et al. 2003; Penagos et al. 2004), and satisfied the pitch constancy criterion, and the criterion for the elimination of peripheral phenomena such as
distortion products on the basilar membrane. Although this region was differentially responsive to noise and to pitch, it could not satisfy the pitch specificity criterion because there were no conditions that contained a feature that was not present in the pitch stimulus. Results from this study did not provide evidence for increasing response to stimuli with increasing pitch salience, so it did not satisfy the criterion for covariation with pitch salience. From this, we conclude that pitch-specific responses may be only sensitive to the presence of pitch, and do not necessarily depend on pitch salience. Implications for further research are that the removal of this criterion from the definition of a pitch center should be considered.

The third experiment (Chapter 6) explored the second aim of this thesis, which was to determine the effect of non-pitch features on the pitch response. Many previous neuroimaging studies that proposed lateral HG as the probable site for a pitch center did so on the results obtained from a specific type of pitch-evoking stimulus. The stimulus in question is IRN, a stimulus that has been a popular choice for researchers in the field of pitch perception due to misconceptions that the salience of the pitch percept could be manipulated without affecting any other fundamental feature of the stimulus. Whilst it is true that more repetitions of the iterative process required to create IRN increases its pitch salience, the comparison of IRN with Gaussian noise does not constitute a controlled comparison. Recent evidence demonstrated that the iterative process used to generate IRN introduces slowly-varying spectro-temporal fluctuation whose fluctuation depth also increases with increasing number of iterations (Hall and Plack 2009; Barker et al. in preparation a; Barker et al. in preparation b). The third experiment introduced a novel auditory stimulus, IRNo, that was created by processing IRN to remove the pitch. Results from this study indicated
that the IRN-related response in auditory cortex was significantly reduced when IRNo was used as a comparison rather than Gaussian noise. This result suggests that non-pitch features within a pitch-evoking stimulus (such as slowly-varying spectro-temporal fluctuations in IRN) can have a significant effect on the pitch response, and stimuli that contain more than one auditory feature should use an appropriate comparison in order to reduce the contribution of non-pitch features to the pitch-related response.

The final experiment that makes up this thesis (Chapter 7) addressed the third aim of this PhD, which was to determine whether or not there was an area of auditory cortex that satisfied all four criteria for a pitch center. This experiment made use of the novel IRNo stimulus from the previous experiment as well as IRN, unresolved harmonic complexes and Gaussian noise to create a factorial design to investigate the two effects of pitch and modulation. Additionally, this experiment used different levels of pitch salience for the two pitch stimuli in order to test all four criteria to determine the most likely candidate for a pitch center. Once again, a pitch-responsive region was identified in PT. There was also a response to slowly-varying spectro-temporal fluctuations in this region, but there was no interaction between the two features and so it was concluded that the responses for pitch and for fluctuation were independent. Furthermore, the magnitude of the pitch response was greater for high-salience pitch stimuli than for low-salience pitch stimuli, but there was no evidence for a salience effect for fluctuation depth for IRNo. This region partially satisfied the pitch specificity criterion for a pitch center; there was a response to fluctuation in the pitch region but the two effects were not inter-dependent, so the use of an appropriately matched control stimulus could reduce or possibly eliminate the effects of non-pitch features.
The criterion for elimination of peripheral phenomena was met by filtering the pitch stimuli into a spectral region that contained only unresolved harmonics, and by adding a low-pass noise masker to mask low-frequency distortion products. The pitch-related response from Chapter 7 was common to the unresolved harmonic complex and the IRN, and was located close to the pitch response found for different pitch stimuli such as Huggins pitch (Chapter 4; Hall and Plack 2009), pulse trains (Chapter 5), pure tones, resolved harmonic stimuli and wideband stimuli (Hall and Plack 2009). Overall, the results for this experiment tentatively suggest a region in PT, close to the anterior border of Te 1.0 and Te 1.2, fits the profile for a human pitch center.

Implications and Directions for Further Research

There has been a great deal of interest in pitch perception recently, but results from neuroimaging research have so far been rather inconsistent. One of the conclusions of this thesis is that some of these inconsistencies might be due to methodological differences. The results reported in this thesis could therefore help to provide a gold-standard method for neuroimaging studies of pitch perception. Specifically, results from the experiments presented here suggest that the following paradigm design elements should be applied:

1. To increase sensitivity to pitch, a ‘continuous stimulation’ paradigm should be used whereby successive pitch stimuli are interspersed with noise to maintain a stable envelope and reduce the effect of multiple energy onsets to the pitch-related response.
2. A control condition that takes into account all of the non-pitch features (such as modulation for IRN) should be used to increase confidence that the residual response can be attributed to pitch.

3. A range of pitch-evoking stimuli with different spectral and temporal characteristics should be used to ensure that the pitch-related response is not stimulus specific.

4. To improve external validity and sensitivity in fMRI studies of pitch, analyses can be restricted to a region that encompasses peaks of pitch-related activity from these, and from previous studies.

It is apparent that the majority of pitch research has used low-frequency pitch stimuli to examine pitch responses within auditory cortex. However, for a site to be considered a pitch center, it would have to show a similar response to stimuli with high and with low f0s. Thus far, the stimuli used in these experiments have typically had an f0 of 250 Hz or less (e.g. Griffiths et al. 1998; Bendor and Wang 2005; Hall and Plack 2009; García et al. 2010, Barker et al. in preparation a, b). Thus it is suggested that a further criterion is added to the original pitch center criteria set out by Hall and Plack (2009) – a pitch value constancy criterion whereby the pitch response must be present for stimuli with a range of f0s.

Results from three of the experiments presented in this thesis (Chapters 5, 6 and 7) provide somewhat inconclusive evidence for a sensitivity to salience in pitch-responsive regions of auditory cortex. Therefore, further research is necessary to determine whether or not the pitch response is sensitive to pitch salience. To address this question, it is suggested that future research is undertaken using different pitch-
evoking stimuli that are matched in terms of (varying) pitch salience, such as simple and complex Huggins pitches and unresolved harmonic complexes filtered into different frequency regions and with different f0s. Based on the results presented here and on the results from previous pitch experiments, a revised list of criteria for a proposed pitch center to meet is as follows:

1. Pitch selectivity: it must be selectively responsive to pitch and not to a closely matched acoustic stimulus that does not evoke a pitch percept
2. Elimination of peripheral phenomena: the response must remain when the contribution from peripheral phenomena are removed
3. Pitch stimulus constancy: it must respond to all pitch stimuli, regardless of the physical generators of the pitch percept
4. Pitch value constancy: it must respond to all pitch stimuli, regardless of the f0.

Furthermore, corresponding evidence from different modalities (such as PET, EEG, MEG and electrode studies) would indicate how robust the sensitivity to salience is within the pitch-responsive region. Although there are inherent difficulties in comparing results across modalities, a consistent response across neuroimaging studies would increase confidence that the response is genuine.
References


