EXAMINING THE EVIDENCE FOR A PITCH CENTRE IN HUMAN AUDITORY CORTEX: A MULTI-METHOD APPROACH

KARIMA SUSI

A thesis submitted in partial fulfilment of the requirements of Nottingham
Trent University for the degree of Doctor of Philosophy

June 2016

DECLARATIONS

I declare that whilst registered as a candidate for the University's research degree I have not been a registered candidate or enrolled student for any other award of the University or other academic professional institution.

I declare that no material contained in the thesis has been used in any other submission for an academic award.

I declare that this research has been subject to ethical review and received ethical approval from the following Research Ethics Committees (College Research Ethics Committee (Nottingham Trent University) and Medical School Research Ethics Committee (University of Nottingham)).

I also declare that I have not deviated from the terms of the ethical approval issued by these Committees.

ACKNOWLEDGEMENTS

“It always seems impossible until it’s done” (Nelson Mandela). I would like to take this opportunity to thank everyone who made this thesis *possible*.

I thank first my supervisors Prof. Deb Hall, Dr Andrew Dunn and Dr Preethi Premkumar. Thank you all so much for your enduring patience, guidance, and enthusiasm over the years. I extend an extra special thank you to Prof. Deb Hall who has been my supervisor since my undergraduate degree. Thank you for always being on hand to offer your time, profound knowledge and remarkable ideas, for elucidating the method of fMRI, and for teaching me the skills needed to embark on my chosen career path. I will never forget your support!

I would also like to thank everyone who has been involved in various aspects of the experiments reported throughout this thesis. Thanks to all the technical staff at Sir Peter Mansfield Magnetic Resonance Centre (University of Nottingham) and Nottingham Trent University (especially Ben Thompson). Thanks to Dr Sue Francis and Dr Karen Mullinger from the Sir Peter Mansfield Magnetic Resonance Centre for generously providing their specific expertise in MR methodology (i.e., T2* mapping used in Chapter 4). For providing the stimuli used for the fMRI experiment (Chapter 5), I thank Prof. Andrew Oxenham (University of Minnesota). Thanks to Dr Daphne Barker (University of Manchester) for providing contrast images of her data (Barker et al., 2011; Chapter 5). A big thank you to Prof. Chris Plack (University of Manchester) for providing his psychophysical expertise and adapting his bespoke Earlab programme that allowed me to generate the stimuli used throughout Chapters 7, 8 and 9. For providing his advice and MATLAB code to calculate d’ for dual pair designs (Chapter 7), I thank Prof. Christophe Micheyl (University of Minnesota). Thank you to Prof. Thom Baguley (Nottingham Trent University) for providing guidance and an open-access R code to plot the appropriate confidence intervals for repeated-measures designs. To all the researchers in my field that I have had the pleasure of meeting at various meetings and conferences, thank you for taking an interest in my work and for your invaluable feedback. To the participants who were kind enough to volunteer (especially those who returned for more despite knowing what to expect), your enthusiasm was a much-needed breath of fresh air – I thank you all!

For keeping me partially sane throughout this process, I thank my family, my extended family and all my friends (especially Becky and Vikki) for providing infinite encouragement, laughter and, most importantly, good food and wine!

Last but not least, I dedicate this thesis to the most important people in my life: Mum, Dad, Tarik, Yasmine, Benjamin and Alastair. Your continual support makes everything I endeavour to do in the future seem *possible*. I will love you all forever.

TABLE OF CONTENTS

ABBREVIATIONS

2AFC	Two Alternative Forced Choice
2D	Two-Dimensional space
2I-2AFC	Two Interval-Two Alternative Forced Choice
3D	Three-Dimensional space
A1	Auditory Area 1 (corresponds to non-human primary auditory cortex)
ABR	Auditory Brainstem Response
AD	Analogue-to-Digital
ADC	Analogue-to-Digital Converter
ANOVA	ANalysis Of VAriance
AFNI	Analysis of Functional NeuroImages
Ag/AgCl	Silver/silver chloride
B0	Static magnetic field
BESA	Brain Electrical Source Analysis software
BOLD	Blood-Oxygen-Level Dependent
cm	CentiMetre
CMS	Common Mode Sense (active electrode)
CSF	CerebroSpinal Fluid
d’	sensitivity or discriminability index D-prime (used in signal detection theory)
dB	DeciBel
DC	Direct Current
DRL	Driven Right Leg (passive electrode)
DVD	Digital Versatile Disc
EEG	ElectroEncephaloGraphy
EOR	Energy Onset Response
EPI	Echo-Planar Imaging
ERP	Event-Related Potential
F0	Fundamental frequency
F	F statistic from an F-test

FA	False Alarm
FDR	False Discovery Rate
fMRI	Functional Magnetic Resonance Imaging
FWE	FamilyWise Error
FWHW	Full Width at Half Maximum
GA	Grand Average
GLM	General Linear Model
H	Hit
HG	Heschl’s Gyrus
HL	Hearing Level
Hz	HertZ
IC	Inferior Colliculus
IPC	Inferior Parietal Cortex
IRN	Iterated Ripple Noise (also known as RIN: Regular Interval Noise)
ISI	Inter-Stimulus Interval
kHz	KiloHertZ
LN	Natural Logarithm
LSD	Least Significant Difference
LQ	Laterality Quotient (Oldfield, 1971)
M_{age}	Mean Age
MEG	MagnetoEncephaloGraphy
MGB	Medial Geniculate Body
MHz	Mega HertZ
mm	MilliMetre
MNI	Montreal Neurological Institute
MR	Magnetic Resonance
ms	MilliSecond
N	Number of (e.g., participants or voxels)
nAm	Nano Amp Metres
OP	Parietal Operculum (Eickhoff et al., 2006)

p	Probability value or statistic used for testing a statistical hypothesis
P1, N1, P2	Transient deflections of the auditory event-related potential
PCs	Proportion of Correct responseS
PET	Positron Emission Tomography
POR	Pitch Onset Response
POR1	Pitch Onset Response for the first ‘adaptor’ pitch stimulus
POR2	Pitch Onset Response for the second ‘probe’ pitch stimulus
PP	Planum Polare
PT	Planum Temporal
R	Rostral area (corresponds to non-human primary auditory cortex)
RF	RadioFrequency
RFX	Random effects analysis
ROI	Region Of Interest
RT	Rostral Temporal area (corresponds to non-human primary auditory cortex)
s	Seconds
SD	Standard Deviation
SE	Standard Error
SENSE	SENSitivity Encoded
SNR	Signal-to-Noise Ratio
SOA	Stimulus Onset Asynchrony
SPL	Sound Pressure Level
SPM	Statistical Parametric Mapping
SPSS	Statistical Package for the Social Sciences
STG	Superior Temporal Gyrus
STS	Superior Temporal Sulcus
SVC	Small Volume Correction
T	Tesla
t	T statistic from a t-test
T1	Decay constant longitudinal magnetisation
T2	Decay constant transverse magnetisation

Te1.0	Central Heschl’s gyrus area (Morosan et al., 2001)
Te1.1	Medial Heschl’s gyrus area (Morosan et al., 2001)
Te1.2	Lateral Heschl’s gyrus area (Morosan et al., 2001)
Te3	An area of non-primary auditory cortex overlapping with parts of PP and PT (Morosan et al., 2001)
TE	Time to Echo (or Echo Time)
TR	Time to Repeat
V1	Primary visual cortex
V4	Visual area 4 (one of the visual areas in the extrastriate visual cortex)
x, y, z	Translational parameters used in fMRI for sagittal, coronal and axial planes, respectively
z	z-score
η_p^2	Partial Eta Squared
µV	Microvolts

PUBLICATIONS & CONFERENCE PRESENTATIONS

Hall, D. A., & Susi, K. (2015). Hemodynamic imaging of the auditory cortex. In G. G. Celesia & G. Hickok (Eds.), Handbook of Clinical Neurology (Vol. 129, pp. 257–275). http://doi.org/10.1016/B978-0-444-62630-1.00015-9

Susi, K., & Hall, D. A. (2012, September). *Optimizing echo time for high resolution auditory functional magnetic resonance imaging (fMRI).* Poster presented at the British Society of Audiology Annual Conference, Nottingham, UK.

Susi, K., Hall, D. A., Dunn, A., Premkumar, P., & Oxenham, A. (2012, September). *Re-examining the evidence for a representation of pitch salience in human auditory cortex.* Poster presented at the British Society of Audiology Annual Conference, Nottingham, UK.

Susi, K., Hall, D. A., Dunn, A., Premkumar, P., & Oxenham, A. (2013, February). *Re-examining the evidence for a representation of pitch and pitch salience in human auditory cortex.* Poster presented at the 38th Annual MidWinter Meeting for the Association for Research in Otolaryngology (ARO), Baltimore, USA.

Susi, K., Hall, D. A., Dunn, A., Premkumar, P., & Plack, C. (2014, September). *Searching for pitch invariant representations in auditory cortex.* Paper presented at the British Psychological Society Cognitive Psychology Section Annual Conference, Nottingham, UK.

LIST OF FIGURES

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Chapter 9

Figure 9.3. Mean d’ for pitch and timbre tasks across conditions for Experiment 12.

Figure 9.4. Mean d’ for conditions collapsed across pitch and timbre tasks for the pitch congruence and timbre congruence interaction term for Experiment 12.

LIST OF TABLES

Chapter 3

Chapter 5

Chapter 7

Chapter 8

Chapter 9

ABSTRACT

EXAMINING THE EVIDENCE FOR A PITCH CENTRE IN HUMAN AUDITORY CORTEX: A MULTI-METHOD APPROACH

Karima Susi

Doctor of Philosophy

NOTTINGHAM TRENT UNIVERSITY

This PhD used a combination of psychophysical, functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) methods to evaluate the evidence for a pitch centre within auditory cortex according to the four pitch criteria: 1) pitch selectivity, 2) pitch constancy, 3) covariation with pitch salience, and 4) accounting for confounding factors, that were described in an article by Hall and Plack (2009). An fMRI study re-examined pitch criteria 1, 3 and 4 using stimuli and a subtractive study design informed by Penagos et al. (2004), but extended this work by addressing some of their limitations. Results indicated that the representation of pitch compared to noise is widely distributed across auditory cortex, while the evidence for an effect of pitch salience was questionable given that the weak pitch salience condition was not significantly different from matched noise at a group level. These findings raise concerns regarding fMRI's sensitivity to pitch salience effects in the context of high individual variability. An ERP 'adaptation' study evaluated pitch criteria 1, 2 and 4 using pitch and timbre stimulus parameters that had been previously matched for discriminability, and sequences either varied in pitch, timbre or both across listeners. Findings from both sensor and source-based analyses suggested that pitch responses may be influenced by timbre (i.e., non-invariant); although further research is required. Thus, evidence failed to support the notion of pitch constancy at the level of the auditory cortex. Further studies using psychophysical listening paradigms continued this work, and again seemed to confirm a lack of evidence for pitch constancy. Reaction times and accuracy data demonstrated that timbre changes (non-target stimulus) interfered with listener's ability to discriminate pitch (target stimulus), and vice versa. Overall, these convergent results suggest that there is no modular representation of pitch (pitch centre), but rather pitch processing sites are distributed throughout multiple areas of primary and non-primary auditory cortex and are seemingly non-invariant to other stimulus parameters related to its perception (e.g., timbre). Under this assumption, the spatio-temporal model of pitch perception may best describe the neural mechanism underpinning pitch perception. Several recommendations are made to address challenges to interpretation identified throughout this PhD, which are likely to guide further research in this area.

Chapter 1. Thesis overview

Pitch is generally defined as “that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale” (American Standards Association, 1960), or somewhat more broadly, “that attribute of sensation whose variation is associated with musical melodies” (Plack, Oxenham, & Fay, 2005, p.1) Why do auditory researchers care so much about pitch? Pitch is arguably one of the most important perceptual attributes of any auditory stimulus. Pitch plays a fundamental role in the perception of music and speech, which are vital for defining, differentiating and communicating in our complex acoustic environment. Evaluation of the pitch research to date has shown that despite over 100 years of research on the psychophysics of pitch, and several decades of research on the neural basis of pitch, we still know relatively little about how we actually perceive pitch. This may be frustrating, but is also exciting in that the way it provides opportunities for researchers to make new major breakthroughs in the field. Unanswered questions include:

- What is the neural code which extracts pitch information from the incoming auditory signal?
- How is pitch represented perceptually in the central auditory system?
- Where in the auditory pathway does a ‘unified’ pitch representation first occur?
- What populations of neurons are involved (and are these modular or distributed)?
- Are these populations of neurons invariant to other attributes of sound such as timbre or spatial location?

The topic of interest for my PhD considers some of these unanswered questions, particularly questions about the nature of those neural populations (modular or distributed) and whether they are responsive to other attributes of sound, such as timbre or spatial location. My approach to studying the perceptual mechanisms of pitch and the underlying coding mechanisms combines both behavioural and neuroimaging methods (i.e., electroencephalography, EEG; functional magnetic resonance imaging, fMRI). This ‘three pronged’ attack enables me to examine questions about neural coding at the same time as gaining increased confidence in the interpretation of results; if findings are convergent across multiple methods.

This thesis builds on previous work published by the Nottingham group in collaboration with Prof. Plack, University of Manchester (e.g., Barker, Plack, & Hall, 2011; Garcia, Hall, & Plack, 2010; Hall & Plack, 2009). Hall and Plack (2009) defined four criteria for establishing the existence of a ‘pitch centre’ dedicated for coding this perceptual property. These have been used to inform and guide the design of my own PhD experiments and so they are described in detail below.

1.1. Pitch selectivity

The first criterion relates to pitch selectivity, where the cortical response to a pitch-evoking stimulus should be greater and distinct from a spectrally matched noise control baseline. The only difference between the baseline and the pitch stimulus should relate to the ‘pitch’. As illustrated in a number of neuroimaging studies (e.g., Hall & Plack, 2009; Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lütkenhöner, 2003) this criterion can be assessed experimentally by contrasting cortical responses to ‘pitch > noise’. Using fMRI as an example, if a region is involved in coding pitch rather than sound in general, it would be expected that significantly differential activity for pitch would occur.

1.2. Pitch constancy

The second criterion relates to pitch constancy, whereby cortical responses to pitch should be the same irrespective of their temporal, spectral, binaural, and perceptual characteristics (e.g., pitch value, timbre, loudness, spatial location; see Bizley, Walker, Silverman, King, & Schnupp, 2009; King & Nelken, 2009; Walker, Bizley, King, & Schnupp, 2011a; Walker, Bizley, King, & Schnupp, 2011b). Subsequently this criterion can be assessed experimentally by manipulating the timbre or spatial location of pitch-evoking sounds and observing the cortical responses. Given that the contrast between pitch and noise may reflect differential activity relating to a perceptual difference (i.e., timbre, loudness, spatial location), it is advisable to replicate experiments with a range of pitch-evoking stimuli, where possible (e.g., Huggins pitch, resolved, unresolved; see Hall & Plack, 2009).

1.3. Covariation with pitch salience

The third criterion relates to covariation with pitch salience. Pitch salience or strength refers to how distinctive the pitch percept is. Cortical sensitivity to pitch salience is best assessed using a parametric design (e.g., see Barker et al., 2011). Similar to loudness, pitch-selective neurons have been found to increase their neural firing rate in marmoset auditory cortex as a function of salience (Bendor & Wang, 2005, 2010). Human fMRI studies are inconclusive, but one study that used well-controlled stimuli did report to find a pitch salience representation in anterolateral Heschl’s gyrus (HG; Penagos, Melcher, & Oxenham, 2004). A number of recent electrophysiological experiments have found that the latency and amplitude of pitch-related responses are influenced by the salience of pitch stimuli, and these correlate to behavioural responses (e.g., Krishnan, Bidelman, Smalt, Ananthakrishnan, & Gandour, 2012; Krumbholz et al., 2003). However more recent fMRI evidence has failed to find a definitive representation of pitch salience in auditory cortex (e.g., Barker et al., 2011), casting doubt on whether fMRI is sensitive enough to detect small changes relating to pitch salience.

1.4. Accounting for confounding factors

The fourth criterion, relates to avoiding explanations of pitch-related activity that may be attributable to confounding factors such as cochlear distortion products and the use of an appropriate baseline control. Study designs should ensure that the only difference between pitch and baseline conditions relates to the pitch, especially when relying on subtraction analysis. Many neuroimaging studies have used a particular type of pitch stimulus known as iterated ripple noise (IRN). IRN is created by iteratively adding delayed broadband noise (Yost, 1996). The process gradually builds up temporal regularity in order to evoke a pitch stimulus with a frequency of 1/delay. IRN has become a popular choice of stimulus because the greater the number the iterations, the more salient the pitch percept is.

These studies have several limitations in their interpretation of the results ascribing the responses specifically to pitch perception. First, the synthesis of IRN either includes pitch information corresponding to the low-frequency harmonics, fails to mask the energy in the signal in this cochlear region, and/or there is uncertainty as to whether the contribution of distortion products can be completely ruled out in some studies (e.g., Bendor & Wang, 2005, 2010; see Abel & Kössl, 2009; Briley, Breakey, & Krumbholz, 2013; McAlpine, 2004; Plack, Barker, & Hall, 2014). For instance, when one wants to filter a 200 Hz complex tone into a 1000-2000 Hz spectral region (i.e., harmonics 5-10), it is important to mask the spectral region encompassing the lower frequency harmonics (including F0, i.e., harmonics 1-4) with noise otherwise these can be reintroduced back into the stimulus percept by distortions in the cochlea, giving a percept of a 200 Hz tone that has been filtered using harmonics consisting of a 200-2000 Hz spectrum (i.e., harmonics 1-10; Pressnitzer & Patterson, 2001). Second, the synthesis of IRN introduces slow spectro-temporal modulations in the noise (Plack et al., 2014), which is unrelated to pitch and is not present in the noise control stimulus (see Barker, Plack, & Hall, 2012, 2013; Hall & Plack, 2009; Steinmann & Gutschalk, 2012). Although it has been suggested that the cortical responses reported in anterolateral HG for IRN stimuli may be driven by spectro-temporal modulations present in the stimulus (e.g., perceived as a timbre change), these longer–term modulations are not likely to affect transient EEG and magnetoencephalography (MEG) responses because the responses set in milliseconds (ms) after the stimulus onset and do not give the longer-term modulations time to unfold (Briley et al., 2013). This has been recently confirmed by Steinmann and Gutschalk (2012) who found that only sustained fMRI and MEG theta responses were sensitive to slow spectro-temporal fluctuations present in IRN. Nevertheless, these results highlight the inherent limitations of using subtractive methodology and the need to precisely match baseline controls to conditions of interest.

1.5. Aims of PhD

The main aim of my PhD was to evaluate the evidence for a pitch centre within auditory cortex, focusing on the four pitch criteria described above, using a combination of psychophysical, fMRI and EEG approaches. I was especially interested in exploring whether the neural representation of pitch is modular or distributed, and whether this representation is invariant to other stimulus features, such as timbre, at the level of auditory cortex.

1.6. Thesis Overview

Chapter 2 describes some of the key concepts underpinning pitch research.

Chapter 3 is primarily comprised of an original manuscript submitted in contribution to an invited book chapter which was co-authored with my PhD co-supervisor (see Appendix for the final version). I contributed to approximately 50% of the writing of Hall & Susi (2015). This chapter provides a comprehensive and focused reference on the neuroscience of hearing and the associated neurological diagnosis and treatment of auditory disorders, and is primarily aimed at students and researchers who are not experts in the field.

Chapter 4 describes a short series of feasibility studies, Experiments 1, 2, and 3, aimed at determining the optimal echo time (TE) and voxel resolution for the auditory fMRI experiment described in Chapter 5.

Chapter 5 describes a pitch-salience rating task (active listening), and an fMRI passive listening experiment (Experiments 4 and 5, respectively) which evaluated pitch criteria 1, 3 and 4. These experiments investigated the neural representation of pitch and pitch salience using harmonic complex tones that were either resolved (strongly pitch salient) or unresolved (weakly pitch salient). I used the same stimuli and a subtractive study design informed by Penagos et al. (2004), but extended this work by addressing some of their limitations. These limitations are discussed in Chapter 5. Contrasts of interest included sound > silence, pitch > noise, strong pitch salience > weak pitch salience, and weak pitch salience > matched control noise. The location, magnitude and distribution of activity were analysed accordingly. Findings indicated that the representation of pitch compared to noise is widely distributed across auditory cortex, and the evidence for pitch salience was questionable, but was also co-localised to the same regions as the pitch response. The findings raise concerns regarding fMRI's sensitivity to pitch salience effects in the context of high individual variability.

Chapter 6 comprises the event-related potential (ERP) and adaptation methodology chapter which briefly describes the main analysis approaches used for Experiment 10 in Chapter 8. This chapter is much shorter than the corresponding fMRI Chapter 3 because it was written specifically for Experiment 10 and not as a pedagogic chapter for teaching purposes.

Chapter 7 describes of series a feasibility studies aimed at matching discriminability for pitch and timbre stimulus parameters. Experiments 6-8 focused on identifying pitch and timbre parameters that were equally discriminable.

Chapter 8 uses EEG methodology to evaluate the criteria 1, 2 and 4. Experiment 9 used the chosen parameters identified in Chapter 7 to match discriminability for pitch and timbre discrimination tasks across listeners for the corresponding ERP Experiment 10. The findings showed that discriminability between pitch and timbre tasks were well-equated across experiments. Experiment 10 was primarily aimed at investigating pitch constancy in relation to pitch's invariance to timbre, using a novel adaptation design. Here the method and paradigm were changed because, as described in Plack et al. (2014), fMRI is known to be insensitive to neuronal populations which encode pitch in terms of their 1) temporal firing patterns, 2) relative rates of responses at a neuron-to-neuron level, and 3) within neuronal assembles that are distributed among other neurons that encode other sound attributes (e.g., timbre), particularly when using conventional subtraction designs (see Chapter 3). Therefore an ERP adaptation design was employed here as it is a direct measure of neural activity and can be used to address the three points outlined above. Harmonic complex sequences were used which transitioned from noise to two pitch tones (adaptor and probe). Sequences either varied in pitch, timbre or both. Findings suggested that pitch is non-invariant to timbre, and thus failed to find evidence of pitch constancy at the level of the auditory cortex using this experimental approach, although further research is required.

Chapter 9 uses a dual-pair (four-interval same-different) discrimination task to further investigate pitch constancy, more specifically, evidence for pitch invariance to timbre. Experiment 11 used the chosen parameters identified in Chapter 7 to match discriminability for pitch and timbre discrimination tasks across listeners for the corresponding reaction time Experiment 12. The findings showed that discriminability between pitch and timbre tasks were well-equated across experiments. Experiment 12 uses the same four conditions used in Chapter 8, and the same stimuli as that reported in Chapter 7. Findings indicated that timbre changes (non-target stimulus) interfered with listener's ability to discrimination pitch (target stimulus), and vice versa. These results therefore also failed to find evidence for pitch constancy.

Chapter 10 summarises the main thread of the argument regarding the four pitch criteria that has built up through the thesis chapters, highlighting any key areas for discussion and suggested future directions for the field.

Chapter 2. Key concepts in pitch research

This chapter describes some of the key concepts underpinning pitch research to provide a scientific context for my thesis work. Pitch is determined by physical acoustic cues that repeat over time. Pure tones consist of a sinusoidal variation over time, where the repetition rate of a pure tone is identical to its spectral frequency (Plack et al., 2005, p. 8-9). However, pure tones are extremely rare in our acoustic environment, with musical melodies and voiced speech being much more common and consisting of harmonic complex tones.

2.1. Importance of pitch

In music, pitch sequences produce musical melodies and concurrent combinations of pitch give rise to harmonies (Oxenham, 2012). In speech communication, pitch provides cues for sound source segregation (i.e., stream segregation; Darwin & Carlyon, 1995) and identification, allowing us to separate out different sounds that occur at the same time (e.g., two people talking simultaneously), and group together sounds that originate from the same acoustic source (e.g., sentence spoken by same person; Oxenham, 2008). Variations in the pitch contours of vowel sounds can also be used to convey prosodic information in languages such as English (where a rising intonation at the end of a sentence may indicate an interrogative statement), or semantic information in tone languages such as Mandarin (where changing the pitch contour of a word can change its meaning entirely; Hall & Plack, 2009; Plack et al., 2005, p. 2-3).

In our everyday lives, the majority of people with normal hearing are able to perceive music and speech effortlessly, even in the noisiest of places. Whilst such hearing abilities appear effortless, the auditory neuroscience community knows surprising little about how the peripheral and central auditory systems actually achieve such a seamless perception of our acoustic world. Subsequently, gaining a greater understanding of pitch perception, and sound more generally, does not just have scientific or academic importance, but also has several practical translational benefits that may eventually benefit clinical populations. For instance, hearing aids and cochlear implants for hearing loss are generally found to have poor frequency selectivity which evades sufficient coding of harmonic pitch sounds given that they are unable to accurately represent individual harmonics (McDermott, 2004; Oxenham, 2008). The same can be said for individuals with brainstem implants and individuals with pitch disorders, such as congenital amusia. In time, understanding how pitch is represented in normally hearing listeners is therefore likely to aid development and improvements to future auditory prostheses.

2.2. The psychophysics of pitch

The choice of stimuli used in my PhD experiments was guided by knowledge about the physics of sound and psychophysical models of pitch coding. Harmonic complex tones are one of the most common pitch-evoking sounds used in research experiments because they can be well controlled. A harmonic complex tone is a periodic waveform that repeats at the rate corresponding to the fundamental frequency (F0; see Figure 2.1; Oxenham, 2012). Harmonic complex tones are therefore essentially a series of pure tones, in which the frequency of each sinusoidal harmonic component or overtone is a multiple integer of the F0. For example, for a complex tone with an F0 of 100 Hz, the first harmonic is 100 Hz, the second harmonic is 200 Hz, and the third harmonic is 300 Hz etc. Generally, one perceives a harmonic complex tone as a single unified sound because the harmonics are ‘perceptually fused’, for instance, it only takes two consecutive harmonics to evoke a pitch perception (Plack et al., 2014).

Harmonic complex tones are often referred to as either being ‘resolved’ or ‘unresolved’. These distinctions relate to whether the harmonics are individually represented by a single auditory filter (resolved) or not (unresolved) on the basilar membrane. Generally, components up to around the 10th harmonic of a complex tone are said to be resolved (Glasberg & Moore, 1990; Shackleton & Carlyon, 1994). For resolved harmonics, the excitation pattern displays a series of marked peaks at filters with centre frequencies (Oxenham, 2008). Conversely, at increasingly higher frequencies, the bandwidth of filters begins to exceed the spacing between adjacent harmonics (Oxenham, 2008). These harmonics interact with multiple filters and so the peak responses across filters become gradually less distinct and eventually disappear altogether (unresolved). Corresponding waveforms appear extremely complex, due to the interaction effects of multiple harmonics, with a temporal envelope repeating at the F0 rate. Subsequently, tones comprised of resolved harmonics elicit a much stronger pitch perception (i.e., greater in pitch salience) than unresolved harmonics.

A number of pitch perception models of complex pitch perception are based on knowledge of such excitation pattern responses (Cohen, Grossberg, & Wyse, 1995; Goldstein, 1973; Terhardt, Stoll, & Seewann, 1982; Wightman, 1973; see Oxenham, 2008, 2012). The most common pitch coding theories based on computational models which relate to the representation of sound within the auditory nerve, include a ‘place’ code (Goldstein, 1973), a ‘temporal’ code (Licklider, 1951), and more recently a synthesis of the two, resulting in a ‘hybrid’ spatial-temporal code (Loeb, White, & Merzenich, 1983; see Oxenham, 2012 for a review).

Figure 2.1. ***Third party copyright material removed.*** *Schematic illustration of the relationship between acoustic frequency and periodicity and perceived pitch for pure and complex tones. A pure tone (top panels, blue) and a harmonic complex tone (lower panels, red) are shown, which both produce the same pitch (A above middle C or 440 Hz). This figure has been taken from Oxenham, Micheyl, Keebler, Loper, & Santurette (2011).*

2.2.1 Place model of pitch coding

According to the “place” or “rate-place” model (von Helmholtz, 1948), the pitch of a given tone is determined by the specific auditory nerve fibre, and hence which particular place on the basilar membrane, is maximally excited (Plack & Oxenham, 2005). Figure 2.2A shows how for the low frequency tone (top row) the largest vibration along the basilar membrane is at position 1, and corresponding action potentials only occur for auditory nerve fibres connected to position 1. The converse is true for the high frequency tone. Peaks identified in the excitation pattern provide information regarding which frequencies are present and this information can later be combined to calculate the underlying F0 (Oxenham, 2008), for example, by matching spectral information to pre-existing ‘harmonic templates’ (Cedolin & Delgutte, 2005; Cohen et al., 1995; Goldstein, 1973; Schroeder, 1968; Terhardt, Stoll, & Seewann, 1982; Wightman, 1973).

Figure 2.2. ***Third party copyright material removed.*** *A schematic illustration of auditory nerve firing in response to two different acoustic signals. Panels A) and B) depict the pattern of firing along the auditory nerve according to a place or temporal model, accordingly. Adapted from Heeger (n.d.).*

2.2.2 Temporal model of pitch coding

According to the “temporal” model, the pitch of a tone is determined by the timing of action potentials, or spiking rate, that occurs in the auditory nerve fibres (Cariani & Delgutte, 1996; Licklider, 1951; Meddis & O’Mard, 1997; Schouten, 1962). The location of activity along the basilar membrane is less relevant. Neurons fire in synchrony with the phase of the acoustic waveform (Plack et al., 2005, p.11-13), for instance, tones with low frequencies (typically <~4-5 kHz; Oxenham, 2012), spikes are more likely to occur at one phase in a sinusoid cycle, than for another. This is called phase locking because the response of the neuron is locked to a particular phase of stimulation (i.e., vibration of basilar membrane) and the time between pairs of spikes are likely to be multiple integers of the period of the sinusoid (see Rose, Brugge, Anderson, & Hind, 1967; Oxenham, 2008). For instance a 500 Hz pure tone with a 2ms period will produce spikes sequentially separated by 2ms, 4ms, 6ms, 8ms and so on. Figure 2.2B shows how unlike the place code, tones with high and low frequencies evoke responses at both positions 1 and 2, but there is more firing in response to the high frequency tone in the lower panel.

2.2.3 Spatio-temporal ‘hybrid’ model of pitch coding

According to the “spatio-temporal” model, the pitch of a tone is determined by both place and time information (Carlyon, Long, & Micheyl, 2012; Cedolin & Delgutte, 2007, 2010; Larsen, Cedolin, & Delgutte, 2008; Loeb, White, & Merzenich, 1983; Oxenham et al., 2011; Shamma, 1985; Shamma & Klein, 2000). Specifically, as described by Oxenham (2008), this theory posits that, in the auditory nerve, temporal information is used to extract pitch, however it must be presented and encoded in the appropriate place (i.e., tonotopic location) along the basilar membrane (Oxenham, Bernstein, & Penagos, 2004; Shamma, 2004). For a travelling wave in the cochlea, different parts of the basilar membrane will be in different phases of the acoustic waveform sinusoidal cycle, whereby the rate of change in phase is generally rapid near the peak of stimulation. However the specific frequency of the sound stimulus determines the precise locations along the basilar membrane that are in or out of phase with one another. Such patterns in phase differences could then potentially be used by the auditory system to extract F0 (Oxenham, 2008).

2.2.4 Psychophysical evidence for pitch perception models

Psychophysical studies have attempted to provide support for the proposed pitch models. While both appear to play roles in pitch perception, auditory research to date has been unsuccessful in confirming one pitch model over another as providing a complete explanation of pitch perception (Oxenham, 2008; Oxenham et al., 2011). For instance, the precision of the temporal code deteriorates as it travels through the auditory system (Plack & Oxenham, 2005). Once the signal has reached the inferior colliculus the maximum level of phase locking is approximated to be around a few hundred Hz (rather than a few thousand; Plack et al., 2005). This casts doubt on a purely temporal code.

In summary, resolved harmonics may be represented by a place or a temporal code (or both). Unresolved harmonics can only be explained by a temporal code because the long term excitation pattern provides no cues (Schouten, 1962; see Plack et al., 2014). It is still unclear what type of code is used to represent pitch in the central auditory system (Oxenham, Micheyl, & Keebler, 2009; Oxenham et al., 2011).

Over the last few decades, the temporal model has been favoured over the place model. A number of psychophysical studies have shown that high-numbered unresolved harmonics elicit a much less salient pitch than low-numbered resolved harmonics (Bernstein & Oxenham, 2003; Houtsma & Smurzynski, 1990; Oxenham, 2012). Moreover, unresolved, but not resolved harmonics, are susceptible to phase distortions caused by acoustics of the room which interfere with the waveform of these complex tones (Qin & Oxenham, 2005;

Oxenham 2012). Recent findings from Oxenham et al. (2011) have called into question a purely temporal code in favour of a place or spatial-temporal code. Oxenham and colleagues were able to evoke a clear pitch percept corresponding to an F0 of 1.2 kHz using tones whose harmonics components were above the frequency range limit of phase locking (i.e., >5 kHz). This implies that high frequencies at least cannot be represented by a purely temporal code or that phase locking extends much higher than previously thought. However, high-numbered harmonics that are too close together to be represented by a particular place on the basilar membrane can still evoke a musical pitch, suggesting that a purely place code cannot be used in such instances.

2.3. Hierarchical framework for conceptualising pitch coding

The perception of pitch coding can be conceptualised in the framework shown in Figure 2.3: i) Sensation of the physical signal and its transduction into neural impulses, ii) Perception of sound attributes, and ii) Interpretation of the sound (i.e., cognition). This section explains this framework in more detail, paying particular attention to pitch perception.

Figure 2.3. Hierarchical framework for conceptualising pitch coding.

2.3.1 Pitch sensation and transduction

The challenge for the auditory system is to sense the external sound energy from both ears, receive, transform and combine the information into meaningful neural signals and interpret it so that it can be used to guide future behaviour (Nelken, 2008).

The sounds we hear in our everyday environment enter and travel through our ear canal as a physical stimulus before being transduced into the electrical signal that is carried and perceived by the central auditory system. The major hearing sense organ responsible for this is the cochlea within the inner ear, where the sound energy is transduced into neural signals that pass through the auditory nerve. This section briefly explains those processes and explains how they are relevant for pitch coding.

The acoustic waves are captured by the pinna and travel through the ear canal causing vibration at the ear drum. Sound waves are transformed into pressure variations in the cochlea fluid by the ossicles (three tiny bones) within the middle ear. The cochlea consists of two membranes, the Reissners membrane and the basilar membrane. The different frequency components of a given sound cause different parts of the basilar membrane to vibrate. Each place on the basilar membrane has a characteristic frequency which it maximally responds to (i.e., a series of auditory filters; optimal characteristic frequency excitation extends throughout auditory pathway). The cochlea can therefore be thought of as a band-pass filter, where high frequency sounds excite areas towards the base of the basilar membrane, and low frequency sounds excite areas towards the apex of the basilar membrane. This forms the basis of the place code where frequency-to-place, or tonotopic mapping, is preserved throughout the auditory pathway, from the cochlea to the auditory nerve and up to at least the primary auditory cortex (see Figure 2.4).

Figure 2.4. ***Third party copyright material removed.*** *Schematic diagram illustrating the hierarchy of sound processing. This diagram represents the ascending pathways only, and has been taken from Nelken (2008).*

The Organ of Corti is a specialised structure containing hair cells that is located on the inner surface of the basilar membrane of the cochlea. The excitation vibrations in the Organ of Corti causes a ‘shearing’ motion (up and down) between the basilar and tectorial membranes, where tiny hair cells (stereocilia) of the basilar membrane sway side-to-side at the same rate, causing depolarisation in the inner hair cells leading to a release of neurotransmitters that facilitate action potentials along the auditory nerve (Plack, 2005, p.72-73). The design and formation of stereocilia are vital for transduction. Each bundle of stereocilia is arranged in a staircase like formation, where each row is organised in ascending height. Each stereocilia contains actin filaments where the positive end is at the ‘tip’ of the structure, and the converse for the negative end. Rows of stereocilia are connected to one another by filamentrous structures known as ‘tip links’. These tip links run from the top end of a given stereocilia to its taller neighbour, and resemble a ‘slinky’. When there are stretched towards scala media (i.e., towards the outside of the cochlea), this causes cation selective channels in the membrane of the sterocilia to open causing potassium (K+) and some calcium (Ca2+) ions to flow into the hair cell, causing the electric potential of the hair cell to become more positive (known as depolarisation because typical resting state of cell is -45 mV). Depolarisation causes a neurotransmitter to be released in the gap (synapse) between the stereocilia and the auditory nerve, and causes action potentials or firing in spiral ganglion neurons (neural spiking). When sterocilia are bent in the opposite direction, these ion channels remain closed and therefore depolarisation does not occur. This is where physical pressure waves are transduced into electrical impulses that travel along the auditory nerve fibres (transduction). Each auditory nerve fibre represents a particular place on the basilar membrane, so each nerve fibre also has a characteristic frequency it is most sensitive to (i.e., place code, where nerve fibres at centre of the auditory nerve represent low frequency sounds, and nerve fibres in periphery represent high frequency sounds).

Although a given sound stimulus will cause the sterocilia to bend from side to side, the electric potential change in hair cells only occurs when the stereocilia are bent in the correct direction (towards the outside of the cochlea), and therefore the spiral ganglion neurons will only fire during a particular *phase* of the acoustic waveform (i.e., particular phase in the vibration of the basilar membrane), which is known as phase locking. For any given sound, the peak of activity in the nerve fibre occurs for onset of the stimulus, and then declines with time.

2.3.2 Pitch perception

Listeners experience a unified percept of pitch, irrespective of whether the stimulus contains resolved or unresolved harmonics, and irrespective of the neural code used in its transduction. This section reviews evidence for perceptual invariance indicating that there must be an abstract ‘higher level’ representation somewhere in the auditory system.

2.3.2.1 Models of pitch perception - supporting evidence

A pitch percept is only evoked when the periodicity is between approximately 30 Hz to ~4000/5000 Hz (Krumbholz, Patterson, & Pressnitzer, 2000; Oxenham, 2012; Pressnitzer, Patterson, & Krumbholz, 2001). It has been well established that even when the lowest harmonic component (i.e., the F0) is removed or masked by noise in a stimulus, the tone will evoke the same pitch percept (Fletcher, 1924; Licklider, 1954; Pressnitzer & Patterson, 2001; Schouten, 1940). This is known as the ‘missing fundamental’ phenomenon (see Figure 2.5C), and suggests that based on information obtained from higher numbered harmonic components, the auditory system can extract and reinstate F0 (e.g., harmonic templates; Houtsma & Goldstein, 1972; Oxenham et al., 2011).

Figure 2.5. ***Third party copyright material removed.*** *Schematic representation of pitch invariance. For instance Panel C illustrates the ‘missing fundamental phenomenon’ where the same pitch percept (e.g., 200 Hz) is evoked when the lowest harmonic (F0; present in Panels A and B) is absent. All panels also depict the relationship between pitch and timbre, where one sound can have the same pitch but a different timbre or spectral frequency composition (A, B and C), or vice versa (C and D). Panels C and D depicts stimuli that have different pitch but similar timbre because they have similar spectra. Panel E: Different spectral distributions of tones played on a violin and piano which share the same F0. Panels A-D have been adapted from Langner, Sams, Heil, & Schulze (1997), and Panel E has been adapted from Shamma (2004).*

Tones also evoke a common pitch percept even when they are composed of markedly different physical and perceptual characteristics (Plack et al., 2014). For example, separated harmonic information delivered to the two ears can be combined into a unified percept (Houtsma & Goldstein, 1972). For instance, if you present one ear with a 400 Hz pure or complex tone, and the other with a 600 Hz pure or complex tone, this will evoke a pitch percept of a 200 Hz tone (Deutsch & Roll, 1976). If pitch is encoded by place alone than how can two different sounds be combined to yield a single pitch percept? This result indicates that harmonics must somehow be combined physiologically at least at, or above, the level of the superior olivary complex (Plack et al., 2014).

Pitch can also be evoked using two sounds presented to each ear respectively that do not contain any pitch information. This is known as binaural or dichotic ‘Huggins pitch’ (Cramer & Huggins, 1958). It is created by presenting the same random signal (broadband noise) to each ear except for a narrow frequency band in which the noise is different across the two ears (decorrelated; Gockel, Carlyon, & Plack, 2011). Combining correlation patterns between the two ears means that listeners report hearing a musical pitch within the centre of the narrow frequency band (Akeroyd, Moore, & Moore, 2001). Gockel et al. (2011) showed that a pitch can be evoked using two Huggins pitch harmonics (e.g., decorrelated bands with centre frequencies of 600 and 800 Hz, evoke a 200 Hz pitch). Furthermore, they also demonstrated it is possible to combine a Huggins pitch harmonic with a conventional complex tone harmonic to produce a pitch, suggesting a combination of harmonics is likely to occur at the level of the inferior colliculus or later. Their findings cast doubt on a purely temporal model because, as mentioned earlier, higher numbered harmonics with frequencies above around 2 Hz are poorly represented at this stage in the auditory system.

Pitch perception also appears to have some perceptual invariance with other perceptual attributes of sound, such as timbre and loudness, such as two tones will evoke the same pitch percept as long as the F0 is the same or masked, irrespective of the timbre or loudness (Langner et al., 1997; McDermott, Lehr, & Oxenham, 2008). Timbre “allows one to distinguish among tones having the same pitch, loudness and duration” (American National Standards Institute, 1994). Timbre refers to the ‘quality’ of a tone and is determined by the spectral envelope (distribution of energy over frequency); see Figure 2.5, Panels C and D (Seither-Preisler et al., 2007; Shamma, 2004). Consequently, a piano and violin can play the same pitch, but each musical instrument has a particular timbre. This allows one to distinguish that the same note played on each musical instrument has been generated by different sources. Generally speaking, individuals find it very difficult to ignore timbre changes, and this has also been demonstrated to be equally the case for musicians (and non-musicians alike) when participants are asked to discriminate small changes in pitch when the timbre of the pitch tones dramatically varies (Moore & Glasberg, 1990; Borchert, Micheyl, & Oxenham, 2011). Although pitch and timbre are independent sound cues, given that two tones can sound very different despite sharing the same pitch (F0) illustrates how timbre may also be inherently related to the neural mechanisms responsible for encoding pitch (Langner et al., 1997; McDermott et al., 2008), and place (spectrum) and temporal (F0) information may be important for providing listeners with a unified pitch percept. It also raises the question as to whether there is a unified module in the brain that is responsible for such invariant representations. This possibility is discussed further in the following section.

2.3.2.2 Neurophysiological evidence for pitch invariance and the spatio-temporal model

Precisely where in the central auditory system information is combined to form a pitch is still unknown. Any of the ascending brainstem and midbrain nuclei are potential candidates. Animal electrophysiological recordings have located neurons in auditory cortex which appear to be selective to F0, irrespective of the stimulus presented. Using single-unit recording, a seminal study by Bendor and Wang (2005; see also 2010) investigated the neural representation of pitch responses in marmoset auditory cortex. Through presenting pure tones and missing fundamental harmonic complex tones they identified a number of pitch-selective neurons in the anterior lateral border of area A1 (primary auditory cortex; see Figure 2.6). This led them to suggest that perhaps a purely pitch-specific mechanism for pitch perception exists, where there are specific neurons that represent pitch regardless of its non-pitch features. Human neuroimaging investigations have attributed analogous pitch-specific neurons to a region of the human auditory cortex called lateral Heschl’s gyrus (HG; e.g., Briley, Breakey, & Krumbholz, 2013; Butler & Trainor, 2012; Griffiths, Büchel, Frackowiak, & Patterson, 1998; Griffiths, Uppenkamp, Johnsrude, Josephs, & Patterson, 2001; Gutschalk, Patterson, Rupp, Uppenkamp, & Scherg, 2002; Gutschalk, Patterson, Scherg, Uppenkamp, & Rupp, 2004; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002; Penagos et al., 2004; Puschmann, Uppenkamp, Kollmeier, & Thiel, 2010; Ritter, Dosch, Specht, & Rupp, 2005).

Figure 2.6. ***Third party copyright material removed.*** *Pitch-selective coding in marmoset auditory cortex. (A) shows the distribution of sound-evoked responses, where pitch selective neurons (black squares) can be seen in border between primary auditory cortex (area A1) and belt region (area R) of marmoset monkey (Bendor & Wang, 2005).*

2.3.3 Pitch cognition

Cognition refers to the mental manipulation of pitch which supports a range of listening tasks such as the recognition of familiar melodies, discriminating changes in pitch contour over time, and perceiving speech.

Whereas pitch perception typically involves bilateral areas of A1, tasks requiring higher-level pitch processing typically involve non-auditory regions and may be more lateralised. This has been described by some as a processing hierarchy (Nelken, 2008); shown in Figure 2.4. The human auditory regions showing spatial layout and terminology used for regions of interest throughout this thesis are shown in Figure 2.7. FMRI findings have allowed auditory neuroscientists to map the auditory hierarchy, particularly using pitch processing tasks. For instance, Warren & Griffiths' (2003) fMRI study investigated the brain regions involved in processing pitch and spatial location changes in pitch sequences using IRN stimuli with fixed and variable F0. Sound sequences with changing pitch activated non-primary auditory regions (lateral HG, anterior planum temporale (PT; immediately posterior to HG), planum polare (PP; anterior to HG). Sound sequences with changing spatial location specifically activated more posteriomedial parts of PT. This dissociation within parts of the belt and parabelt area of auditory cortex (Figure 2.4) provides some evidence for distinct processing mechanisms for pitch sequences with changing pitch and spatial location in humans (Warren, Uppenkamp, Patterson, & Griffiths, 2003; Zatorre, Belin, & Penhune, 2002).

Figure 2.7. ***Third party copyright material removed.*** *Panel 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 gray region, and how belt and parabelt regions might be organised is shown in dark (belt) and light gray (parabelt) shading. Panel B: Microanatomical structure of Heschl’s gyrus (see Morosan et al., 2001). This figure has been adapted from Hall and Barker (2012).*

Patterson et al. (2002) found that while fixed pitch sequences activated lateral HG when compared with spectrally matched noise, melodic-type pitch sequences differentially activated areas extending beyond HG, namely anterior activation in PP and superior temporal gyrus (STG). They also found evidence suggesting that melodic sequences should engage right hemisphere processing more than left (right hemisphere advantage for melody). They claimed that this is evidence in favour of a pitch processing hierarchy where low-level pitch perception engages regions in and near to primary auditory cortex, while more complex pitch listening engages regions further away across auditory cortex.

2.4. The modularity of pitch perception

Fodor’s theory of modularity (1983) is the most well-known modular theory of perception and cognition, and argues that certain psychological processes are ‘self-contained’ or modular. Specifically modular systems must fulfil certain properties, namely, domain specificity, encapsulation, mandatory operation (automaticity), inaccessibility to consciousness, speed, shallow outputs, fixed neural localisation, and characteristic breakdown patterns. It is clear that some of these criteria are similar to, and likely underpin, the pitch criteria identified in Chapter 1. The question of modularity has been well-studied in the visual system. There are purported to be a number of distinct neural regions that encode different perceptual visual features. For instance, in the primary visual cortex (V1) there are specialised neurons responsible for detecting the orientation of edges (Hubel & Wiesel, 1962), while V4 is the main visual colour processing area (Zeki, 1983). The evidence for a modular organisation in the auditory system has been less well substantiated. In some respects, frequency can be considered as having a modular organisation in primary auditory cortex (i.e., tonotopy; Humphries, Liebenthal, & Binder, 2010). Nevertheless, evidence relating to a modular organisation of other sound attributes, such as pitch, still remains highly debated (Bendor, 2012). The following section describes neuroimaging support for localised responses to pitch in human auditory cortex.

2.4.1 ‘Modular’ sustained pitch-sensitive responses in lateral HG?

A number of human positron emission tomography (PET), fMRI, EEG and MEG studies have localised sustained pitch responses to a brain region in lateral HG (e.g., Hall, Barrett, Akeroyd, & Summerfield, 2005; Griffiths et al., 1998, 2001; Patterson et al., 2002; Penagos et al., 2004; Puschmann et al., 2010; Seither-Preisler, Patterson, Krumbholz, Seither, & Lütkenhöner, 2006; Steinmann & Gutschalk, 2012). Many of these studies have used IRN as a pitch-evoking stimulus. PET and fMRI studies that attribute pitch activity to lateral HG typically report that this region is maximally activated for pitch versus other sounds such as a spectrally matched noise control (e.g., Griffiths et al., 1998, 2001; Patterson et al., 2002; Puschmann et al., 2010).

2.4.2 ‘Modular’ onset pitch-sensitive responses in lateral HG?

The energy onset of any sound stimulus creates a large transient neural response that can mask more subtle stimulus-specific transient responses (e.g., Garcia et al., 2010). For example, an MEG study by Krumbholz et al. (2003) used the ‘continuous stimulation paradigm’ to present a novel sound sequence that progressed from spectrally matched noise to an IRN stimulus (see Chapter 6 for more detail). This allowed them to identify and isolate a pitch-onset response (POR) which occurred approximately 150ms after pitch onset from the N100m component evoked for the energy onset of any sound stimulus. The authors localised PORs to the medial portion of HG which they argued was in support of HG being a ‘pitch centre’. Chait, Poeppel, and Simon (2006) also conducted an MEG study using the continuous stimulation design, and found that two different types of pitch stimulus (i.e., Huggins pitch and tone in noise) evoked similar pitch-onset responses with a latency of ~150ms after stimulus onset in HG. The authors postulate that this is further evidence in favour of an all-inclusive ‘pitch centre’ (also see Butler & Trainor, 2012; Ritter et al., 2005). An intracranial recording study by Schönwiesner and Zatorre (2008), that used depth electrodes, have confirmed separate neural sources for noise and pitch onsets, with pitch onsets generally being more consistently associated with activity in lateral HG.

A more recent fMRI study by Steinmann and Gutschalk (2012) has shown that sustained responses can reflect a combination of evoked activity to the energy onset and pitch onset as well as sustained activity to a long duration steady-state pitch sound (24 seconds; s). Greater sustained activity was localised in lateral HG for IRN versus noise, but not for regular click trains versus jittered click trains or silence. This not only casts doubt on the reliability of IRN as a pitch-evoking stimulus but also on HG as a pitch centre. A true pitch centre should encode pitch regardless of other stimulus characteristics. This was not the case. A somewhat different interpretation could be given to the MEG data collected in the same participants. Sustained field responses evoked by IRN and click trains produced similar pitch responses which were consistently localised to lateral HG. Time-frequency analysis of the MEG data revealed that disparity in activation between stimulus types in fMRI was related to theta band activity. Specifically sustained fMRI responses and MEG theta activity were both associated with slow-moving stimulus fluctuations unrelated to pitch, suggesting that fMRI activation in auditory cortex may be much more sensitive to slowly varying stimulus fluctuations than to the pitch. An additional interpretation is that this might contaminate pitch-related responses (i.e., like those associated with the creation of IRN). Conversely, theta activity relating to slow stimulus fluctuations can be eliminated from evoked MEG or EEG responses. However, like Steinmann and Gutschalk (2012) point out, problems choosing an appropriate pitch stimulus are not confined to IRN. For example when using a harmonic complex sound, fluctuations in the noise baseline, used for subtraction designs, may reduce sensitivity for pitch-specific activity. Therefore this work highlights the importance of avoiding the use of confounding parameters of the pitch stimuli when using any neuroimaging technique, but is perhaps especially important for fMRI studies.

2.4.3 Sensitivity to pitch salience as evidence for ‘modular’ organisation in lateral HG?

Sensitivity to pitch salience can be demonstrated by an increase in activation as a function of the salience of the pitch. Neuroimaging findings for sustained pitch responses as a function of salience are rather inconsistent. Griffiths et al. (1998, 2001) respective PET and fMRI studies reported greater activation for IRN stimuli with increasing temporal regularity (and hence pitch salience). Sensitivity to pitch salience for transient pitch responses has been found for MEG onset responses to pitch for different types of pitch-evoking stimuli (e.g., IRN: Krumbholz et al., 2003; Click trains: Gutschalk et al., 2002). For instance, an MEG study by Krumbholz et al. (2003) not only localised POR to medial HG (adjacent to anterolateral HG) but also found that the latency and amplitude of the POR for IRN stimuli varied with the pitch strength (i.e., number of iterations). Stimuli with greater pitch salience (i.e., greater number of iterations) evoked a more robust POR (with shorter latencies) compared to stimuli with a weaker pitch salience. Similarly an EEG study by Krishnan et al. (2012) found that the magnitude of cortical pitch responses increased and behavioural pitch discrimination performance improved as a function of increasing pitch strength (i.e., increasing temporal regularity of the IRN stimuli). However, using jittered pulse trains and unresolved harmonic complex tones with different levels of phases to create three levels of pitch salience (i.e., strong, medium, and weak), Barker et al. (2011) found no evidence for differential fMRI activation anywhere in auditory cortex that could indicate a sensitivity to pitch salience (but see Barker et al., 2013).

The most intriguing fMRI evidence for pitch salience comes from Penagos and colleagues (2004) who, unlike some previous studies, used well-controlled stimuli that accounted for temporal regularity or periodicity, by keeping this fixed for each condition (i.e., the number of waveforms that repeated over time were the same for all conditions), but whose pitch salience differed across conditions. They used harmonic complex tones, where low (80-95 Hz) and high F0s (240-285 Hz) were either filtered into high (1200-2000 Hz) or low (340-1100 Hz) spectral regions to create three strong pitch salience (resolved) conditions and one weak pitch salience (unresolved) condition (i.e., strong condition 1 (low F0 and low spectrum), weak condition 2 (low F0 and high spectrum), strong condition 3 (high F0 and low spectrum), and strong condition 4 (high F0 and high spectrum). Two control conditions (5 and 6) were created using white noise (12 kHz) that were filtered into high and low spectral regions. Stimulus sequences for each condition were 32s in duration, and tones were presented as 300ms bursts (including 10ms onset and offset cosine ramps), where the F0 for each tone burst within a given condition varied randomly (e.g., 80-95 Hz for low F0) every 1.67s, within a background of continuous Gaussian noise (to mask distortion products). Sound conditions were presented randomly, and each sound condition was interleaved with a period of silence which lasted the same duration (classical ‘on-off’ stimulation paradigm, but with continuous Gaussian noise running in the background). They explored differential effects across conditions in cortical and subcortical brain regions, using a 1.5 T fMRI system coupled with cardiac gating techniques (see Chapter 3). Through contrasting conditions with high and low pitch salience, they found no differential activation for pitch salience in the primary auditory cortex or in subcortical areas (i.e., inferior colliculi and cochlea nuclei). Nevertheless, they did find that the majority of cortical activity across the 5 listeners was greater for strong versus weak pitch salience conditions in anterolateral HG (i.e., nonprimary auditory cortex). These effects were still significant after controlling for F0 and frequency and therefore provided some support for the role of lateral HG as a pitch centre. This result is questionable, however, because of the relatively few subjects scanned, the high inter-subject variability and the rather lenient correction statistics used. Chapter 5 re-examines the question of pitch salience using a study design informed by Penagos et al. (2004).

There are a number of differences between published studies, mostly relating to the type of stimuli used and imaging method (e.g., Griffiths et al., 2010; Hall & Plack, 2009; Penagos et al., 2004). It may not be that sensitivity to pitch salience does not exist, but rather the sensitivity to detect such differential activation may be reduced or insufficient for current neuroimaging techniques and corresponding designs. It would be informative if studies opted to: 1) employ at least two types of pitch stimulus when undertaking their research, which would address the criterion of pitch constancy, and/or 2) auditory neuroscientists could decide on a ‘gold standard’ pitch stimulus that should be included in all future investigations, which would address the reliability of findings.

2.5. Pitch perception as a distributed process

So far my review highlights findings that lend support to the HG as being a ‘pitch centre’. More recently however a shift towards a distributed explanation of pitch coding has emerged (Bizley et al., 2009; Griffiths et al., 2010; Hall & Plack, 2009; Kumar et al., 2011; Walker et al., 2011a). A number of experimental findings provide evidence supporting distributed pitch-sensitive representations in auditory cortex. Hall and Plack (2009) was the first fMRI study to evaluate the evidence for a pitch centre (criteria described in Chapter 1) in auditory cortex, by measuring pitch-related activation arising from a number of different pitch stimuli (i.e., pure tone, wideband complex tone, resolved harmonic complex tone, unresolved harmonic complex tone, Huggins pitch, and IRN), across 2 experiments. They also obtained psychophysical measures of pitch discriminability (and therefore pitch salience) for each stimulus condition, in the same listeners, using a Two-Interval Two-Alternative-Forced-Choice (2I-2AFC), two-up, one-down, adaptive procedure that targeted 70.7% performance (task: in which interval was the pitch higher?), and compared these with fMRI measures of the pitch response for each stimulus condition. The first experiment explored fMRI and psychophysical measures for all stimulus conditions except IRN, in 16 listeners. The second experiment was a follow up experiment which explored the activation effects and psychophysical measures of IRN, with and without the use of a low-pass noise masker, in 9 listeners. Hall & Plack (2009) used random effects analysis, as well as incidence “probability” mapping to explore between-subject fMRI activation. Results from experiment 1 revealed that activity was most consistently found posterior to HG, in PT; however, they failed to identify a single pitch center that was common to all pitch stimuli across listeners. Instead, there was a wide variability in the pattern of cortical responses across multiple areas of primary and nonprimary auditory cortex, with no auditory area fulfilling all of the pitch criteria. This led them to conclude it may be too presumptuous to accept HG as a definitive pitch centre, and PT is more likely to be involved in pitch processing, and a more likely candidate for a pitch centre (see Figure 2.8; see also Barker et al., 2011; Garcia et al., 2010; Steinmann & Gutschalk, 2012). Nevertheless, the evidence is more consistent with the distributed processing perspective. Findings from single-unit recordings from ferret auditory cortex also report no evidence in favour of a specialised cortical field for pitch, but rather multiple non-invariant and overlapping areas of primary and non-primary auditory cortex sensitive to pitch, timbre and spatial location (Bizley et al., 2009; Walker et al., 2011a).

Hall and Plack (2009) found no correlation between psychophysical discrimination (differing levels of pitch salience) and fMRI bold activation, highlighting a potential disparity between pitch perception measured behaviourally and that measured using fMRI. In experiment 2, they found that activation for IRN stimuli, with and without a low pass masker, was localised to both HG and PT; however, when no low pass masker was used, IRN activity was centred on lateral HG. This finding, in conjunction with findings from experiment 1 and spectrograms showing the stimulated cochlea response to IRN, led them to conclude that IRN might contain features unrelated to pitch, notably spectro-temporal fluctuations due to increasing pitch salience, that have contributed to earlier studies reporting activation predominantly in lateral HG.

Figure 2.8. ***Third party copyright material removed.*** *Incidence maps illustrating the localised pattern of results from Hall & Plack (2009). Panel A) Depicts the location of the lateral HG (left) and the distribution of activity for all pitch conditions (right), i.e., how many people (N = 1-25) had that particular voxel activated. Panel B) Depicts the widely distributed stimulus-specific pitch-related activity for each of the 5 stimuli. The location of concordance across subjects varies for each stimulus. Note that HG is defined by a yellow border and PT is defined by a white border.*

It is important to note that many human studies tend to show activity in lateral HG *and* anterior PT (e.g., Barrett & Hall, 2006; Hall, Barrett, Akeroyd, & Summerfield, 2005; Hall, Edmondson-Jones, & Fridriksson, 2006; Penagos et al., 2004; Plack, Barker, & Hall, 2014). One explanation for this can be provided by referring to the findings from Steinmann & Gutschalk (2012). Recall earlier that their results implied fMRI responses may be *more* sensitive to slow-varying fluctuations present within pitch stimuli. In this regard, it could be argued that slowly varying spectral-temporal modulations present in IRN evoke greater fMRI responses in HG, but the ‘true’ underlying activity relating to the pitch of the stimulus is observed in the non-primary area typically observed as PT, but to a lesser extent due to such responses being contaminated by these spectro-temporal modulations. However, it should be noted that confounds introduced by using IRN may not necessarily be observed when using relatively short stimulus sequences, like those typically adopted for EEG or MEG (Briley et al., 2013).

2.6. Summary and concluding remarks

Overall, there is still much debate surrounding the representation of pitch throughout the auditory pathway. There are a number of psychophysical models purported to describe how pitch is represented in the auditory nerve (i.e., place, temporal or a hybrid spatio-temporal model). Furthermore, whether the neural representation of pitch is organised in a modular manner (i.e., pitch centre), or involves a pattern of activation widely distributed across auditory cortices, is still highly controversial. Animal studies have demonstrated pitch selective neurons that are tuned to F0 instead of spectrum in primates, but conversely ferret studies have shown that neurons throughout several auditory cortical areas are non-invariant to the pitch value (F0), spectrum (timbre) and spatial location. This discrepancy is also apparent in human studies, with both transient and sustained pitch responses being localised to the purported human pitch centre, lateral HG. Others argue in favour of the possibility of PT being a more likely pitch centre, if any, although the concept of a pitch centre is still highly contentious given the variability in pitch responses across different types of stimuli, methods and species.

Chapter 3. Third party copyright material removed.

See Hall, D. A., & Susi, K. (2015). Hemodynamic imaging of the auditory cortex. In G. G. Celesia & G. Hickok (Eds.), Handbook of Clinical Neurology (Vol. 129, pp. 257–275). http://doi.org/10.1016/B978-0-444-62630-1.00015-9

Chapter 4. Optimising voxel resolution and TE parameters for auditory fMRI

4.1. Introduction

The review of fMRI methodology (Chapter 3) described the important scanning parameters that can influence the quality of the fMRI data and sensitivity to detecting brain activity of interest. These were summarised in Table 3.1. This chapter considers those parameters described in Table 3.1, for the fMRI studies conducted in Chapters 4 and 5.

Some of the hardware parameters were pre-determined across all scanning sessions; a Philips 3 T Intera Achieva whole-body MR scanner, which was equipped with a 32-parallel channel sensitivity encoded (SENSE) head coil. These were available at the Sir Peter Mansfield Magnetic Resonance Imaging Centre, where the new 32-channel SENSE head coil boasted accelerated image acquisition (reduction in total scan by a factor up to number of coil elements, i.e., 32), improved SNR and facilitated higher spatial resolution compared to more conventional 8-channel phased array coils. Because the 32-channel SENSE head coil was new at the time of scanning, it had not been formally tested for fMRI studies of central auditory processing prior to this study. Scanning across all sessions was also completed with a fixed bandwidth of acquisition (details not recorded).

Two other scanning parameters were chosen based on previous auditory fMRI experience; time to repeat (TR) values (required for sparse imaging), and number of acquisitions. TR varied slightly across the different studies reported here depending on the number of acquisitions required within a given time period.

Scanning parameters that were yet to be decided were voxel resolution and TE, because I did not have the benefit of previous experience to decide accordingly. The impact of voxel resolution and TE parameters were introduced and described in Chapter 3. The appropriate choice of voxel size can have an important impact on SNR and often involves a trade-off between ability to detect activity (BOLD contrast sensitivity) and ability to detect activity that is specifically localised within the gray matter (spatial specificity). It was not clear whether a high ($2mm^3$) or more standard ($3mm^3$) voxel resolution would yield the best balance between BOLD contrast sensitivity and spatial specificity in the auditory ROIs, nor what was the optimal TE that would best contribute to this result.

This chapter outlines a series of feasibility studies to optimise voxel resolution and TE, before proceeding to conduct the main fMRI study of pitch processing. To summarise, the aims of Experiments 1-3 were:

Objective 1) Determine optimal TE for scans acquired at a high voxel resolution ($2mm^3$) by evaluating a range of possible TEs during auditory activation (Experiment 1) and at rest (Experiment 2 and 3).

Objective 2) Determine optimal voxel resolution for scans acquired at $2mm^3$ and $3mm^3$ using the same range of TE values (Experiment 3).

4.2. Common methodology

The studies were approved and performed in accordance with the Medical School Research Ethics Committee, University of Nottingham (ref: I12012012, Pitch Salience Coding). Scanning was conducted at the Sir Peter Mansfield Magnetic Resonance Imaging Centre, University of Nottingham. All scanning was performed on a Philips 3 T Intera Achieva whole-body MR scanner equipped with a 32-channel SENSE head coil. All scanning data were converted from raw format (PAR/REC) into image format (IMG/HDR) using bespoke software. Further image quantification and analysis was conducted using MATLAB v7.10.0 (The Mathworks, Natick, MA), MRIcro (www.mccauslandcenter.sc.edu/mricro/) and SPM8 (www.fil.ion.ucl.ac.uk/spm/).

4.3. Experiment 1: examining optimal TE for auditory fMRI

A short pitch experiment was conducted to establish at which TE value (TE 40, 50 or 60ms) provided the best BOLD contrast sensitivity and spatial specificity for images acquired using a $2mm^3$ voxel resolution.

A $2mm^3$ voxel resolution was chosen for three reasons. First, a quick single T2*-weighted scan at voxel resolutions $1.5mm^3$, $2mm^3$ and $3mm^3$ revealed that $3mm^3$ provided rather poor tissue contrast which failed to clearly distinguish gray mater from white matter. Second, there was no visible benefit from using $1.5mm^3$ over $2mm^3$ resolution, and third, $1.5mm^3$ would have increased time for data acquisition.

TEs 50 and 60ms were chosen because a recent auditory fMRI study (Paltoglou et al., 2009) conducted using the same 3 T MR scanner equipped with an 8-channel SENSE head coil and high voxel acquisition (1.5 x 1.5 x 2.5mm) reported an optimal TE of 60ms in auditory cortex at rest. However, later identification of a data processing error which may have inflated estimates led them to conclude and use a TE of 55ms in their fMRI study of frequency selectivity. TE 40ms was mainly chosen so that I could also directly compare high resolution (2mm^3) activation in the current experiment with comparable data acquired using more conventional imaging parameters (3mm^3 resolution with a TE of around 36-40ms according to a number of previous studies; see Garcia et al., 2010; Warren et al., 2003).

4.3.1 Methods

4.3.1.1 Participants

One right-handed (Oldfield, 1971) female participant (#01: 41 years old) was tested. She reported no history of neurological, psychological, hearing impairment, use of medication or substance misuse. Hearing thresholds for participant #01 were within normal range (≤ 25 dB HL) for the audiometric frequencies tested (250 to 8000 Hz). The participant gave written informed consent prior to scanning.

4.3.1.2 Stimuli and task

The functional study involved passive listening to sounds presented at approximately 88 dB SPL (measured by post-experiment calibration, see Section 5.4.2). The participant was instructed to attend to the sounds and remain alert. The sound condition, which was taken from Penagos et al. (2004), was a 32s sequence of harmonic complex tone bursts that varied in pitch and were embedded in a background of continuous noise. Condition 1 from Penagos et al. (2004) was used in this experiment because the fMRI and ratings task described in Chapter 5 used the exact same stimuli from Penagos et al. (2004), and the results from this chapter would ultimately inform the voxel resolution and TE used in Chapter 5. The same stimuli from Penagos et al. (2004) was chosen because this was a highly influential paper at the time, and has been one of the only studies to find encouraging evidence of pitch salience in lateral HG using harmonic complex tones, that appropriately controlled for various stimulus confounds, such as temporal regularity, pitch and spectrum; most pitch salience studies have been limited by their use of IRN (see Section 1.4). Nevertheless, results from studies that don’t use IRN are very inconsistent, which is one of the reasons that warrants replication of Penagos et al. (2004) in Chapter 5 (see Section 5.1.1).

It is important to note that the assumption here, and throughout the experimental chapters in this thesis, was that any pitch stimulus should evoke auditory processing areas or a pitch centre, and therefore it was not deemed necessary to use a series of different pitch-evoking stimuli and/or differing F0s. Therefore for the purposes of this chapter, only one stimulus condition was used. Details of sound Condition 1 are described in Section 2.4.3 (i.e., 300ms tone bursts that varied randomly in pitch within a low F0 range (80-95 Hz) and band-pass filtered into a low spectral region (340-1100 Hz; see Figure 4.1)).

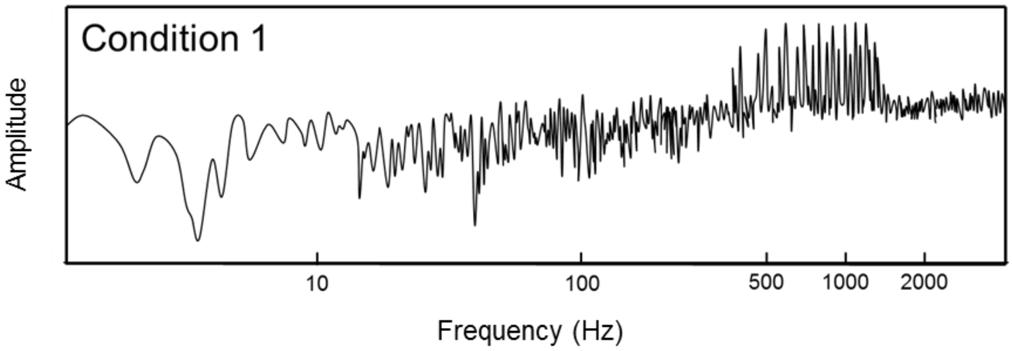

Figure 4.1. Fast Fourier Transform (FFT) of the tone burst, shown on a logarithmic Hz scale.

For each scanning run, participant #01 was presented with 8 repeats of this sound condition interleaved with 8 repeats of a silent ‘baseline’ condition in an on-off stimulation paradigm. Both sound and silent conditions were presented for 32s, with a 2s inter-stimulus interval (ISI). A custom-built MR compatible system was used to deliver distortion-free sound using high-quality electrostatic headphones (Sennheiser HE60 with high-voltage amplifier HEV70). The headphones had been specifically modified for auditory fMRI (i.e., acoustically padded) to attenuate ambient scanner noise by approximately 30 dB.

Imaging protocol

Each of the functional runs acquired 66 T2*-weighted gradient echo echo-planar imaging (EPI) scans. Scans were positioned centrally and angled parallel to the STG (see Figure 4.2). A single saturation band (rest slab thickness = 43mm, power = 2) was placed in the anterior position (front of head, outside FOV) to reduce aliasing and Nyquist artefacts (ghosts) from the eyes (Figure 4.2). This is because the eyes contain a high proportion of water which can cause Nyquist artefacts to arise because of the phase differences between gradients. These artefacts can manifest as reproductions of the eyes that have been shifted by half the field of view, for example a reflection of the eyes would appear around the middle of the brain in the STG. To avoid aliasing, the maximum Nyquist frequency to be identified is equal to one-half of the sampling rate. Each scan had 20 (2mm^3 data) or 36 slices (3mm^3 data) and no inter-slice gap. The matrix size for all 2mm^3 scans was 168 x 168 x 40mm (reconstructed voxel size = 1.75 x 1.75 x 2mm, reconstructed matrix size = 96 x 96 x 20mm). The matrix size for all 3mm^3 scans was 80 x 80 x 36mm. Additional scanning parameters included a flip angle = 90˚ and a SENSE factor = 2 (to reduce image distortions). To reduce the impact of background scanner noise on stimulus evoked BOLD changes, the time between successive scans (TR) was set to 8.5s (see Hall et al., 1999) and a SofTone factor of 2 was used to further attenuate scanner noise by approximately 9 dB (Hall and Plack, 2009). Each run lasted approximately 10 minutes, with the total session running approximately 60 minutes.

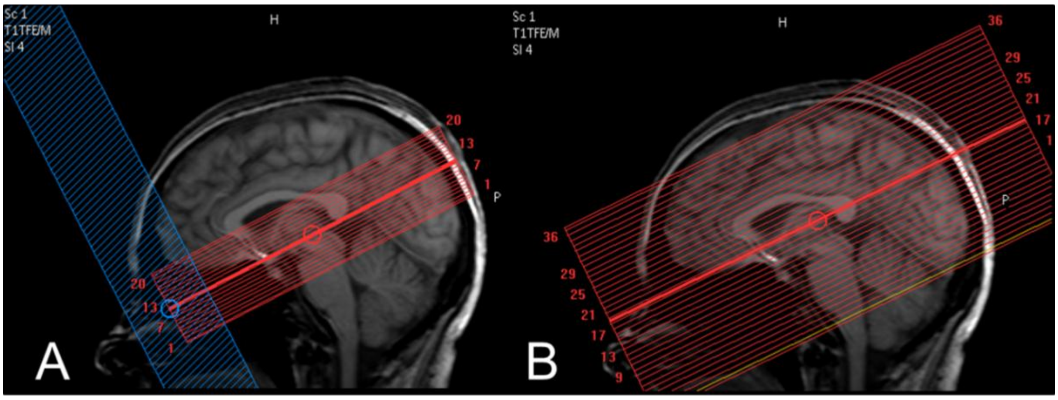

Figure 4.2. Anatomical scans showing the orientation (parallel to superior temporal plane) and slices implemented in Experiments 1 and 5. Panel A: Position of scans for 2mm^3 scan is shown in red. Note that because the field of view at this resolution is smaller than the head, a saturation band (depicted as a blue rectangle) was placed at the anterior edge of the scan to reduce aliasing and Nyquist artefacts. Panel B: Position of scans for 3mm^3 scan is shown in red. Note that because the field of view at this resolution extends beyond the head, no saturation bands were required. Position of scans 1.5mm^3 prior to study 1 are not shown, but were the same as A).

The functional study comprised four runs, manipulating voxel resolution and TE respectively: (Run 1) 2mm^3 and 40ms, (Run 2) 2mm^3 and 50ms, (Run 3) 2mm^3 and 60ms, and (Run 4) 3mm^3 and 40ms.

4.3.1.3 Analysis of fMRI runs

Analysis of fMRI time series data typically requires a lot of spatial pre-processing and statistical analysis (for full description, see Section 3.5). Each functional run was pre-processed and analysed separately using SPM8. Not all the analysis steps reported in Section 3.5 were necessary, and therefore only the following steps were undertaken:

Spatial pre-processing:

- Reorientation of the functional scans to ensure that they were all displayed in the same 3 dimensional brain space (i.e., in the same x, y and z plane) as the standardised MNI brain template (Evans et al., 1993).
- Realignment to correct for any spatial mismatch of functional scans in the time series that occurred due to head movement during the experimental session. To reduce the

influence of head movement within and between runs (Friston et al., 1995) a rigid body transformation is applied to all images in the time series so that they are realigned to match a single reference image (i.e.,. the middle scan, scan #33). Acceptable amounts of translational and rotational head movements were limited to 3mm and 3 degrees, respectively. These guidelines were taken from a comparable study (Paltoglou et al., 2009).

- Spatial smoothing involves applying a Gaussian filter to the data, removing high frequency information and helping to increase SNR and sensitivity within the auditory cortex (Friston et al., 2000). The smoothing kernel parameters were 3.5 x 3.5 x 4mm full width at half maximum (FWHM) for the 2mm^3 dataset, and 6 x 6 x 6mm FWHM for the 3mm^3 dataset.

Analysis of auditory activity:

As previously described (Section 3.5.2.1), GLM estimation was used to model the BOLD signal change separately for each functional run. A design matrix specified which scans were acquired during the sound condition using a coding of ‘1’ (present) and ‘0’ (absent). The baseline condition was implicitly specified in the model. To eliminate low frequency artefacts, such as physiological fluctuations due to breathing and heartbeat, a high-pass filter cut-off of 136 s was applied. The appropriate value for the high-pass cut-off was calculated using the general rule of thumb formula based on Nyquist sampling theorem (Huetell et al., 2009): two times the TR period multiplied by the max number of scans between two repeated presentations of a condition (i.e., 2*8.5*8 = 136). The sparse sampling fMRI sequence used a relatively long TR (8.5 s) which ultimately meant the high-pass cut-off was also correspondingly high (over 2 minutes). The trade-off of using sparse sampling is that a high filter does not remove much of the extremely low frequency temporal noise in the data. Despite the potential for more noise to be included in the signal, I did not anticipate that this would negatively impact the activation results to a significant degree. Using a fixed-effects analysis (Friston et al., 1999), the effect of ‘pitch sounds > silence’ was computed with a one-sample T-test ($p < .001$, uncorrected, extent threshold = zero) separately within each functional run. As mentioned earlier, I had two criteria for determining the optimal TE for auditory fMRI: 1) BOLD contrast sensitivity, and 2) spatial specificity. The first stage of analysis allowed evaluation of BOLD contrast sensitivity by creating statistical activation maps for each run. This analysis is performed on each and every voxel but was not corrected for multiple comparisons here because it was judged to be too stringent for these pilot data. To facilitate further comparison between TEs for the 2mm^3 functional runs, all three TEs were modelled collectively so patterns of activation using pairwise comparisons could be directly compared between runs ($p < .001$, uncorrected, extent threshold = zero; see Figure 4.3). This allowed evaluation of spatial specificity.

4.3.2 fMRI results

4.3.2.1 Objective 1: Determining optimal TE from visual inspection of the BOLD fMRI activity maps

Visual inspection of the four maps for sound-related activation (Figure 4.3A-D) revealed that scans acquired at 2mm³ using TE 40ms seemed to produce a relatively large spread and greatest magnitude of BOLD activity within auditory cortex. The peak of activity with the greatest *t* statistic was localised in the right hemisphere (x = 42, y = -20, z = 2mm, *t* = 19.77; see Figure 4.3A).

For the other two runs acquired at 2mm³, TE 50ms yielded greater activity (peak at x = 41, y -20, z = 2mm, *t* = 9.94) than TE 60ms (peak at x = 41, y =- 21, z = 0, *t* = 9.12), but the spread of activity for TE 60 was visibly larger and more distributed. Note the smaller *t* statistic values for these two runs (see Figure 4.3B and 4.3C).

Scans acquired using a 3mm³ voxel resolution (TE 40ms; Figure 4.3D) showed the largest spread of activity approximately located in auditory cortex (peak at x = 43, y = -25, z = 4mm, *t* = 17.28). The peak *t* statistic value for the 3mm³ (TE 40ms) scans was slightly lower than those obtained for the corresponding 2mm³ TE 40ms run.

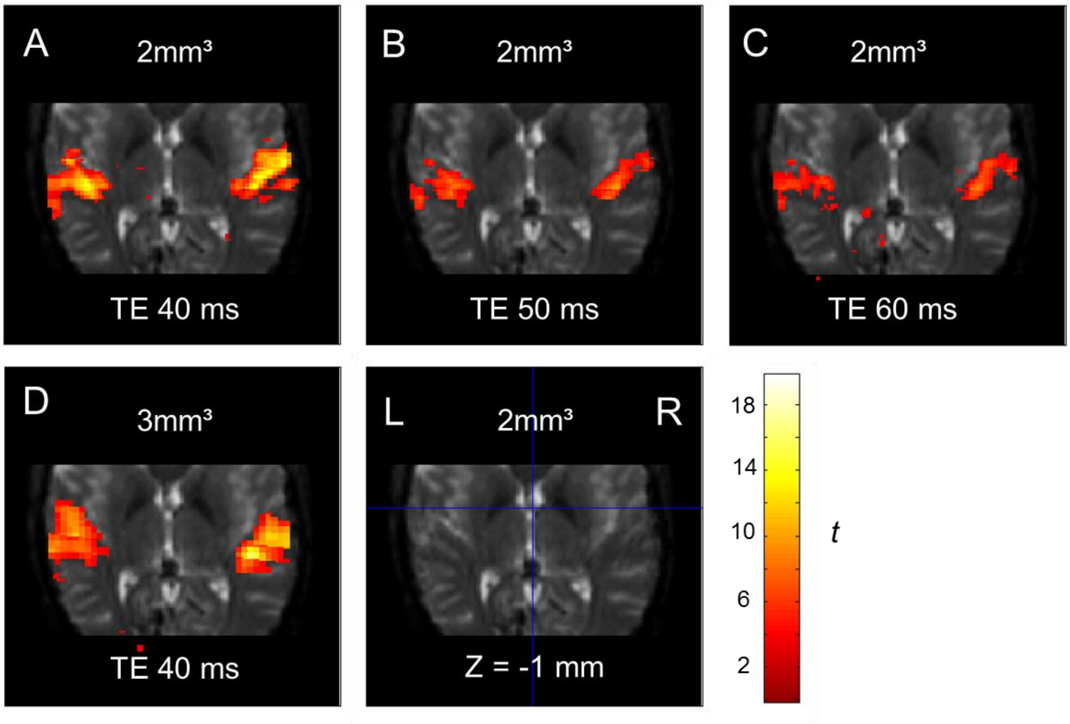

Figure 4.3. Activation maps for 'sound versus silence' contrasts (p < .001, uncorrected) for each functional run overlaid onto a T2-weighted functional scan for 2mm³ TE 40ms (x =*

0.32, y = 0.02 and z = -1). A = 2mm^3 TE 40ms (peak cluster size = 1062 voxels), B = 2mm^3 TE 50ms (peak cluster size = 429 voxels), C = 2mm^3 TE 60ms (peak cluster size = 607 voxels), and D = 3mm^3 TE 40ms (peak cluster size = 401 voxels). Note neurological convention left = left, which is also the case for all figures in this chapter unless otherwise stated. Coloured bar represents range of t statistic values.

The ‘sound versus silence’ pairwise comparisons for each of the three different TE values assessed at 2mm^3 are shown in yellow in Figure 4.4. The pairwise comparisons between the three different TE values assessed at 2mm^3 are also shown below in red. For example, panel A shows where there is statistically greater activity in the TE 40ms and TE 50ms runs (red), relative to the underlying activity for TE 50ms ‘sound versus silence’ contrast alone (yellow). Although the two TE runs detect broadly equivalent activity across auditory cortex, there is significant differential activity (40>50ms, shown in red) localised to the right lateral part of HG, most likely in both gray and white matter. Personal communication with physicist and fMRI expert Dr Sue Francis (Sir Peter Mansfield Resonance Centre, University of Nottingham, 12th April 2012) confirmed that the peak of activity from the TE 40 run was likely to be attributable to nearby lying veins (using specialist software). Other studies using gradient echo acquisitions also report that pial signal and its veins give the strongest BOLD contrast sensitivity (e.g., Zhao, Wang, & Kim, 2004).

The direct contrast between 40 and 50ms is displayed in panel B, where statistically greater activity in the TE 50ms run compared to TE 40ms run is shown in red. Although there are a few voxels that reach statistical significance, this comparison yields no differential activity in auditory cortex and therefore does not represent any functional importance.

Figures 4.4C and 4.4D show the direct comparisons between the 50ms and 60ms TE runs. Both panels show that although there are some voxels that reach statistical significance, these are scattered across the slice and lie primarily outside auditory cortex (blue arrows in Figure 4.4 highlight examples). Such differential activation is most likely attributable to random noise (i.e., false positive errors) given that we did not correct for multiple comparisons here, and therefore have no functional relevance. Specifically comparing the spread of activity between ‘sound versus silence’ and TE 60ms > TE 50ms in panel C, revealed that, relative to the TE 50ms run, TE 60ms exhibited significantly greater ‘random noise’ activity outside of auditory cortex. This is further illustrated in panel D, where the direct contrast reveals comparable less activity outside of the auditory cortex for the TE 50 ms compared to TE 60ms run.

Given that the likely sources of differential activity for TE 40ms (veins) and TE 60ms (random noise) lie outside auditory cortex implies that the TE 50ms run had the greatest spatial specificity to auditory activation overall.

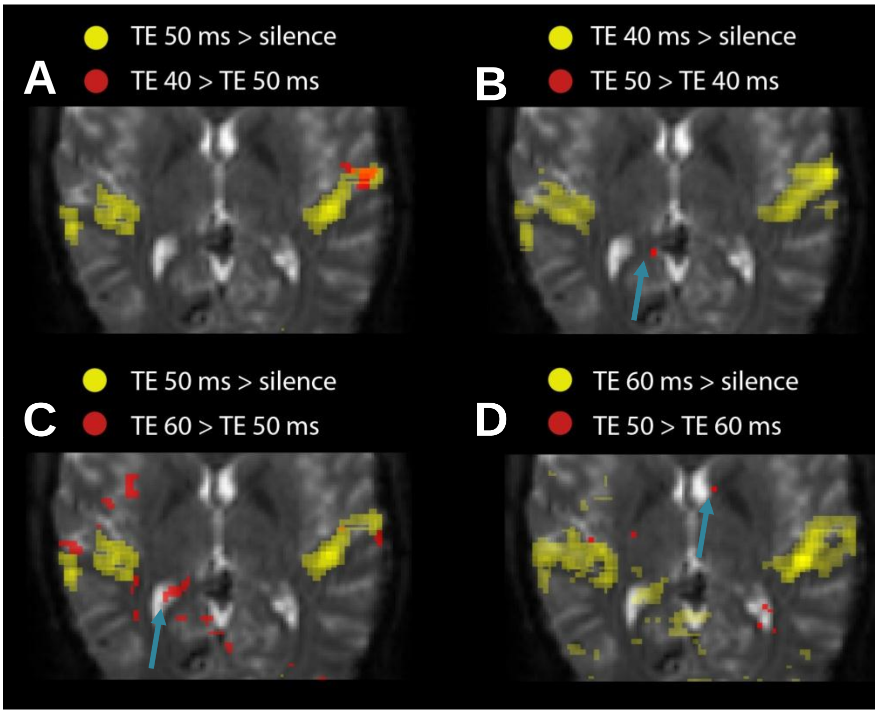

Figure 4.4. Difference maps for 'sound versus silence' contrasts from direct pairwise comparisons between TE values. Difference maps are shown in red (p < .001, uncorrected). For comparison, the underlying effect of a single 'sound versus silence' contrast measured for one TE value is shown in yellow. Activity is overlaid onto the corresponding slice of a T2 functional scan. The figure conforms to neurological convention where left = left. Blue arrows depict random noise.*

4.3.3 Summary

Overall these results show that the TE 40ms run detects the most significant sound-related activity in the auditory cortex when using a 2mm³ voxel resolution (hence greatest BOLD contrast sensitivity), but the peak of activity in the right is most likely coincidental with a large draining vein than with gray matter neural activity (hence reduced spatial specificity). This evidence therefore suggests that TE 50ms might be a better choice of TE with a 2mm³ voxel resolution, given that this TE value had the second highest BOLD contrast sensitivity and greatest spatial specificity. Nevertheless, further validation work is needed to lend weight to this decision.

4.4. Experiments 2 and 3: calculating optimal TE in human auditory cortex at rest

As discussed in Section 3.2.3, the greatest BOLD signal changes are known to occur when TE is approximately equal to T2* in a given ROI as these parameters are influenced by the tissue properties of the brain regions being scanned. Methods are available for quantifying the average T2* value for tissue within a particular ROI in order to provide an objective estimate of the optimal TE. Experiments 2 and 3 achieved this by acquiring a series of 2mm^3 T2*-weighted functional scans at a series of TE values (Experiment 2: 26-50ms, and Experiment 3: 30-80ms) in two participants. A wide range of TE values were used (26-80 across Experiments 2 and 3) to obtain a better estimate of the T2* values across our ROIs, whereby using more data points enabled a more reliable estimate of the noise window (by providing a better measure of the standard deviation).

Experiments 2 and 3 involved a partial factorial design in which conditions were not repeated identically across the two participants, or across voxel resolutions. Experiment 2 assessed a relatively small range of TE values (26-50ms) used in published auditory fMRI studies (Garcia et al., 2010; Warren et al., 2003) with a 2mm^3 voxel resolution. This is because Experiment 1 results showed that this resolution had the greatest BOLD contrast sensitivity. Experiment 3 assessed a much broader range of TE values (30-80ms) in order to re-estimate T2*, across 2mm^3 and 3mm^3 resolutions. Plotting the MR signal across a broader range of TE values provides a more reliable estimate of T2*, and hence more weight was placed on the results of Experiment 3 in deciding the optimal TE. Two different analysis methods were used; manual calculation of T2* relaxation values and T2*-weighted mapping.

4.4.1 Methods

4.4.1.1 Participants

The participant for Experiment 2 was a right-handed (Oldfield, 1971) female (#02: 24 years old). She reported no history of neurological, psychological, hearing impairment, use of medication or substance misuse. Hearing thresholds for participant #02 were identified to be within normal range (≤ 25 dB HL) for standard audiometric frequencies (250 to 8000 Hz). The participant gave written informed consent prior to scanning. Participant (#01) that took part in Experiment 1 also participated in Experiment 3 (see Section 4.3.1.1 for details).

4.4.1.2 Stimuli and task

All measures were made at rest and so throughout the scanning session each participant was only required to lie still in the bore of the MR scanner in the absence of any sound stimulation.

4.4.1.3 Imaging protocol

Scanning sessions for participants #01 and #02 were conducted on different days and lasted approximately 60 minutes each. The imaging protocol for Experiments 2 and 3 varied slightly, but unless otherwise stated the scanning parameters were the same as those reported in Experiment 1.

For Experiment 2, a series of 26-sliced T2*-weighted EPI gradient echo scans using a TR = 8 s were acquired at $2mm^3$ across a range of TEs (26, 30, 35, 40, and 50ms). The acquired voxel and matrix size for Experiment 2 was the same as Experiment 1, but the reconstructed voxel and matrix size was slightly different (1.5 x 1.5 x 2mm and 112 x 112 x 20mm, respectively). This does not affect study results or interpretation. Two sets of scans were acquired for each TE and were averaged for T2*-weighted mapping only.

For Experiment 3, a series of T2*-weighted EPI images were also acquired at $2mm^3$ and $3mm^3$ voxels resolutions across TEs (30, 40, 50, 60, 70, and 80ms) using the same TR as Experiment 2. Again, two sets of scans were acquired for each TE and were averaged for T2*-weighted mapping.

4.4.1.4 Analysis of T2*-weighted images

Spatial pre-processing

Co-registration was implemented for Experiment 3 only, ensuring both $2mm^3$ and $3mm^3$ datasets were spatially comparable. Co-registration maximised the mutual information between the $2mm^3$ and $3mm^3$ scans by adjusting the $2mm^3$ scans to fit the same $3mm^3$ brain space (MNI; Evans et al., 1993). This meant the same ROI could be implemented during analysis in order to compare T2* values accordingly. For the manual calculation of T2* relaxation values, coregisteration of the $2mm^3$ to $3mm^3$ TE series was conducted prior to T2* quantification. For T2*-weighted mapping, coregisteration of the TE data series from $2mm^3$ to $3mm^3$ was conducted after both T2*-weighted maps had been created

Quantifying the T2 relaxation values and calculating optimal TE in different auditory ROIs*

MRIcro software was used to create ROIs which were used to select specific voxels for MR signal quantification and analysis. In Experiment 2, three ROIs within the auditory cortex were manually drawn based on the T2*-weighted functional scan at TE 26ms for participant #02 (see Figure 4.5).

T2* values in three different types of auditory ROIs were quantified. ROI #1 was restricted to the gray matter along bilateral HG (391 voxels) to ensure that any potential contribution of veins was excluded from the T2* estimates. Wald's (2012) recent review discusses the importance of reducing pial vein contamination by limiting the analysis to voxels which did not directly encroach the pial border. ROI #2 was larger and included both gray and white matter around HG and the anterolateral parts of PT (729 voxels). ROI #3 was larger still and included the STG (4559 voxels). Within each ROI, a bespoke MATLAB script was used to calculate mean signal intensities across the TE series.

There was no specific directional hypothesis about how the T2* relaxation values and optimal TE would vary across these different focal regions. However, based on the previous published evidence, it would be expected that pitch-related activation should map most closely onto ROI #2 around HG (see Hall and Plack, 2009). Hence, more weight was placed on the results for ROI#2 in deciding the most optimal TE.

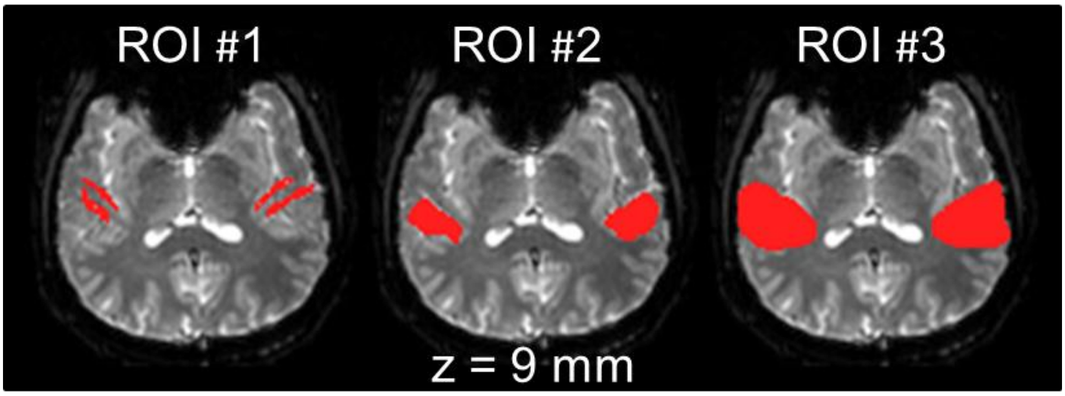

Figure 4.5. Illustration of the three ROIs overlaid onto a slice acquired for participant #2 (TE = 26ms, central slice z = 9mm, scan 1). Display follows radiological convention, left = right.

For Experiment 3, a new ROI was manually drawn because participants #01 and #02 had visible individual differences in cortical morphology. For quantification of T2* relaxation, only one ROI was used because it specifically focused around HG as did previous maps of pitch-related activation (see Hall and Plack, 2009). The ROI was similar in size (559 voxels) and shape to ROI #2 in Experiment 2 and extended across the gray and white matter. The ROI was manually drawn from the 2mm^3 scan at TE 30ms that had been co-registered to 3mm^3 average scan (TE 30ms; see Figure 4.6).

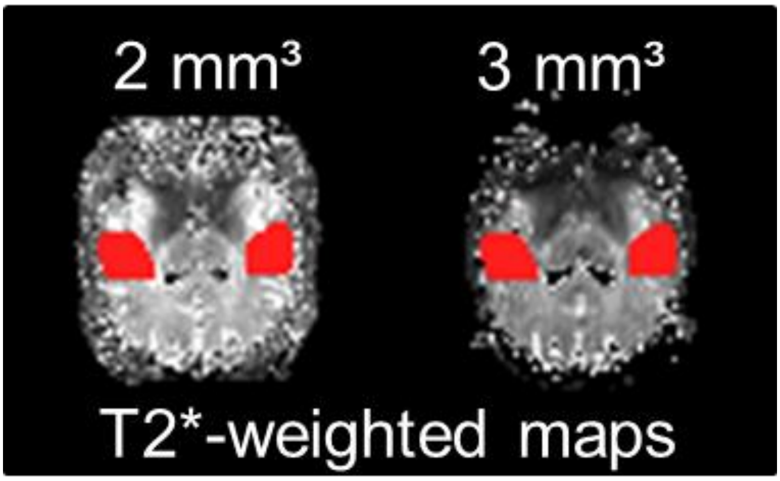

Figure 4.6. Graph illustrating 2mm³ functional data that were co-registered to the 3mm³ data so that the same ROI could be applied. Display follows radiological convention, left = right. Two slices are chosen through the ROIs.

Estimating T2 relaxation values: manual calculation*

Recall that the T2* relaxation value in a given ROI is approximately equal to the optimal TE. To estimate the T2* relaxation values (i.e., where $S(TE) = S_0 e^{-t/T2^*}$) we used the following formula: T2* = -1/gradient ~ optimal TE = $\ln S_0(-TE/T2^*)$, where S refers to the signal, and S_0 refers to the signal strength or intensity immediately following the RF pulse. For any given brain region, the T2* relaxation value is the reciprocal of the slope of the regression line (y = ax+b; where a is the gradient and b is the intercept) plotted through the natural logarithm (LN) of the mean signal intensity within the ROI (see Results, Section 4.4.2).

Estimating T2 relaxation values: T2*-weighted mapping*

Advice from Drs Francis and Mullinger at the Sir Peter Mansfield Magnetic Resonance Imaging Centre suggested using an alternative ‘automated’ approach to estimate T2* relaxation values across ROIs. This involved fitting T2* curves to a given data series to obtain a weighted map of T2* first, and then averaging over each ROI. This method computes T2* separately for each voxel, concatenating the data series to fit a linear regression to the plot of the LN of the MR signal collapsed across TEs (Clare et al., 2001). T2*-weighted mapping is considered to be better at estimating optimal TE, than previously described manual calculations, for two main reasons. First, it weights data at earlier TEs (e.g., 30ms) higher than at later TEs (e.g., 80ms). This takes into account the fact that the MR signal at later TEs may be approaching or superimposed by the noise window because it has much lower intensities as a result of the decay in transverse magnetisation. Second, once the T2*-weighted maps are created they can be easily thresholded to exclude voxels with a T2* and/or noise intensity above a given value, to account for the unwanted contribution from CSF or noisy data. This prevents CSF and noise dominating, and hence skewing, the T2* average measures.

Only a handful of auditory fMRI studies have implemented this TE optimisation method, because it typically requires specialist expertise and software to complete (see Clare et al., 2001). In this case, Dr Francis used some in-house software to create a single T2*-weighted map for each TE data series acquired in Experiments 2 and 3. I was then able to apply thresholds to exclude selected voxels and ROI masks to calculate the optimal TE for a given ROI. Both T2*-weighted maps were thresholded to exclude voxels with a noise intensity value of above 500. Furthermore, T2* values greater than 200 (Experiment 2) or 150 (Experiment 3) were set to zero to account for CSF. In each case, the cut off for CSF was manually selected according to the dynamic range of the voxel intensities across the whole head.

4.4.2 Results

4.4.2.1 Objective 1: Determining optimal TE from manual calculations of T2* relaxation estimates

The MR signal intensity decreased with increasing TE in a linear manner (Figure 4.7A). The data were highly replicable across scans 1 and 2 in each time series and did not vary markedly across the different auditory ROIs. The R^2 values obtained for the linear regressions were extremely high in all cases indicating a good fit to the data.

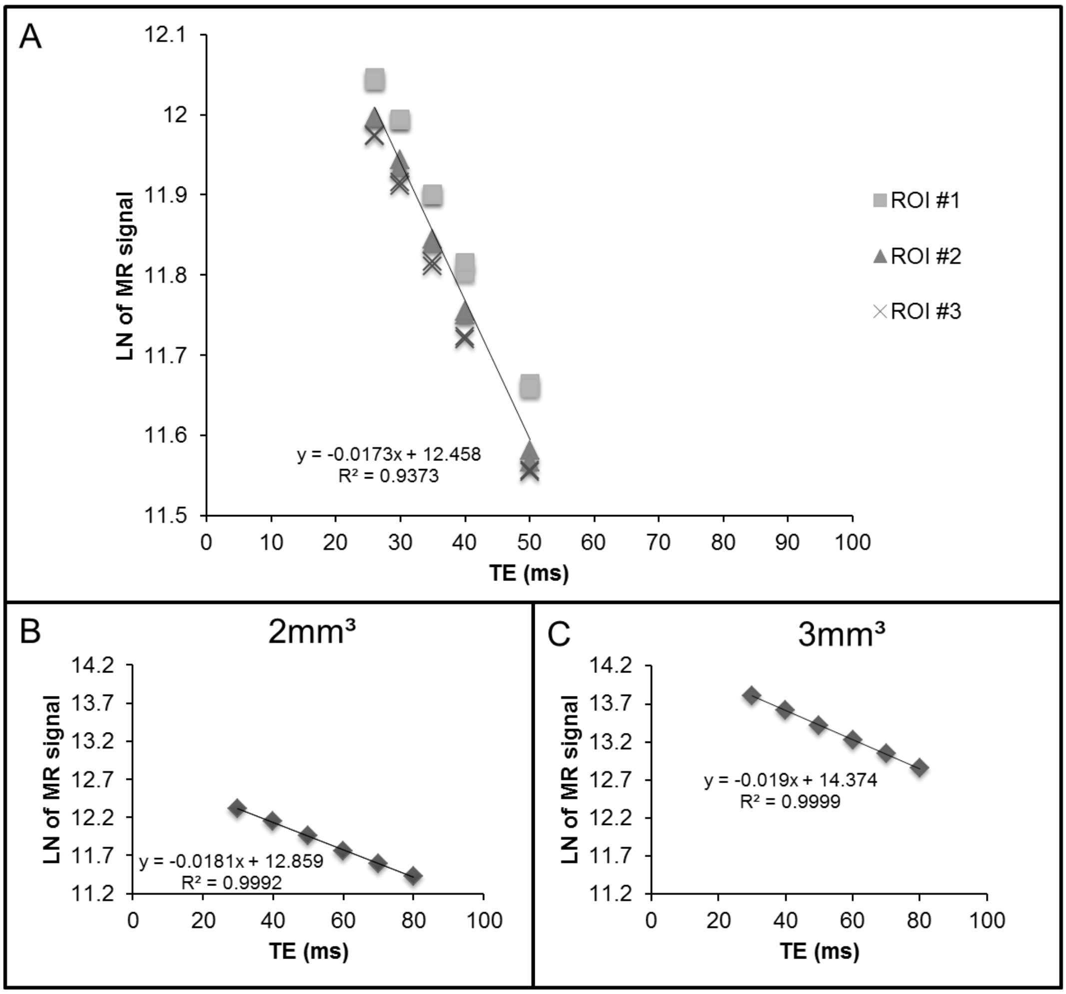

Figure 4.7. Graph showing the manual calculation of optimal TE across ROIs in Experiment 2 and 3. Panel A depicts 2mm^3 data series from Experiment 2 plotted over TE for scans 1 and 2 and for each ROI. Panels B and C respectively illustrate the co-registered 2mm^3 and 3mm^3 data series from Experiment 3 for one ROI plotted as a function of TE. For each panel, the line of best fit and R^2 (goodness of fit) has been plotted through LN of the mean signal intensity across the range of TEs.

For Experiment 2, the estimated T2* relaxation values for the different ROIs were 61.35ms (ROI #1), 56.18ms (ROI #2), and 56.50ms (ROI #3), with an average fit of T2* across all three ROIs of 57.80ms (see Figure 4.7A). Given that most weight should be placed on the findings from the ROI based around HG and the anterolateral parts of PT (ROI#2 for Experiment 2, and the ROI for Experiment 3), these results for the manual calculation support an optimal TE of around 56ms for data acquired at a high 2mm^3 resolution. This figure is based on averaging results across ROIs (i.e., ROI#2 for Experiment 2, and results from the only 2mm ROI in Experiment 3) i.e., 56.50+55.25/2 = 55.87. Note that these results are similar to the optimal TEs reported by Paltoglou et al. (2009; i.e., TE 55-60ms).

4.4.2.2 Objective 1: Determining optimal TE from T2*-weighted mapping

Experiment 2 ($2mm^3$) results confirmed those from the manual calculation, with an optimal TE of around 58ms. For Experiment 2, when ROIs #1-3 were applied to the T2*-weighted map, they yielded a T2* mean estimate of 63.46ms, 60.17ms, and 59.59ms, respectively (average = 61.07ms; data not shown). For Experiment 3, T2*-weighted maps gave a T2* mean estimate of 54.96ms and 53.15ms, for $2mm^3$ and $3mm^3$ voxel resolutions respectively (data not shown). Again, the range of estimated optimal TE was quite narrow across the three ROIs.

4.4.2.3 Objective 2: Determining optimal voxel size from calculations of T2* relaxation estimates

Again, the MR signal intensity decreased with increasing TE in a linear manner (Figure 4.7A). The data were highly replicable across scans 1 and 2 in each time series and did not vary markedly across the different auditory ROIs. The R^2 values obtained for the linear regressions were extremely high in all cases indicating a good fit to the data. For Experiment 3, in the single auditory ROI, the estimated T2* relaxation values for the $2mm^3$ series was 55.25ms, and for $3mm^3$ series it was 52.63ms (see Figures 4.7B and 4.7C, respectively). Again, the range of estimated optimal TE was quite narrow across the two voxel resolutions.

4.5. Conclusion

To arrive at one optimal TE, findings from Experiments 1-3 were drawn together to ensure I chose a TE that optimised BOLD contrast sensitivity whilst still maintaining spatial specificity. An intermediate TE value of 50ms was chosen for the subsequent fMRI study, as this value fell approximately in the middle of TE 40 and TE 57ms. This decision was primarily driven by the auditory fMRI results in Experiment 1 showing good BOLD activation that was less influenced by veins and random noise.

Chapter 5. Re-examining the neural representation of pitch and pitch salience in auditory cortex

5.1. Introduction

This chapter describes Experiments 4 and 5 which were particularly interested in investigating the third pitch criterion, ‘covariation with pitch salience’ (see Section 1.3), in addition to the first pitch criterion of pitch selectivity. This is because there has been variable support for pitch salience sensitivity in cortical and subcortical areas of the auditory system (see Section 2.4.3; Barker et al., 2011, 2013; Bendor & Wang., 2010; Griffiths et al., 2010; Hall & Plack, 2009; Krumbholz et al., 2003; Norman-Haignere, Kanwisher, & McDermott, 2013; Penagos et al., 2004; Kumar et al., 2011; Soeta, Nakagawa, & Tonoike, 2005). Despite Penagos et al. (2004) using well-controlled stimuli, there are a number of notable limitations to the original study that need to be addressed. The fMRI work in the present chapter describes attempts to address some of these limitations to re-examine the evidence for pitch salience sensitivity in the auditory cortex.

5.1.1 Conflicting evidence for a representation of pitch salience

Penagos et al. (2004) conducted a seminal fMRI study to investigate whether there is a neural representation of pitch salience in human auditory cortex and/or midbrain. Determining this is important because it has been long associated as an essential variable for improving listener’s ability to make pitch discriminations (Houtsma & Smurzynski, 1990). To date, it is one of very few fMRI studies that have successfully addressed pitch salience using non-confounding stimuli, such as IRN, that were well controlled for F0 and spectral dimensions, as well as temporal regularity. Specifically, they used harmonic complex tones with low or high F0s that were filtered into low and high frequency bands to create strong (resolved harmonics) and weak (unresolved harmonics) pitch salience conditions, but where the temporal regularity was always fixed (i.e., the number of waveforms that repeated over time were the same for all conditions). The study purported to find a region in the anterior non-primary auditory cortex (i.e., the anterolateral HG) sensitive to pitch salience. These effects were still significant after controlling for F0 and frequency differences. This region has been long implicated in the processing of pitch or a possible ‘pitch centre’ (Bendor & Wang, 2005, Bendor & Wang, 2010; Griffiths et al., 1998, 2001; Patterson et al., 2002; Krumbholz et al., 2003; Kumar et al., 2011; Norman-Haignere et al., 2013; Puschmann et al., 2010; Warren et al., 2003), providing some support that neurons processing pitch and pitch salience are located within the same focal region.

fMRI work (previously described in Section 2.5) by Hall and colleagues have challenged this claim (see Hall & Plack, 2009; Barker et al., 2011), failing to find any evidence supporting a representation of pitch salience in auditory cortex, but rather finding evidence for distributed pitch processing over a number of auditory areas (but see Barker et al., 2013; Norman-Haignere et al., 2013). This was further corroborated by Barker et al. (2011), who investigated whether fMRI activity covaried as a function of pitch salience. Pitch salience was parametrically varied along three levels (strong, medium and weak pitch salience) for two different types of pitch-evoking stimuli (i.e., pulse trains and unresolved complex tones). These differing levels of pitch salience were confirmed in psychophysical measures in the same 16 listeners, using a 3 alternative forced-choice, two-down, one-up, design, with an adaptive procedure that targeted 70.7% performance (task: in which interval was the pitch higher?). They found no evidence of a neural representation of pitch salience in auditory cortex (i.e., a decrease in activity was observed for increasing pitch salience). They did, however, localise the majority of pitch related responses to PT.

More recent fMRI evidence from Norman-Haignere et al. (2013) found that pitch-sensitive regions in more anterior areas of auditory cortex (i.e., anterolateral HG and PP) showed greater pitch-related activation to resolved harmonic compared with unresolved harmonic complex tones, providing evidence in favour of pitch salience sensitivity.

It should be noted that there are a number of differences between these fMRI studies, mostly relating to the type of stimuli used. For instance, Hall & Plack (2009) used a similar pitch-evoking stimuli (resolved and unresolved) but in a fixed pitch sequence, attributing the majority of pitch related activity to PT, and not HG. Barker et al. (2011) also used a fixed pitch sequence (reported above). Findings from Patterson et al. (2002) suggest that the location of pitch related activity may be influenced by the stimulus presentation context. Specifically they reported more anterior pitch activity in PP and STG when melodic pitch sequences were used compared with fixed pitch sequences. Penagos et al.'s (2004) stimuli were more melodic where the pitch varied across a narrow range (as were Norman-Haignere et al., 2013), compared with Hall & Plack (2009) and Barker et al. (2011), which might be one explanation for the differential patterns of activation and inconsistencies within the literature (see limitations of cognitive subtraction, Section 3.4.1). Specifically, given that Penagos used melodic sequences, more anterior activation might have been observed, however this was not addressed in their analyses.

5.1.2 Limitations of Penagos et al. (2004)

Despite Penagos et al.'s (2004) promising findings, there are notable grounds that motivate replication of their study. The study obtained imaging data from a small number of participants (i.e., 5 listeners), which did not permit more rigorous RFX group analysis. As a guide, Desmond & Glover (2002) found that for a liberal threshold of $p < .05$, approximately 12 subjects are required to achieve 80% statistical power at the single voxel level for typical activations (however this sample size is contentious; see Button et al., 2013; Friston, 2012; Friston et al., 1999). A second weakness of the previous work was that they used relatively unconventional fMRI analysis techniques that displayed individual activation maps on inflated brain views for each listener, which further obviated any form of group-based statistics. High inter-subject variability was also evident, where exact locations purportedly sensitive to pitch salience were considerably different across listeners. Although this is a problem that arises in many fMRI studies (see Bendor, 2012; Griffiths & Hall, 2012), the lack of group-based statistics makes it impossible to determine the degree of commonality in activation across listeners. Although individual analysis provides valuable information about fixed-effects within participants, group data are necessary to address a number of more generalisable hypotheses. Moreover, for individual subjects the statistical threshold was extremely lenient and thus made their findings much more susceptible to type 1 errors. Lastly, no perceptual ratings of pitch salience were obtained for their stimuli to confirm their perceptual pitch salience (subjective or otherwise).

Subsequently, a pitch salience rating task (Experiment 4) and an fMRI study (Experiment 5) aimed to re-examine the neural representation of pitch and of pitch salience in auditory cortex by using the same stimuli and a study design informed by Penagos et al. (2004), in collaboration with one of the original authors (Prof. Andrew Oxenham). To address some of the aforementioned methodological issues surrounding the original paper, novel aspects of the current experiments included recruiting a large number of participants (Button et al., 2013), and thus allowing sufficient data to conduct more conventional group-based analyses (e.g., ROI analysis, and incidence ‘probability’ mapping), as well as using corrected statistics (where appropriate), a higher field scanner and a 32-channel head coil for improved SNR, and optimised TE. Although the fMRI experiment constituted the main study in this chapter, subjective ratings of pitch salience were obtained beforehand in the same listeners for the same stimuli.

5.1.3 Objectives and hypotheses

The research objectives and hypotheses were as follows:

Objective 1) Where in auditory cortex is there greater activity for pitch-evoking sounds than for noise controls? To address the pitch criterion 1 (pitch selectivity) there should be significantly greater auditory cortical activity (and higher behavioural ratings) for pitch-evoking sounds than for their spectrally matched noise control.

Objective 2) Is there sensitivity to pitch salience located in auditory cortex, and it is co-localised with the pitch-related response? To address pitch criterion 3 (covariation with pitch salience) there should be significantly greater auditory cortical activity (and higher behavioural ratings) for strongly pitch salient than weakly pitch salient conditions.

Objective 3) In addition to replication of the same questions that motivated Penagos and colleagues, the impact of the melodic sound sequence was also explored (see Patterson et al., 2002). Given that the sound sequences used contained random changes in pitch (which are inherently more melodic) does this stimulus context promote a right hemisphere preference? Melodic sequences should engage right hemisphere processing more than the left hemisphere, and activity should extend more anteriorly. For this objective, analysis was restricted to only looking at those auditory brain regions reported by Patterson et al. (2002), specifically the lateral HG, PP and PT.

5.2. General methods and materials

5.2.1 Listeners

A total of 18 right-handed listeners (#01-#18) participated in both the behavioural (Experiment 4) and fMRI (Experiment 5) phases of the experiment (9 male, 9 female; M_{age} = 23.94, *SD* = 5.98, age-range = 19-42 years). The laterality quotient (LQ; Oldfield, 1971) for listeners was +93% (Decile R.8), where values between +75 and +100 indicated extreme right handedness. A medical and musical screening questionnaire, the Edinburgh Handedness Inventory (Oldfield, 1971) and a hearing test assessed the eligibility of participants (described in the experiment-specific procedure Section 5.3.2 below). These three screening measures were also used throughout experiments described in Chapters 8 and 9. All participants reported normal or corrected-to-normal vision, and had clinically normal hearing (≤25 dB hearing thresholds for audiometric frequencies between 250-8000 Hz). No listeners reported a history of hearing, neurological, and/or psychological impairment, nor use of psychiatric medication or substance misuse. Four listeners (#05, #07, #08 and #16) were formally musically trained on the clarinet, drums, keyboard and piano, achieving grades 3-7 across 5-25 years. Seven listeners (four of whom were not musically trained; #02, #05, #08, #09, #13 #14, and #16) reported informal musical experience on the bass guitar, guitar, keyboard and piano, gained over 2-3 years.

Nine subjects (#19-27, not reported above) were excluded or withdrew from Experiment 5. One listener was excluded from the fMRI study because their hearing thresholds were too high (i.e., ≥25 dB HL). The other listener was excluded because they had metal body adornments that could not be removed for scanning. The remaining seven listeners, although eligible, failed to arrange an appropriate date to participate in the next stage (withdrawn). Listeners gave written informed consent, and the experiment was approved and performed in accordance with the College Research Ethics Committee (ethics code 2011/46), Nottingham Trent University, and the Medical School Research Ethics Committee, University of Nottingham (ethics code I12012012).

5.2.2 Stimuli

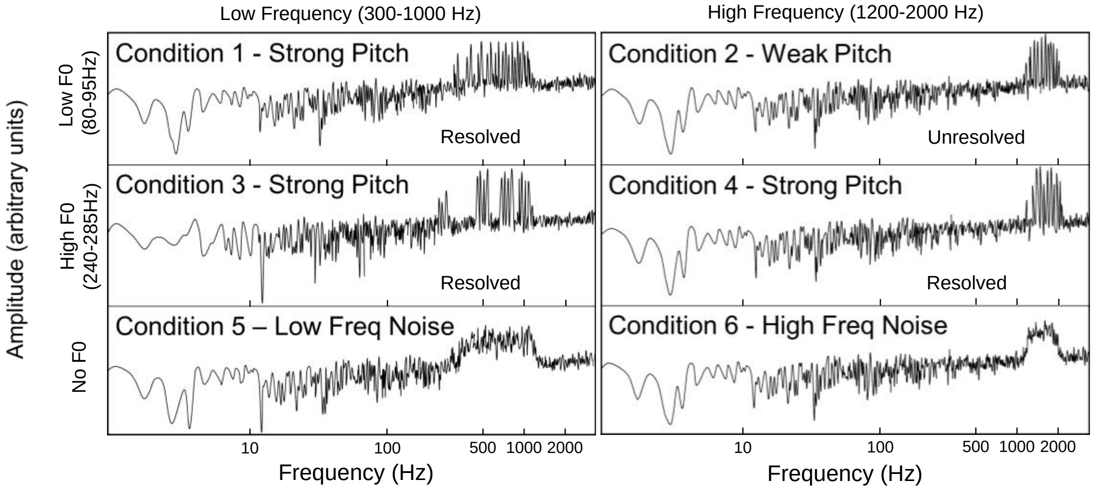

Figure 5.1. A Fast Fourier Transform of the six sound conditions used for Experiments 4 (active listening ratings task) and 5 (fMRI passive listening task), shown on a logarithmic scale.

The stimuli and conditions for the ratings task for Experiment 4 and the fMRI listening task for Experiment 5 were identical to those used in Penagos et al. (2004; see Figure 5.1). Pitch conditions were comprised of a 32s sequence of harmonic complex tone bursts that varied in pitch and were embedded in a background of continuous noise. The four pitch conditions were created by filtering harmonic complex tones into high and low spectral regions to create stimuli with strong pitch salience (conditions 1, 3, and 4), and weak pitch salience (condition 2). Complex tones used in conditions 1 and 2 (resolved and unresolved, respectively) were created by band-pass filtering harmonic tones with F0s between 80 and 95 Hz into low (340-1100 Hz) or high (1200-2000 Hz) spectral regions, respectively (see Penagos et al., 2004). Complex tones used in conditions 3 and 4 (both resolved) were created by band-pass filtering harmonic tones with F0s between 240 and 285 Hz into the same low and high spectral regions, respectively (see Penagos et al., 2004). This yielded four conditions that varied in resolvability and hence perceived pitch salience (3 strong, 1 weak), but were controlled for temporal regularity (i.e., the number of waveforms that repeated over time were the same for all conditions). The pitch conditions were therefore comprised a series of tonal sounds that gave rise to a random melody sequence. Low frequency Condition 1 approximately (depending on combination of F0s) included predominantly resolved harmonics 4, 5, 6, 7, 8, 9, 10, 11, and 12. Low frequency Condition 3 approximately (depending on combination of F0s) included resolved harmonics 2, 3 and 4. High frequency Condition 2 included approximately (depending on the combination of F0s) unresolved harmonics 15 16, 17, 18, 19, 20, 22, 22, 23, 24 and 25. High frequency Condition 4 included approximately (depending on the combination of F0s) resolved harmonics 5, 6, 7 and 8. To control for spectral differences between pitch conditions, two additional noise conditions (no pitch salience; 12 KHz wide-band white noise) were filtered into the same low and high spectral regions as the pitch conditions (conditions 5 and 6, respectively). Both pitch and noise conditions were presented as 300ms bursts (with 10ms cosine onset and offset ramps) at a rate of 1.67 per s within a 32s sound sequence. A continuous Gaussian wide-band background noise masker was embedded in all of the pitch stimuli at approximately 20 dB above their masked threshold to mask distortion products (estimated by Penagos et al. (2004) from their pilot psychophysical data) (fulfilling the fourth pitch criterion, ‘accounting for confounding’ variables; see Section 1.4). All sounds were originally sampled at 24 kHz, but were resampled to 20.05 KHz for stimulus presentation. All filtering used a 256 point Hanning window in the frequency domain, and cut-off frequencies were determined by the point at which the filter gain was attenuated by 6 dB (see Penagos et al., 2004).

5.2.3 Procedure and Design

All 18 participants took part in both screening and behavioural (Experiment 4) and fMRI listening (Experiment 5) tasks. Participants’ completion of Experiment 4 always preceded Experiment 5, and generally on a different day (i.e., between 2-4 weeks depending on participant availability). All participants for the current studies, as well as for all subsequent experiments, were recruited via Nottingham Trent University’s online Research Participation Scheme or via email. Both experiments took approximately 2 hours in total. Participants were awarded psychological research credits based on their length of participation (i.e., awarded 1 credit per 10 minutes of participation, where appropriate).

5.3. Experiment 4: Ratings of pitch salience

5.3.1 Stimuli

The same sound stimuli and conditions were used in Experiment 4 as for fMRI sequences in Experiment 5 and Penagos et al. (2004). This was to ensure that the sound context was the same. However, to aid listeners in the ratings task, reference stimuli were created in Adobe Audition (version Cs6, Adobe Systems Incorporated) using the same high and low F0s as in Experiment 5 and Penagos et al. (2004). The pitch reference stimuli were therefore comprised to match the pitch conditions as closely as possible (i.e., a series of tonal sounds that gave rise to a random melody sequence). For the low F0 reference sound, F0s between 80-95 Hz were used (i.e., 80, 82.5, 85, 87.5, 90, 92.5, and 95 Hz). For the high F0 reference sound, F0s between 240-285 Hz were used (i.e., 240, 245, 250, 255, 260, 265, 270, 275, 280, and 285 Hz). Tones were 300ms in duration (with 5ms onset and offset ramps), and were randomly combined together and interleaved with 300ms of silence (i.e., 10 random repetitions of nominal F0s and 10 repetitions of silence) to create two melodic-like sequences (i.e., 6 s in duration) which were comparable to the experimental stimuli but had no background noise (i.e., one low F0 reference sound, and one high F0 reference sound).

5.3.2 Procedure and Design

Experiment 4 was completed in a sound attenuated booth at Nottingham Trent University. Part-way through testing, the booth become no longer available due to on-going building works, and so a small number of participants had to be tested in a quiet room at Nottingham Trent University. This also extended to all studies that followed the present chapter. However, ambient noise was not anticipated to affect the results for the rating task as listening was conducted at suprathreshold levels. Nevertheless audible thresholds during screening may have been affected by ambient noise, but not to a significant degree. All psychophysical measures were conducted on 17 inch MacBook Pro laptop computer using a Windows 7 professional operating system. Listeners were positioned approximately 60 cm from the centre of the computer screen. A high-fidelity sound card delivered stimuli binaurally using Sennheiser HD-280 circumaural headphones. Sound level was measured prior to testing using a Brüel & Kjær 4231 Sound Calibrator, affixed with a Brüel & Kjær 2250 Sound Level Meter, Brüel & Kjær 4153 artificial ear, and Brüel & Kjær 4192 Half Inch Microphone. Sound pressure levels were calibrated for the MacBook Pro laptop computer using the “LAF” setting, which simulates hearing in a real listener whilst presenting a 1 kHz tone at maximum amplitude, and measuring dB sound pressure level (SPL) outputs for the left (117.3 dB SPL) and right (118.6 dB SPL) side of the headphones. The sound level on the laptop was then attenuated to present sounds over headphones at an overall level of 70 dB SPL.

Participants first completed an experimental consent form as well as three screening measures. The ‘medical and musical screening’ questionnaire asked a series of questions relating to health and hearing history, as well as previous musical experience. The Edinburgh Handedness Inventory (Oldfield, 1971) assessed listener’s handedness and degree of right handedness, whereby LQ values between +75 and +100 indicated extreme right handedness. Participants who did not meet any of the exclusion criteria identified from the screening questionnaires then completed a hearing test. A hearing test ensured listener’s hearing thresholds were within normal range (≤25 dB HL). Hearing screening reported here, and throughout the thesis, were conducted using bespoke software known as Earlab (i.e., not a clinical audiometer; provided by Prof. Chris Plack). Specifically, the hearing test required participants to listen to two observation intervals (time periods) and indicate in which interval they heard a sound, pressing buttons 1 or 2 accordingly. Each ear was tested separately at audiometric frequencies between 250 and 8000 Hz.

The main purpose of Experiment 4 was to obtain perceptual ratings in the same subjects who participated in the following fMRI study and to confirm that the ratings corresponded to the expectation of Penagos et al. (2004) that condition 2 was perceived as a weak pitch, and conditions 1, 3 and 4 were a strong pitch. Participants were firstly asked to listen to a reference sound for a given block to give them an indication of what a highly salient pitch stimulus sounded like. Reference sounds represented 10 on the rating scale (the most salient pitch stimulus). For conditions 1, 2 and 5 the corresponding low F0 reference sound was used (block 1), and for conditions 3, 4, and 6 the corresponding high F0 reference sound was used (block 2). Participants were then required to listen to the three experimental sounds for a given block and individually rate how salient the pitch was relative to the reference sound for the corresponding F0. Participants recorded their responses on the answer sheet provided by placing a tick under the number that corresponded to how salient they thought the experimental sequences were, from 0 (i.e., no pitch) to 10 (i.e., very salient pitch). Participants listened to a random version of each experimental condition, where each stimulus was presented three times (i.e., total of six blocks) to obtain an average rating for each condition. Participants always completed blocks 1 and 2 in the same alternating order (i.e., 1, 2, 1, 2, 1, and 2). Experiment 4 took approximately 1 hour in total for each listener. Upon completion, participants were granted the appropriate research credits.

5.3.3 Statistical Analysis

The average rating for each condition was subjected to 2 x 3 repeated measures ANOVA. The first factor related to the spectral property of the pitch stimuli (Frequency: low or high frequency band), whilst the second factor related to the pitch percept (Pitch: low pitch, high pitch, no pitch). To determine an effect of pitch salience, whether the interaction term between spectral frequency and pitch was significant was of particular interest here. Based on the hypotheses outlined previously, it was expected that conditions that had a more salient pitch (conditions 1, 3 and 4) would be rated higher than conditions that had a weaker pitch salience (condition 2), or no pitch salience at all (conditions 5 and 6), fulfilling the ‘covariation with pitch salience’ (Objective 2) and ‘pitch selectivity’ criteria (Objective 1). A significance level alpha criterion of $p < .05$ was used. All post hoc paired sample t-tests were Bonferonni corrected to control for multiple comparisons. If Mauchly's test of sphericity was violated a Greenhouse-Geisser correction was used.

Here, and throughout this thesis, difference-adjusted Cousineau-Morey 95% confidence intervals were calibrated, specifically for the repeated-measures ANOVA designs used, so that the absence of an overlap accurately reflected a confidence interval for a difference between two means (see Baguley, 2012).

5.3.4 Results

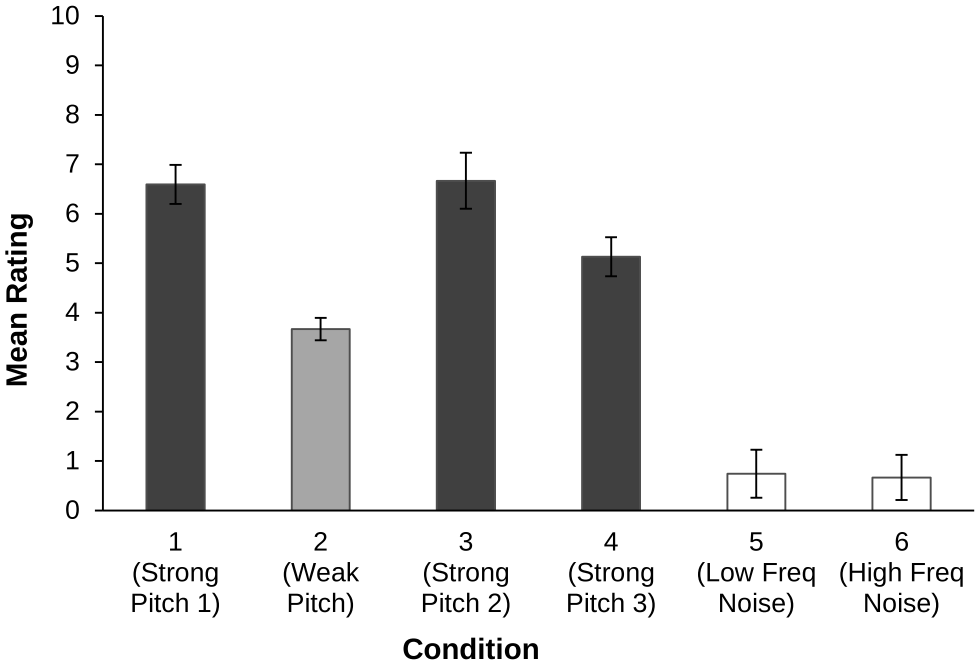

Figure 5.2. Mean subjective ratings of pitch salience for conditions 1-6 across all listeners. Dark gray bars depict strong pitch salience conditions (conditions 1, 3, and 4), light gray bars depict the weak pitch salience condition 2, and white bars depict low and high frequency noise conditions 5 and 6. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

The main objective for Experiment 4 was assessed by the interaction term from the 2 x 3 ANOVA (see Figures 5.2 and 5.3). There was a significant interaction between frequency and pitch, *F*(1.49, 25.31) = 25.99, *p* < .001, partial eta squared (η_p^2) = .605. As a rule of thumb, η_p^2 values .01, .08, and .14 reflect small, medium and large effects, respectively (see Richardson, 2011). Planned comparisons (paired sample t-tests) revealed that listeners did rate condition 2 (*M* = 3.67, *SD* = 1.23) as weaker in pitch salience than condition 1 (*M* = 6.59, *SD* = 1.94), condition 3 (*M* = 6.67, *SD* = 2.13), and condition 4 (*M* = 5.13, *SD* = 1.73); *t*(17) = 10.47, *p* < .001, *t*(17) = 6.56, *p* < .001, and *t*(17) = 5.48, *p* < .001, respectively. This indicates that there is a perceptible difference in pitch salience between the conditions, satisfying the ‘covariation with pitch salience’ criterion (Objective 2).

Furthermore, there was no significant difference in ratings between strong pitch salience conditions 1 and 3 (low F0 conditions), *t*(17) = -0.25, *p* = .808, but there was a significant difference between strong pitch salience conditions 1 and 4 and conditions 3 and 4; *t*(17) = 4.80, *p* < .001] and *t*(17) = 3.84, *p* < .001, respectively. There was also no significant difference between subjective ratings for low and high frequency noise condition 5 (*M* = 0.74, *SD* = 0.81) and condition 6 (*M* = 0.67, *SD* = 0.70), *t*(17) = 0.57, *p* = .579, suggesting that the difference in ratings between pitch conditions was much more marked. All pitch conditions (1-4) were significantly different from their matched noise counterparts (conditions 5 or 6), $t(17) = 11.49$, $p < .001$, $t(17) = 9.55$, $p < .001$, $t(17) = 9.70$, $p < .001$, and $t(17) = 9.40$, $p < .001$ respectively, satisfying the 'pitch selectivity' criterion (Objective 1).

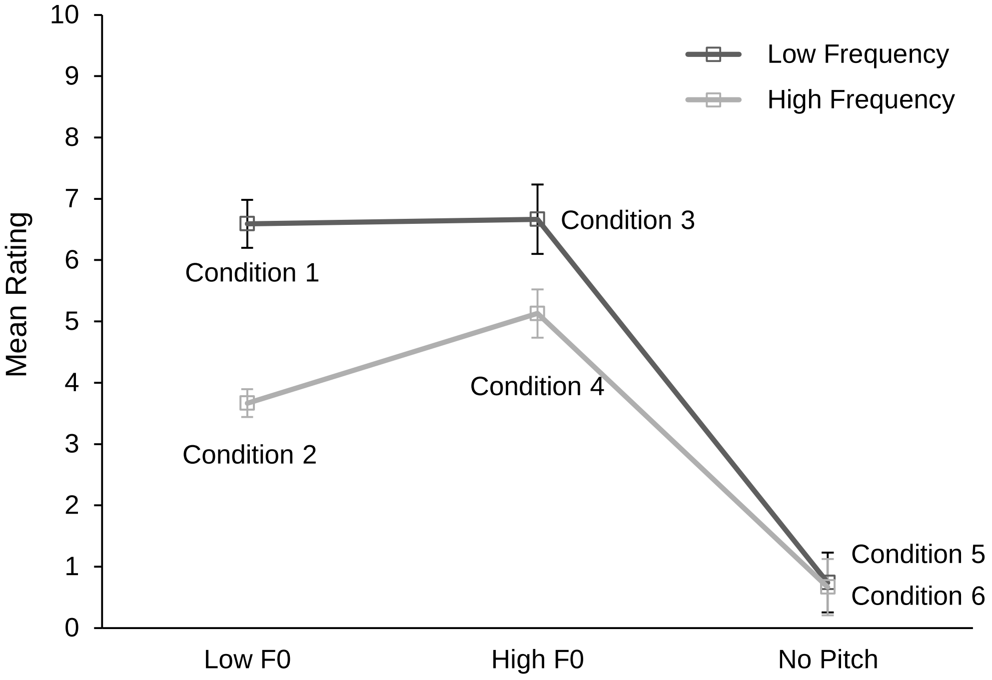

Figure 5.3. Interaction plot for frequency and F0 (pitch). Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

The ANOVA also revealed a significant main effect of frequency, $F(1, 17) = 69.11$, $p < .001$, $\eta_p^2 = .803$, where listeners rated low frequency sounds (conditions 1, 3, and 5) significantly higher in pitch salience ($M = 4.67$, $SD = 3.40$), than high frequency sounds ($M = 3.15$, $SD = 3.15$).

There was also a significant main effect of pitch, $F(1.33, 22.61) = 106.79$, $p < .001$, $\eta_p^2 = .863$, where subjective ratings for conditions with a low F0 ($M = 5.13$, $SD = 2.07$) or a high F0 ($M = 5.90$, $SD = 1.09$) were significantly different from each other ($p < .01$) and also greater ($p < .001$) than noise conditions with inherently no pitch ($M = 0.70$, $SD = 0.05$). Pairwise comparisons revealed that the 'no pitch' sounds were rated significantly lower in pitch salience compared to the other sounds, which was expected given that these sounds did not contain a pitch. Typically the no pitch condition was given a rating of 0 or 1.

5.3.5 Conclusion

Overall, listeners rated all pitch conditions as higher in pitch salience compared to noise, satisfying the pitch selectivity criterion for these stimuli (Objective 1). Listeners also rated the weak pitch salience condition as lower in salience compared to the other strong pitch salience conditions, satisfying the criterion for pitch salience sensitivity (Objective 2). It appears that subjectively at least, listeners could discern between weak pitch and matched noise as well as weak pitch salience and strong pitch salience.

5.4. Experiment 5: fMRI passive listening task

5.4.1 fMRI protocol

Eligible participants from Experiment 4 were invited to undergo the fMRI passive listening task comprising Experiment 5, at the Sir Peter Mansfield Resonance Centre, University of Nottingham. All scanning was performed using the same hardware as described in Chapter 4, namely the Philips 3 T Intera Achieva whole-body MR scanner. Functional data were acquired over three sequential scanning sessions over three separate days.

There were six different versions of random melody sequences generated for conditions 1-4. Noise conditions (conditions 5 and 6) were represented by only one version of each because only one version was used in the original Penagos et al. (2004) study, and I wanted to keep the stimulus conditions exactly the same. Sounds were presented using a quasi-continuous stimulation paradigm (see Krumbholz et al 2003, Section 2.4.2). The design of the current study gave careful consideration to optimise the statistical reliability of within-single subject inferences about condition-specific effects (e.g., sufficient number of data points per condition). Sixteen different stimulus lists were created and sequentially assigned to each listener. For a given list, each stimulus version was presented at least once, each sound condition was presented 8 times, and an additional ‘silent’ condition (32s of silence) was included and presented at least 8 times for a given listener (totalling seven conditions). Conditions were presented quasi-randomly but the same condition was never repeated twice in a row (to avoid adaptation effects). Each sequence lasted 32s and had an ISI of 2 s (34 s in total). Within each 34s sequence, a scan was acquired every 8.5 (TR) s (i.e., 4 times per sequence), totalling 32 scans per condition (totalling 224 scans per listener, plus 5-9 additional scans at the beginning of the session to account for the active noise cancellation system and stabilising of the MR signal). The active noise cancellation system aims to reduce unwanted ambient scanner noise by introducing an additional signal that is appropriately matched (i.e., antiphase) to effectively ‘cancel’ out the unwanted signal (see Section 3.3.1.2). However during testing the active noise cancellation system did not appear to be working properly for the first two listeners (#01 and #02), and therefore was discarded for the following 16 listeners.

A T1-weighted image (1mm^3 resolution, matrix size = 256mm x 256mm, 160 sagittal slices, TE = 3.9ms, and TR = 8.5ms) comprised the high-resolution anatomical scan on which to overlay functional analyses. The anatomical scan also helped to position the functional scans centrally on HG, include the entire STG, avoid inclusion of the eyes (which can cause

Nyquist artefacts) and position the saturation band (avoid aliasing for EPI functional scans; see Section 4.3.1.2, Figure 4.2).

A maximum of 233 T2*-weighted functional images were acquired for each listener using a gradient echo EPI sequence (acquired voxel resolution $2mm^3$; reconstructed 1.75 x 1.75 x 2mm; acquired matrix size = 84 x 84mm; reconstructed 96 x 96mm), 20 oblique-axial slices oriented parallel to the supra-temporal plane with no inter-slice gap, optimised TE = 50ms, flip angle = 90˚, SENSE factor = 2, and dynamic stabilisation to reduce signal drift over time). A single saturation band (rest slab thickness = 43mm, power = 2) was placed in the anterior position (front of head and perpendicular to EPI scans) to reduce aliasing and Nyquist ghosts from the eyes (see Section 4.3.1.2, Figure 4.2). An additional image using the same scanning parameters as the functional scans (with the exception of TE = 30ms, 70 slices) was acquired to help re-orient fMRI images during the pre-processing stage (see Section 3.5.1).

5.4.2 Sound Presentation Protocol

Functional scanning used a modified pulse sequence (i.e., SofTone factor 2) to reduce the impact of the ambient scanner noise on patterns of stimulus-evoked auditory activation and further attenuated the background scanner noise level by approximately 9 dB (Hall & Plack., 2009). Furthermore, scans were collected at regular intervals using a clustered acquisition mode; scans were clustered into a 1902ms acquisition time with a TR of 8.5 s, known as sparse imaging (Edmister, Talavage, Ledden, & Weisskoff, 1999; Hall et al., 1999; see Section 3.3.1.1). Conditions were presented predominantly in the quiet periods between each scan.

Sounds were delivered using custom-built MR compatible system delivered distortion-free sound using high-quality electrostatic headphones (i.e., Sennheiser HE60 with high-voltage amplifier HEV70) that had been specifically modelled with no ferromagnetic components to be safe for use in fMRI. The headphones had also been specifically modified for auditory fMRI (i.e., acoustically padded) to further attenuate ambient scanner noise by approximately 30 dB. Sound level was measured prior to testing using a Brüel & Kjær 4231 Sound Calibrator, affixed with a Brüel & Kjaer 2250 Sound Level Meter, Brüel & Kjær 4153 artificial ear, and Brüel & Kjær 4192 Half Inch Microphone. Sound pressure levels were calibrated for the MR compatible system using the “LAF” setting, which simulates hearing in a real listener, whilst presenting a 1 kHz tone at maximum amplitude and measuring dB SPL outputs for the left (99.6 dB SPL) and right (98.3 dB SPL) side of the headphones. Once the electrostatic headphones had been calibrated, the dB SPL output for the six sound conditions across headphone sides was measured using the “LAS” slow broadband scale. The average sound level measured across left and right sides was 88 dB SPL. The sound level on the MR sound presentation system was therefore rescaled and attenuated by -16 dB (using the MR sound presentation software) to avoid clipping, and have an overall root mean square sound level of approximately 72 dB SPL over headphones. This was purposefully chosen to be 3 dB SPL more than that presented in the Penagos et al. (2004) scanner setup (doubling in sound intensity) because piloting revealed that sounds were too quiet in the MR scanner when presented at 69 dB SPL over headphones (which is the sound level at which Penagos and colleagues presented stimuli).

Listeners were instructed to listen to the sounds being presented to them over the headphones but were not required to perform any task. Experiment 5 lasted approximately 1 hour in total for each listener, with functional scans being acquired in 35 minutes, and anatomical scans being acquired within 5 minutes. Upon completion, participants were granted the appropriate research credits, if applicable, and also sent a Digital Versatile Disc (DVD) of their anatomical brain scan.

All scanning data was converted from raw format (PAR/REC) into image format (IMG/HDR) using a bespoke programme known as ‘ptoa’. The following fMRI pre-processing, analysis and image generation were conducted primarily using SPM8 (www.fil.ion.ucl.ac.uk/spm/), which is supported by MATLAB R2010a Student v7.10.0 (The Mathworks, Natick, MA). MRIcro/MRIcron (Neuroimaging Informatics Tools and Resources Clearinghouse), and Adobe Photoshop CS6 (Adobe Systems Incorporated) software programmes were used to generate some images after the initial analysis had been carried out in SPM.

5.4.3 fMRI pre-processing

The anatomical and functional data were pre-processed following conventional procedures in SPM8 (e.g., realignment to smoothing) as described in Section 3.5.1. Slight differences in the pre-processing of images related to:

- Re-orientation: This was conducted by reorienting the more detailed functional 70 slice TE 30 image to match the MNI EPI (functional) template, and then reorienting the gradient echo EPI functional images to the same values. The 70 slice TE 30 functional image (reference) was then co-registered with the experimental functional images (source) to allow for better spatial normalisation later. The anatomical images were re-oriented to match the MNI T1 (anatomical) template as normal.
- Re-alignment: Using a rule of thumb, movements did not exceed a translation of 2mm and a rotation of 2 radians, because functional scans were acquired at a voxel size of 2mm^3. Pre-processing did not include the optional ‘slice timing correction’ as described in Section 3.5.1, however six multiple regressors were included in the GLM comprising subsequent fMRI analysis to correct for temporal differences/residual error variability between slices (acquired from the realignment translational parameters of x, y, z, and rotational parameters of pitch, roll and yaw, see Section 3.5.1, Figure 3.8).
- Co-registration: The first step in co-registration involved co-registering the 70 slice TE 30 functional (reference) image with the anatomical (source) image. The 70 slice

functional image was used instead of the mean functional image to maximise mutual information (i.e., more data points for SPM to estimate and coregister) and enabled better normalisation.

- Smoothing: The kernel used during spatial smoothing was different depending on the nature of the subsequent analyses. A smoothing kernel of 4mm^3 at FWHM was used for individual-based analysis (i.e., incidence maps, and conjunction analysis). A larger smoothing kernel of 8mm^3 FWHM was used for group-based analyses (i.e., group-wise activation maps and ROI analysis) as a trade-off between spatial resolution and inter-subject variability.

Normalised anatomical images for each individual were averaged to create one group-averaged anatomical image that could be used to overlay functional analyses.

5.4.4 fMRI analysis

Image analysis was conducted using GLM estimation and procedures described in Sections 3.5.2.1 to 3.5.2.5, some of which describe an example derived from the present fMRI study. A design matrix was firstly created for each participant, and individually reviewed to confirm it was correct at both individual (first level specification) and group level (second level specification). When estimating the GLM, a high–pass filter of 476 s was applied to the fMRI time series to alleviate the effects of variable amplifier gain and scanner drift during image acquisition (low frequency artefacts).

The GLM was used to create SPM activation maps in MNI brain-space. This approach performs a voxel-wise analysis of statistical contrasts of interest to determine whether there is consistent activity across the individual or group. Through carrying out an F test (non-directional test) and/or T-test (directional test), activity in each voxel in the normalised brain scan can be tracked across scans, and the β-values compared to create activation images.

Where F tests were conducted in SPM8 (e.g., group-wise activation described below) at the second level specification stage, a 2 x 3 repeated measures ANOVA was conducted. The first factor related to the spectral property of the pitch (frequency band: high, low), whilst the second factor related to the pitch percept (pitch; high pitch, low pitch, no pitch). Individual paired sample *t*-tests for each condition (1-6) > silence were used to create the 2 x 3 ANOVA in SPM8. ‘Offline’ analysis was also conducted in SPSS for ROI analysis described below; a 2 x 6 repeated measures ANOVA. The first factor related to the laterality of activation (hemisphere; left and right), while the second factor related to the condition (conditions 1-6). Main effects of interest from the F tests (non-directional) included:

1. The main effect of frequency (low frequency conditions 1, 3 and 5 ≠ high frequency conditions 2, 4, and 6).

2. The main effect of pitch (low F0 conditions 1, 2 ≠ high F0 conditions 3, 4 ≠ no pitch noise conditions 5, 6).
3. The interaction term between frequency and pitch (and corresponding post hoc t-tests).
4. The main effect of hemisphere (left hemisphere ≠ right hemisphere).
5. The main effect of condition (condition 1 ≠ condition 2 ≠ condition 3 ≠ condition 4 ≠ condition 5 ≠ condition 6).
6. The interaction term between hemisphere and condition (and corresponding post hoc t-tests).

Where t-tests were performed at an individual (first level specification), group-level analyses (second level specification), or offline in SPSS (e.g., post hoc paired sample t-tests for ROI analysis), the main statistical contrasts of interest were:

7. Sound > Silence (conditions 1-6 > silence)
8. Pitch > Noise (conditions 1-4 > conditions 5-6 (OR conditions 1 > 5, conditions 2 > 6, conditions 3 > 5 and conditions 4 > 6, for ROI analysis))
9. Strong Pitch Salience > Weak Pitch Salience (conditions 1, 3, & 4 > condition 2 (OR conditions 1 > 2, conditions 3 > 2 and condition 4 > 2, for ROI analysis))
10. Weak Pitch Salience > Matched Noise (condition 2 > condition 6)
11. To control for frequency and F0 differences between stimuli and ensure that they had no influence on the patterns of activity (see Section 5.4.5.1 or Penagos et al. (2004) for more details); control for spectral differences: conditions 3 > 4 and conditions 4 > 3; control for F0 differences: conditions 1 > 3 and conditions 3 > 1.

In SPM8, some of the above t-test contrasts of interest were embedded as planned comparisons in the 2 x 3 ANOVA, or conducted as part of a planned paired sample t-test comparison (i.e., for ROI analysis). Throughout the results section the number (#*N*) reported above for each effect/contrast of interest for a given F test or t-test is provided.

For the present fMRI study, robust group-level analyses were particularly important because they provided a measure of inter-subject variability; something which Penagos and colleagues (2004) did not address in their study. Four approaches were employed to provide different statistical information regarding the nature of pitch-related activation, namely group-wise activation, incidence mapping, ROI analysis, and comparing activation with a previous pitch salience study (i.e., Barker et al., 2011). The use of four different types of analysis necessitated the use of different statistical thresholds and different corrections for multiple comparison methods where appropriate (see Section 3.5.2.4). The specific details for each analysis approach are therefore described in the corresponding sections below.

Group-wise activation from the SPM analysis

Statistical group analysis was performed using a mixed effects model (see Section 3.5.2.2). Fixed effects analysis of the above contrasts of interest was computed first for single-subject analysis (first level specification) and then a RFX analysis was conducted across all 18 participants (second level specification). Specifically, contrast images (t-tests) were first computed for each participant, and then subjected to a one-sample t-test or an F-test (ANOVA) across the group.

An RFX analysis was conducted by specifying a 2 x 3 repeated measures ANOVA at the second level specification stage. The first factor related to the spectral property of the pitch (Frequency band: high, low), whilst the second factor related to the pitch percept (Pitch; high pitch, low pitch, no pitch). From this, activation from each contrast (whether t-test or from an F test) was separately mapped onto MNI brain-space and a table was created displaying information about all the peak supra-threshold voxels, z-values, xyz coordinates, corrected and uncorrected p-values (at voxel and cluster level), as well as number of activated voxels per cluster, for each main effect, interaction and/or contrast (previously identified above and embedded within the 2 x 3 ANOVA). Only the top two clusters (corresponding to left and right hemispheres) of activated voxels were selected. The top three peaks of activity per cluster which survived FDR correction (at the peak voxel level $p < .05$, zero voxels extent threshold, SVC for auditory cortex which included all ROIs; Section 3.5.2.4, see Table 5.1, and ‘ROI analysis’ section below)) were displayed where applicable (see Table 5.2). However, it should be noted that some peaks of activity did not survive FDR-correction ($p > .05$), but were significant at the $p < .001$ uncorrected voxel-level and so are included in Table 5.2 for illustrative purposes, but are denoted by an asterisk.

In order to obtain the probable (%) locations of significant clusters of activity for contrasts of interest, a SPM-based software programme, known as the Anatomy Toolbox v1.8 (see Eickhoff et al., 2005), was used to compare statistical images with cytoarchitectonic maps of the human brain. Using probability maps for a given contrast of interest, I was able to determine the likelihood of a given cluster of significant activation being located in, or ‘assigned’ to, a labelled area(s), for example ‘cluster 1 is located in Te1.0 (central HG) with a 70% probability’. This information is therefore also included in Table 5.2. One main drawback of using the template brain is that the toolbox did not have some brain regions specified, such as PT or PP, with activity in some areas being labelled as STG or unassigned completely. In addition, because it is a probabilistic atlas, activity could also be assigned to more than one brain region at a time (not mutually exclusive). Although we can use the ROI analysis (described later) to explore activation across conditions within PT or PP, because these regions are not defined in the toolbox we were unable to obtain corresponding probability values for clusters within these regions. In addition, inherent high inter-subject variability increases the likelihood of mapping activity to brain regions that do not actually correspond to the functional activity observed (Brett, Johnsrude, & Owen, 2002). The localisation information should therefore be interpreted with caution.

Incidence maps

The RFX analysis described above can be problematic if the mean activation is weak and the variability across subjects is high (Thirion et al., 2007). Given the inherent individual variability in group-averaged fMRI data, I employed the same approach that has been used previously within the Nottingham group to create incidence maps to visually plot the distribution of activity for the contrasts of interest (i.e., planned paired sample t-tests) across the group of listeners (see Hall & Plack, 2009; Hall & Griffiths, 2012). This was achieved by summing individual binary maps (p < .01, uncorrected; zero voxel extent threshold) obtained from each subject to form a group activation map which was overlaid onto a group-averaged anatomical image (see Figure 5.2). This meant that the number (and corresponding percentage) of listeners showing activity within a voxel for a given contrast could be examined. Therefore a less stringent threshold and t-test was used here to explore the distribution of sound (contrast of interest #07) and pitch-related (contrast of interest #08) activity across listeners. Slices were chosen based on co-ordinates from the most significant peak identified for the pitch-related paired sample t-tests, and are encompassed within the incidence maps (i.e., peak in right hemisphere). Central HG (area Te1.0), medial HG (area Te1.1), lateral HG (area Te1.2), PP and PT ROIs are outlined on the incidence map for reference (see also Figure 2.7, Section 2.3.3).

Regions-of-interest

ROI analysis was performed across a number of sub-regions in auditory cortex (see Section 3.5.2.5). Unlike the SPM contrasts described above, ROI analysis has the advantage of being able to quantify the magnitude and evaluate the overall activation within a given region for a stimulus condition, rather than at the voxel or cluster level across the whole brain over time. Five regions were chosen for analysis, which included three subdivisions of HG, namely central HG (area Te1.0), medial HG (area Te1.1) and lateral HG (area Te1.2), PT and PP (see Figure 2.7, Section 2.3.3). Each ROI acted like a mask image in which individually averaged β-values were extracted. The three HG masks were based on cytoarchitectonic probability maps created by Morosan et al. (2001) and were the same masks as those used by some researchers at the Nottingham group (e.g., Barker et al., 2011; Garcia et al., 2010). The PT mask was initially based on the morphological details provided by Westbury, Zatorre, & Evans (1999) who quantified the variability in PT. PP was also initially based on morphology and constructed in-house by the Nottingham group by tracing the outline of the anterior portion of the STG using a group-averaged normalised anatomical image. During specification, any ROIs with overlapping voxels were attributed to the area with which they had the highest probability of membership, therefore ensuring all ROI masks were mutually exclusive (i.e., not overlapping). This need for mutual exclusivity was the main difference between the ROIs and the Anatomy Toolbox, where a given voxel could be assigned to multiple brain areas at any one time with differing probabilities. Where SVC is used here and throughout this refers to the summed image of these five ROIs, unless otherwise stated (i.e., Te1.0, Te1.1, Te1.2, PP and PT).

After the initial ROI analysis had been conducted, a decision was made to subdivide the ROIs PP and PT. For PT, this decision was driven by finding that the majority of consistent sound and pitch-related activity was located in the anterolateral parts of PT in the initial incidence maps (not shown), revealing that the majority of the posterior part of PT was not activated by sound. Given that a high-number of voxels unresponsive to sound was being included in the average, this essentially diluted any potential statistically significant effects across conditions for this ROI. Activation in PP was at floor level and was similarly attributed to the inclusion of voxels unresponsive to sound. To circumvent these problems, both PT and PP ROIs were subdivided to only include voxels that were active for sound. This was achieved by using the 'Sound vs. Silence' contrast ($p < .001$, uncorrected, SVC) to create a mask image that was then applied to the given ROI. Table 5.1 shows the original size of all five ROI masks, as well as the reduced sizes for PT (left hemisphere: -64%; right hemisphere: -57%) and PP (left hemisphere: -96%; right hemisphere: -92%). It is appreciated that there are problems associated with defining ROIs from the data (i.e. circularity; see Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). However given that the PT region is known to be highly variable (Westbury et al., 1999) and that both PT and PP ROIs were partly-based on morphological details, this was not envisaged to be a major problem. It might be the case of PT, the voxels that did not respond to sound may have been activated by active listening conditions.

Table 5.1
Voxel size (N) of the five ROIs (and % reduced for revised ROIs PP and PT) for the left and right hemisphere

	Voxel size (*N*)	
ROI	Left hemisphere	Right hemisphere
Te1.0 (central HG)	377	459
Te1.1 (medial HG)	433	312
Te1.2 (lateral HG)	230	267
Original PP	1087	1667
Revised PP	47 (96% reduction)	127 (92% reduction)
Original PT	4510	3472
Revised PT	1608 (64% reduction)	1479 (57% reduction)

A MATLAB batch script was applied to extract the region-averaged β-values for each stimulus condition, for each listener, and for each ROI mask. For a given ROI there were corresponding left and right hemisphere masks that were run separately (as shown in Table

5.1). The output of this extraction process was a text file that included the mean standardised β-values (arbitrary values) of the response size for a given listener, condition and ROI hemisphere that could be input into a spread sheet to produce tables and graphs to compare activation across the six experimental conditions for each ROI.

A 2 x 6 repeated measures ANOVA was conducted separately for each region. The first factor related to the laterality of activation (hemisphere; left and right), while the second factor related to the condition (conditions 1-6). A significance level alpha criterion of $p < .05$ was used. Post hoc t-tests were collapsed across hemispheres (unless otherwise stated) for ROIs central HG, medial HG, lateral HG, PT, and PP because there was no significant interaction between hemisphere and condition. Planned comparisons and post hoc paired sample t-tests were Bonferonni corrected to account for FWE associated with multiple comparisons. If Mauchly's test of sphericity assumption of sphericity was violated a Greenhouse-Geisser correction was used.

Comparing activation with a previous pitch salience study

Imaging data obtained from Barker et al. (2011; see paper for details), who used fixed pitch stimuli (unresolved harmonic complex tones), was compared to the current study’s findings to explore any differences in pitch and pitch-salience related activation patterns, which may possibly relate to the context of pitch stimuli. Contrasts carefully selected ($p < .05$, uncorrected) from Barker et al. (2011) for comparison included:

1) ‘Schroeder phase + Random phase > Gaussian Noise’, which was comparable to ‘Pitch > Noise’ contrast (contrast of interest #08) in the current experiment.
2) ‘Schroeder phase > Random phase’, which was comparable to ‘Strong > Weak Pitch Salience’ contrast (contrast of interest #09).

Contact was made with Daphne Barker to obtain electronic versions of these contrast images thresholded at $p < .05$ and $p < .001$, so a comparative assessment could be conducted. A less stringent threshold at the $p < .05$ level was used for the contrast images because the data taken from Barker et al. (2011) did not show much pitch-related activation at the $p < .001$ level. Slices were chosen according to peak coordinates identified in SPM8 for the 2 x 3 ANOVA (see Table 5.2) and paired sample t-tests (not embedded in the 2 x 3 ANOVA).

5.4.5 Results

5.4.5.1 Objective 1: fMRI findings for pitch selectivity

Group-wise activation from the SPM analysis

Probability maps were used to localise the peaks of activity from the 2 x 3 repeated measures ANOVA in SPM8 (see Table 5.2). For sound (conditions 1-6) > silence (contrast of interest #07), there was a significant cluster in the left hemisphere (3693 voxels) with a peak of activity at coordinates x 50, y -16, z 6mm, $t(102) = 32.25$, $p < .05$ FDR-corrected and SVC, and a significant cluster in the right hemisphere (3351 voxels) with a peak of activity at coordinates x 40, y -26 z 6mm, $t(102) = 28.30$, $p < .05$ FDR-corrected and SVC. Overall activation was widespread, with multiple auditory areas being responsive to sound > silence, namely lateral HG, central HG and STG. However across hemispheres, the central HG was most consistently assigned as being attributable for significant peaks of activity for sound > silence.

The effect of pitch (F0; conditions 1-4) > noise (conditions 5-6; contrast of interest #08) was statistically significant for a cluster of activity (304 voxels) in the right hemisphere, with a peak at x 50, y 4, z -8mm, $t(102) = 5.52$, $p < .05$ FDR-corrected and SVC. There was also a cluster of significant activity in the left hemisphere (106 voxels), with a peak at x -66, y -14, z 4mm, $t(102) = 4.35$, $p < .001$, uncorrected. For this hemisphere, peak voxels did not survive FDR-correction and SVC. Nevertheless the results showed that activation to pitch was greater than noise, and that pitch-related activity was widespread across subdivisions of auditory cortex (e.g., lateral HG, central HG and STG), indicating that multiple auditory areas were more sensitive to pitch than sound in general.

There were no significant voxels ($p > .05$, FDR-corrected and SVC) found for the main effect of frequency (contrast of interest #01; see Table 5.2).

Not shown in Table 5.2, but as an additional control for frequency effects (contrasts of interest #11) also included the t-test comparison between condition 3 vs. condition 4 (see Penagos et al., 2004), as these were the only conditions that could control for low and high spectral differences between stimuli (respectively) without being different in F0 or pitch salience (resolvability). This contrast also revealed no suprathreshold voxels ($p > .05$). To control for F0 differences, condition 3 vs. condition 1 were compared (see Penagos et al., 2004) as these were the only conditions that could control for high and low F0 between stimuli without differing in spectra or pitch salience (resolvability). This t-test comparison also revealed no suprathreshold voxels. Comparing conditions 3 vs. 4 (frequency control) and conditions 3 vs. 1 (F0 control) were also employed as appropriate controls during ROI analysis (see ROI analysis sections below).

Table 5.2

Significant clusters of activity, information relating to the peak voxels and locality from the 2-by-3 repeated-measures ANOVA specified in SPM8, for the main effects, interaction and contrasts of interest (#N)

	Left hemisphere							Right hemisphere						
	Peak	Peak coordinates (x, y, z)	Z-score	Voxel-level *p*-value*	Location	Probability (%)	Cluster size (voxels)	Peak	Peak coordinates (x, y, z)	Z-score	Voxel-level *p*-value*	Location	Probability (%)	Cluster size (voxels)
Sound > Silence (#7)	1	-50 -16 6	Inf	.001	Te1.0 Te1.2	70 40	3693	1	40 -26 16	Inf	.001	OP 1 OP 2 Te1.1	50 40 30	3351
	2	-38 -28 16	Inf	.001	OP 1 Te1.1	60 60		2	48 -16 10	Inf	.001	Te1.0 OP1	50 40	
	3	-44 -24 10	Inf	.001	Te1.0 OP1	80 20		3	42 –22 10	Inf	.001	Te1.1 Te1.0	50 40	
Main effect of Frequency (#1)	1	-54 -16 8	3.17	.916*	Te1.0 Te1.2	40 30	3	1	62 6 -2	3.12	.916*	-	-	1
Main effect of Pitch (#2)	1	-54 -14 6	4.07	.104*	Te1.2 Te1.0	50 30	89	1	50 4 -8	5.08	.005	OP 4	10	224
	2	-66 -14 4	4.00	.104*	Te3	90		2	62 4 0	4.67	.017	IPC	10	
	3	-56 -10 4	3.99	.104*	Te1.2 OP 4 Te1.0	40 30 20		3	56 -8 6	3.18	.874*	Te1.0 OP 4 Te1.2	50 40 40	
Frequency x Pitch Interaction (#3)	No suprathreshold voxels							No suprathreshold voxels						
Pitch > Noise (#8)	1	-66 -14 4	4.16	.077*	Te3	90	106	1	50 4 -8	5.15	.003	OP 4	10	304
	2	-56 -10 4	3.79	.170*	Te1.2 OP 4 Te1.0	40 30 20		2	62 4 0	4.76	.010	IPC	10	
	3	-68 -22 8	3.39	.429*	Te3	10		3	-	-	-	-	-	
Strong > Weak (#9)	1	-52 -16 8	4.49	.061*	Te1.0 OP4 OP1	50 30 20	261	1	62 4 0	4.35	.061*	-	-	190
	2	-52 -2 0	3.90	.222*	Te1.2 OP4	40 20		2	50 6 -8	3.72	.292*	-	-	
	3	-66 -12 4	3.61	.292*	Te3	60		3	52 0 -2	3.61	.292*	-	-	

Notes: Here significance refers to those clusters of activated voxels which survive FDR correction for multiple comparisons ($p < .05$) at the peak voxel level. Asterisk (*) denotes peaks that did not survive FDR-correction, but were significant at an uncorrected level of $p < .001$. Peak voxels within those clusters are reported for local maxima more than 4mm apart within left and right hemispheres, respectively. Te3 = area of non-primary auditory cortex overlapping with parts of PP and PT (see Morosan et al., 2001); OP1-OP4 = Parietal Operculum (see Eickhoff, Schleicher, Zilles, & Amunts, 2006); IPC = Inferior Parietal Cortex.

Incidence maps

Using paired sample t-test contrasts the incidence map (Figure 5.4, Sound > Silence, row 1, contrast of interest #07) depicted that response to sound was much more consistent than to pitch (maximum overlap 18/18 listeners in some areas), with activity observed in the majority of primary and non-primary auditory areas, and in all of our five ROIs (central HG, medial HG, lateral HG, PP and PT). The most significant peaks for this contrast were identified in SPM in the left hemisphere at coordinates x -50, y -16, z 6mm, $t(17) = 15.17$, $p < .05$ FDR-corrected and SVC, and for the right hemisphere at coordinates x 40, y -16, z 16mm, $t(17) = 12.65$, $p < .05$ FDR-corrected and SVC.

Response to pitch > noise (see Figure 5.4, row 2, contrast of interest #08) revealed less extensive activity across listeners (compared to sound) with a maximum overlap in approximately 13/18 listeners (~72%) across left and right auditory cortices (both primary and non-primary areas). Pitch-related activation was widely distributed and evident in all ROIs, however activation was most consistently observed in HG and PT. Significant peaks for pitch > noise were identified in SPM in the right hemisphere at coordinates x 58, y -6, z 4mm, $t(17) = 9.12$, $p < .05$ FDR-corrected SVC, and in the left hemisphere at coordinates x -56, y -18, z 4, $t(17) = 8.15$, $p < .05$ FDR-corrected and SVC.

Data presented in rows 3 and 4 of Figure 5.4 are described later in Section 5.4.5.2, as these findings relate to the second experimental objective, pitch salience.

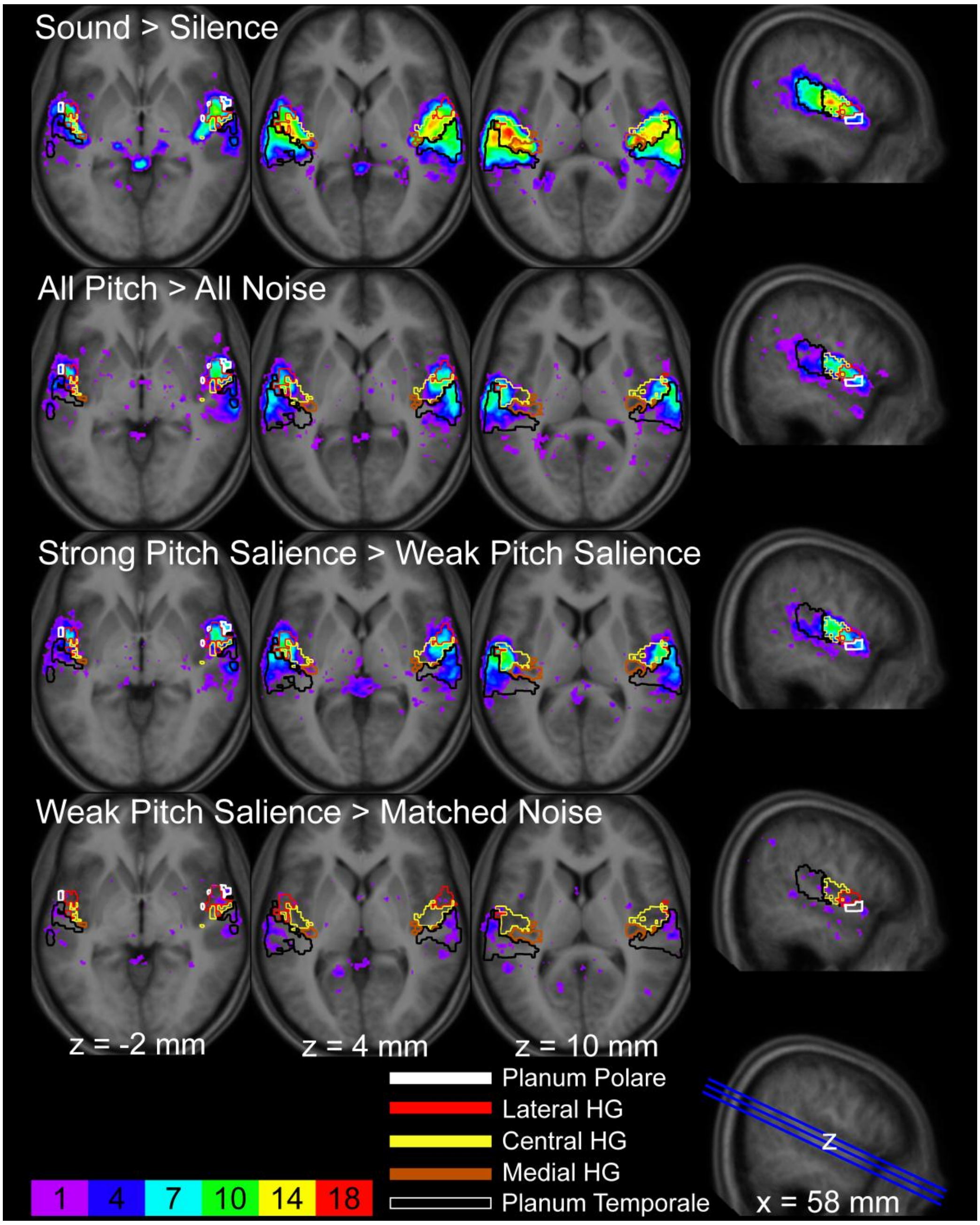

Figure 5.4. Incidence maps showing the consistency of activation across 18 listeners for the four contrasts of interest (#07: Sound > Silence, #08: All Pitch > All Noise, #09: Strong Pitch Salience > Weak Pitch Salience, and #10: Weak Pitch Salience > Matched Noise; p < .01, uncorrected) across three oblique-axial slices (z = -2, 4 and 10mm) and one right hemisphere sagittal slice (x = 58mm) of auditory cortex. The colour bar (range 1-18) indicates the number of listeners showing brain activity overlap (i.e., how many listeners had that voxel as active for a given contrast; red = high spatial concordance, purple = low spatial concordance). Activation has been smoothed and overlaid onto a group averaged anatomical image conforming to neurological convention (left = left). ROI outlines have been overlaid onto contrasts to aid interpretation.

ROI analysis

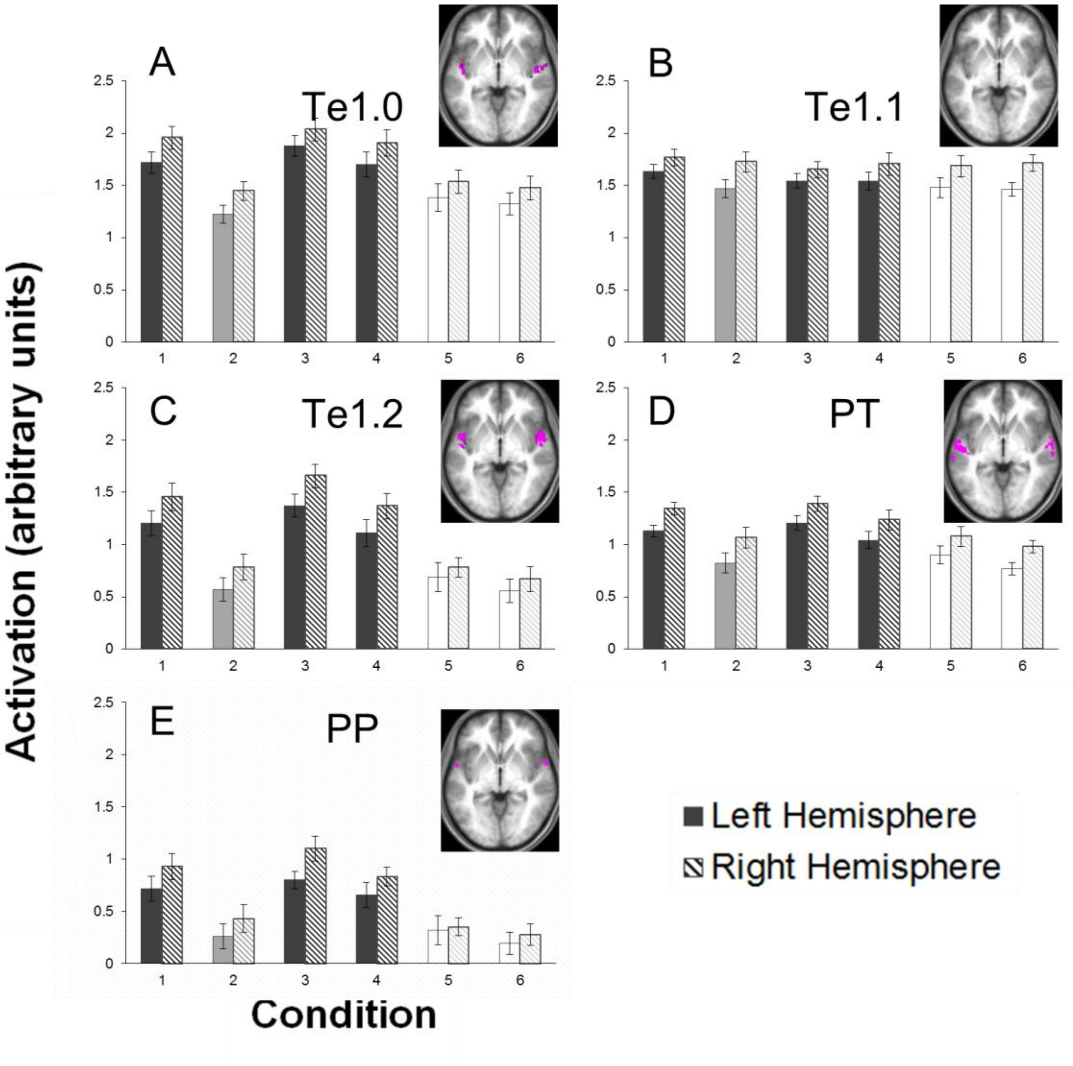

Figure 5.5. Group ROI analysis depicting BOLD activation for sound conditions (1-6) > the implicit 'no sound' baseline. Activation is plotted separately for left and right hemispheres for the five ROIs (panels A-E; with within-subjects Cousineau-Morey 95% confidence intervals plotted). Panel A depicts central HG (area Te.1.0), Panel B depicts medial HG (area Te.1.1), Panel C depicts lateral HG (area Te.1.2), Panel D depicts PT, and Panel E depicts PP. These five ROIs have been overlaid onto a group-averaged anatomical image (shown in purple) oblique-axial slice (z = -2mm) for illustration purposes conforming to neurological convention (left = left). Bars have been grayscale colour-coded to aid interpretation. Solid coloured bars indicate the left hemisphere; striped bars indicate the right hemisphere. Dark gray bars (solid and striped) indicate strong pitch salience (conditions 1, 3 and 4). Medium gray bars (solid and striped) indicate the weak pitch salience (condition 2). White and light gray (solid and striped) bars indicate noise (conditions 5 and 6).

β-values were used to conduct a separate two (hemisphere; left and right) by six (conditions; 1-6) repeated measures ANOVA for each of the five ROIs.

Central HG (area Te1.0)

A significant main effect (#04) of hemisphere, *F*(1, 17) = 5.08, *p* = .380, η_p^2 = .230, indicated significantly greater activation across conditions in the right hemisphere (*M* = 1.73, *SD* = 0.65) compared with the left hemisphere (*M* = 1.54, *SD* = 0.53).

There was also a significant main effect (#05) of condition, *F*(5, 85) = 14.45, *p* < .001, η_p^2 = .459. For Te1.0 (see Figure 5.5), post hoc t-tests revealed that all strong pitch salience (conditions 1, 3 and 4; individual contrasts of interest #08) produced significantly greater activation than their spectrally matched noise (conditions 5 or 6) across hemispheres, *t*(17) = -3.97, *p* = .001, *t*(17) = 4.62, *p* < .001, and *t*(17) = 4.50, *p* < .001, respectively for conditions 1, 3 and 4). Therefore responses to pitch were greater than to noise.

There was no significant interaction (#06) between hemisphere*condition; *F*(3.26, 55.46) = 0.78, *p* = .519, η_p^2 = .044.

Activation was comparable across all three strong pitch salience conditions (*p* > .05) indicating that previously described controls for frequency differences (i.e., condition 3 vs. 4) and F0 differences (i.e., condition 3 vs. 1) between stimuli were satisfied. The weak pitch salience condition (2), was not significantly different from its spectrally matched noise (condition 6); *t*(17) = -0.85, *p* = .409. In summary for Te1.0, although all strong pitch salience conditions fulfilled the first pitch criterion relating to pitch selectivity (Objective 1), the weak pitch salience condition did not.

Medial HG (area Te1.1)

As shown in Figure 5.5, activation across conditions was comparable for this ROI. There was no significant main effect of hemisphere, *F*(1, 17) = 3.69, *p* = .072, η_p^2 = .178, or condition, *F*(5, 85) = 0.70, *p* = .626, η_p^2 = .039, and no interaction between hemisphere*condition, *F*(2.64, 44.92) = 2.00, *p* = .134, η_p^2 = .105. Therefore responses to pitch were not significantly different from noise. In summary for Te1.1, none of the pitch conditions fulfilled the first pitch criterion relating to pitch selectivity (Objective 1). Because of these null results, it was not necessary to analyse data from medial HG with respect to Objective 2.

Lateral HG (area Te1.2)

There was a significant main effect of hemisphere, *F*(1, 17) = 12.13, *p* = .003, η_p^2 = .416, where greater activation occurred across conditions in the right hemisphere (*M* = 1.12, *SD* = 0.55) compared with the left hemisphere (*M* = 0.92, *SD* = 0.44).

There was also a significant main effect of condition, *F*(5, 85) = 26.88, *p* < .001, η_p^2 = .613. For Te1.2 (see Figure 5.5), post hoc t-tests revealed that all strong pitch salience (conditions 1, 3 and 4) produced significantly greater activation (*p* < .05) than their spectrally matched noise (conditions 5 or 6) across hemispheres, *t*(17) = 5.33, *p* < .001, *t*(17) = 7.94, *p* < .001, and *t*(17) = 6.21, *p* < .001, respectively for conditions 1, 3 and 4. Therefore responses to pitch were greater than to noise.

There was no significant interaction between hemisphere*condition, *F*(2.09, 35.53) = 2.01, *p* = .147, η_p^2 = .106, but see Section 5.4.5.3 regarding further exploration of possible laterality effects within this ROI.

Activation was comparable across all three strong pitch conditions (*p* > .05) indicating that previously described controls for frequency differences (condition 3 vs. 4) and F0 differences (condition 3 vs. 1) between stimuli were satisfied. Again the weak pitch condition (2) was not significantly different from its spectrally matched noise condition (6); *t*(17) = 0.66, *p* = .519. In summary for Te1.2, although all strong pitch salience conditions fulfilled the first pitch criterion relating to pitch selectivity (Objective 1), the weak pitch salience condition did not. This is the same pattern as in Te1.0.

PP

There was no significant main effect of hemisphere; *F*(1, 17) = 3.85, *p* = .066.

There was a significant main effect of condition, *F*(5, 85) = 19.20, *p* < .001, η_p^2 = .530. For PP (see Figure 5.5), post hoc t-tests revealed that all strong pitch salience (conditions 1, 3 and 4) produced significantly greater activation than their spectrally matched noise (conditions 5 or 6) across hemispheres, *t*(17) = 4.51, *p* < .001, *t*(17) = 6.27, *p* <.001, and *t*(17) = 5.15, *p* < .001, respectively for conditions 1, 3 and 4). Therefore responses to pitch were greater than to noise.

There was no significant interaction between hemisphere*condition, *F*(3.18, 54.00) = 2.49, *p* = .067, η_p^2 = .128, but see Section 5.4.5.3 regarding further exploration of possible laterality effects within this ROI

Activation was comparable across all three strong pitch conditions (*p* > .05) indicating that previously described controls for frequency differences (condition 3 vs. 4) and F0 differences (condition 3 vs. 1) between stimuli were satisfied. The weak pitch condition (2), was not significantly different (*p* > .05) from its spectrally matched noise condition (6); *t*(17) = 1.24, *p* = .232. In summary for PP, although all strong pitch salience conditions fulfilled the first pitch criterion relating to pitch selectivity (objective 1), the weak pitch salience condition did not. Again, this is the same overall pattern as the other ROIs.

PT

A significant main effect of hemisphere, *F*(1, 17) = 7.67, *p* = .013, η_p^2 = .311, revealed greater activation occurred across conditions in the right hemisphere (*M* = 1.18, *SD* = 0.53) compared to the left hemisphere (*M* = 0.98, *SD* = 0.40).

There was also a significant main effect of condition; *F*(5, 85) = 10.82, *p* < .001, η_p^2=.389. For PT (see Figure 5.5), post hoc t-tests revealed that all strong pitch salience (conditions 1, 3 and 4) produced significantly greater activation than their spectrally matched noise (conditions 5 or 6) across hemispheres, *t*(17) = 3.66, *p* = .002, *t*(17) = 3.77, *p* = .002, and *t*(17) = 3.77, *p* = .002, respectively for conditions 1, 3 and 4. Therefore responses to pitch were greater than to noise.

There was no significant interaction between hemisphere*condition, *F*(2.72, 46.30) = 0.88, *p* = .450, η_p^2 = .049, but see Section 5.4.5.3 regarding further exploration of possible laterality effects within this ROI.

Activation was comparable across all three strong pitch conditions (*p* > .05), indicating that previously described controls for frequency differences (condition 3 vs. 4) and F0 differences (condition 3 vs. 1) between stimuli were satisfied. The weak pitch condition (2), was not significantly different (*p* > .05) from its spectrally matched noise condition (6); *t*(17) = 1.02, *p* = .320. In summary for PT, although all strong pitch salience conditions fulfilled the first pitch criterion relating to pitch selectivity (objective 1), the weak pitch salience condition did not. Again, this is the same pattern as Te1.0.

Summary of ROI analysis for Objective 1

In summary (see Table 5.3) there was a significant overall effect of pitch in four auditory ROIs, fulfilling the first pitch criterion relating to pitch selectivity (Objective 1), but not in medial HG (area Te1.1). However, this was driven by the strong pitch salience conditions. There was no significant effect of pitch for the weak pitch salience comparison.

Table 5.3

Summary of pitch selectivity across ROIs (individual contrasts of interest #08; pitch > noise)

	Pitch selectivity (Objective 1)	
ROI	Strong pitch salience	Weak pitch salience
Te1.0	✓	✗
Te1.1	✗	✗
Te1.2	✓	✗
PP	✓	✗
PT	✓	✗

Note: PP and PT ROIs refer to the revised versions, as defined in Table 5.1.

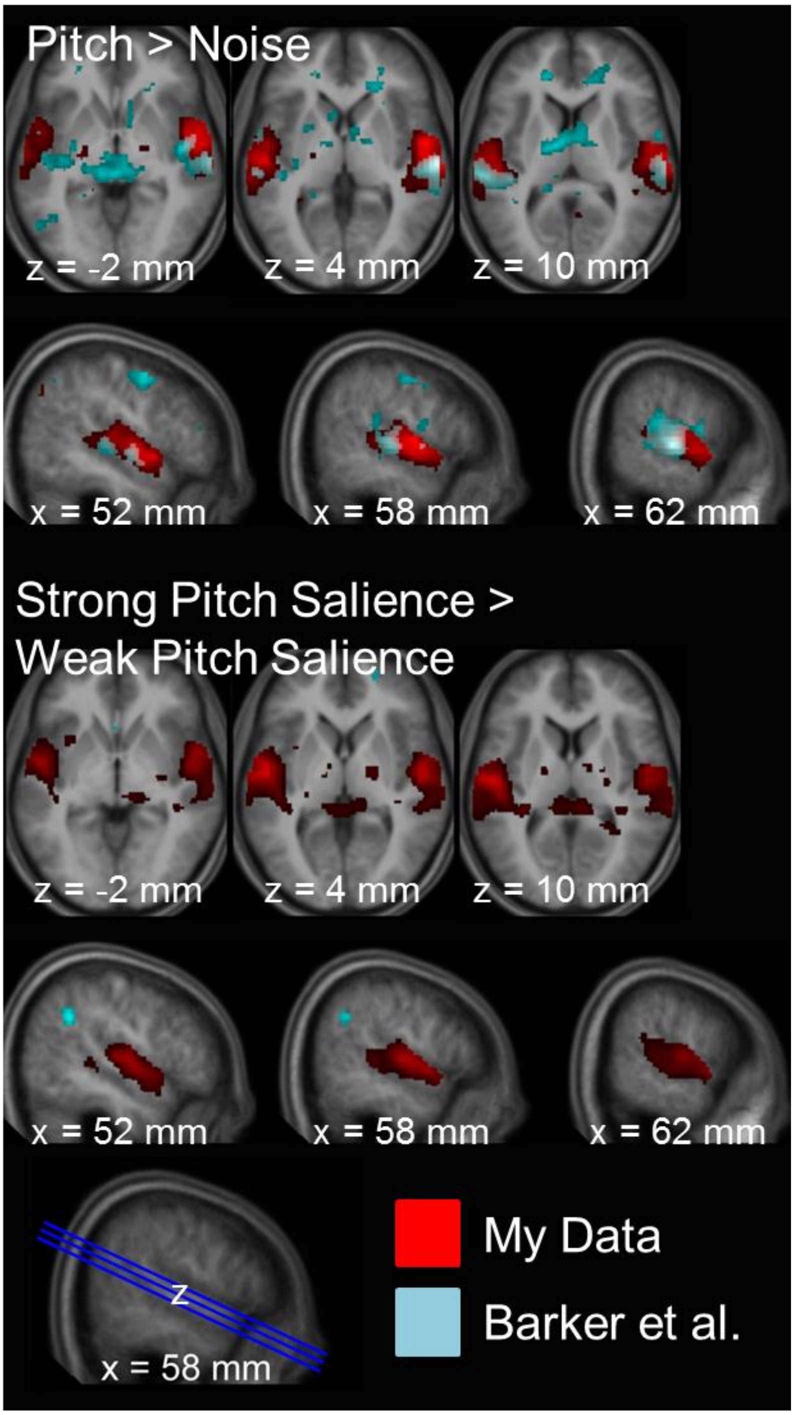

Figure 5.6. Activation maps (p < .05, uncorrected) taken from the current study (shown in red) and Barker et al., 2011 (shown in cyan) for comparable contrasts Pitch > Noise (#08) and Strong Pitch Salience > Weak Pitch Salience (#09). Activation has been smoothed and overlaid onto three oblique-axial slices (z = -2, 4 and 10mm) and three sagittal slice (x = 52, 58 and 62mm) of a group averaged anatomical image conforming to neurological convention (left = left).

As shown in Figure 5.6, activation for pitch > noise for the current experiment was generally more widespread, covering multiple primary and non-primary areas of auditory cortex. Visual inspection indicated that activation for pitch > noise for Barker et al. (2011) was more confined to the PT, but did overlap with activity for the current experiment in posterior parts of PP and HG.

Summary of evidence for pitch selectivity (Objective 1) from the four analysis approaches

Overall activation to sound is much more widespread across auditory cortex and consistent across listeners, than is activation to pitch. However findings from the current study indicate that there are multiple primary and non-primary auditory regions that selectively responded to pitch over noise, namely central and lateral HG, PP and PT. Although activation for all strong pitch salience conditions was significantly greater than for noise, this was not the case for the weak pitch salience condition. This finding was therefore interrogated further in other types of analyses that follow by exploring the weak pitch salience > matched noise contrasts.

5.4.5.2 Objective 2: fMRI findings for sensitivity to pitch salience (and it is co-localised with the pitch response)?

Group-wise activation from the SPM analysis

The effect of strong pitch salience > weak pitch salience (contrast of interest #09) for both hemispheres did not survive FDR-correction for multiple comparisons (see Table 5.2) after SVC, but was significant at an uncorrected level of $p < .001$, with a cluster of activity (261 voxels) in the left hemisphere, peaking at x -52, y -16, z 8mm, $t(102) = 4.73$, $p < .001$, uncorrected. There was also a cluster in the right hemisphere (190 voxels), peaking at x 62, y 4, z 0mm, $t(102) = 4.58$, $p < .001$, uncorrected. Activation for strong pitch > weak pitch (contrast of interest #10) occurred across multiple subdivisions of auditory cortex (e.g., lateral HG, central HG and STG). These results seem to suggest that there was some differential pitch-salience related activity that was co-localised to the same regions that was previously identified for pitch > noise (contrast of interest #08). Given that the previous ROI analysis (see Section 5.4.5.1) indicated the weak pitch salience condition was not significantly different from matched noise for any of the ROIs, this contrast was explored in SPM again. Findings revealed no suprathreshold voxels. This null result was therefore further explored in the incidence map.

Incidence maps

The incidence map (see Figure 5.4, row 3) revealed activation across subjects for strong pitch salience > weak pitch salience (contrast of interest #09) had a maximum overlap in approximately 15/18 listeners (~83%) across left and right auditory cortices (both primary and non-primary areas). Activation was most consistently observed in central and lateral HG and anterolateral parts of PT. The most significant peaks for this contrast were identified by SPM in the left hemisphere at coordinates x -54, y -14, z 6mm, $t(17) = 11.93$, $p < .05$ FDR-corrected and SVC, and in the right hemisphere at coordinates x 56, y -14, z 4mm, $t(17) = 8.38$, $p < .05$ FDR-corrected and SVC. A maximum of ~50% of listeners showed pitch salience-related activity overlapping with the pattern of pitch-related activity involving central and lateral HG, PP and PT. This result seems to suggest a sensitivity to pitch salience which is co-localised to the same four ROIs which showed a pitch-related response in Section 5.4.5.1. This does provide some evidence in favour of pitch constancy (see Section 1.2).

However, the ROI analysis described in Section 5.4.5.1 revealed that the weak pitch salience condition did not exhibit any statistically significant differential activation compared to spectrally matched noise (contrast of interest #10). In other words, condition 2 failed to satisfy the first pitch criteria of pitch selectivity (Objective 1). This contrast was further explored using incidence maps here. It is possible that pitch salience was co-localised to the same regions identified in Section 5.4.5.1 because the strong pitch salience conditions are driving the pitch-related response (see Figure 5.4, rows 2 and 3). Unlike the group-wise activation observed using SPM8, the incidence map (see Figure 5.4, row 4) did show some significant activation for a maximum of 5 listeners (~28%) for weak pitch salience > matched noise (#10). This was the least consistent activation observed across all contrasts and involved fewer auditory regions (i.e., lateral HG and PT). Although the group-based analyses (activation maps and ROI analysis) failed to find a significantly differential activation for weak pitch salience > matched noise (contrast of interests #10), the incidence map did indicate 5 listeners who did display differential responses in favour of the weak pitch salience condition.

This finding prompted individual-level conjunction analysis of ‘Strong > Weak’ (contrast of interest #09) AND ‘Weak > Matched Noise’ (contrast of interest #10) on the activation data, which would reveal any brain regions that show a parametric change in relation to pitch salience. The conjunction revealed no suprathreshold voxels when using a stringent threshold ($p < .05$, FWE-corrected). As shown in Figure 5.7, when a more lenient threshold ($p < .001$, uncorrected) was used, 14/18 (~78%) listeners showed a small number of voxels of mutual activation across left and right hemispheres of auditory cortex, suggesting that there are some voxels, across primary and non-primary auditory cortex, that were truly responsive to pitch salience in the parametric sense (see Section 3.4.3). Activity occurs across the same auditory regions as identified in Objective 1 for pitch selectivity, providing evidence for pitch constancy (Section 1.2). Figure 5.8 shows that when an even more lenient threshold was used ($p < .05$, uncorrected), all 18 listeners showed pitch salience-related activation across auditory cortex.

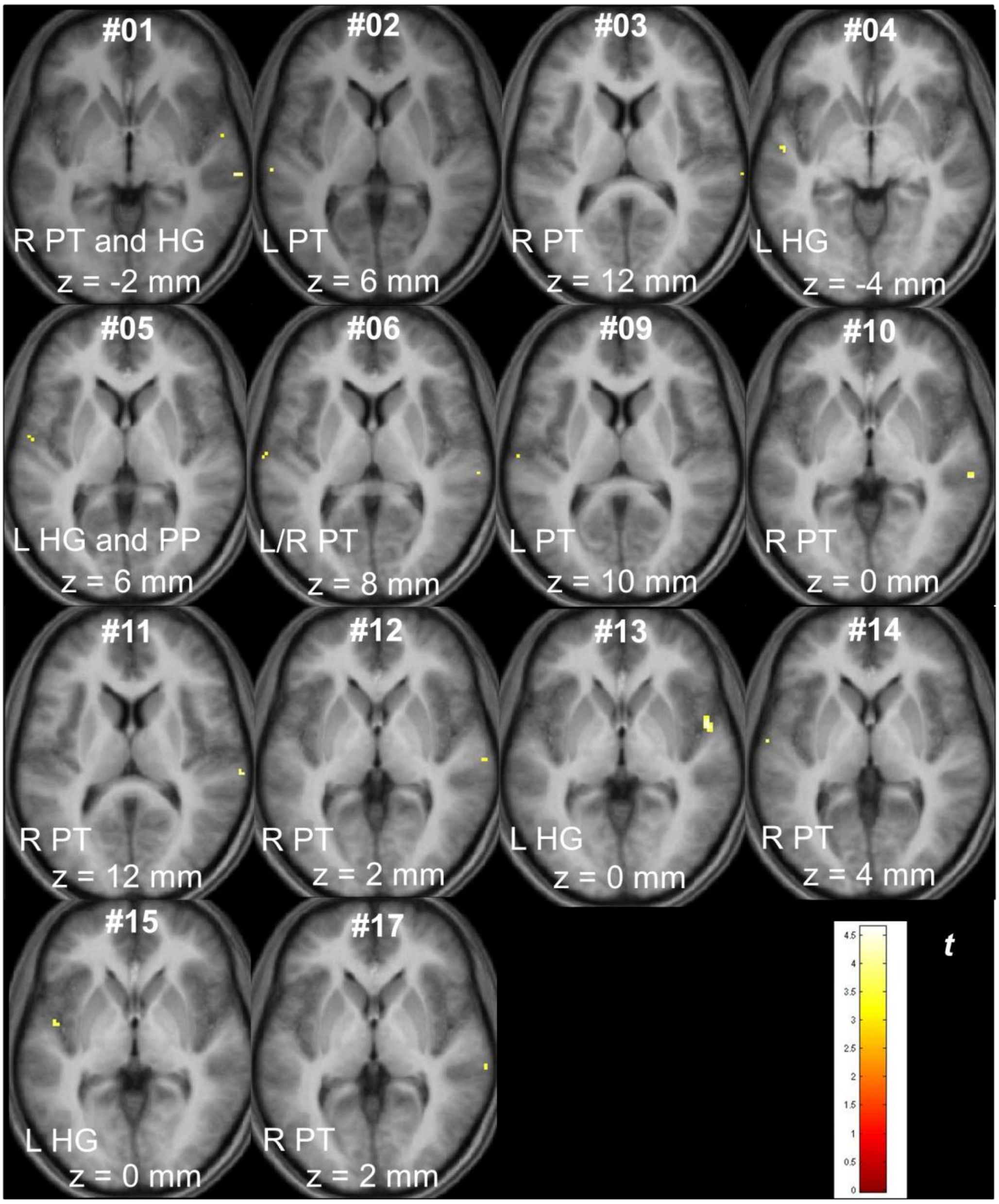

Figure 5.7. Individual activation maps for the Strong > Weak' (#09) AND 'Weak > Matched Noise' (#10) conjunction for the 14 listeners (#01-#06, #09-#15, and #17) who showed significant activation at the p < .001, uncorrected level. Activation has been overlaid onto the group averaged anatomical oblique-axial plane (z) that had the greatest peak of activity for a given listener. Approximate ROI locations are denoted and the activation maps conform to neurological convention left = left.

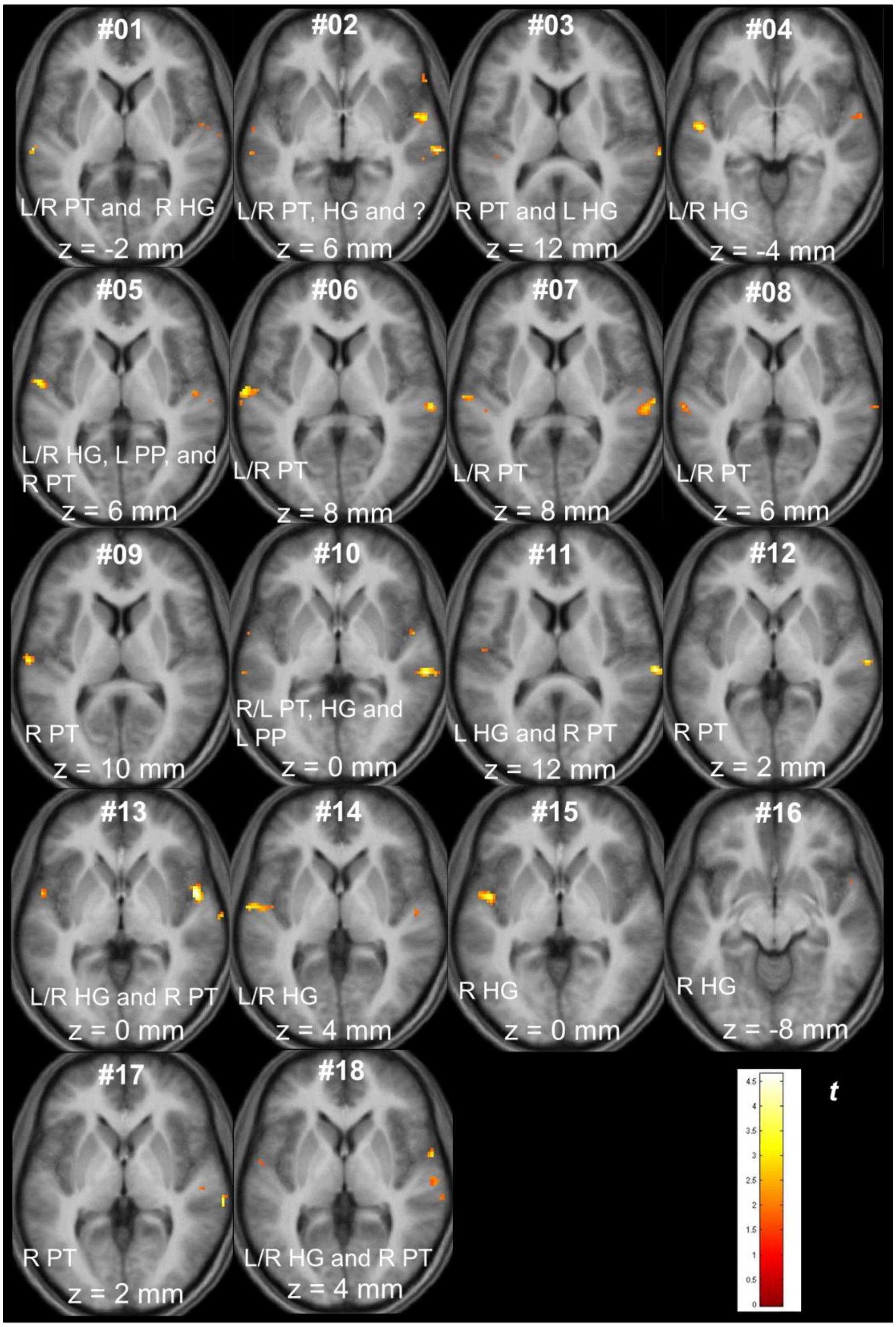

Figure 5.8. Individual activation maps for the Strong > Weak' (#09) AND 'Weak > Matched Noise' (#10) conjunction for all listeners (#01-#18) who showed significant activation at the p < .05, uncorrected level. Activation has been overlaid onto the group averaged anatomical oblique-axial plane (z) that had the greatest peak of activity for a given listener. Approximate ROI locations are denoted and the activation maps conform to neurological convention left = left.

ROI analysis

Table 5.4

Summary of pitch salience across ROIs (individual contrasts of interest #09; strong > weak)

	Salience (Objective 2)		
	Strong pitch salience > Weak pitch salience (#09)		
ROI	Condition 1 vs. condition 2	Condition 3 vs. condition 2	Condition 4 vs. condition 2
Te1.0	✓	✓	✓
Te1.1	(Not conducted)	(Not conducted)	(Not conducted)
Te1.2	✓	✓	✓
PP	✓	✓	✓
PT	✓	✓	×

Note: PP and PT ROIs refer to the revised versions, as defined in Table 5.1.

For central HG (see Figure 5.5), post hoc t-tests (Bonferroni corrected) revealed that all strong pitch salience (conditions 1, 3 and 4) produced significantly greater activation (p < .05) than the weak pitch salience (condition 2). T-test statistics were $t(17) = 5.87$, $p < .001$, $t(17) = 7.70$, $p < .001$, and $t(17) = 4.54$, $p < .001$, respectively for conditions 1, 3, and 4. This was also found for lateral HG (area Te1.2), where t-test statistics were $t(17) = 5.86$, $p < .001$, $t(17) = 8.27$, $p < .001$, and $t(17) = 4.44$, $p < .001$, respectively for conditions 1, 3 and 4. For PP, all strong pitch conditions also produced significantly greater activation compared to the weak pitch condition, where T statistics were $t(17) = 4.29$, $p < .001$, $t(17) = 6.21$, $p < .001$, and $t(17) = 3.62$, $p < .001$, respectively. For PT, only condition 1, $t(17) = 4.00$, $p = .001$, and condition 3, $t(17) = 4.81$, $p < .001$, were significantly greater than the weak pitch condition. These results suggest that there is sensitivity to pitch salience throughout primary and non-primary areas of auditory cortex, seemingly satisfying the third pitch criterion relating to sensitivity to pitch salience (Objective 2). However, as discussed in Section 5.4.5.1, although all strong pitch salience conditions fulfilled the first pitch criterion relating to pitch selectivity (Objective 1), the weak pitch salience condition did not. Condition 2 may therefore not be sufficient to drive reliable pitch-related activation in an fMRI experiment.

Directly comparing activation with a previous pitch salience study

Activation for strong pitch salience > weak pitch salience (#09) for my data (see Figure 5.6) was comparable to the pitch > noise contrast (#08) for my data, whereby activation was widespread and covered multiple primary and non-primary areas of auditory cortex. Comparatively, activation for broadly the same contrast for Barker et al. (2011) was not visible on any oblique-axial slices and only evident in areas outside STG in the sagittal slices. This is consistent with the fact that Barker et al. (2011) did not report any pitch-salience related activation anywhere in auditory cortex for their stimuli.

Summary of evidence for pitch-salience sensitivity (Objective 2) from the four analysis approaches

At initial face value, the results support pitch salience sensitivity. The results showed that there were differential responses to strong and weak pitch salience conditions, across multiple areas of auditory cortex. These results were co-localised to pitch-related response identified for Objective 1, which also provides some evidence of pitch constancy, see Section 1.2. However the weak pitch salience condition did not satisfy the first objective (pitch selectivity), because it failed to drive any detectable activation at the group level, and only occasionally at the individual level (i.e., *N* = 5 listeners in incidence map, and *N* = 14-18 for conjunction analysis) when lenient and uncorrected statistics were used. This has called into question the reliability of the initial interpretation. Subsequently these results need to be approached with caution.

5.4.5.3 Objective 3: fMRI findings for high-level 'cognitive' responses to melodic sound sequences?

Incidence maps

The sagittal view in the right hemisphere shown in Figure 5.4 (rows 1 and 2) indicates a more anterior-ward pattern of activation for pitch > noise (contrast of interest #08) than for sound > silence (contrast of interest #07). Pitch-related activity extended more anteriorly towards PP in approximately 39% of listeners. The oblique-axial view for this contrast (Figure 5.4, row 2) also depicts a larger area of greater consistency on the right hemisphere (maximum overlap in approximately 61% of listeners across the three oblique-axial slices) compared with the left hemisphere, and a smaller area of pitch-related activity (maximum overlap in approximately 61% of listeners) which was more lateralised towards PP compared to the sound > silence contrast (i.e., z = -2mm oblique-axial slice).

ROI analysis

ROI analysis is the only way to directly test Objective 3 because it enables a direct statistical comparison of sides. As described briefly in Section 5.1.3, only the laterality issue in regions where Patterson et al (2002) reported melodic vs. fixed pitch context effects were explored (see Table 5.5 below). Therefore although there was no significant interaction between hemisphere*condition for lateral HG, PP or PT (effect of interest #06; see Section 5.4.5.1), planned comparisons paired sample *t*-tests for effect of interest #06 were conducted to examine any laterality effects across pitch conditions (see Figure 5.5; e.g., left hemisphere condition 1 vs. right hemisphere condition 1, left hemisphere condition 2 vs. right hemisphere condition 2 etc.). It was expected that there would be a greater right hemisphere preference for strong pitch (conditions 1, 3 and 4), than for weak pitch (condition 2), given that the melody is more prominent for these conditions.

Lateral HG (area Te1.2)

Paired sample *t*-tests revealed activation for 2 out of 4 of the pitch conditions (conditions 2 and 4) was significantly greater in the right hemisphere compared to the left hemisphere after Bonferonni correction, *t*(17) = -3.65, *p* = .002, and *t*(17) = -3.77, *p* = .002, respectively. There was no significant difference in activation across hemispheres for noise conditions, indicating that a right hemisphere dominance was isolated to pitch-related conditions.

PP

Paired sample t-tests (Bonferroni corrected) revealed no statistically significant differential activity (*p* > .05) between left and right hemispheres across all six conditions.

PT

Paired sample t-tests conducted showed 1 out of 4 of the pitch conditions (condition 2) was significantly greater in the right hemisphere compared to the left hemisphere after Bonferonni correction; *t*(17) = -3.42, *p* < .003. There was no significant difference in activation across hemispheres for noise conditions.

Table 5.5

Summary of right hemisphere preference for melodic sequences across restricted ROIs (paired sample t-tests for effect of interest #06)

	Right hemisphere preference (Objective 3)			
	Left condition n vs. right condition n (#06)			
ROI	Strong pitch condition 1	Strong pitch condition 3	Strong pitch condition 4	Weak pitch condition 2
Te1.2	✗	✗	✓	✓
PP	✗	✗	✗	✗
PT	✗	✗	✗	✓

Note: PP and PT ROIs refer to the revised versions, as defined in Table 5.1.

Directly comparing activation with a previous pitch salience study

Interestingly, pitch-related activation in the current study appears to extend more anteriorly in both hemispheres towards lateral HG and PP than that previously reported by Barker et al. (2011), as shown in Figure 5.6 for the pitch > noise contrast. This is evident across both oblique-axial and sagittal slices. Whilst it suggests that slightly different patterns of pitch related activity (i.e., more anterior) might be observed for melodic pitch compared to fixed pitch stimulus contexts, it does not support the notion of a right hemisphere preference.

Summary of evidence for laterality effects for melodic sound sequences from the four analysis approaches

The findings provide partial support a right hemisphere preference, but perhaps what was most surprising was that there was a right hemisphere preference for the weak pitch condition in which the melody would be more difficult to perceive than for the strong pitch conditions. Moreover, the findings failed to support a greater anterior focus of activity in right than in left auditory cortex since PP showed null results for all pitch conditions. However the activation data was more anterior than that reported by Barker et al. (2011). The findings are not compelling in favour or against a right hemisphere preference for melody and the emergence of a pitch processing hierarchy. It is important to highlight that the ROI analysis indicated greater activation for the right hemisphere for all stimulus conditions (including noise conditions) and therefore wasn’t specific to pitch. Subsequently, it might be reasonable to suggest that because the noise was also pulsed this created a type of melody context. Further research is therefore needed to explore these further.

5.5. Discussion

The current experiments used the same stimuli and a study design informed by Penagos et al. (2004) to re-examine pitch-related and pitch-salience related effects by addressing some methodological limitations identified in the original study. To recap, the main experiment of interest (Experiment 5) examined the evidence for pitch selectivity (Objective 1), pitch salience sensitivity (Objective 2) and a right hemisphere preference implying a pitch processing hierarchy (Objective 3), using four main analysis approaches. The subjective pitch salience rating task (Experiment 4) examined the first two objectives by getting listeners to rate the stimuli across the six conditions. Listeners could discern between weak pitch and matched noise, as well as weak pitch salience and strong pitch salience, satisfying both Objectives 1 and 2 on a behavioural level.

5.5.1 Evidence of multiple regions of auditory cortex selective to pitch over noise

Pitch-related activity was widely distributed and localised to central HG, lateral HG, PP and PT, confirming previous results (Barker et al., 2011; Bizley et al., 2009; Garcia et al., 2010; Hall & Plack, 2009; Griffiths et al., 2010; Staeren, Renvall, De Martino, Goebel, & Formisano, 2009). This result broadly supported Penagos et al (2004), but highlights how pitch effects were not restricted to the anterolateral end of HG. This finding appears contrary to studies that propose a pitch centre (e.g., Bendor & Wang, 2005, 2010; Griffiths et al., 1998, 2001; Krumbholz et al., 2003; Patterson et al., 2002; Puschmann et al., 2010). Instead, the representation of pitch appears to be widely distributed. Griffiths & Hall (2012) plotted the individual maps of pitch activation originally from Hall & Plack (2009), and illustrate large variability across individual listeners. Individual’s exhibited highly variable activation in a number of sites, including PT, PP, superior temporal sulcus and inferior frontal gyrus. High spatial variability of the pitch response across listeners exemplifies the risks of generalising from only a small number of participants.

5.5.2 Is fMRI able to detect small or variable pitch-related activations for a weak pitch stimulus?

Activation for pitch and pitch-salience activity was seemingly comparable and co-localised to the same regions as the pitch response. This result appears to contradict previous studies from the Nottingham group (i.e., Barker et al., 2011; Hall & Plack., 2009; but see Barker et al., 2013). However, both studies used fixed pitch stimuli so were more susceptible to adaptation effects. It is important to note, however, the credibility of this apparent pitch salience effect is called into question by the observation that activation for the weak pitch salience condition was broadly comparable to the matched noise condition at the group level (i.e., not significantly different). This equivalent activation for weak pitch and matched noise was also observed by Penagos and colleagues, but was not specifically discussed. This condition failed to fulfil the criteria for pitch selectivity at the group-level, but did fulfil this criteria for some listeners at the individual-level. Specifically, the incidence map and individual conjunction analyses revealed differential activity (5/18 and 14-18/18 listeners, respectively) for pitch > noise, this was not significant across the group. This might have also been because the incidence maps and individual conjunction analyses reported here used a lenient uncorrected statistics of $p < .001$ and/or $p < .05$, whilst the group-based activation analysis did not. Alternatively it may also be reflective of the spatial variability across individuals (see Griffiths and Hall, 2012). However, in the subjective ratings pitch task, listeners reliably rated the weak pitch condition as significantly weaker in salience than the strong pitch salience conditions, but greater in pitch salience than noise. This means that perceptually at least, there was a perceptual difference in pitch salience, but this was not reflected or detectable in the fMRI signal.

This result is contrary to what one would expect given that a number of EEG and MEG studies have found that stimuli with stronger pitch salience (namely IRN) evoke larger pitch onset responses (PORs) that have shorter latencies (see Krishnan et al., 2010, Krishnan et al., 2012; Krishnan & Plack, 2011; Krumbholz et al., 2003; Kumar et al., 2011). Evidence from intracranial recordings using IRN stimuli have also found that the magnitude of both evoked responses and induced gamma power increased as a function of pitch salience throughout all areas of HG (Griffiths et al., 2010). Krishnan and colleagues (2010; 2012) have proposed several reasons that might explain the disparity in findings between fMRI and EEG, MEG and behavioural experiments in identifying representations of pitch salience in cortical and subcortical areas of the auditory pathway. Namely, that using fMRI to investigate representations of pitch salience is confounded because subcortical and cortical representations of pitch or pitch salience have a much finer temporal resolution (i.e., ms) than is afforded by fMRI (i.e., s; Krishnan et al., 2012). fMRI is much more sluggish than other neuroimaging methods, and is not a direct measure of neural activity (see Chapter 3). The incidence maps confirm a wide distribution of pitch activity across the 18 subjects. It might not be that there is no representation of pitch salience in the midbrain or auditory cortex, but instead that the fMRI methodology used hitherto is not sensitive enough to detect changes in response sensitivity to more detailed features of an auditory stimulus (Ernst, Verhey, & Uppenkamp, 2008; Krishnan et al., 2012) particularly on a group level, thus reflecting a type 2 error. This coupled with high inter-subject variability (see Griffiths and Hall, 2012) would make it extremely difficult to isolate such activity. As proposed by Barker et al. (2011), who failed to find an area sensitive to pitch salience, fMRI may only be able to reveal auditory areas which are maximally responsive to the presence or absence of pitch, rather than the perceptual pitch salience per se.

5.5.3 No strong evidence of a right hemisphere preference, but some evidence for more anterior activation dependent on the stimulus context (melodic pitch vs. fixed pitch)

Findings reported for Experiment 5 did not show compelling evidence in favour of a right hemisphere preference or emergence of a pitch processing hierarchy, since PP showed null results for all pitch conditions. However comparison of my data with Barker et al. (2011) did confirm that activation in my study was more anterior, thus suggesting that the context of the pitch stimuli used (i.e., melodic or fixed) can affect the distribution of pitch-related activity observed. These results support other passive listening experiments which have also reported that stimuli which vary in pitch (melodic type sequences), show more anteriorly distributed representations than compared to fixed pitch stimuli (Patterson et al., 2002; Warren & Griffiths, 2003). However more research is still needed to confirm whether presenting pitch signals as random melody sequences promotes involvement of the right anterior temporal lobe, providing support of a pitch processing hierarchy, as reported in Patterson et al. (2002).

5.5.4 Alternative explanations for results

5.5.4.1 Are pitch responses only related to the melody (context) of the pitch percept?

One alternative explanation for the fMRI results that could be argued is whether the pitch responses were driven by the pitch of the stimuli, or just the melody of the pitch (i.e., context). Although the pitch > noise (contrasts of interest #08) had a melody difference (i.e., melody > no melody), the strong > weak contrast (#09) did not (i.e., both included a melody). This contrast still gave rise to differential activation even when the melody was present for both stimulus conditions (strong and weak). If the results were driven by melody alone, I would have not expected to see a significant result for this contrast. Subsequently it seems reasonable to assume that the activation was attributable to pitch coding (pitch-specific) effects and not melody (context) coding alone.

5.5.5 Summary

Overall the results reported here satisfied the first pitch criterion of pitch selectivity, but provided important convergent evidence in favour of a distributed representation of pitch in auditory cortex, which is contrary to a designated pitch centre. Although a representation of pitch salience was found (for some listeners at least) seemingly fulfilling the third pitch criterion of covariation with pitch salience, these results should be taken with caution until further work has confirmed whether or not fMRI methods are sensitive enough to detect pitch salience effects. It seems vital that this future work involves exploring the use of more advantageous experimental and analysis approaches that can be used to tease apart any pitch salience effects using fMRI, and this should be carried out in conjunction with other methods (e.g., electrophysiological and behavioural).

Chapter 6. ERP and adaptation methodology

6.1. Introduction

This chapter briefly describes the ERP and adaptation methodology used in Chapter 8. Please note that, given that this chapter has been specifically written to describe the methodology used in Chapter 8 alone, it is much shorter than the fMRI methodology described in Chapter 3, which was written for more general purposes.

6.1.1 What is EEG and ERPs?

Electroencephalogram, or EEG, uses multiple electrodes placed on the scalp and amplifies the electrical activity of the human brain (signal) to graphically plot changes in voltage between two different electrode locations over time (Luck, 2005, p. 3). Since its inception by Hans Berger in 1929, EEG has grown in popularity amongst researchers and clinicians alike, mainly because it noninvasively measures direct neural activity with excellent temporal resolution (i.e., ms; unlike fMRI, see Chapter 3) by recording the electromagnetic fields generated by certain neuronal populations. For instance, through averaging, it is possible to extract neural responses to specific sensory, cognitive or motor events relating to specific experimental conditions or events that are time-specific, and these are known as event-related potentials, or ERPs (Sanei & Chambers, 2007, p. 127). ERPs, therefore, provide a continuous measure of covert, as well as overt, processing of stimuli of interest, making it possible to determine which stage(s) of mental processing are affected by different independent variables that have been manipulated.

The electric potential from a single neuron is considered far too small to be detectable by EEG. Therefore, the EEG signal is generally considered to reflect the summation of thousands or millions of synchronised inhibitory and excitatory postsynaptic potentials in the cortex of the brain, which fire instantaneously and are spatially aligned, and mainly confined to the dendrites and cell body of neurons (Kropotov, 2009, Part I, p. 11-19; Sanei & Chambers, 2007, p. 4-8). It is postulated that if an excitatory transmitter is released from the presynaptic terminals, this causes positive ions to flow into the postsynaptic neuron, creating a net negative extracellular voltage in the other parts of the neuron (depolarisation), which ultimately yields a small electrical dipole between the soma (body of neuron) and apical dendrites (neural branches). The dipoles from large populations of neurons summate and the resulting voltage is measurable at the scalp (Kropotov, 2009, Part I, p. 11-19; Sanei & Chambers, 2007, p. 4-8). Multiple dipoles can be measured using an equivalent current dipole, although some dipoles can partially cancel each other out if they are more than 90 degrees from each other, and completely cancel each other out at a 180 degree difference (Luck, 2005, p. 31). Pyramidal cells are aligned perpendicularly to the surface of the cortex and fire together, forming an extracellular cortical dipole layer, and therefore activity from these cells are considered to be the most likely source of the EEG signal that is measurable at the scalp (Luck, 2005, p.31). The EEG signal is primarily generated by the large, vertically oriented pyramidal neurons located in cortical layers II, III, V, and VI (Kropotov, 2009, Part I, p. 18). The EEG signal can also be modified by the current flow or conductive properties of the tissues (e.g., brain, CSF, skull and scalp) between the electrical source and the recording electrode site on the scalp, the electrode itself, and the orientation of the cortical generator of the recording electrode (known as volume conduction; Kropotov, 2009, Part I, p.147-148). Subsequently, EEG has poor or undefined spatial sensitivity (two-dimensional projection of a three-dimensional reality) because it is not possible to accurately localise the source of ERP components using sensor-based EEG information alone, known as the inverse problem (Kropotov, 2009, Part I, p. 148); but source localisation can help to solve this problem (see Section 6.2.4).

The spectral content or rhythmic activity of the EEG signal can also be used to measure different brain waves that are subdivided into different frequency bandwidths, such as alpha (8-13 Hz), beta (13-31 Hz), delta (>4 Hz), theta (4-8 Hz) and gamma waves (32+ Hz; Kropotov, 2009, Part I, p. 2-3). These are associated with specific locations and cognitive states (e.g., alpha activity is typically observed in posterior regions when someone is closing their eyes or at rest; Kropotov, 2009, Part I, p. 3-4).

Given that auditory perception is remarkably fast, EEG is a particularly useful method for investigating auditory processing. For example, EEG has been used to measure early Auditory Brainstem Responses or brainstem Frequency Following Responses (Krishnan et al., 2012; Krishnan & Plack, 2011), motion onset responses in auditory space perception (Kreitewolf, Lewald, & Getzmann, 2011), selective attention to sound location and pitch perception (Dergerman et al., 2008), and even oscillatory EEG activity in patients with chronic tinnitus (Moazami-Goudarzi, Michels, Weisz, & Jeanmonod, 2010). More recently, EEG has been used in combination with adaptation methodology to investigate frequency (Briley & Krumbholz, 2013; Lanting, Briley, Sumner, & Krumbholz, 2013), spatial location (Briley, Kitterick, & Summerfield, 2012; Briley & Summerfield, 2014), and pitch processing in auditory cortex (Briley et al., 2013; see Section 6.3).

6.2. EEG data acquisition and analysis

EEG data acquisition and analysis broadly involves attaching electrodes to the scalp of participants to pick up the EEG signal, filtering and amplifying the signal so it can be stored as a set of discrete voltage measurements on the computer, removing various artifiacts that can contaminate the EEG signal (e.g., eye blinks) before averaging the data to extract the ERPs from the overall EEG for each condition, and then applying various signal processing techniques (e.g., digital filters) to remove noise and isolate specific ERP components. The following sections describe the procedures used in Chapter 8.

6.2.1 EEG data acquisition

The section highlights the main procedures used for recording ERPs using a 64-channel Active-Two acquisition system (BioSemi; see Figure 6.1), sampled at 2048 Hz, with a bandwidth of 417 Hz and digitised at 24-bit, and describes the approaches used to analyse data from Experiment 10.

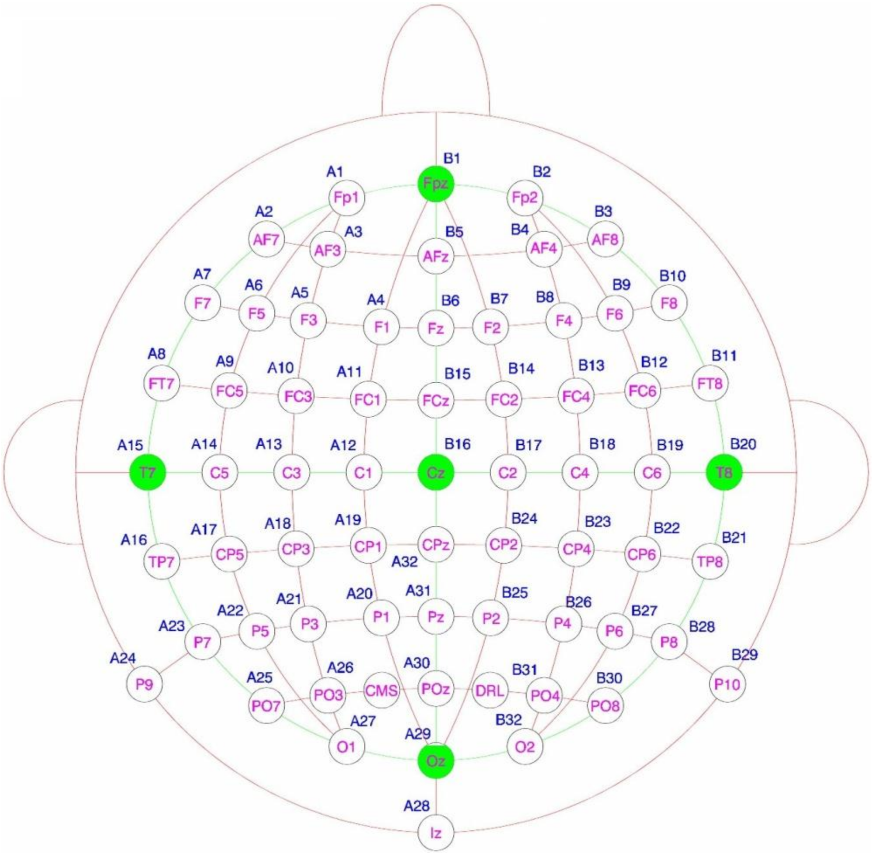

Figure 6.1. Schematic of the BioSemi 10-20 montage with the 64 electrode positions labelled plus two additional grounding electrodes Common Mode Sense (CMS) and Driven Right Leg (DRL).

Various craniometric measurements were obtained (longitudinal: nasion-to-inion; lateral: left (PAL) and right (PAR) pre-auricular points). The half-way point of these two measurements is known as the vertex. The specialised BioSemi head-cap is comprised of 64 pin-type Active electrodes placed according to the standard 10-20 arrangement. This system refers to the actual distance between electrodes, either 10% or 20%, and is based on the relationship between the location of a given electrode and the underlying area of the cerebral cortex. Eight additional silver/silver chloride (Ag/AgCl) electrodes that were positioned at F9, F10, F11, F12, T9, T10, back of neck and chin, to provide greater coverage of the lower part of the head surface to aid reconstruction of topographic voltage maps and facilitate source localisation. The head-cap was fitted so that the Cz electrode fell directly on the vertex and so the Fpz, Fp1 and Fp2 electrodes are on the 10% midline from the nasion. Signa (conductive) gel was to fill the electrode holders before the Active-Two electrodes were pushed in to the corresponding electrode holders. Given that high electrode impedances can be tolerated, the system can be used without having to prepare the skin in the conventional manner (i.e., alcohol and/or abrasive scrubbing of the skin). Skin-electrode impedances were kept below 5 kOhms throughout data acquisition, which is common good practice (Luck, 2005, p.119).

Signals were amplified using ActiveTwo (BioSemi) Analogue-to-Digital-box (AD-box) channel which consists of a low noise Direct Current (DC) coupled post-amplifier with an anti-aliasing filter. The digital outputs of the Analogue-to-Digital Converter (ADC) are digitally multiplexed and sent to the recording personal computer via a single optical fibre without compression or data reduction. The receiver then converts the optical data coming from the AD-box to an USB2 output, which in turn delivers the data to the recording personal computer which can be viewed online using an acquisition system. The BioSemi system typically replaces the ‘ground’ electrode used in conventional systems with two separate electrodes known as a Common Mode Sense (CMS) active electrode and a Driven Right Leg (DRL) passive electrode (“BioSemi EEG reference,” n.d.). These two electrodes (see Figure 6.1) form a feedback loop which drives the average potential for the subject (CMS voltage) as close as possible to the ADC reference voltage in the AD-box. With this BioSemi system, every electrode or a combination of electrodes can be used as a ‘reference’, with the choice being selectable in the given acquisition or analysis software used. Incoming signals from the Active-Two channels are monitored online and saved (in BDF format) using a complete acquisition program called ActiView (v. 6.05, BioSemi). Therefore all data collected uses the average reference from all connected electrodes. The incoming data was then low-pass filtered at 100 Hz, and high-pass filtered at 0.16 Hz. before offset values and any visible problems (e.g., excessive alpha waves, slow voltage shifts, amplifier saturation, muscle and heart activity) were manually checked by ‘eyeballing’ the data before running the experiment.

Participants were seated approximately 60 cm from the stimulus presentation monitor. To correct for signal artefacts, data collection began with an artefact recording experiment. Specifically, a 10 minute artefact correction experiment was run to record muscle movements which can contaminate the signal, notably eye blinks and horizontal eye-movements (i.e., horizontal electro-oculograph). Participants had to perform these movements when instructed to do so by the experimenter. Next, a 26 minute passive pitch listening task was employed. Here participants were instructed to remain alert whilst listening to the sounds sequences and watching a silent cartoon DVD. A passive task was implemented as task-dependent effects can alter the sensitivity of neurons to the stimulus attributes of pitch (see Alho et al., 2013; Walker et al., 2011b). A ‘paradigm’ file was used to define the conditions and trials at the same time as the data recording.

Once data collection was completed, the data were analysed using Brain Electrical Source Analysis Software (BESA Research v. 5.3.7, Gräfelfing), although a number of different EEG analysis software packages are available. The following sections describe pre-processing and analysis steps involved when using BESA to analyse Experiment 10 (described in Chapter 8), where there were three main stages: 1) Pre-processing, 2) Sensor-based analysis, and 3) Source-based analysis, which are described below.

6.2.2 Pre-processing

Pre-processing prepared the data for analysis by correcting and rejecting artefacts, filtering and averaging. It involved a number of conventional steps described in the BESA manuals for data pre-processing and analysis (available online at http://www.besa.de/downloads/training-material/tutorials/):

i) Low-pass and high-pass filtering the data for artefact correction.

The raw EEG data were high-pass (0.5 Hz ~ 6 dB/octave, forward shift) and low-pass filtered (35 Hz ~ 24 dB/octave, zero-phase shift) to aid manual identification of blink and horizontal electro-oculograph artefacts only.

ii) Artefact correction for blinks and horizontal electro-oculograph

Artefact correction aims to extract unwanted signals like blinks, horizontal electro-oculograph or external noise from the data, while leaving all brain activity of interest as undisturbed as possible. To achieve this, artefact and brain topographies must be separated because they are typically spatially correlated. For artefact correction without distortion, it is better to create a model or spatial description of the brain topographies to be retained, rather than define and reject artefact topographies (Scherg, Berg, & Hoechstetter, 2010). Blinks and horizontal electro-oculograph artefacts were removed following the principle components analysis procedure proposed by Ille, Berg, and Scherg (2002). This is an ‘adaptive’ method which models both artefact topographies and underlying brain activity. Artefact data collected prior to Experiment 10 (see in Chapter 8) was used to individualise artefact correction for each participant. Blinks and horizontal electro-oculograph topographies individually explained between 94.1% and 99.8% of the total variance.

iii) Re-referencing to the average reference

Whenever artefact correction is applied in BESA, the data is automatically re-referenced to the average reference.

iv) Segmenting trials into epochs

A ‘paradigm’ file was used to define the conditions and trials at the same time as the data recording. The paradigm file was then used to define the conditions and trials in BESA, which segments the data into time epochs ranging from the onset of the first stimulus (e.g., noise stimulus) and a given period after the onset of the last stimulus (if more than one stimulus event in the sequence; e.g., pitch stimulus). Trials were segmented into an epoch window of 2520ms, which began from the onset of the first noise stimulus to 500ms after the onset of the last stimulus (POR2; probe).

v) Baseline correcting epochs

Segments are usually baseline corrected to a given pre-stimulus interval. Handy (2005, p. 39) recommends using a pre-stimulus intervals greater than 200ms because shorter baselines are more sensitive to residual voltage fluctuations which have a negative effect on ERP amplitudes. Segments were baseline corrected to a 300ms pre-stimulus interval for averaging.

vi) Artefact scanning and artefact rejection

The BESA system allows you to apply low- and high-pass filters for the artefact correction scan only so you can essentially evaluate the quality of the data, and reject trails prior to averaging. These digital filters can sometimes help to improve artefact detection, and function by supressing the frequencies that are attributeable to noise and artefacts that are not of interest (Sanei & Chambers, 2007, p. 18). Low-pass filters attenuate high frequencies and pass low frequencies, whereas high-pass filters attenuate low frequencies and pass high frequencies. The high-pass filter (i.e., 0.2 Hz) was kept on for averaging because this can have a negative effect on relatively short data epochs if applied after averaging (Hoechstetter, Berg, & Scherg, 2010). Unlike the high-pass filter, low-pass filters (i.e., 35 Hz, 24 dB zero phase) do not need continuous data for these filters to work properly and can be applied at any time. The low-pass filter was therefore switched off for averaging, and applied after averaging.

It should be noted that artefact scanning and averaging was conducted on the EEG data on two separate occasions, mainly to increase the number of trials accepted for averaging and thus to improve SNR. The BESA system allows you to apply low- and high-pass filters for the artefact correction scan only so that you can evaluate the quality of data before averaging. The low-pass filter was not applied on the first attempt and thus meant that many trials were rejected from averaging, and/or electrodes needed to be set to ‘bad’ or interpolated. The BESA manual advises that it is however generally best practice to change the data as little as possible (e.g., avoid interpolation unless absolutely necessary, try not to interpolate sensors on edges). The artefact scan indicates any noisy channels that could either be set to bad or interpolated. Any trials that had maximum amplitude greater than 120 µV, gradient jumps of greater than 75 µV, and/or a low signal of less than 0.01 were rejected from averaging (although users can manually override this process and accept reject trials for averaging if desired). On the first attempt 29.55% (2364/8000) of trials were rejected. On the second attempt only 10.25% (820/8000) of trials were rejected across the 20 initial subjects, thus confirming my motivation to have to re-run this step.

vii) Individual averaging across epochs

Averaging was conducted separately for each subject to obtain the averaged ERP waveform files for each condition (collapsed across trial types) per subject. Artefact correction (i.e., blink and horizontal electro-oculograph topographies) were switched off for averaging because they can distort brain topographies. Individually averaged ERP waveforms for each condition per participant were then low-pass filtered (i.e., 35 Hz ~ 24 dB/octave, zero-phase shift). These were inspected according to the expected cortical responses to noise, adaptor and probe stimuli. After inspecting individual ERP waveforms, two subjects (#29 and #36) who had participated in both studies were excluded from Experiment 11 during the ERP analysis stage. They both failed to exhibit typically large and distinguishable cortical responses to noise (i.e., EOR).

viii) Grand averaging

A grand average (GA) waveform collapsed across the remaining 18 participants for each condition was also compiled. This was then be low-pass filtered (i.e., 35 Hz ~ 24 dB/octave, zero-phase shift). The GA served as a visual aid for the overall pattern of results, as well as a guide for latency ranges for later peak-to-peak analysis and dipole model fitting epochs.

6.2.3 Sensor-based analysis

Sensor-based analysis for Experiment 10 described in Chapter 8 involved manual quantification of ERP components of interest. The source-based analysis also involved the same quantification procedure, but after the source waveforms had been obtained (see source-based analysis section below).

6.2.3.1 ERP components of interest

The onset of any sound stimulus elicits a cascade of ERP responses, whereby there is an onset response, a sustained field and an offset response (Poeppel & Hickok, 2015, p. 248-249). ERP waveforms consist of a number of positive (denoted by P prefix) and negative (denoted by N prefix) voltage deflections in amplitude. The most common components of interest in auditory ERP studies relate to transient responses, such as the robust and large evoked components, known as the P100, N100 and P200, and are described below in more detail.

The P100 is typically observed as a frontally positive deflection which peaks around 50ms after the onset of a sound stimulus (Luck, 2005, p. 39). Generators of the P100 are thought to be located in the central part of HG, in primary auditory cortex (Butler & Trainor, 2012). The P100 is usually followed by the N100.

The N100 is the auditory component that is the most prominent and the one that has been the most extensively studied. The N100 is typically observed as a frontally negative deflection which peaks around 100ms after the onset of a sound stimulus, and is thought to have multiple subcomponents and neural generators located in secondary auditory areas, such as lateral HG and PT (Butler & Trainor, 2012; Näätänen & Picton, 1987). The N100 is considered to reflect stimulus representation or simple feature detection, such as a change in the energy or physical properties of the stimulus (Näätänen & Picton, 1987). Stimulus repetition is also known to diminish the amplitude of N100 (i.e., adaptation; Näätänen & Picton, 1987).

The P200 is a positive deflection that peaks at approximately 200ms; however, little is known about the neural generators of P200, although these are probably different to N100 generators (Lanting et al., 2013). Generally, the P200 generators are found to be more anterior in auditory cortex than N100 (see Lütkenhöner & Steinsträter, 1998; Krumbholz et al., 2003), and suggested to involve mechanisms that involve stimulus feature evaluation (Behroozmand, Korzyukov, & Larson, 2012). Lanting et al. (2013) found that the P200 amplitudes are more strongly influenced by adaptation effects (something which was confirmed in my data reported in Chapter 8).

The N100 can be evoked by the onset of any sound stimulus, regardless of its physical or perceptual properties. This can contaminate pitch responses, because areas of the brain that responds to noise onset in general will saturate the cortical signal of interest (Garcia et al., 2010; Krumbholz et al., 2003; Näätänen & Picton, 1987; Seither-Preisler, Krumbholz, Patterson, Seither, & Lütkenhöner, 2004). Krumbholz et al. (2003) used MEG to employ a novel stimulus paradigm using IRN stimuli which varied in pitch strength (increasing pitch salience as a function of increasing iterations; 2, 4, 8, 16 and 32). This was known as the ‘continuous stimulation paradigm’ and involved presenting a sound sequence which transitioned from a spectrally matched white noise (2000ms) to a pitch stimulus (1000ms). The energy-onset response (EOR) is an automatic response which can be evoked by any sound stimulus, and is comprised of components N100-P200 (large negative deflection at around 100ms after stimulus onset, followed by a large positive deflection at around 200ms after onset). The continuous stimulation paradigm therefore allowed the EOR to be isolated from the transition response that is observed for pitch specifically, known as the pitch-onset response (POR). This is because the initial segment of noise evokes only the noise onset components and is then followed by the pitch-eliciting segment. Generally the time windows for EORs peak between 70-120ms for the N100 (Bosnyak, Eaton, & Roberts, 2004; Seither-

Preisler et al., 2004; Shahin, Roberts, Miller, McDonald, & Alain, 2007), and 170-220ms (Lütkenhöner & Steinsträter, 1998; Seither-Preisler et al., 2004; Shahin et al., 2007) for the P200, after stimulus onset. PORs are broadly determined by the most negative peak occurring between 100-200ms (Seither-Preisler et al., 2004), but have been found to typically peak around 150ms after pitch stimulus onset (Chait et al., 2006; Krumbholz et al., 2003). Krumbholz et al. (2003) conducted three experiments, where in the first two experiments the noise preceded the pitch stimulus, whereas in the third experiment the pitch stimulus preceded the noise. They found that the POR was only evoked for noise-to-pitch stimulus transitions, and not for pitch-to-noise transitions. The authors located the source of the POR in the medial HG.

The sensitivity and consistency of POR has been studied across a number of MEG studies that have utilised the continuous stimulation paradigm (e.g., Chait et al., 2006; Gutschalk et al., 2004; Krumbholz et al., 2003; Ritter et al., 2005; Seither-Preisler et al., 2004, 2006). These studies suggest that the POR offers an excellent window for studying early cortical representations of pitch (Krishnan et al., 2012). For example, a follow-up study by the same authors (i.e., Seither-Preisler et al. 2004) further investigated the POR, as well as whether its neural generators were similar to that of EOR. They used stimuli similar to Krumbholz et al. (2003) except that the duration of noise segments varied from 500 to 4000ms (i.e., 500, 1000, 2000 and 4000ms). The duration of the IRN pitch stimulus was always 1000ms, so the total length of the stimulus sequences (for noise and pitch) were either 1500, 2000, 3000 or 5000ms. They found prominent PORs in all subjects when the latency of the noise segments was 1000-4000ms. When the 500ms noise segment was used, a POR was only found in four out of seven listeners. The latency of PORs did not appear to be affected by noise duration, but the amplitude increased as a function of increasing noise duration. Ritter et al. (2005) also managed to evoke the POR using sounds sequences which transitioned between two IRN pitch stimuli.

Chait et al.'s (2006) study involved presenting listeners with pure tones embedded in noise as well as Huggins pitch stimuli. They found PORs for both these types of pitch-evoking stimuli, thus confirming the pitch constancy of POR. This was mainly because the inclusion of binaural Huggins pitch stimuli (where two noises are presented to each ear and it is the combination of noises that creates a pitch percept) provided support that similar pitch extraction mechanisms are involved. Gutschalk et al. (2007) also measured MEG responses to regular and irregular click trains and reported similar results as Krumbholz et al. (2003).

The POR is therefore thought to reflect synchronised cortical activity related to pitch, and depends on specific features, such a F0 or salience (Chait et al., 2006; Krumbholz et al., 2003; Seither-Preisler et al., 2006). Some MEG studies have also found that the amplitude of the POR response increases and the latency decreases as a function of pitch value and salience (Krumbholz et al., 2003; Seither-Preisler et al., 2003; Soeta, Nakagawa, & Tonoike, 2005), further suggesting POR's involvement in pitch processing. This illustrates why it is important to match stimulus parameters for discriminability, which can be considered as one indicator of stimulus salience (see Chapter 7).

It should be noted that the POR, although visually similar to, it is probably not the same as the N100-P200 elicited by the onset of sound because it only occurs for the perception of an emerging pitch in an on-going sound sequence (Poeppel & Hickok, 2015, p. 249), although it is likely that POR shares some neural structures involved in the EOR (Seither-Preisler et al., 2004). Some have postulated the POR and N100 have partially overlapping neural generators (Seither-Preisler et al., 2004, 2006), but Krumbholz et al. (2003) has argued that the neural generators of N100m and POR are functionally independent. Schönwiesner and Zatorre (2008) also confirmed separate neural sources for noise and pitch onsets using depth electrodes, with pitch onsets being localised most consistently to lateral HG. Some have also suggested the POR and P200 have more similar neural generators than POR and N100. One hypothesis is that the POR and P200, whilst having different polarities and different latencies, are actually generated from the same location but in different cortical layers (Seither-Preisler et al., 2004). However, it is impossible to determine the precise location of MEG sources with absolute certainty. Nevertheless, in Experiment 10 the temporal windows for N100 and P200 components were used to guide analysis in order to obtain a negative and positive peak-to-peak measure for PORs (like EOR), and are subsequently referred to throughout as the N100 and P200 of the POR response, for simplicity.

Sound sequences from Experiment 10 consisted of three stimulus events (i.e., noise, pitch stimulus 1 (adaptor) and pitch stimulus 2 (probe), which were measureable in the ERPs as EOR, POR1 and POR2. For each participant across conditions, the EOR, POR1, and POR2, were individually computed (see Section 6.2.3.2).

6.2.3.2 ERP quantification

ERPs are measured in terms of units or amplitudes, specifically microvolts (µV) for sensor-based analysis, and dipole moment (i.e., strength) in nano amp metres (nAm) for source-based analysis. Latencies (time point in ms) for a given component of interest (e.g., POR) are also used for both sensor and source-based analyses. A peak amplitude measure quantifies the time point (i.e., peak latency) a given component reaches its maximum amplitude (Handy, 2005, p.38).

Peak-to-peak measure

A peak-to-peak measure was used for Experiment 10, which measures one peak (N100) relative to an adjacent peak (P200). Peak-to-peak difference measures were computed separately for each listener and condition by taking the peak amplitude for each component of interest (i.e., N100 and P200) comprising each of the EOR, POR1 and POR2 responses, and inputting these into an excel spread sheet. These values were then used to calculate peak-to-peak measures for EOR, POR1 and POR2 per listener per condition.

This peak-to-peak approach has been used by several other studies that have measured stimulus selectivity of adaptation in ERPs (recently, Briley et al., 2013; but see Näätänen & Picton, 1987 for a review). This is because the N100 and P200 deflections have opposite polarities and partly overlapping time courses which mean they might partially cancel each other out (Makeig, Jung, Bell, Ghahremani, & Sejnowski, 1997; Näätänen & Picton, 1987). Using a peak-to-peak, rather than a baseline-to-peak measure subsequently avoids this potential cancellation effect. Furthermore, the peak-to-peak measure remains free from residual noise, DC shifts, and other confounding artefacts that might exist for the pre-stimulus baseline (see Handy, 2005, p.40; Luck, 2005, p. 237; Picton et al., 2000).

All sensor-based peak-to-peak analysis was conducted at the Cz electrode because the GA waveforms had demonstrated that the Cz sensor displayed the most robust responses to the noise, adaptor and probe stimuli. It is acknowledged that analysing the data from multiple electrode sites would have possibly made the data more robust; however Shahin, Bosnyak, Trainor, & Roberts (2003) used Cz and T8 electrode sites but only found large N1 and P2 responses for Cz. Auditory papers also tend to report data from the Cz vertex as it is known to be highly responsive to auditory-evoked stimuli (e.g., Briley et al., 2013; Krumbholz et al., 2003). Therefore for the purposes of Experiment 10, and because source analysis was conducted which used the data from all electrodes, it was felt that analysis for one electrode site for the sensor-based analysis would be sufficient. Voltage distribution maps were also the most negative and positive at the top of the head for N100 and P200 (Figure 6.2) for the Cz electrode site, respectively. This was also true for the source-based analysis (not shown).

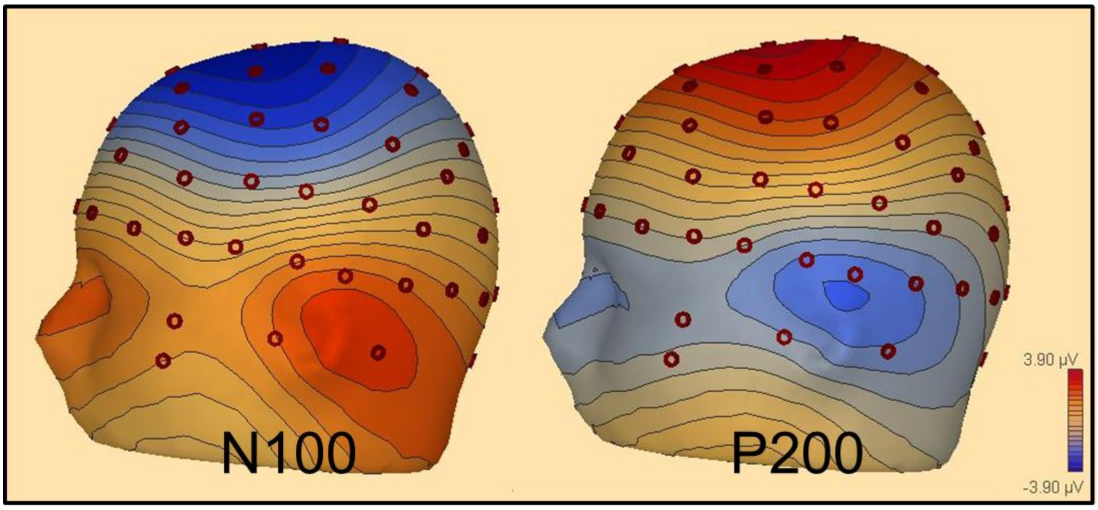

Figure 6.2. A representative example of the topographical maps observed for N100 and P200 components across subjects and conditions taken from the GA sensor-based data for one condition.

The peak-to-peak amplitudes for POR1 and POR2 for each listener and condition were then used to calculate the amount of adaptation occurring from POR1 (adaptor) to POR2 (probe), by taking the P200-N100 amplitudes of the probe (i.e., peak amplitude of P200 minus peak amplitude N100), and subtracting this from the P200-N100 amplitudes of the adaptor (see Figure 6.3). The pre-stimulus interval baseline corrected the EOR waveforms and so an absolute difference in peak amplitudes could be taken. The POR responses for the adaptor and probe are far from the pre-stimulus interval baseline and appeared to drift slightly in relation to the baseline. This is probably because the ERP source and sensor-based waveforms were of long duration. However, digital filtering did appear to improve the situation. In this case, an absolute peak amplitude measure was therefore not appropriate (Luck, 2005, p.237). Nevertheless, rather than using a baseline to peak measure (which may have not been appropriate to use), a relative peak-to-peak measure was obtained (see Figure 6.3, and Briley et al., 2013). This also helped to reduce possible contributions arising from the P200 waveform, given that N100 and P200 have overlapping time courses (Bosnyak, Trainor, & Roberts, 2003; Näätänen and Picton 1987; Makeig et al. 1997). It is important to note that given that for the ‘same pitch, same timbre’ condition the response was expected to be near zero, only measurements from the largest two peaks in the appropriate time epoch were taken.

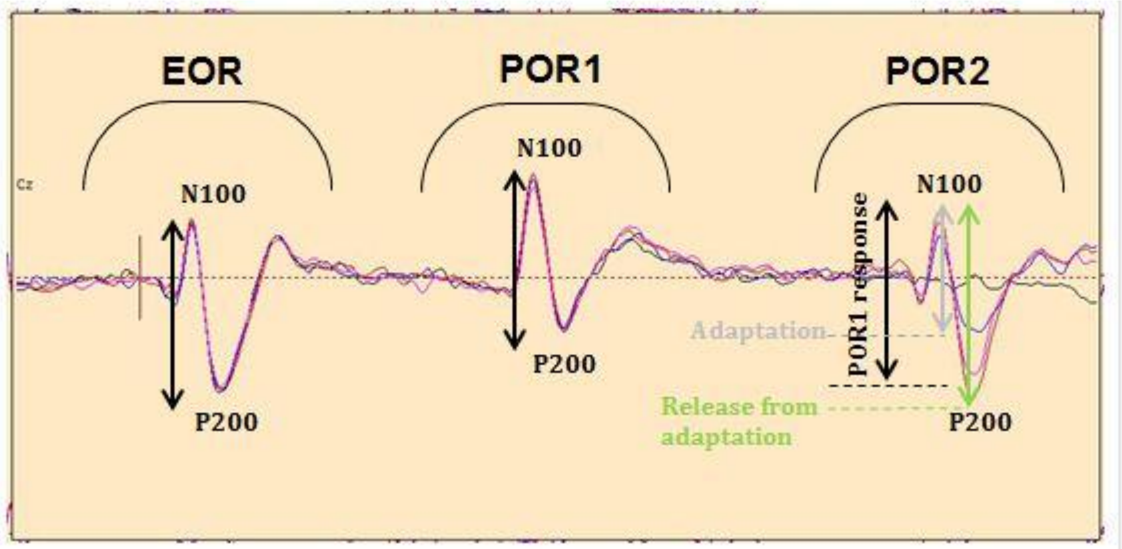

Figure 6.3. Illustration based on grand mean data from Cz of how the peak amplitudes for noise (EOR), adaptor (POR1) and probe (POR2), as well as the amount of adaptation from POR1 to POR2 were calculated for sensor (and source-based analyses). Black vertical arrows depict peak-to-peak amplitudes, but for illustration the peak-to-peak vertical arrow that appears next to POR2 responses represents the POR1 response for a given condition. The gray vertical arrows for POR2 relative to this black vertical arrow highlights adaptation (POR2 relative to POR1) for the 'blue' condition; smaller response for POR2 than for POR1 (for reference compare the gray dashed horizontal line with black dashed horizontal line). The green vertical arrows (and dashed) highlights release from adaptation (POR2 relative to POR1) for 'red' condition; greater response for POR2 than for POR1 (for reference compare the green dashed horizontal line with the black dashed horizontal line). Note that negative voltage is plotted upwards at this electrode site.

6.2.4 ERP localisation (source-based analysis)

ERPs have excellent temporal resolution but lack the spatial specificity afforded by other methods, such as fMRI. Given that currents spread laterally when they encounter resistance from the scalp, ERPs generated in one brain region may lead to substantial voltages at different parts of the scalp. Therefore it is important to use approaches that attempt to alleviate the inverse problem. The inverse problem refers to trying to determine the locations and orientations of dipoles by using the observed distribution of voltage over the scalp (Handy, 2005), when there are infinite set of dipoles that could give rise to the observed voltage distribution. Several approaches have been proposed (see Luck, 2005, p. 269), but here only the 'equivalent current dipole' approach to source localisation in BESA is discussed (see Luck, 2005, p.271-278), as this was the programme used for conducting source localisation in Experiment 10.

i) Using the Grand Average ERP waveform to derive the source model

A source model is typically created based on the GA data. This is because the GA model explains the data variance in all conditions and across all subjects, and therefore can be used to calculate the source activity in each subject across all experimental conditions (Hoechstetter et al., 2010). The average source location of the GA functions as a good model for each individual data set because a change in source location of 1-2 cm has little effect on the temporal course of source waveforms (Hoechstetter et al., 2010). Fitting individual data sets usually results in a large uncertainty in source location, especially in depth, due to the poorer SNR, so BESA recommends using a GA source model to stabilise the individual solutions. BESA advise three different ways of how to achieve a robust source model (see Hoechstetter et al., 2010) and was applied to the procedure used for Experiment 10 described in Chapter 8. Specifically, the source model should use a model with fixed sources and individual orientations (locations from the GA; Hoechstetter et al., 2000). Individual orientations were permitted in order to optimise the source model for all subjects given that the optimal orientation for each subject will vary and is dependent on the cortical folding, gyral anatomy and functional representation of each subject.

ii) Fitting procedure for discrete source analysis

When a discrete source analysis model is used, this relies on the assumption that each equivalent current dipole represents an extended brain region. Each dipole separates and mutually contrasts dipole activities with minimal cross talk, but requires users to manually fit dipoles and select appropriate fit intervals.

Two equivalent current dipoles were placed approximately in the left and right hemispheres of the auditory cortex (Morosan et al., 2001) using a four shell ellipsoidal volume conductor as a head model; this is a common procedure used for auditory ERP studies (e.g., Briley et al., 2013). These dipoles were then be used to estimate source locations of the ERPs for the probe responses across conditions and participants. Therefore, dipole orientations were fitted to average POR2 probe responses across conditions (during GAing an “all conditions” condition had been created). The resulting source model acted as a spatial filter for each condition and participant per hemisphere, which spatially restricts all conditions. Dipoles were subjected to a symmetrical constraint because both left and right auditory cortices were assumed to be approximately symmetrical. Given that individual data are generally much less stable than GA data, this symmetrical constraint also meant that erroneously placed dipoles were avoided. However orientations were unconstrained to account for individual variability in cortical folding, etc. The fitting window was chosen to encapsulate the POR2 response (0-300ms; 1810-2110 in the stimulus sequence ms). The POR2 probe response was chosen as the fit interval because it contained the adaptation-related activity of interest (pitch invariant/non-invariant responses). The fitting procedure auto-fits both sources to selected/highlighted fit interval. For example, in Briley et al. (2013)

dipole orientations were fitted to average probe responses across conditions; the same procedure employed in Experiment 10. The fitting procedure for Experiment 10 revealed two equivalent and symmetrical dipoles located in x = 40.2, y = -14.2, and z = 16.4mm (see Figure 6.4; approximately located in medial HG like Krumbholz et al., 2003). The residual variance of 12.57% revealed a good model fit. The resulting GA solution was used as a master model.

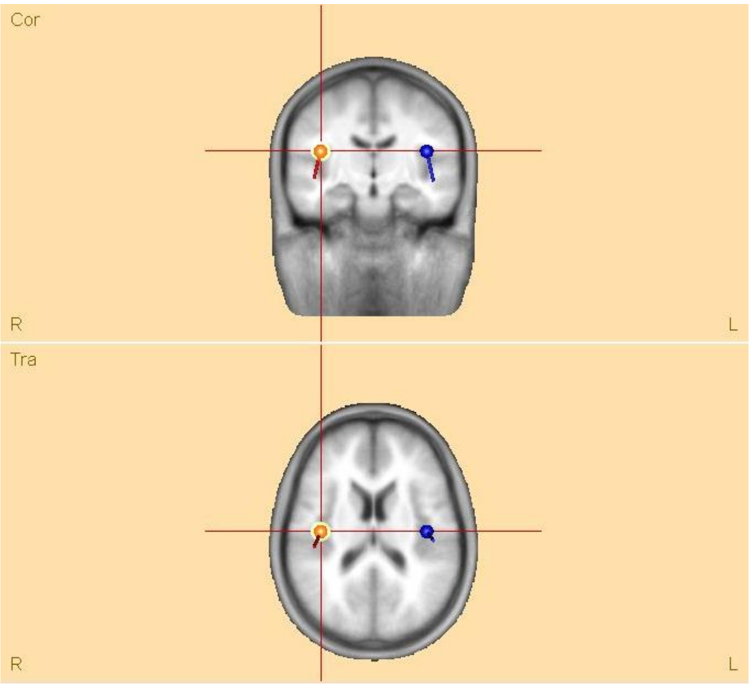

Figure 6.4. Dipoles fitted during source analysis. Display follows radiological convention, left = right.

iii) Running the batch script

A batch script was created and applied to the GA master model to all the individual data sets, orienting the sources individually and extracting the individual source waveforms for each condition. Unlike sensor-based analysis, artefact corrected data should not be loaded in the source analysis window. Instead, artefact correction should be turned off, and the individually saved artefact topography files (saved during artefact correction in the sensor-based analyses) should be individually appended to each subject's source model solution whilst running the source analysis batch script. This is because the use of already artefact corrected data can distort source activities. Using this 'optimising method' (Scherg et al., 2010) instead meant that correct source localisation could be obtained whilst taking account of blink and horizontal electro-oculograph artefact topographies in the averaged data segments, so that further statistical analysis can be performed.

The resulting source waveforms were then averaged together before conducting peak-to-peak quantification of the EOR, POR1 and POR2, as well as the calculation of the amount of adaptation from POR1 to POR2 were undertaken by following the same procedures described for the sensor-based analysis.

6.2.5 Statistical Analysis

The peak-to-peak amplitudes for EOR, POR1 and amount of adaptation for POR2 relative to POR1 (for both sensor and source-based analyses) were subjected to statistical analysis using a 2-by-2 repeated-measures ANOVA. The first factor reflected pitch congruence from adaptor to probe (same, different), and the second factor reflected timbre congruence (same, different). ANOVA is the most dominant statistical approach for analysing ERPs. ANOVA assumes that the data are normally distributed, and that there is compound symmetry and sphericity, specifically that the variances of the differences between all conditions are identical (homogeneity of variance) and the correlations between them are also equal (homogeneity of covariance). However, these assumptions are typically violated by the majority of ERP studies (Luck, 2005, p.258). For instance, the assumption of homogeneity of covariance is violated in ERP studies because the data from nearby electrodes tend to be more correlated than data from more distant electrodes. Nevertheless, given the robustness of this analytic technique, mild to moderate violations of normality do not typically influence type I error rates (Luck, 2005, p.258). However, for repeated measures designs, when violations to sphericity occur, a Greenhouse-Geisser correction should be used (Luck, 2005, p. 259). For all ANOVA testing an alpha criterion of $p < .05$ was used to test for statistical significance. All planned and post hoc paired sample t-tests were Bonferonni corrected to account for FWE associated with multiple comparisons. If Mauchly's test of sphericity assumption of sphericity was violated a Greenhouse-Geisser correction was used.

6.3. Combining ERP with an adaptation design

The following sections describe ERP methods for adaptation designs (as in Chapter 8). The ubiquitous phenomenon of adaptation was broadly discussed in Section 3.4.4, where it has been reported by numerous studies, across methods, species and senses, that when a stimulus is repeated, neural activity is reduced (see Grill-Spector et al., 2006, for a review). This section describes adaptation and the underlying neurophysiological basis for adaptation in a bit more detail, as well as reviewing how ERP and adaptation methodology that has been used to investigate auditory processing, and in particular, pitch constancy (i.e., Briley et al., 2013).

Adaptation has been widely used in behavioural and neuroimaging studies to probe the functional properties of sensory neuronal populations in humans. This is based on the assumption that processing is more efficient for a repeated stimulus feature. For example, when the adaptor and probe are identical, there will be adaptation which is known as a reduction in the cortical or behavioural response (e.g., reduced ERP amplitudes, reduced fMRI activation, or shorter reaction times). However, when the adaptor and probe differ along a given stimulus feature, if neurons are selective for the feature along which the adaptor and probe differed, the different probe will recruit new unadapted neurons to process the probe change, and subsequently show a release in adaptation compared to when the probe is identical. The first human fMRI study to use adaptation was conducted by Grill-Spector et al. (1999). They found that posterior object-selective regions of lateral occipital cortex showed sensitivity to many object transformations including size and position; however, more anterior object-selective regions were to some extent invariant to size and position, but sensitive to object rotation or illumination. In the auditory system, adaptation has mostly been studied at single-neuron level and is purported to play a role in change detection, such as that indexed by mismatch negativity (Jääskeläinen et al., 2004; Nelken & Ulanovsky, 2007). Nevertheless, the adaptation paradigm relies on temporal rather than spatial properties of the neural signal, so is particularly suited to ERP studies. EEG and MEG studies typically reflect changes in the amplitude and/or synchrony of ERPs and/or local field potentials (LFPs) caused by transmembrane currents in large numbers of neurons (Grill-Spector et al., 2006).

Adaptation offers a unique approach for examining questions relating to cortical responses in auditory cortex that cannot be addressed using conventional methods (i.e., subtraction). However, little is known about the neural mechanisms underlying this phenomenon. To account for repetition-related reductions (adaptation) in neural activity, three models have been proposed and discussed in Grill-Spector's (2006) review paper, namely facilitation, sharpening and fatigue models. The fatigue model is based on the assumption that the amplitude of firing of stimulus-responsive neurons decreases, the sharpening model assumes fewer neurons respond, and the facilitation model assumes the latency and or duration of neural activity is shortened (see Grill-Spector et al., 2006). The Sharpening model supported by Desimone (1996) and Wiggs and Martin (1998) suggests that repetition results in sparser representations of stimuli. Importantly it is neurons that code features irrelevant to the identification of that stimulus that exhibit adaptation. This is because the repetition-related changes are considered to be part of a learning process whereby representations (e.g., tuning curves) are sharpened and the distributed representation becomes sparser resulting in fewer responsive neurons in total (Grill-Spector et al., 2006). This sharper representation means that the tuning curves become narrower, the neurons become more sensitive to change, and this may ultimately lead to faster processing (Grill-Spector et al., 2006). A key difference between the fatigue and sharpening models is that according to the sharpening model neurons that are optimally tuned to a stimulus that is repeated should show little or no response reduction, whereas according the fatigue model neurons that are optimally tuned should exhibit maximal response reduction (Grill-Spector et al., 2006).

Only the fatigue model is based on the assumption that neurons that respond most strongly to the adaptor stimulus will be most adapted by it, and so the amount of adaptation from adaptor to probe determines the overlap and selectivity between neuronal populations (Grill-Spector et al., 2006). The fatigue model has recently been confirmed in an auditory EEG adaptation study (Briley & Krumbholz, 2013). They attempted to investigate whether adaptation in the auditory cortex is caused by neural fatigue/reduction in neural responsiveness, or by sharpening of the neural tuning of the adaptor stimulus. They used pure tones where the frequency separation between the adapter and probe stimuli and the stimulus onset asynchrony (SOA) varied. In their first experiment they examined the relationship between the degree of specificity of adaptation as a function of frequency separation and the rate of decay of adaptation with increasing SOA. They used one adaptor that preceded the probe, and this was either the same frequency or varied by 1/6, ½ or 1.5 octaves above the probe frequency (~1 kHz). The SOA varied in doublings from 125-1000ms (i.e., 125, 250, 500 and 1000ms). Through fitting the data with population models of neural fatigue and neural sharpening, the results from experiment one showed independence between adaptation specificity and decay rate, and therefore fitted the fatigue model much better. The second experiment enabled them to measure adaptation specificity after multiple presentations of the adaptor stimulus (i.e., two or three adaptors, mimicking the oddball paradigm), but the SOA remained fixed at 500ms. They found that multiple adaptors did lead to more adaptation overall, but this was specific to the adapting frequency (i.e., the adaptation tuning curves were much sharper for multiple adaptors, compared to single adaptor).

Using adaptation, investigators are enabled to search for suppression of cortical responses in neuronal populations to a repeated pitch value, regardless of the stimulus with which the pitch is associated (Griffiths & Hall, 2012). The adaptation technique can therefore be used to effectively address questions relating to pitch constancy because stimulus manipulations are not confined to one acoustic feature, and can reveal neuronal populations that are invariant or non-invariant to multiple acoustic features independently of the stimulus context. Consequently, studies using ERP adaptation designs to investigate auditory processing have increased in the last few years. For example, recent EEG investigations have proven fruitful in exploring frequency (Briley & Krumbholz, 2013; Lanting, Briley, Sumner, & Krumbholz, 2013), spatial location (Briley, Kitterick, & Summerfield, 2012; Briley & Summerfield, 2014) and pitch tuning in auditory cortex (Briley et al., 2013). Lanting et al. (2013) investigated the temporal properties of adaptation in late auditory-evoked components. The first experiment used a single adaptor, and the SOA and ISI varied between adaptor and probe. The adaptor and probe were the same frequency (i.e., 1 kHz), and the duration of both was fixed to 100ms in Experiment 1A, but the adaptor duration varied in Experiments 1B and 1C (100-975ms). The SOAs could either be doublings from 125ms to 1000ms (125, 250, 500 and 1000ms; in Experiments 1A and 1B) or stayed the same in Experiment 1C (1000ms). ISI varied from 25-900ms (25, 150, 400 and 900ms).

They found greater recovery of adaptation with increasing ISI, but adaptation increased with increasing adaptor duration (i.e., SOA). They claim that this finding suggests that adaptation is caused by on-going, rather than onset, responses to the adaptor, and recommended that future studies use long adaptors and short ISIs between adaptor and probe to maximise the likelihood of finding stimulus-specific adaptation effects. In their second experiment, which used multiple adaptors, they found that adaptation decayed quickly after the adaptor was repeated.

Briley et al. (2013) also conducted a study that explored how pitch is represented in the auditory cortex. Specifically they wanted to determine whether adaptation of auditory cortical responses is selective to pitch, and if so, does the representation of pitch reflect the physical or perceptual characteristics of pitch. Physical aspects of pitch relate to the repetition rate (i.e., single monotonic dimension ranging from low to high). Perceptual aspects of pitch can be described along two dimensions, namely pitch height (single monotonic dimension reflecting the octaves in which given notes can reside) and pitch chroma (cyclical; accounts for similarity of cycle notes across different octaves). They used IRN complex tones that were similar to most musical instruments or voiced speech to create adaptor and probe stimuli which varied in repetition rate (i.e., pitch). They also used pure tones that differed in frequency to compare the results for different pitch-evoking stimuli.

Briley et al. (2013) hypothesised that if pitch is represented by its physical dimension then the probe stimulus should increase monotonically with increasing pitch separation from adaptor to probe. Nevertheless, if pitch is represented in terms of the perceptual dimensions of pitch height and chroma, the adaptation effects of the probe should be nonmonotonic with a dip at octave pitch separations. They also predicted that if pitch processing neurons are invariant to the spectrum or timbre of the stimuli, the pattern of adaptation effects for pure tones should be similar as the results for complex (i.e., demonstrating evidence for pitch constancy).

In Briley and colleagues study, the long adaptor stimulus (1500ms) was always immediately followed by a short probe stimulus (250ms). For the probe stimuli, the IRN stimuli had a repetition rate of 125 Hz and 500 Hz, and pure tones had a frequency with the same nominal values. For the adaptor stimuli, repetition rate or frequency was varied relative to that of the probe (i.e., pitch separations varied between 0.5, 1.0 and 1.5 octaves across experiments). Both stimulus types were presented in a continuous background of masking noise.

As mentioned previously, the analytic approaches used by Briley et al. (2013) were used as a guide for the analysis of Experiment 10 described in Chapter 10, and are detailed in Section 6.2.3. They found that for IRN stimuli, adaptation effects were nonmonotonic, whereby adaptation was greater (i.e., the probe response was smaller), when the adaptor and probe were separated by octave, than half an octave or 1.5 of an octave. They claim that this suggests that a particular note and its octave share more overlap in their neural representations than a note and its half octave. These EEG adaptation effects therefore mirrored the cyclicality of the pitch chroma dimension and the authors argue that this suggests that there is a representation of pitch chroma in the auditory cortex. Results from source analysis of IRN responses localised the source of this pitch chroma representation to somewhere between the anterior or lateral primary auditory cortex. In contrast, pure tones showed a monotonic increase with increasing pitch between the adaptor and probe, suggesting that pure tones have different neural generators to IRN stimuli (i.e., non-invariant to timbre), and therefore provided evidence against the idea of pitch constancy or a pitch center. The source of pure tone responses was located in medial HG.

6.4. Summary

In summary, EEG provides a continuous measure of electrical activity recorded at the scalp and plots changes in voltage between two different electrode locations over time. A number of preprocessing steps are required to extract the ERPs relating to specific sensory, cognitive or motor events relating to experimental conditions. Once these ERPs have been extracted for each condition and listener, a number of source and sensor-based analytical steps can be undertaken, before statistical analyses is employed to test the null hypothesis. Most EEG and MEG studies of auditory processing have concentrated on examining specific auditory-evoked components, such as the P100, N100, P200, EOR and POR, using conventional subtraction designs. However, following advancements in the visual modality, adaptation methodology has now grown in popularity in the auditory domain. This is because it enables researchers to investigate questions relating to perceptual invariance. Recent auditory ERP methodology combined with adaptation has been used to examine to frequency, spatial location, and importantly, pitch. This has led to novel insights into their neural representations, and it is hopeful that future efforts in this regard will yield the same.

Chapter 7. Identifying parameters for pitch and timbre to equate perceptual salience

7.1. Introduction

This chapter describes how the pitch and timbre parameters were chosen and matched to create the stimuli used to investigate pitch invariance to timbre for the experiments reported in Chapters 8 and 9 (ERP Experiment 10, reaction time Experiment 12). Pitch invariance or pitch constancy is defined as the ability to discriminate a pitch regardless of the spectral, temporal and binaural characteristics. Appropriate selection of these parameters is vital given that changes in one non-target or task-irrelevant perceptual dimension (e.g., timbre) are known to have detrimental effects on the discrimination of the other dimension of interest (e.g., pitch; Borchert et al., 2011; Steele & Williams, 2006). Identifying stimulus parameters that yield equivalent discrimination performance across pitch and timbre tasks ensures that the perceptual salience of both parameters is well matched. This is based on the assumption that discrimination thresholds provide an objective surrogate marker for perceptual salience. Throughout, I will use the terms discriminability and perceptual salience interchangeably, as they both relate to a signal being discriminable from noise. Discriminability is an important control, especially for examining pitch invariance to timbre, given that both behavioural and neurophysiological findings can be influenced by the salience of the stimulus parameters used (see Krumbholz et al., 2003; Krishnan et al., 2012; see my own data in Chapter 5). However, such attempts at controlling perceptual salience or discriminability, prior to human investigations of pitch invariance, are seldom seen (if at all) in the pitch coding literature to date.

Three experiments (Experiments 6, 7 and 8) were employed to identify the stimulus parameters for pitch and timbre that yielded matched discrimination performance across pitch and timbre tasks. Materials and methods, results, and conclusions across Experiments 6-8 are discussed in turn, highlighting the cumulative evidence which led to the final parameters being chosen, and the evidence that pitch and timbre was appropriately matched before undertaking Experiment 10 and 12. First, a brief literature review of how irrelevant changes in timbre or pitch affect pitch or timbre discrimination respectively, the importance of matching discriminability when searching for pitch invariant representations in the human auditory cortex, as well as the objectives and criteria for equating discrimination performance, is given.

7.1.1 Timbre changes affect pitch discrimination and vice versa

Sounds are multidimensional, varying across physical and perceptual dimensions. Humans and non-humans alike are able to demonstrate perceptual constancy, for example matching the pitch of a musical note, irrespective of the instrument playing it. Pitch discrimination is achieved effortlessly when all other feature dimensions remain fixed. Typically psychophysical experiments vary one dimension with the assumption that the parameter being manipulated corresponds to a unique sensory attribute or perceptual dimension (Dai & Micheyl, 2012). F0 difference limens are less than 1% for tones containing resolved harmonics. For example, Moore's (1974) study reported F0 difference limens of less than 0.2%, however in reality, this is rarely the case because changes in one dimension, often induce changes in other dimensions (see Dai & Micheyl, 2012). For instance, differences in spectral region (timbre) may cause tones with the exact same F0 to sound like they have a different pitch (i.e., pitch shifts; Chuang & Wang, 1978; Singh & Hirsh, 1992; Vurma, 2014) Furthermore, as demonstrated in the following sections, pitch discrimination is often impaired when another irrelevant dimension, like timbre, is varied across intervals (and vice versa).

7.1.2 Timbre changes impair pitch discrimination

Most of the work conducted in this area has focused on the timbre effects on pitch perception, with studies rarely exploring the effects of both parameters on one another. Pitch is always associated with timbre. Timbre is also an important perceptual attribute of sound that “allows one to distinguish among tones having the same pitch, loudness and duration” (ANSI, 1994), and is vital for speech and music perception (e.g., allows us to group together belonging to the same source, such as same person talking or the same instrument playing a melody) . As mentioned previously, timbre is specifically determined by the distribution of energy over harmonics (i.e., spectral envelope), but perceptually, timbre refers to the unique ‘quality’ of a sound (see Town & Bizley, 2013 for a detailed discussion about timbre perception). As demonstrated in Figure 2.5 (Chapter 2), a piano and a violin can play a tone with the same pitch (at the same loudness and duration), but each musical instrument has a particular timbre which allows one to distinguish that the same notes played has been generated by different musical sources. Given that humans and non-humans show pitch constancy, some researchers have argued that there are pitch selective neurons or a putative pitch centre which codes the F0 of a sound regardless of other perceptual characteristics (Bendor & Wang, 2005; Krumbholz et al., 2003). This pitch constancy criterion for a pitch centre is more rigorously addressed in Chapters 8 and 9, but first, the following sections illustrate how, on a perceptual level at least, changes in one dimension affect the perception of the other. For example, listeners are not able to discriminate task-relevant pitch and timbre changes as effectively when task-irrelevant differences are introduced for timbre and pitch dimensions, such as by using non-overlapping harmonics to create a timbre difference or changing the F0 between intervals.

A growing number of psychophysical studies have found that timbre changes impair pitch discrimination performance on psychophysical tasks (e.g., Borchert et al., 2011; Krumhansl

& Iverson, 1992; Micheyl & Oxenham, 2004; Moore & Glasberg, 1990; Warrier & Zatorre, 2002). Moore & Glasberg (1990) conducted a series of pitch discrimination (F0 difference limens) studies in which they manipulated timbre differences of complex tones by using overlapping and/or non-overlapping harmonics. When referring to overlapping harmonics, this indicates that two given complex tones shares all (or some) harmonic numbers in common (e.g., tone 1 harmonics = 1, 2, 3, 4, 5, 6; tone 2 harmonics = 1, 2, 3, 4, 5, 6), whereas non-overlapping harmonics indicate that all (or some) of the harmonic numbers are different (e.g., tone 1 harmonics = 1, 4, 5, 8, 9 and 12, and tone 2 harmonics = 2, 3, 6, 7, 10, 11). Complex tones can also have a mixture of both overlapping and non-overlapping harmonics (e.g., partially overlapping, where harmonics 1-6 are overlapping and harmonics 7+ are non-overlapping; tone 1 harmonics = 1-6, 7, 9, and tone 2 harmonics = 12, and 1-6, 8, 10, 11). Two research questions that were of particular interest to the current chapter were, i) whether non-target timbre differences (i.e., different harmonics/spectra) impair the ability to discriminate pitch (Experiment 1), and ii) how good is pitch discrimination when the two tones comprise no common harmonics (different timbre), versus when the tones have harmonics in common (same timbre; Experiment 3). They used an adaptive Two-Interval Two-Alternative-Forced-Choice (2I-2AFC) task using a range of nominal F0s (43.75-400 Hz). A 2I-2AFC usually involves presenting a tone in one interval, followed by another tone presented in another interval. The participant is usually required to decide in which interval the pitch was different (higher or lower). An adaptive psychophysics task refers to presenting trials based on a staircase method (up/down rule) based on the listener’s performance, for example if they get more trials correct the harder the task becomes (i.e., signal decreases). Thresholds are obtained at targeted accuracy levels (see Section 7.4.3.1). There is only one pitch change the participant needs to identify, such as ‘was the pitch higher in the first or second interval?’. Any other non-target dimension changes, such as simultaneous timbre change occur across both intervals. The output is typically given as a geometric threshold metric or percentage correct. For example, a F0 difference limen threshold of 1% for a 200 Hz tone indicates that the F0 of the other tone needs to be at least 202 Hz for the listener to be able to detect a difference in pitch. Lower F0 difference limen thresholds or greater accuracy indicate listeners are better at discriminating that dimension.

Moore & Glasberg (1990) asked listeners to discriminate which interval contained the higher F0. As shown in Figure 7.1, across intervals tones could either have: condition 1) all harmonics in common, to produce the same timbre (e.g., same harmonics 1, 2, 3, 4, 5, 6), condition 2) some harmonics in common, which creates a small timbre difference across intervals (e.g., partial overlap yields interleaved harmonics 1-6, 7, 9 and 12, and 1-6, 8, 10, 11), and condition 3) no harmonics, in common to produce a somewhat larger timbre difference (e.g., non-overlapping harmonics 1, 4, 5, 8, 9 and 12, and 2, 3, 6, 7, 10, 11). Harmonics for condition 3 were either low (1-12) or high (7-17).

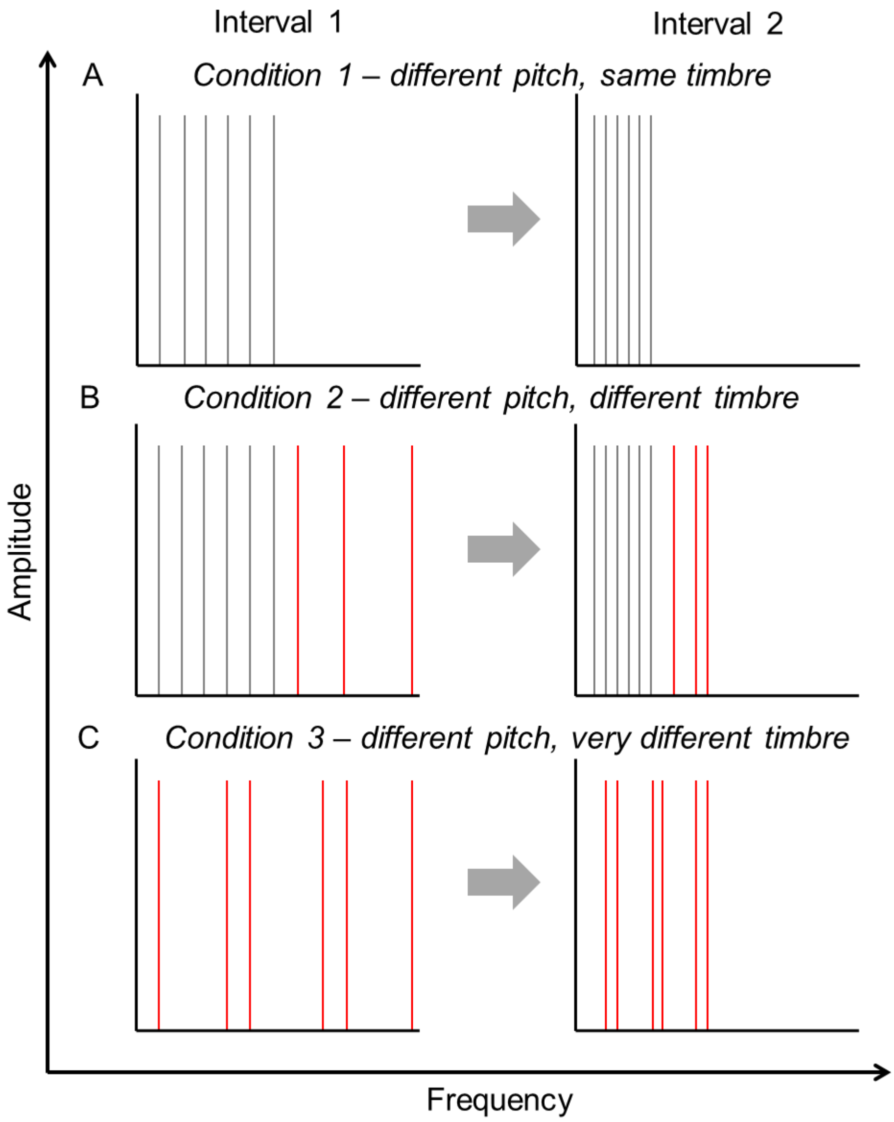

Figure 7.1. Schematic diagram illustrating how the harmonics (vertical bars) were manipulated to create three timbre conditions for the F0 task in Moore & Glasberg (1990). Panel A: Condition 1; both tones have all harmonics in common (different pitch, same timbre). Panel B: Condition 2; both tones have some of the harmonics overlapping and some are non-overlapping (different pitch, different timbre). Panel C: Condition 3; for both tones all the harmonics are non-overlapping (different pitch, very different timbre). In this example, interval 1 always contained the higher F0 target. For illustration purposes gray bars = overlapping harmonics, and red bars = non-overlapping harmonics.

In experiment 1 the harmonics were either completely overlapping (condition 1) or partially overlapping (condition 2). They kept the first six harmonic components overlapping for all conditions in this experiment because this information provides the greatest cue for the residue pitch of the sound (Plomp, 1967; Ritsma, 1967; Moore, Glasberg, & Shailer, 1984; Moore, 1985), but also used upper harmonics as far as the 10/11th as these are considered to be individually resolved by the cochlea and therefore create a more salient pitch percept.

These upper harmonics were either absent or overlapping (condition 1) or non-overlapping (condition 2). Overall performance in experiment 1 monotonically improved with increasing F0. However, F0 difference limen's were larger (approx. 1.6% F0 difference limen) when the upper harmonics were non-overlapping (i.e., different pitch, different timbre; condition 2), compared to when the upper harmonics were overlapping and/or absent (approx. 0.5-1.2%; different pitch, same timbre; condition 1). Moreover, thresholds were actually better for condition 1 trials in which harmonics above 6th were used. These results suggest that timbre variation has a detrimental effect on pitch discrimination performance, despite irrelevant timbre differences being generated by non-overlapping harmonics above the 6th. It might be possible that timbre affects pitch discrimination regardless of which harmonics are manipulated. Despite common upper harmonics leading to improved performance, having some partially overlapping harmonics undoubtedly helped listeners with the task to some degree, as the timbre differences created were relatively small. Tones which share no overlapping harmonics at all would have produced much greater timbre differences, and potentially making it much more difficult to discriminate F0. This is likely why the F0 difference limen thresholds reported in Moore & Glasberg's paper were slightly better than other studies that have used complex tones with no harmonics in common (e.g., Borchert et al., 2011; Faulkner, 1985; Micheyl & Oxenham, 2004; Ritsma, 1963). Nevertheless, it is important to note that when greater starting F0 differences were used (i.e., 5-10% of 400 Hz) performance did improve for all conditions, which suggests that if the F0 cue is large enough then the detrimental effects of timbre differences can be diminished to some degree.

In experiment 3, Moore and Glasberg (1990) investigated the pitch discrimination of tones that have no harmonics in common (different pitch, different timbre; condition 3) and all harmonics in common (different pitch, same timbre; condition 1). To create a moderate timbre difference for condition 3, interleaved harmonics were used. Similarly, F0 difference limens also did improve with increasing F0 value (e.g., F0 difference limens were smaller for the 400 Hz condition than for 50 Hz condition). Additionally, F0 difference limens were larger (approx. 1.5-4% F0 difference limen) for tones with non-overlapping harmonics (condition 3) than for tones with the same harmonics (approx. 0.5% F0 difference limen; condition 1). Thresholds were also worse in experiment 3, compared with those reported in experiment 1 (where tones were comprised of some common harmonics). This might be because interleaved harmonics created a somewhat larger timbre difference which listeners may have found more difficult to ignore. However the authors also reported large inter-subject variability, with listeners being affected by timbre differences to varying degrees. For example, one listener (BM), performed equally well (F0 difference limens less than 1%) across the majority of conditions (at F0s 100 and 200 Hz), which suggests that residue pitch discrimination for certain individuals, and at certain stimulus parameters, can be equivalent for conditions in which in the timbre is the same (overlapping harmonics) and conditions in which the timbre is different (non-overlapping harmonics). However it should be noted that BM is the first author and a very experienced psychophysical test participant.

Moore & Glasberg’s (1990) findings that timbre differences impair pitch performance have been confirmed by more recent evidence (e.g., Borchert et al., 2011; Micheyl & Oxenham, 2004; Warrier et al., 2004). Although the majority of psychophysical studies of pitch use 2I-2AFC designs (Green & Swets, 1966), Borchert et al. (2011) used a dual-pair design (Rousseau & Ennis, 2001) to examine pitch discrimination performance for simultaneous and sequential tones, whilst varying timbre in a series of experiments. Dual-pair designs are a modified version of the same-different paradigm ‘embedded’ within a 2I-2AFC design (Noreen, 1981). This is also known as the 4-Interval AX (4IAX) paradigm, or the 4-interval same-different paradigm (Creelman & Macmillan, 1979; Noreen, 1981; Rousseau & Ennis, 2001; see Micheyl & Messing, 2006 for an overview). They conventionally involve presenting a pair of tones in each of the two observation intervals and assume that the listener makes four observations on each trial (Micheyl & Messing, 2006). In one interval, the two tones differ in at least one dimension (e.g., different pitch). In the other interval, the two tones do not differ along that same dimension (e.g., have the same pitches). However it is possible to manipulate both intervals along another dimension (e.g., different timbre). Unlike *m*AFC methods (where m is number of intervals), dual-pair designs (and similar) are advantageous because they do not require listeners to specify the direction of the dimension difference between the two intervals (e.g., higher and/or lower), as listener’s are only required to indicate in which interval (first or second) the pair of tones differed along the relevant dimension (Micheyl & Dai, 2009). Listeners are likely to find this an easier task since they do not have to focus on the specific direction of the dimension difference, just indicate where one occurs (Rousseau & Ennis, 2001). This is important because a number of studies have shown that some listeners find identifying pitch direction changes much more difficult than detecting pitch changes alone (e.g., Johnsrude, Penhune, & Zatorre, 2000; Semal & Demany, 2006; Tramo, Cariani, Koh, Makris, & Braida, 2005).

Borchert et al. (2011) used the dual pair design and asked listeners to decide in which interval a F0 difference occurred between tones. They used 200 Hz (+- 0.5, 1.2 and 4.5 semitone difference) complex tones which were filtered into low (0-700 Hz) and high (1150-3500 Hz) spectral regions, to create perceptually large timbre differences. Overall they found that listeners were significantly worse at discriminating two spectrally non-overlapping tones when they were presented sequentially (i.e., the two tones within each observation interval are presented immediately one after the another with no overlap; sensitivity or discriminability index D-prime (d’) = 1, approx. 3.5% F0 difference limen) compared to simultaneously (i.e., the two tones presented within each observation interval have the same onset and offset; approx. 1.5% F0 difference limen). Large mistuning (F0 difference of 4.5 semitones) did slightly improve sequential performance, but was still relatively poor. They concluded that poorer performance in sequential condition was likely to be due to large spectral differences between tones, as opposed to masking or memory decay.

It has also been documented that it is difficult for listeners to detect pitch changes when there are large changes in timbre (Labuschagne & Hanekom, 2013). Studies using large timbre differences often report influences on pitch perception (e.g., Singh & Hirsh, 1992; Russo & Thompson, 2005; Vurma & Ross, 2006). However Labuschagne & Hanekom (2013) used smaller spectral parameter adjustments and found that pitch balancing was not required. Subsequently it was important for the current study to keep timbre changes as small as possible in order to avoid any perceived pitch changes. The bigger the timbre difference the larger the interference on performance so we would expect there to be a monotonic improvement in performance as a function of decreasing timbre difference.

These findings indicate that for listeners to better discriminate pitch, it is important to create a timbre difference that is not too large. Given that complex tones with no common harmonics have very different timbres (i.e., Borchert et al., 2011), it was decided that the current Experiments 6-8 would employ the partially overlapping and interleaved convention used by Moore & Glasberg (1990) as this created a smaller timbre difference. However at this moment it was unclear what should be the optimal combination of overlapping and non-overlapping harmonics. This was determined in the series of feasibility studies reported here.

7.1.3 Pitch changes impair timbre discrimination

Timbre invariance refers to the ability to discriminate tones that have different F0s but have been generated by the same musical instrument. Only a handful of studies have investigated how F0 changes affect timbre discrimination (e.g., Handel & Erickson, 2001; Pitt, 1994; Steele & Williams, 2006). Steele & Williams (2006) were interested in the bandwidth of timbre invariance to F0 changes, and examined listener’s ability to discriminate the timbre of pairs of musical sounds (i.e., horn and bassoon) whilst the pitch was varied within a 0-2.5 octave range (in 0.5 steps). Pairs of tones could differ in pitch, timbre or both. Listeners had to discriminate whether or not the tones had the same pitch and same timbre, as well as provide an indicator of how confident they were in their responses. Overall musicians were more accurate (89.8%) than non-musicians (61.8%). They found that non-musicians could only accurately discriminate timbre when the pitch separation difference was below one octave. However when the pitch separation increased, performance dropped to chance level above one octave. Although musicians’ performance also declined, this was to a much lesser extent, performing at approximately 80% accuracy at 2.0 and 2.5 octave separations. Nevertheless, musicians and non-musicians showed similar error patterns. Judgments for smaller octave separations (e.g., 0.5 octaves) were more often misjudged as the same instrument when it was different, and conversely, larger separations (e.g., 2.5 octaves) were more often misjudged as a different instrument when it was the same. Both groups also misjudged identical pitches as being different when they were played on different musical instruments. The authors suggest that these findings indicate that increasing octave separations led to a loss of timbre invariance. However, the tolerance of timbre invariance in musicians appears broader than that for non-musicians. Importantly the authors acknowledged that the shape of this bandwidth might change depending on other factors, such as musical experience, instruments used and stimulus parameters. Steele & Williams' (2006) results not only demonstrate that pitch changes impair timbre discrimination, but also that musical training can alleviate these effects to a certain degree. Although overall musicality does not seem to make a significant improvement in listener's ability to do such tasks (but see Zarate, Ritson, & Poeppel, 2013), training appears to be important for learning how to separate pitch and timbre dimensions. It might be that individuals who have exposure to musical training regarding pitch and timbre dimensions have developed better listening skills to be able to separate out these dimensions (Pitt, 1994). Therefore it seems reasonable to suggest that if listeners are given sufficient training for a given set of stimulus parameters, they may be able to reliably discriminate pitch and timbre.

7.1.4 Summary of the evidence

Although the precise interactions between pitch and timbre are unknown, overall cumulative evidence highlights that: 1) both timbre and pitch differences impair pitch and timbre discriminability, respectively, 2) pitch discriminability is better when overlapping and partially overlapping harmonics are used compared with completely non-overlapping harmonics, 3) timbre may not significantly impair pitch discrimination for some listeners, at certain parameters, 4) using larger F0s improves pitch discrimination performance, 5) training individuals might help facilitate discriminability to some degree, and 6) large timbre differences not only impair discrimination but may also lead to pitch shifts.

7.2. Importance of equating discrimination performance for pitch and timbre parameters

Given that both F0 and frequency can affect the discriminability of pitch and timbre, it was important to match discrimination performance so that the perceptual salience of these two dimensions was approximately equal. This is vital for quantifying neurophysiological responses to pitch and timbre since it was already described in Section 2.4.3 how perceptual salience can affect pitch response magnitude or amplitude (e.g., Griffiths et al., 1998, 2001; Gutschalk et al., 2002; Krishnan et al., 2010; Krishnan et al., 2012; Krumbholz et al., 2003; Penagos et al., 2004; Seither-Preisler et al., 2003; Soeta et al., 2005).

Studies examining cortical responses to pitch do not always obtain both neurophysiological and psychophysical measures in the same listeners (e.g., Penagos et al., 2004), so they are unable to associate cortical responses with behavioural results. Despite the body of evidence described previously, recent human (Briley et al., 2013) and non-human studies

(Bizley et al., 2009) examining neurophysiological findings to pitch and timbre dimensions have also failed to control their stimuli for perceptual salience, which may have influenced their results. For instance, if the timbre of a stimulus is more salient than the pitch, neuronal populations may appear more selective or sensitive to that acoustic feature (e.g., as shown in Bizley et al., 2009).

The studies presented here describe the first attempt to determine stimulus parameters for pitch and timbre where the discriminability or perceptual salience is well matched. Given that timbre impairs pitch discrimination performance and vice versa, it is of primary importance to identify pitch and timbre stimulus parameters that led to equivalent performance on both pitch and timbre tasks, specifically, when the non-target dimension is varying from interval to interval (i.e., ‘different pitch, different timbre’ conditions; see Table 7.1). It was assumed that if listeners could discriminate pitch and timbre for these more difficult conditions, then they would be able to easily discriminate conditions in which only pitch or timbre targets were present, whilst the non-target dimension was the same between intervals (i.e., ‘different pitch, same timbre’ and ‘same pitch, different timbre’ conditions). These conditions correspond to those implemented across experiments reported in Chapters 8 and 9 in which there is a difference in pitch, timbre or both. Experiments 6-8 (and corresponding Experiments 9 and 11) only required listeners to judge a difference in pitch or timbre, therefore the ‘same pitch, same timbre’ condition reported for the ERP Experiment 10 in Chapter 8 and the reaction time Experiment 12 in Chapter 9, could not be implemented here.

Table 7.1
Conditions for pitch and timbre tasks

	Task	
Condition	Pitch task	Timbre task
Condition 1	Different pitch, different timbre	Different pitch, different timbre
	(i.e., One target and one non-target dimension changes)	(i.e., One target and one non-target dimension changes)
Condition 2	Different pitch, same timbre	Same pitch, different timbre
	(i.e., One target change only)	(i.e., One target change only)

7.3. Objectives

Experiments 6-8 aimed to equate discrimination performance between pitch and timbre tasks by:

- Determining potential stimulus parameters for pitch (i.e., F0) and timbre (i.e., minimum number of overlapping harmonics) differences that yields broadly equivalent discrimination performance for both pitch and timbre tasks for ‘different pitch, different timbre’ conditions for one experienced listener (Experiment 6).

- Re-assessing those parameters again in two naïve listeners to finalise the most appropriate parameters (Experiment 7).
- Ensuring listeners could be trained on the pitch and timbre tasks for all pitch and timbre conditions and trial types, and achieve comparable levels of performance across listeners (Experiment 8).

The two requirements for determining stimulus parameters for matching pitch and timbre discriminability were that the overall performance for each condition for each pitch and timbre task was, 1) comparable and 2) had an approximate d’ prime of ≥2. A d’ criterion of 2 was chosen to evaluate the suitability of pitch and timbre stimulus parameters because this equates to approximately 85% accuracy and indicated a good level of discriminability.

7.4. General methods and materials

The following sections describe the general methodology and materials for Experiments 6-8, 9 and 11. Experiment-specific details are presented in their corresponding sections.

7.4.1 Listeners

Three listeners (two right-handed, #02 and #09, one left-handed, #28) volunteered to participate in one or more of the psychophysical Experiments 6-8 (1 male; M_{age} = 24.67, *SD* = 0.58, age-range 24-25 years). Listener #02 served as an experienced listener to test a wide range of different stimulus combinations in Experiment 6. The most optimal results from Experiment 6 then informed the partial subset of stimuli that naïve listeners #09 and #28 would discriminate in Experiment 7. In Experiment 8, the most optimal parameters identified in Experiment 7 were then fully tested by listeners #02 and #09. Participant #28 was musically trained on the piano to grade 8. Listeners #02 and #09 reported no formal musical experience, but did however report 2 years of self-taught musical skills on the guitar, piano and keyboard. All participants reported normal or corrected-to-normal vision, and had clinically normal hearing (≤25 dB HL between 250-8000 Hz). No listeners reported any history of hearing, neurological, and/or psychological impairment, and use of psychiatric medication or substance misuse. Listeners gave written informed consent, and the study was approved and performed in accordance with the College Research Ethics Committee’s guidelines, Nottingham Trent University (ethics no. 2011/46).

7.4.2 Stimuli

Presentation and timing of stimuli and trials were controlled using a bespoke MATLAB toolbox (The Mathworks, Natick, MA) called EarLab (provided by Professor Chris Plack).

EarLab enables users to digitally synthesise a variety of pitch stimuli. Wideband harmonic complex stimuli were chosen because they generate a more salient and reliable pitch percept compared to IRN, particularly narrow bandwidth IRN (Barker et al., 2012; 2013; Hall & Plack, 2009). All stimuli were 200ms in duration (with 10ms onset and offset ramps) and were low-pass filtered at 2 kHz to ensure all stimuli were resolved up the $9^{th}/10^{th}$ harmonic. A noise masker was not used because all stimuli were matched to activate the same gross spectral region, helping to maintain a stable envelope and prevent cochlear distortions (fulfilling the fourth pitch criterion, ‘accounting for confounding’ variables; see Section 1.4).

The stimulus parameters used across Experiment 6-8 are shown in Table 7.2. The starting F0 was either 200 Hz or 220 Hz, and the F0 difference (%) between the two tones in each pair was manipulated to varying degrees depending on the experiment (between 5%-20%). Given that small F0 differences and large timbre differences can make stimulus discrimination more difficult (see Introduction), the numbers of pitch and timbre differences were kept within a reasonable but narrow range to avoid excessive hours of testing. A smaller range of F0 differences was also used to ensure that harmonics that were not meant to overlap did not overlap. Subsequently the starting F0 was always 200 Hz in Experiments 6 and 7. The varying timbre differences used across studies were digitally synthesised using overlapping and non-overlapping harmonic filter bands (similar convention used in Moore & Glasberg, 1990). The default was set so that each tone presented would always have the first harmonic number in common (i.e., lowest overlapping component was 1), but the non-overlapping harmonics could be manipulated to cause changes in the perceived timbre. The lowest possible non-overlapping harmonic was 2. Across studies, to create a small but perceivable timbre difference, the lowest non-overlapping harmonic component was 2, 3, 5 and/or 7. For instance, to create two tones required for a given pair, each with a different timbre, setting the lowest non-overlapping component to 3 would mean for one tone the harmonic components would be 1, 2, 3, 6, 7, and 10, and for the other tone would be 1, 2, 4, 5, 8, and 9.

Table 7.2

All possible pitch and timbre stimulus parameter combinations used for discrimination tasks across Experiments 6, 7 and 8 (both adaptive and/or method of constant paradigms)

		Discrimination Task	
Stimulus dimension and parameters		Pitch	Timbre
Pitch	Starting F0 (Hz)	200 or 220	200 or 220
	Nominal starting difference F0 difference (%)	5, 10, 15, or 20 (target dimension)	5, 10, 15, or 20 (non-target dimension)
Timbre	Lowest overlapping harmonic	1	1
	Nominal starting difference Lowest non-overlapping harmonic	2, 3, 5 or 7 (non-target dimension)	2, 3, 5 or 7 (target dimension)

Note: The same parameters were used for both pitch and timbre discrimination tasks in order to isolate parameters that were equivalently discriminable across both tasks. The target for one dimension was always coupled with a non-target dimension from the other dimension. For example, for either task a listener could have been presented with a 200 Hz starting F0, with a 5% F0 difference where the lowest overlapping harmonic is 1 and the lowest non-overlapping harmonic is 3. The F0 difference (%) for the non-target interval for a pitch task was always zero (i.e., same F0 repeated) and lowest non-overlapping harmonic for the non-target interval for a timbre task was always repeated.

Sounds were generated and presented in EarLab with a DVD quality sampling rate of 48 kHz and a 16-bit resolution and was calibrated to Sennheiser HD-280 headphones to present stimuli with an overall sound level of 70 dB SPL. Sound level was measured prior to testing using a Brüel & Kjær 4231 Sound Calibrator, affixed with a Brüel & Kjær 2250 Sound Level Meter, Brüel & Kjær 4153 artificial ear, and Brüel & Kjær 4192 Half Inch Microphone. Sound pressure levels were calibrated for the MacBook Pro laptop computer using the “LAF” setting which simulates a real listener, whilst presenting a 1 kHz tone at maximum amplitude and measuring dB SPL outputs for the left (117.3 dB SPL) and right (118.6 dB SPL) side of the headphones. The average dB SPL (118 dB SPL) was then input into the Earlab code to calibrate the HD 280 headphones accordingly. The sound level on the laptop was then attenuated to present sounds over the headphones at an overall level of 70 dB SPL.

7.4.3 Procedure and Design

All Experiments 6-8 were completed in a quiet booth at Nottingham Trent University. All studies were run on a 17 inch MacBook Pro laptop computer using a Windows 7 professional operating system. Listeners were positioned approximately 60 cm from the centre of the computer screen.

EarLab was used to implement a dual-pair design (Rousseau & Ennis, 2001) in all studies. Two 200ms tones were presented in each interval with an ISI of 100ms and an inter-trial interval of 500ms. For either the pitch or timbre task, participants had to make a decision by indicating in which interval the two tones in a pair were different, according to the relevant dimension (pitch or timbre target). Responses were made by pressing button “1” or “2” accordingly. For both pitch and timbre tasks, listeners were required to make a judgement within intervals to decide in which interval the target dimension was the same or different. Any non-target within pair changes were considered as the source of interference on performance. In the standard version of the dual-pair design, there are two basic stimuli known as A and B, which can combined to yield eight possible stimulus sequences (e.g., AA-AB, AA-BA etc.; Micheyl & Messing, 2006). However in the current studies, a modified version of this paradigm was used where only two possible stimulus sequences were employed for simplicity. The interval in which the target appeared was always random but equally likely, but the target was always the second tone presented in any given pair (i.e., AB-AA, AA-AB).

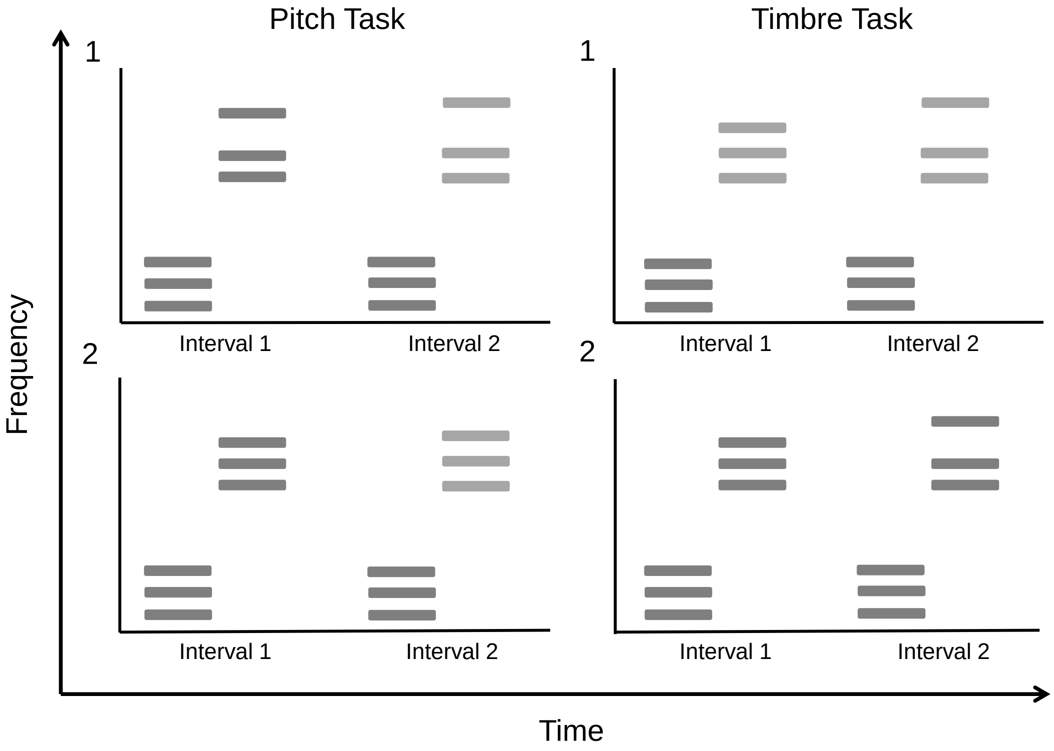

Figure 7.2. Schematic diagram showing the conditions used in Experiments 6-8 (and corresponding Experiments 9 and 11 in Chapters 8 and 9, respectively). Each subcomponent denotes an interval (or pair). Increased harmonic spacing and lighter coloured bars indicate a higher F0. Missing components indicate different timbre where lowest non-overlapping harmonic is 3 (i.e., harmonics 1, 2, and 3 or 1, 2, and 4). In this example the second pair was always the target interval and only information for four harmonics is shown. Top row denoted 1) Depicts ‘different pitch, different timbre’ conditions. Bottom row denoted 2) Depicts ‘different pitch, same timbre’ or ‘same pitch, different timbre’ conditions depending on the target dimension.

As shown in Figure 7.2, participants listened to two pairs of spectrally overlapping harmonic complex tones, with overlapping and non-overlapping harmonics. In condition (1), shown in the top row of Figure 7.2, there was only one interval (or pair) where the target differed, whilst the non-target dimension varied within each interval in the same manner ('different pitch, different timbre' conditions). In condition (2), shown in bottom row of Figure 7.2, there was only one interval where the target differed, whilst the non-target dimension stayed the same between and within intervals ('different pitch, same timbre' and 'same pitch, different timbre' conditions). Condition 1 was used in Experiments 6 and 7. Conditions 1 and 2 were used in Experiment 8.

In the pitch task, the question was 'In which interval was there a pitch difference?' In the timbre task, the question was 'In which interval was there a timbre difference?' Feedback was provided after each trial, where a green bar indicated a correct response and a red bar represented an incorrect response. For both pitch and timbre tasks, each condition was completed in blocks which made it easier for listeners to complete the task. To keep the task for both studies as similar as possible and not affect discriminability, each condition was completed in blocks but the order of which task (pitch or timbre) completed first was randomised to eliminate any order effects. Each block took approximately 4 minutes to complete. If applicable, participants were awarded psychological research credits based on their length of participation (i.e., awarded 1 credit per 10 minutes).

Experiments 6-8 used a mixture of adaptive and constant stimuli methods. The specific details of trial types for adaptive and method of constant stimuli designs varied as follows:

7.4.3.1 Adaptive method

The adaptive method design used a 2-down-1-up adaptive procedure targeted at 70.7% accuracy (Levitt, 1971). On the first trial the pitch or timbre difference was set as the nominal starting difference as indicated in Table 7.2. For every two correct responses, the pitch/timbre difference decreased for the subsequent trial, whilst for every incorrect response the pitch/timbre difference increased. The percentage difference increased or decreased by a factor of two for the first four reversals, and 1.414 for the final 12 reversals. In each block, trials continued until 16 reversals (ascending to descending and vice-versa) had been recorded. The pitch and timbre thresholds were taken as the geometric means of the frequency difference limens (F0 difference limen % and lowest non-overlapping harmonic, respectively) of the last 12 reversals. Unless stated otherwise, listeners completed three blocks for each pitch and timbre combination to obtain more reliable threshold estimates. The pitch/timbre threshold was taken as the average (%) of the three runs completed for a given pitch and timbre combination.

7.4.3.2 Method of constant stimuli

For the method of constant stimuli design, there were 49 trials in each block. The method of constant stimuli presents the same stimulus parameters on every trial. The order of completion of pitch and timbre tasks, were always randomised across participants to eliminate order effects. However the order of the respective conditions within a task was completed in a fixed order (see Training below). Unless otherwise stated, listeners completed three blocks for each pitch and timbre combination to obtain more reliable estimates of percentage correct responses for sensitivity estimates, and pitch/timbre discrimination performance was taken as the average (% correct) of the runs completed for a given pitch and timbre combination.

7.4.3.3 Training

Training was completed before and during the experimental session as required. Training was not only necessary to ensure that the salience of pitch and timbre were well equated, but neurophysiological evidence has also shown that auditory components, such as N100 and P200, are enhanced in musicians compared to non-musicians (Shahin, Roberts, Chau, Trainor, & Miller, 2008). We therefore wanted to ensure that all listeners were familiar with the stimuli, and equally good at discriminating pitch and timbre. It was also envisaged that this training would help to enhance ERP amplitudes across listeners. Listeners were given approximately 1-2 hours of training before their final results were recorded. Although Experiments 6 and 7 only employed the ‘different pitch, different timbre’ conditions (see Figure 7.2, top row), training for all experimental tasks always began on the easier conditions where only the relevant dimension was specifically manipulated and the non-target dimension stayed the same (e.g., pitch task: ‘different pitch, same timbre’ condition). This allowed listeners to become familiar with identifying the relevant cue. A specific procedure for training and equating pitch and timbre discrimination for the more difficult ‘different pitch, different timbre’ conditions was developed, particularly for more naïve listeners like #09 and #28. If discrimination difficulties on the timbre task arose, the timbre difference was decreased making the task more difficult. Specifically the lowest non-overlapping harmonic was changed to 7 for two blocks, then 5 for another 2 blocks, and then the original blocks were presented. The idea behind this technique was that by gradually increasing the timbre difference (i.e., 7 to 5 to 3), the timbre cue would be greater and thus more easily identifiable and participants could use this to aid them when the blocks became gradually easier again. Although this approach may seem counterintuitive, initial pilots showed that this method worked much better than when the tasks were made easier and then gradually made more difficult again. Therefore, based on similar logic, if difficulties in discriminating pitch arose, the F0 difference was increased (e.g., 20% for two blocks, 15% for another two blocks), until the original blocks were presented again.

7.4.4 Analysis

Accuracy data were analysed using signal detection theory (Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999; Wickens, 2002). Signal detection theory is typically employed when two possible stimulus types need to be discriminated, such as ‘same/different’ (signal and noise stimuli; Stanislaw & Todorov, 1999). Listeners are assumed to make decisions based upon information derived from two distributions, a ‘signal present’ distribution (i.e., presence of a difference) which represents an increase to background noise level caused by the introduction of a stimulus, and a ‘signal absent’ distribution (i.e., absence of a difference), representing a background level of noise. Signal detection theory uses d’ as a discriminability or sensitivity index, and relates to the standardised difference (z units) between the means of the internal ‘signal present’ and ‘signal absent’ distributions of evoked activity by the two stimuli that need to be discriminated. A correctly identified signal is termed as a ‘hit’, whilst misidentifying a noise signal as a *hit,* is a called a ‘false alarm’ (Wickens, 2002). D-prime quantifies a participant’s ability to discriminate the signal (hits; H) relative to the noise (false alarms; FA) for a given trial [e.g., d’ = z(FA) – z(H)]. Larger d’ values indicate greater sensitivity or ability to distinguish signals from noise (i.e., greater distance between the means for the signal present and noise distributions, and less overlap), whereas d’ values near zero indicate an inability to distinguish signals from the noise, with performance reflecting chance-level (i.e., distance between the means of the two distributions is much smaller).

d’ value is a common metric which corresponds to proportions of correct responses (PCs; see Micheyl & Messing, 2006). In the experiments reported here and throughout, d’ values relate to the number of times a participant identifies the correct target interval (H rate; selects the interval that contains a pitch/timbre change). For example, if a participant selected the correct interval for a pitch change on 90/100 trials, their PCs would be 0.90 (H rate, with a corresponding FA rate of 0.10). The PC value would then be used to calculate the corresponding d’ value for a given task.

d’ values were used instead of the raw PCs obtained for each condition. This is because PCs are susceptible to floor and ceiling effects, and are often not comparable with PCs measured in other experiments using different psychophysical paradigms (e.g., 2AFC, see Borchert et al., 2011). Furthermore, sensitivity and response bias typically confound most performance measures (e.g., PCs, hit rates, false alarms). However, d’ is considered to be unaffected by response bias as long as assumptions regarding normality and variance are met (see Stanislaw & Todorov, 1999). D-prime was therefore calculated for all performance data acquired in Experiments 6-8, (as well as for Experiment 9 in Chapter 8 and Experiment 11 in Chapter 9), because it provided a much more reliable indicator of listener’s sensitivity to pitch and timbre discriminations than percentage correct, and can be used to compare d’ from other studies, where appropriate (see Borchert et al., 2011).

The average PCs per condition was used to calculate *d'* separately for each listener and task using MATLAB. It is extremely important to select correct decision models (i.e., formulas for d') that are appropriate for the specific tasks employed, especially when different designs are used (see Micheyl & Oxenham, 2005). Values of d' corresponding to PCs can be measured in different ways. Strictly speaking, there is only one d' for a given experimental design, which is the one obtained using the optimal observer model. However, given that it was not known *a priori* what decision rule (criterion/strategy) listeners would use and whether it is optimal or not, the figures from both calculations are presented throughout following personal communication with Prof. Christophe Micheyl (personal communication, 16th November, 2012). It was anticipated that listeners would fall somewhere in between sub-optimal and optimal observers (. One relates computing d' for conventional dual-pair designs with 8 possible stimulus sequence options (see Micheyl & Messing, 2006). However, given that the current Experiments 6-8 (and 9 and 11, reported in Chapters 8 and 9, respectively) used a modified version the dual-pair design with only 2 possible sequences, the observer could have taken advantage of this knowledge. This is because the optimal observer uses a decision rule that maximises PC (see Micheyl & Dai, 2009). Consequently the distance between the two distributions will be 2d'/2, which is equivalent to the conventional d' calculation for a yes-no forced choice paradigm, described above (Stanislaw & Todorov, 1999). D' was therefore calculated for PCs between 0 and 1.0 using the standard formula for the yes-no paradigms. However d' corresponding to dual-pair designs was also computed, for PCs between 0.5 and 1.0, using a bespoke MATLAB code provided by Christophe Micheyl (see Micheyl & Messing, 2006). This is because it reflected a situation in which the observer does not take advantage of knowing the order of A and B in the different pair (ideal observer), which was referred to as 'suboptimal observer' because performance would be poorer.

Extreme hit rates and false alarm rates were dealt with using a standard correction procedure (Hautus, van Hout, & Lee, 2009; Stanislaw & Todorov, 1999). Any false alarm rates of 0 were changed to 1/(2*N*), and any hit rates of 1 were changed to 1-1/(2*N*), where *N* was the number of trials for a given condition (Macmillan & Kaplan, 1985). In rare cases, values of d' were negative (see Figure 7.3). Negative values of d' typically arise through sampling error or response confusion (Stanislaw & Todorov, 1999).

The dependent variable related to the d' value for each condition per listener. For a given condition, multiple runs (where applicable) were averaged together to obtain an average d' value for that condition (per pitch and timbre task) for each listener. The averaged d' values for each condition was then averaged over individuals to create a group average that could be used to create figures. Individuals' d' values for a given condition were also input into SPSS to run statistical analyses between pitch and timbre tasks.

Paired sample t-tests conducted on the d’ values for each listener were used to confirm stimulus parameters were equally matched between tasks (i.e., accept the null hypothesis), and were Bonferonni corrected to account for FWE associated with multiple comparisons. Unless stated otherwise, the d’ optimal observer model and suboptimal observer model always yielded qualitatively similar patterns of results and/or the same statistical results. Therefore, although the data for both are reported, only the d’ results and statistical analyses for the optimal observer model are described in detail.

7.5. Experiment 6: Determining potential stimulus parameters for pitch and timbre

7.5.1 Procedure and Design

Experiment 6 used the adaptive method and the method of constant stimuli in separate runs. This study undertook the majority of stimulus parameter combinations shown in Table 7.2. However different pitch and timbre combinations were implemented depending on the method of presentation.

The adaptive method was used in the first instance to obtain a general overview of the thresholds for each pitch and timbre discrimination tasks and likely pitch and timbre parameters to be tested. In the first step, the timbre task was implemented for the range of F0 starting differences (200 Hz +5-20%), but the lowest non-overlapping harmonic was always 2. This enabled exploration of the possible values for the lowest non-overlapping harmonic. In the second step, the pitch task was implemented to confirm that the pitch discrimination thresholds for all four F0 differences (%) at the possible timbre values identified in step 1; both smaller (7 as the lowest non-overlapping harmonic) and larger timbre differences (3 as the lowest non-overlapping harmonic).

The method of constant stimuli was used to systematically and rigorously validate the likely pitch and timbre parameters that had been identified using the adaptive method. The combinations that were implemented for both tasks included a starting F0 = 200 Hz, F0 differences = 5%, 10%, 15% and 20%, and lowest non-overlapping harmonics 3, 5 and 7 to create the timbre difference. Although 5 was not identified as a ‘likely’ timbre value during the adaptive method phase, it was used to ensure a more representative range of timbres differences were explored (i.e., large, medium, and small, respectively).

Listener #02 completed 7056 trials (144 blocks) for Experiment 6, which took approximately 13 hours in total to complete.

7.5.2 Results

7.5.2.1 Adaptive method

The adaptive thresholds from listener #02 for the timbre task indicated that at F0 difference of 5%, the 7th harmonic was the lowest discriminable non-overlapping harmonic. For larger F0 differences (10%, 15% and 20%) the lowest non-overlapping harmonic was approximately 3. These were subsequently used as the lowest non-overlapping values for the adaptive pitch task. For the pitch task, F0 difference limens at an F0 difference of 5% were relatively good at 1.24% with 7 as the lowest non-overlapping harmonic. At greater F0 differences (10%, 15% and 20%), with the lowest non-overlapping harmonic of 3, thresholds were slightly poorer (3.15%, 2.41% and 1.68%, respectively).

7.5.2.2 Method of constant stimuli

Figure 7.3 illustrates d' obtained for each stimulus and task combination. Overall listener #02 was generally better at doing the pitch task than the timbre task. This is evidenced by the greater number of black lines (pitch) with d' > 2, than the gray lines (timbre). It is likely that this was because relatively small timbre differences were created using the overlapping and non-overlapping harmonics, which made the timbre cue generally more difficult to discriminate.

For the timbre task, listener #02 was better at discriminating timbre when the timbre difference was larger (i.e., 3 as lowest non-overlapping harmonic), than when it was smaller (i.e., 5 and 7). This is again evidenced in Figure 7.3 where only the gray lines denoted by a square being d' > 2.

There was a general monotonic decrease in d' as function of increasing F0 difference for all of the lowest non-overlapping harmonic values (3, 5 and 7). For the pitch task, listener #02 was generally better at discriminating pitch when the timbre difference was smaller (i.e., 7 and 5) than when it was larger (i.e., 3). This is evidenced by the black lines denoting circles and triangles yielding the higher d'. However, the only instance when d' for the pitch task showed the expected monotonic increase (i.e., increase in d' as a function of increasing F0 difference), was when the lowest non-overlapping harmonic was 3. For 5 and 7, although majority of d' values were above the criterion of 2, they were non-monotonic

In Panel B of Figure 7.3 , the data point for timbre task at F0 difference 20% with a timbre difference generated by using the lowest overlapping harmonic of 5 is not shown because the percentage correct fell below the minimum d' value criterion the programme could calculate (i.e., 50%).

7.5.3 Conclusion

The red circles in Figure 7.3 denote examples where discriminability was matched across pitch and timbre parameters and d’ was greater than 2. These were for F0 differences at 10% and 15%, with a large timbre difference (i.e., 3 as lowest non-overlapping harmonic). This was true regardless of the d’ measurement model used (optimal or suboptimal). A paired sample t-test performed using d’ values from both optimal and suboptimal models (entering the data into SPSS as two participants as unable to run a t-test on only one participant), confirmed that d’ was equally matched for pitch and timbre tasks for both sets of parameters; $t(1) = -5.68$, p = .111, and $t(1) = -5.78$, p = .109, respectively for F0 differences 10% (pitch task: *M* = 2.17, *SD* = 0.08; timbre task: *M* = 2.46, *SD* = 0.01) and 15% (pitch task: *M* = 2.21, *SD* = 0.07; timbre task: *M* = 2.29, *SD* = 0.05). These two sets of parameters were therefore selected for further testing on two naïve listeners in Experiment 7.

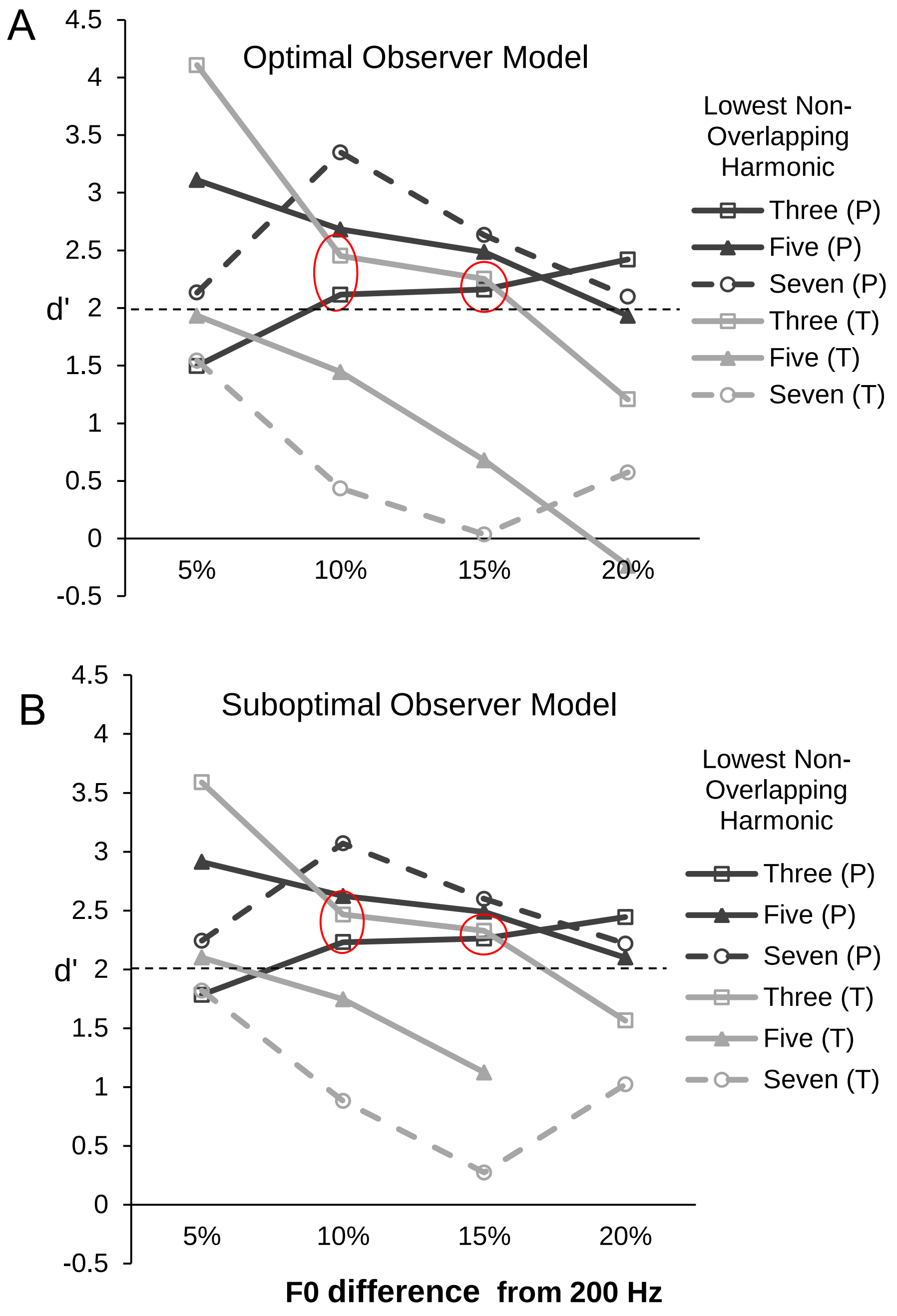

Figure 7.3. Panel A: d' based on the optimal observer model. Panel B: d' based on the suboptimal observer model. D' for the 'different pitch, different timbre' conditions for pitch and timbre tasks at different F0 and timbre combinations for listener #02. Line colour depicts the task type (black: pitch; gray: timbre). The line style and marker type depicts the timbre type and corresponding lowest non-overlapping harmonic. For the lowest non-overlapping harmonic, (P) denotes pitch task, and (T) denotes timbre task. Dotted black line depicts the d' prime criterion threshold of ≥2. These graphic conventions also apply to the other figures that follow. Data is based on one listener so no confidence intervals are shown.

7.6. Experiment 7: Re-assessing stimulus parameters for pitch and timbre in two naïve listeners

7.6.1 Procedure and Design

Experiment 7 used only the method of constant stimuli paradigm. Naïve listeners #09 and #28 were required to complete a subset of pitch and timbre parameter combinations for pitch and timbre tasks, with starting F0 = 200 Hz, F0 differences of 10% and 15%, and with 3 as the lowest non-overlapping harmonic. Listeners completed around 588 trials (12 blocks) each, which took approximately 3-4 hours per listener to complete (including 1-2 hours of training).

7.6.2 Results

After training, the results for individual listeners are shown in Figure 7.4. Both listeners could discriminate the chosen parameters for pitch and timbre tasks relatively well. A paired sample t-test was performed on the averaged d' values across listeners #09 and #28 (averaged data is not shown) for the optimal observer model. The t-tests indicated that, at a F0 difference of 10%, there was no statistically significant difference in d' between the pitch (*M* = 2.32, *SD* = 0.46) and timbre (*M* = 2.52, *SD* = 0.41) tasks, *t*(1) = -5.68, p = .111. For the F0 difference of 15%, there was also no statistically significant difference between the pitch (*M* = 2.73, *SD* = 0.12) and the timbre (*M* = 1.81, *SD* = 0.10) tasks, *t*(1) = 5.98, *p* = .105. However, in terms of seeking matched performance across pitch and timbre, discrimination performance for both listeners was more comparable for the F0 difference of 10% than the F0 difference of 15%. This is demonstrated by d' being greater than 2 on most conditions. This pattern was the same for d' based on the suboptimal observer model.

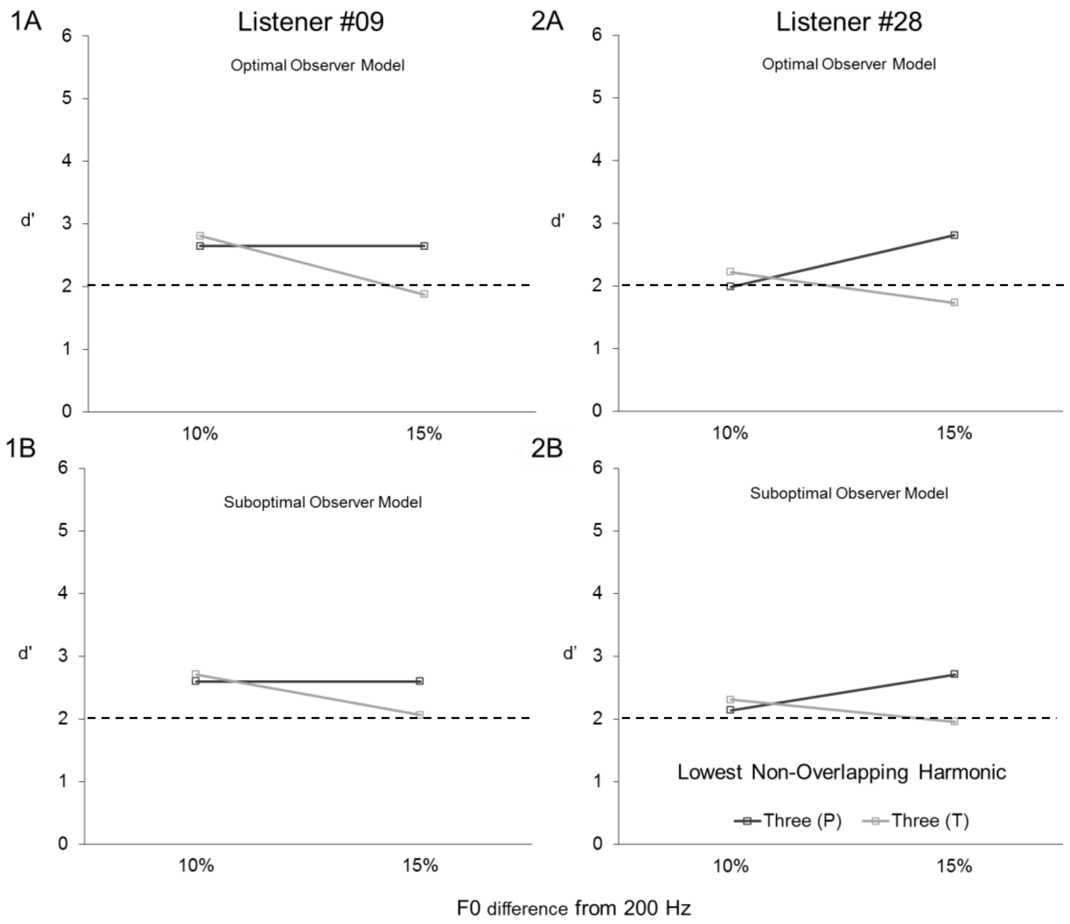

Figure 7.4. D-prime for listeners #09 and #28 for the 'different pitch, different timbre' conditions for pitch and timbre tasks with F0 differences from 200 Hz at 10% and 15% and a timbre difference created using 3 as the lowest non-overlapping harmonic. Panels 1A and 2A illustrate d' are based on the optimal observer model. Panels 1B and 2B illustrate d' are based on the suboptimal observer model. In the key for the lowest non-overlapping harmonic, (P) denotes pitch task, and (T) denotes timbre task. Data is based on individual listeners so no confidence intervals are shown.

7.6.3 Conclusion

Based on these findings an F0 difference of 10% at F0s 200 Hz and 220 Hz and a timbre difference created using 3 as the lowest non-overlapping harmonic were chosen as the most optimal stimulus parameters. These parameters were fully verified in Experiment 8.

7.7. Experiment 8: Verifying choice of stimulus parameters using a counterbalanced design

Listeners #02, #09 and #28 were by now experienced in performing pitch and timbre discrimination for dual-pair designs in which the conditions were always ‘different pitch, different timbre’ (see Figure 7.2, top row). Experiments 8, 9 and 11 consider performance when conditions could also include dual-pairs in which one of the perceptual dimensions remains the same across pairs (see Figure 7.2, bottom row).

7.7.1 Procedure and Design

Experiment 8 used the method of constant stimuli to assess discrimination performance. Listeners #02 and #09 completed the experiment. The starting F0 for each block was either 200 Hz or 220 Hz. Tasks were fully counterbalanced so that the difference in pitch not only went upwards (200 Hz to 220 Hz) but also went downward (220 Hz to 200 Hz (-9.09% F0 difference). This was because frequency change direction thresholds are known to vary across listeners (e.g., Semal and Demany, 2006). As identified previously, the lowest non-overlapping harmonic of 3 was used as to create the timbre difference between tones.

For both pitch and timbre discrimination tasks, trial types for each condition were completed twice, totalling an average of four blocks per condition (8 blocks for each task). Listeners were only permitted to continue to the next condition when performance on their current condition had reached an average accuracy of 85% across the given block runs. Listeners completed approximately 784 trials each for Experiment 8, which took approximately 1-3 hours per listener to complete (including 1-2 hours of training).

Table 7.3

The conditions for pitch and timbre discrimination tasks that were fully implemented in Experiment 8

Stimulus dimension and parameter		Discrimination task							
		Pitch In which interval was there a pitch difference?				Timbre In which interval was there a timbre difference?			
		Condition 1: Different pitch, different timbre		Condition 2: Different pitch, same timbre		Condition 1: Different pitch, different timbre		Condition 2: Same pitch, different timbre	
Pitch	Starting F0	200	220	200	220	200	220	200	220
	F0 difference (%)	10	-9.09	10	-9.09	10	-9.09	0	0
Timbre	Lowest overlapping harmonic	1	1	1	1	1	1	1	1
	Lowest non-overlapping harmonic	3	3	3	3	3	3	3	3

Note: ‘Different’ or ‘Same’ pitch or timbre relates to the stimulus dimension congruence between intervals 1 and 2.

7.7.2 Results

Figure 7.5 shows the average d’ obtained from listeners #02 and #09 for conditions with a target dimension change only (i.e., pitch task: ‘different pitch, same timbre’, timbre task: ‘same pitch, different’ timbre) and conditions with both target and non-target dimension changes (i.e., pitch and timbre tasks: ‘different pitch, different timbre’). As expected, listeners were extremely good and generally better (higher d’) at discriminating conditions when only the target dimension was changing between intervals and not the non-target dimension. When both relevant (target) and irrelevant (non-target) dimensions were changing between intervals, d’ was reduced but still exceeded 2.

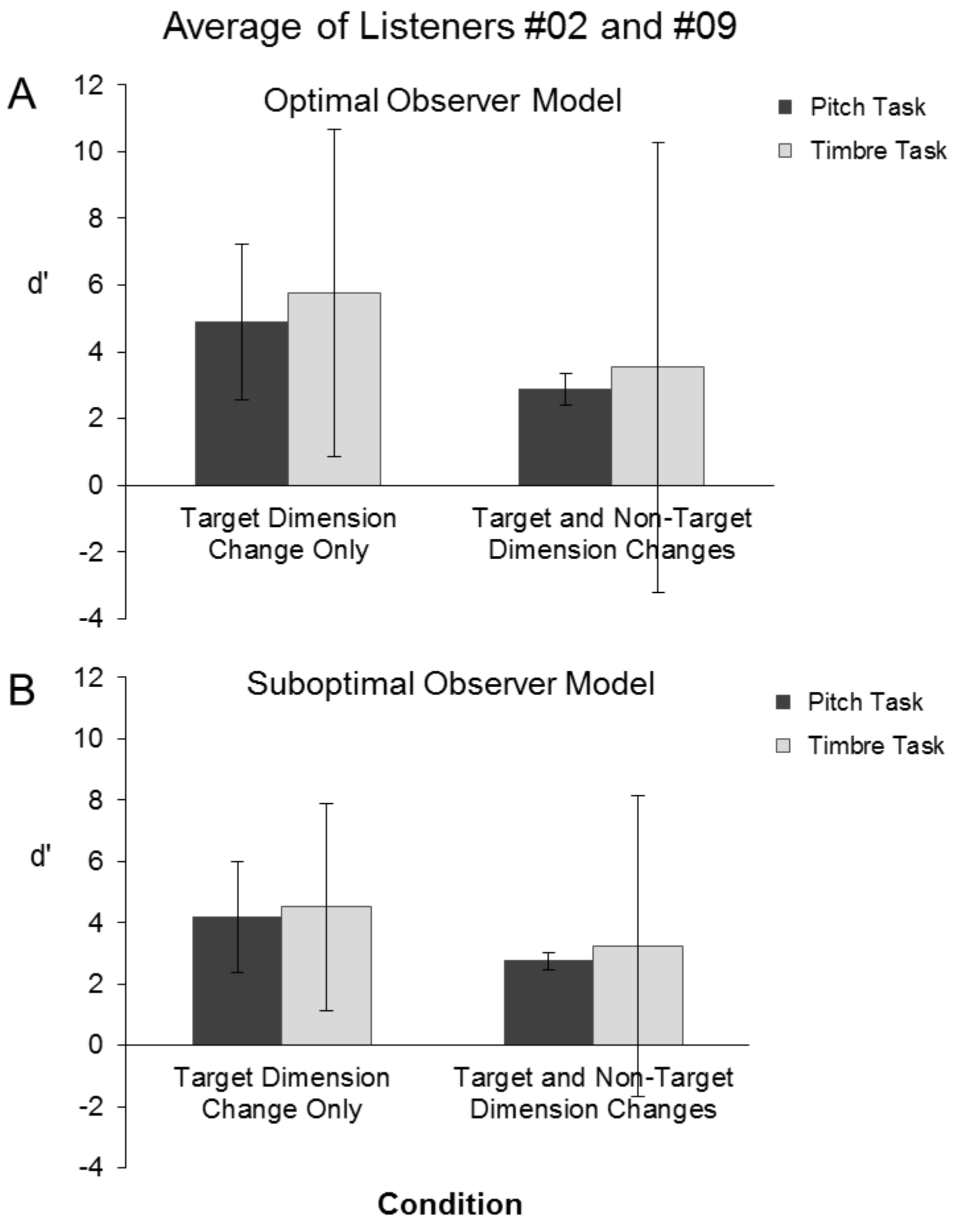

Figure 7.5. Averaged d' for listeners #02 and #09 for target dimension change only conditions and target and non-target dimension change conditions for pitch and timbre tasks. Panel A illustrates d' based on the optimal observer model. Panel B illustrates d' based on the suboptimal observer model. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

Statistical analysis of the d' data for the optimal observer model using paired sample t-tests revealed no statistically significant difference in discriminability between pitch (*M* = 2.87, *SD* = 0.74) and timbre (*M* = 3.53, *SD* = 1.59) tasks for the conditions in which target and non-target dimensions were changing (i.e., 'different pitch, different timbre' conditions; *t*(1) = -1.08, *p* = .474). There was also no statistically significant difference between pitch (*M* = 4.90, *SD* = 0.35) and timbre (*M* = 5.76, *SD* = 0.00) tasks for conditions in which only the target dimension changed i.e., pitch: ‘different pitch, same timbre’ condition, and timbre: ‘same pitch, different timbre’ conditions; *t*(1) = -3.42, p =.181). This pattern was the same for d’ based on the suboptimal observer model. Confidence intervals appear quite large because the data is based on only two listeners.

7.7.3 Conclusion

Overall Experiment 8 confirmed the use of the chosen parameters defined in Table 7.3.

7.8. Summary of Experiments 6-8

The overall aim of the Experiments 6-8 was to determine what stimulus parameters for pitch and timbre tasks were required to give equivalence in discrimination performance. The objective was to eliminate differences in discriminability (and thus perceptual salience) as a potential confound in ERP and reaction time experiments reported in Chapters 8 and 9, respectively. These sequential studies with a small number of listeners tested over many 1000’s of trials validated the choice of a 200 Hz tone with an F0 difference of 10% and a timbre difference created using 3 as the lowest non-overlapping harmonic. These parameters provide the most optimal matched discrimination performance across listeners, where d’ is consistently greater than 2.

Chapter 8. Searching for pitch invariant representations in auditory cortex using a novel ERP adaptation approach

8.1. Introduction

This chapter describes Experiments 9 and 10, where ERP Experiment 10 primarily investigated the second pitch criterion identifying a specialised modular pitch centre: *pitch constancy* (in addition to pitch criteria 1 (pitch selectivity) and 4 (accounting for confounding factors)). Pitch constancy predicts responses that are always sensitive to pitch, despite changes in other acoustic features. Of particular interest here is whether or not a neurophysiological measure of pitch coding, measured using a novel ERP adaptation approach, is influenced by timbre changes. Pitch constancy or invariance has typically been addressed in non-human animal studies (e.g., Bendor & Wang, 2005; Bizley et al., 2009), whose findings may or may not be comparable across species. Pitch constancy has also been studied in humans (often using subtractive methodology), but these have often yielded somewhat inconsistent results (e.g., Barker et al., 2011, 2013; Hall et al., 2009; Chait et al., 2006; Garcia et al., 2010; Puschmann et al., 2010, Penagos et al., 2004; Steinmann & Gutschalk, 2012). The advantages of adaptation designs to investigate stimulus-specific representations have previously been described (Chapters 3 and 6). The current ERP experiment describes one of the first attempts to directly apply this paradigm to investigate pitch invariance to timbre using the same well-matched pitch-evoking stimuli. The experimental approach combined behavioural and electrophysiological methods. Experiment 9 was a behavioural study which sought to equate psychophysical performance for pitch and timbre tasks within subjects, as well as assess between-subject differences in the discriminability between and within pitch and timbre tasks. Experiment 10 was the main study of interest presented in the current chapter which explored pitch constancy using an ERP adaptation design in which pitch and timbre were manipulated across conditions in the same way they had been for Experiment 9.

8.1.1 Neurophysiological evidence for pitch invariance in auditory cortex

In the last 15 years or so there has been an exponential increase in the amount of non-human and human neurophysiological investigations of pitch constancy on a more general level, and more recently, regarding pitch invariance to other perceptual features. Findings from these studies, not only help gain a better understanding of the psychophysical evidence regarding pitch perception and discrimination, but also allows for comparisons between non-human animals and humans in order to build a cumulative picture regarding the neural representation of pitch perception in auditory cortex. This relates to questions surrounding where in auditory cortex pitch perception is encoded, and also the neural mechanisms underlying pitch abstraction. However, both non-human and human neurophysiological evidence regarding pitch representations, and the existence of pitch constancy at the level of auditory cortex remain unresolved (Griffiths, 2012).

8.1.1.1 Non-human studies

Support for and against pitch-constant representations in auditory cortex has been reported in both non-human and human neurophysiological studies. In favour of pitch constancy, a number of non-human studies (predominately in primates) have argued that pitch coding involves dedicated “pitch neurons” that code pitch invariantly of the spectral composition of the stimulus (e.g., Bendor & Wang, 2005, 2010; Fishman, Reser, Arezzo, & Steinschneider, 1998; Schwarz & Tomlinson, 1990; Steinschneider, Reser, Fishman, Schroeder, & Arezzo, 1998). For example, Bendor & Wang (2010) measured single unit responses in marmoset auditory cortex using a variety of pitch-evoking stimuli (e.g., harmonic complex tones with missing F0). They reported pitch selective neuronal populations in primary (A1) and non-primary (R) areas that responded to the pitch value (i.e., F0) regardless of differing physical (i.e., spectrum) and perceptual (i.e., salience) characteristics of the different pitch-evoking stimuli. Interesting, they also found an increase in firing rate of neurons as a function of pitch salience (temporal regularity) within this putative pitch centre (but see Schnupp & Bizley, 2010 for a critique of this study).

Against pitch constancy, more recent animal studies using single-unit recording have reported neuronal sensitivity to pitch and other acoustic features (e.g., Bizley et al., 2009, 2010; Walker et al., 2011a; see Schnupp & Bizley, 2010 for a critique of the idea of invariant pitch neurons). Bizley et al. (2009) investigated the neural representation of pitch, spectral timbre and sound localisation cues in ferret auditory cortex. They presented ferrets with artificial vowel sounds that differed in pitch (F0), timbre (formant frequency) and sound source azimuth (virtual sound source directions), and recorded cortical responses (single units) of individual neurons in five areas encompassing primary and nonprimary areas of ferret auditory cortex. To analyse the data they used a special ‘variance decomposition’ technique, which allowed them to quantify how altering one parameter affected neural responses. They found that most neurons in multiple cortical areas (primary and non-primary) were sensitive to more than one acoustic feature. They found no evidence of a specialised cortical field for pitch, timbre or spatial location. Nevertheless they did report that neurons in some areas were more sensitive to one feature than to others features. For example, primary areas such as A1 were more sensitive to pitch, whilst non-primary areas like the anterior auditory field were more sensitive to timbre. Walker et al. (2011a) also found that the same neurons in ferret auditory cortex could effectively multiplex information about pitch, timbre and spatial location acoustic features within separate temporal windows of their response. Specifically, timbre change information was represented earlier in auditory cortical responses than pitch change information. These findings suggest that neurons responsible for encoding pitch may be non-invariant to timbre changes, and so a given pitch value might need to be encoded by multiple neurons or slightly different neuronal populations depending on the timbre and spatial location of the particular stimulus, rather than by a specialised region of highly-tuned pitch-exclusive neurons or neuronal populations.

8.1.1.2 Human studies

Neurophysiological evidence for pitch invariance to timbre in humans has been limited by the fact that the majority of studies use subtractive methodology. Unlike studies that employ alternative approaches (e.g., single-cell recording in primates, or adaptation designs), subtractive studies cannot effectively examine responses for both stimulus features (see Chapter 3). Subsequently, the majority of human evidence to date is somewhat indirect, and involves comparing across studies that report either pitch and/or timbre patterns and simply inferring whether or not they engage the same brain regions. Overall, evidence is largely against the idea of pitch constancy. For example, a number of fMRI studies have shown that pitch and timbre have overlapping and widely distributed neural substrates, notably in HG, and PT, as well as STG and Superior Temporal Sulcus (STS) areas of auditory cortex (Formisano, De Martino, Bonte, & Goebel, 2008; Warren et al., 2005; Overath et al., 2008; Overath et al., 2010).

Furthermore, a recent ERP study that has employed adaptation methodology has also shown some evidence that pitch is non-invariant to timbre. Specifically, as described previously in Section 6.3, Briley et al. (2013) investigated whether the pitch is represented by its physical dimension (i.e., repetition rate, high or low), or by its perceptual dimensions (i.e., pitch height and/or pitch chroma). Pitch height refers to the octave in which a given note is situated, whilst pitch chroma represents the cycle of notes within a given octave (Briley et al., 2013). The authors compared adaptation effects for pure tones and IRN. Under the assumption of pitch constancy, they should have exhibited a similar pattern of results for both types of stimuli, despite variations in their ‘timbre’. However Briley and colleagues found evidence for a representation of pitch chroma in auditory cortex for IRN, but not for pure tones. This suggests that the two different pitch stimuli activate different neurons in auditory cortex. This therefore did not support claims that there are invariant pitch neurons in auditory cortex. They claim their results support Butler (1972) who found that a pure tone does not adapt a complex tone that was the same pitch but had a non-overlapping spectral composition. However, Briley et al. (2013) examined evidence for pitch constant effects by comparing adaptation effects separately for two different classes of stimuli whereby differences in stimuli could extend beyond just the spectral composition of sounds, rather than comparing whether pitch is invariant to timbre in the same stimuli.

8.1.2 Objectives and hypotheses

Overall, the amount of evidence for and against pitch constancy in auditory cortex is relatively mixed. It is still unclear whether there is evidence for pitch constancy at the level of auditory cortex. The current study therefore aims to address this by examining pitch invariance to timbre by comparing adaptation effects to changes in the same pitch-evoking stimuli (Experiment 10), using stimuli that are well-controlled for inferred perceptual salience (Experiment 9). The main objectives were as follows:

Objective 1) whether adaptation of auditory cortical probe responses are selective to pitch? To address pitch criterion 1 (pitch selectivity; see Section 1.1) there should be no significant differences in amplitude across conditions for noise (EOR) and the adaptor (POR1), but there should be significant differences in amplitude for the probe (POR2) across conditions compared to the adaptor (POR1), specifically most adaptation should occur when the same pitch stimulus is repeated (see Figure 8.1). It was expected that neither the EOR nor the POR1 would evoke statistically significant differences across conditions

Objective 2) if pitch-selective adaptation is observed then, are pitch processing neuronal populations selective to encoding pitch only, or are they sensitive to encoding both pitch and timbre (see Figure 8.1)

If pitch is invariant to timbre (see Figure 8.1A), there should be release from adaptation (less adaptation) for pitch changes (different pitch). A 2-by-2 repeated-measures ANOVA with pitch (same, different) and timbre (same, different), should reveal a significant main effect of pitch, but no significant main effect of timbre and no interaction. This would be consistent with a pitch constant or modular ‘pitch specific’ representation in auditory cortex (addressing pitch criterion 2, pitch constancy; see Section 1.2).

If pitch is non-invariant to timbre (see Figure 8.1B), there should be a release from adaptation for both pitch and timbral changes. A 2-by-2 repeated measures ANOVA with pitch (same, different) and timbre (same, different), should reveal a significant main effect of pitch, a significant main effect of timbre, and a possible pitch and timbre interaction. This would be counter to the idea of a pitch-specific representation in auditory cortex.

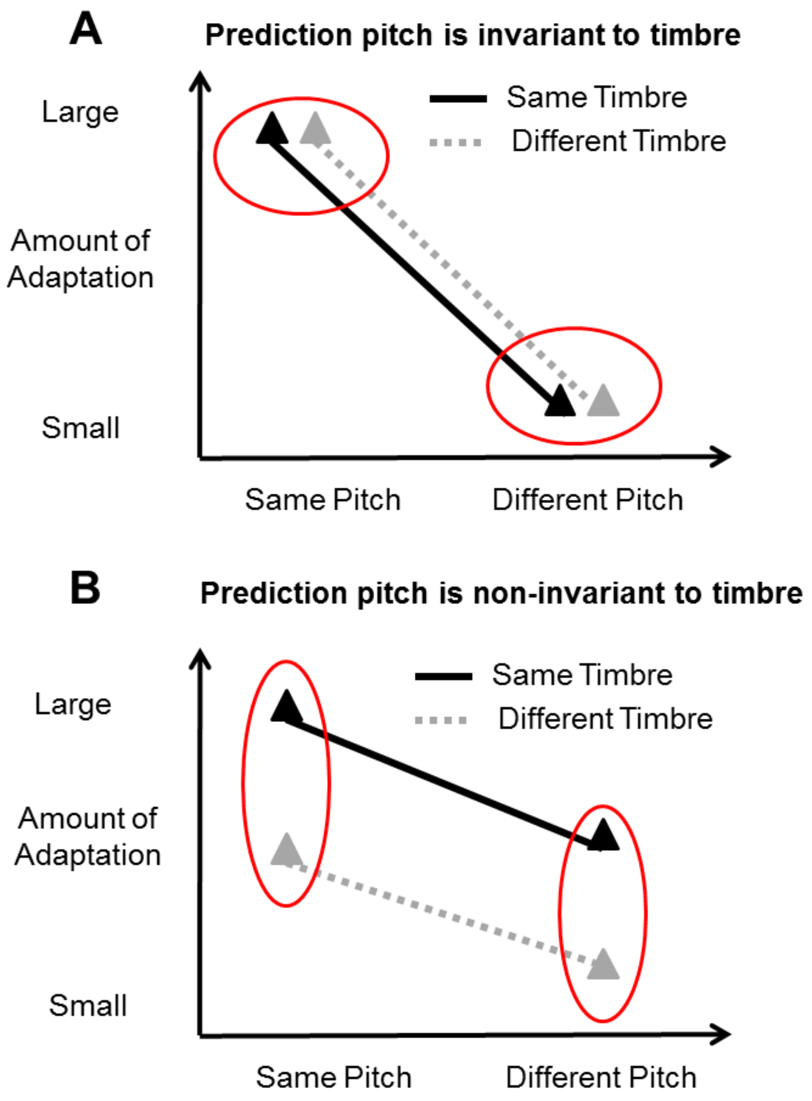

Figure 8.1. Schematic illustration of experimental hypotheses relating to pitch invariance to timbre. See text for details.

The fourth pitch criterion (‘accounting for confounding’ variables; see Section 1.4) was addressed by using stimuli that activated the same gross spectral region, helping to maintain a stable envelope and prevent cochlea distortions (see Section 8.3.2.1).

8.1. General methods and materials

The following sections describe the general methodology and materials for Experiments 9 and 10. Experiment-specific details are presented in their corresponding sections.

8.1.1 Listeners

Twenty right-handed (LQ = +89%, Decile R.7) healthy listeners (#01, #02, #05, #06, #07, #08, #09, #10, #14, #15, #16, #29-#37) volunteered to participate in both the psychophysical (Experiment 9) and ERP (Experiment 11) studies (10 male, 10 female; M_{age} = 23.35, *SD* = 6.03, age-range 18-42 years). All participants reported normal or corrected-to-normal vision, and had clinically normal hearing (≤25 dB HL between audiometric frequencies 250-8000 Hz). No listeners reported a history of hearing, neurological, and/or psychological impairment, and use of psychiatric medication or substance misuse. Seven participants (#05, #07, #08, #14, #16, #31, and #37) were musically trained on the clarinet, drums, guitar, keyboard, piano, and recorder achieving grades 2-7 across 2-20 years. Six listeners (three of whom were not musically trained; #02, #05, #09, #14, #16, #35) also reported informal musical experience on the bass guitar, drums, guitar, keyboard, piano and violin, gained over 2-6 years. Eight listeners (not reported above; #38-#45) did not successfully complete Experiment 9. One participant (#38) withdrew before completion. Seven participants (#39-#45) did not achieve appropriately matched pitch and timbre discrimination performance during the training phase and so were excluded. After inspecting individual ERP waveforms, two subjects (#29 and #36) who participated in both studies were then excluded from Experiment 10 during the ERP analysis stage. They both failed to exhibit typically large and distinguishable cortical responses to noise (EOR; see Section 6.2.3.1). Listeners gave written informed consent, and the study was approved and performed in accordance with the College Research Ethics Committee's guidelines, Nottingham Trent University (ethics no. 2011/46).

8.1.1 Procedure and Design

Participants' completion of Experiment 9 always preceded Experiment 10, and generally on a different day (i.e., between 2 weeks depending on participant availability). Both experiments took place at Nottingham Trent University. All participants were recruited via Nottingham Trent University's online Research Participation Scheme or via email. Participants were awarded psychological research credits based on their length of participation (i.e., awarded 1 credit per 10 minutes of participation, where appropriate).

8.2. Experiment 9: Ensuring pitch and timbre discriminability was matched for ERP Experiment 10

8.2.1 Objective

Experiment 9 was employed to ensure that the salience for pitch and timbre had been equated across listeners in Experiment 10.

8.2.2 Methods

8.2.2.1 Stimuli

The stimuli and trial types in Experiment 9 were the same as those in Experiment 8 (Section 7.7). Also see Section 7.4.2 for details regarding how these were generated and calibrated.

8.2.2.2 Procedure and Design

Experiment 9 followed the same procedure and design as described in Experiment 8 (see Sections 7.7.1 and Section 7.4.3 for more information) and lasted between 2-4 hours depending on individual training needs. The procedure for training and equating pitch and timbre discrimination is described in Section 7.4.3.3. Some participants were invited to complete more than one pitch and timbre discrimination training session before they were assessed the eligibility to complete Experiment 10. It should be noted that participants could be trained only to discriminate pitch and/or timbre for conditions in which the pitch and/or timbre was ‘different’. Therefore there was no training for the ‘same pitch, same timbre’ condition.

8.2.3 Data Analysis

Data analysis for Experiment 9 followed the same procedure as described in Section 7.4.4.

8.2.4 Results

Average d’ data for all 20 listeners (#01, #02, #05-#10, #14, #15, #16, #29-#37) are plotted in Figure 8.2.

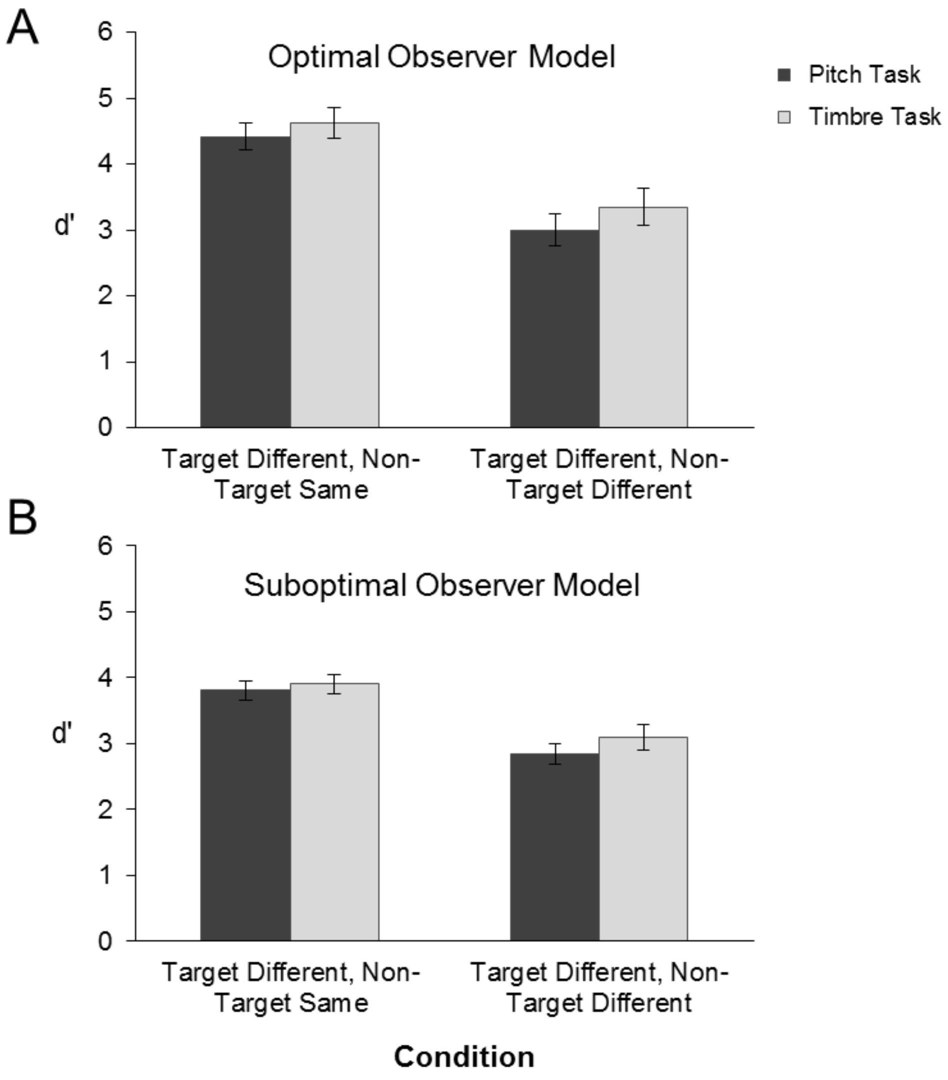

Figure 8.2. Averaged d' across the 20 listeners for Experiment 9 for pitch and timbre tasks for conditions in which the target dimension was different but the non-target remained the same (target different, non-target same), and for conditions in which both the target and non-target were different (target different, non-target different). Panel A illustrates d' based on the optimal observer model. Panel B illustrates d' based on the suboptimal observer model. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

Paired sample t-tests conducted on the d' data for the optimal observer model revealed no significant difference in discrimination performance between pitch and timbre tasks for both 'target different, non-target same' conditions (pitch: $M = 4.42$, $SD = 0.63$; timbre: $M = 4.62$, $SD = 0.82$), $t(19) = -1.07$, $p = .296$, and 'target different and non-target different' conditions (pitch: $M = 3.00$, $SD = 0.63$; timbre: $M = 3.35$, $SD = 0.77$), $t(19) = -1.43$, $p = .169$, indicating that the discriminability of pitch and timbre was appropriately matched across listeners (see Figure 8.2).

Results from the statistical paired sample t-tests indicated that the non-target dimension negatively affected discriminability of the target dimension for both pitch and timbre tasks. As expected, within tasks, listeners were significantly worse at discriminating the target when the non-target dimension also varied from interval to interval (i.e., the ‘target different and non-target dimension different’ condition) for both the pitch tasks (non-target same: *M* = 4.42, *SD* = 0.63; non-target different: *M* = 3.00, *SD* = 0.63), *t*(19) = 7.62, *p* < .001] and timbre tasks non-target same: *M* = 4.62, *SD* = 0.82; non-target different: *M* = 3.35, *SD* = 0.77), *t*(19) = 5.24, *p* < .001, compared to when only the target dimension changed and the non-target dimension remained fixed (i.e., ‘target different, non-target same change’ for pitch and timbre tasks). These results seem to suggest a detrimental effect of the non-target dimension on discriminating the target (signal), and one that is reciprocal for both dimensions (i.e., pitch affects timbre discrimination and vice versa).

Paired sample t-tests conducted on d’ data for the suboptimal observer model confirmed the same qualitative and statistical pattern of results as the optimal model.

8.2.5 Conclusions

Overall listeners who took part in Experiment 9, showed that discriminability for pitch and timbre was well-matched for the stimulus parameters selected. The significant difference within pitch and timbre tasks was expected given the nature of interference timbre has on pitch (and vice versa), supporting previous results (Borchert et al., 2011), and studies who claim an interdependence between pitch and other features related to its perception (Bizley et al., 2009). The group averaged d’ data reported here was better than the expected or that reported for Experiment 8 (see Section 7.7.2), however this is likely to have occurred because some listeners were given more practice trials over a maximum of two sessions (2-4 hours) until they reached criterion, or were excluded.

8.3. Experiment 10: ERP adaptation study

8.3.1 Objectives

To recap the amount of stimulus-specific adaptation refers to the size of the reduction in neural responsiveness to a probe (POR2) when it is preceded by the same or similar adapter (POR1). The extent of this reduction reflects the extent to which the adapter and probe activate similar groups of neurons (i.e., assumes large amount of adaptation = same/similar neuronal populations coding probe, and assumes small amount of adaptation = different neuronal populations coding probe). The current ERP experiment used this adaptation approach to address pitch criterion 1 (pitch selectivity), criterion 2 (pitch constancy) and criterion 4 (accounting for confounding variables).

8.3.2 Methods

8.3.2.1 Stimuli

The ERP stimuli were chosen based on the most optimal stimulus parameters confirmed in Experiments 6-8 (see Sections 7.5-7.8) and using the same stimulus parameters as Experiment 9.

The ERP stimuli were digitally synthesised using a bespoke MATLAB code (The Mathworks, Natick, MA), with a DVD quality sampling rate of 48 kHz and a 16-bit resolution. The resulting pitch and noise sound sequences evoked continuous stimulation within a given trial (e.g., Krumbholz et al., 2003; see Chapter 6). A representative example is given in Figure 8.3. Specifically, each sound sequence transitioned from a silent period (100ms), to white noise stimulus (800ms), a long harmonic adaptor stimulus (1000ms), a short harmonic probe stimulus (500ms), and silence (100ms), with a total trial stimulus duration of 2520ms. Each element in the sequence was butted together with no ramps, and then the whole sequence was low-pass filtered at 2 kHz, meaning that a noise masker was not needed because all stimuli were matched to activate the same gross spectral region, helping to maintain a stable envelope and prevent cochlea distortions (accounting for potential confounding factors, see Chapter 1). The sound sequence was gated to add two 10ms cosine ramps at beginning and the end of sound stimuli (noise and probe segments). The sound level for each segment was originally fixed at 50 dB SPL and the energy across the whole sequence was matched. The long adaptor, short probe, and no ISI between adaptor and probe were chosen in order to maximise auditory adaptation effects as found by Lanting et al. (2013; see Section 6.3 for more details). The noise and adaptor stimuli were a different duration (i.e., 800ms, and 1000ms, respectively), but this was not envisaged to affect the response measures. In Krumbholz et al.’s (2003) continuous stimulation MEG study exploring PORs, sequences contained a 2000ms noise which than transitioned to a 1000ms IRN pitch stimulus. Sequences were gated with periods of silence and onset/offset ramps to prevent audible clicks and truncation.

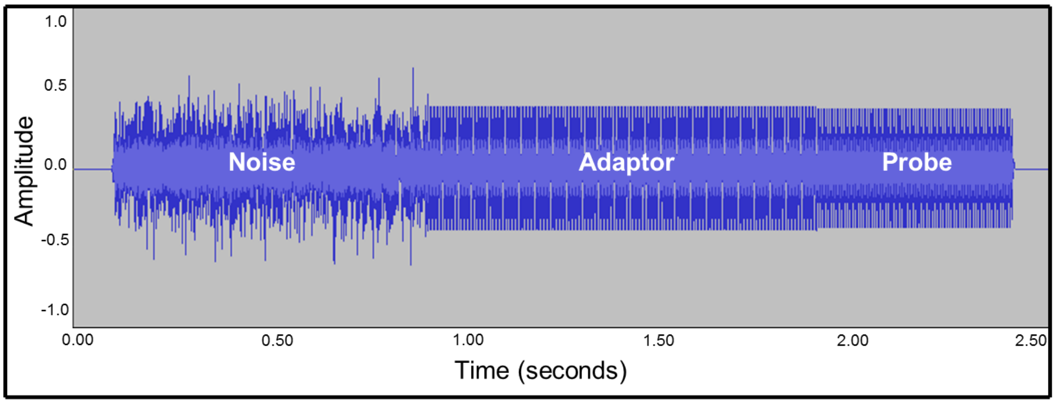

Figure 8.3. A representative example of one of the stimulus sequences used for the ERP Experiment 10. Sound transitioned from noise to pitch-evoking adaptor and then a pitch-evoking probe. Please note that the adaptor and probe pairing shown here have a different F0 and timbre (i.e., 'different pitch, different timbre' condition).

The stimuli and trial types for Experiment 10 are shown in Table 8.1. Four conditions were created by crossing two levels of pitch congruence (same, different) with two levels of timbre congruence (same, different) in a factorial design. Pitch and timbre congruence referred to the pairing of stimulus values between the two pitch evoking stimuli, adaptor and probe. Within each condition, the adaptor-probe pairings yielded four trial types, with a total of 16 trial types across conditions. Twenty-five versions of each trial type (25 x 16) were generated afresh to ensure a new random white noise sample for each trial, equating to 100 trials per condition, and totalling 400 trials. Many trials improve SNR by reducing the residual EEG noise (Kropotov, 2009, Part II, p. 347)

Table 8.1
Conditions used for the ERP Experiment 10

	Timbre congruence (adaptor -> probe)	
Pitch congruence (adaptor -> probe)	Same timbre	Different timbre
	Condition 1	*Condition 2*
Same pitch	200-1 -> 200-1 200-2 -> 200-2 220-1 -> 220-1 220-2 -> 220-2	200-1 -> 200-2 200-2 -> 200-1 220-1 -> 220-2 220-2 -> 220-1
	Condition 3	*Condition 4*
Different pitch	200-1 -> 220-1 200-2 -> 220-2 220-1 -> 200-1 220-2 -> 200-2	200-1 -> 220-2 220-2 -> 200-1 200-2 -> 220-1 220-1 -> 200-2

Note: Timbre congruence indicated by suffixes '1' or '2' indicates the two different types of timbre that were used. '1' refers to the timbre type created using 3 as the lowest non-overlapping harmonics, whilst '2' refers to the timbre type using 4 as the lowest non-overlapping harmonic (see Section 7.4.2).

8.3.2.2 Procedure and Design

The ERP Experiment 10 used a continuous stimulation paradigm to avoid further potential confounds when measuring adaptation effects (see Section 2.4.2). To date, the continuous stimulation paradigm is only typically employed in subtraction studies and therefore does not resolve the problem related to the requirement of an appropriately ‘matched’ baseline (Friston et al., 1996b; Sartori & Umiltà, 2000). However the current experiment resolves the baseline problem because an adaptation design was used (where the feature of interest is present in both the adaptor and probe stimulus) in conjunction with a continuous stimulation paradigm within the same pitch-evoking stimuli. Subsequently, the sound sequences consisted of noise, the adaptor pitch stimulus, and the probe pitch stimulus. As shown in Table 8.1, from adaptor to probe the same stimulus could be repeated (condition 1), or have a different pitch (condition 3), timbre (condition 2) or both (condition 4).

The ERP listening task was conducted in a quiet EEG lab at Nottingham Trent University. Presentation and timing of stimulus trials were controlled using an open source experimental building programme called OpenSesame 0.26 (Mathôt, Schreij, & Theeuwes, 2012) on a 17 inch stimulus presentation personal computer. The 400 trials were presented in a randomised order, and each trial was separated by a 1500ms silent inter-trial interval that randomly varied by 10% (i.e., ± 150ms) to avoid cross-over adaptation effects (Dale & Buckner, 1997; Vidyasagar, Stancak, & Parkes, 2010) and predictability between trials (Cohen & Eretz, 1991). Digital triggers were placed at the beginning of each segment in a given trail to synchronise stimulus presentation with recorded ERP activity.

The sound presentation level in OpenSesame was fixed to ensure stimuli presented binaurally using Sennheiser HD-280 headphones had an overall sound level of 70 dB SPL. Sound level was measured prior to testing using a Brüel & Kjær 4231 Sound Calibrator, affixed with a Brüel & Kjær 2250 Sound Level Meter, Brüel & Kjær 4153 artificial ear, and Brüel & Kjær 4192 Half Inch Microphone. Sound pressure levels were calibrated using the “LAEQuilize” setting which gives a measure of the overall dB SPL over time (i.e., overall dB SPL for the whole trial). Experiment 10 was generally completed by participants on a separate day to Experiment 9, but usually within a maximum of 14 days, and lasted between 2-3 hours in total for each participant (including lab set-up and clean-up times).

8.3.3 EEG data acquisition

ERPs were recorded using a 64-channel Active-Two acquisition system (BioSemi; www.biosemi.com), using the same acquisition procedures as described in Section 6.2. ERP data pre-processing and analysis for Experiment 10 was conducted using BESA Software (BESA Research v. 5.3.7, Gräfelfing), and involved three main stages: 1) Pre-processing, 2) Sensor-based analysis, and 3) Source-based analysis, as described in Section 6.2.

8.3.4 Data pre-processing

Data pre-processing followed the procedure described in Section 6.2.2.

8.3.5 ERP analysis

Experiment 10 used the two sensor-based and source-based analysis approaches described in Sections 6.2.3 and 6.2.4, respectively.

In order to address Objective 1, whether auditory cortical probe responses are selective to pitch, the amplitudes for the noise (EOR) and adaptor (POR1) responses needed to be quantified across conditions.

In order to address Objective 2, whether pitch processing neuronal populations are constant for encoding pitch only, or are sensitive to both pitch and timbre, the difference in amplitude between the response to the adaptor (POR1) and response to the probe (POR2) had to be quantified across conditions, to reveal the amount of adaptation or release from adaptation. The amount of adaptation was measured in microvolts for sensor-based analyses, and nAm for dipole source analysis.

8.3.5.1 Statistical analysis

The peak-to-peak amplitudes for EOR, POR1 and amount of adaptation for POR2 relative to POR1 (for both sensor and source-based analyses) were subjected to statistical analysis using a 2-by-2 repeated-measures ANOVA. The first factor reflected pitch congruence from adaptor to probe (same, different), and the second factor reflected timbre congruence (same, different). For all ANOVA testing an alpha criterion of $p < .05$ was used to test for statistical significance. All planned and post hoc paired sample t-tests were Bonferonni corrected to account for FWE associated with multiple comparisons. If Mauchly's test of sphericity assumption of sphericity was violated a Greenhouse-Geisser correction was used.

8.4. Results

8.4.1 Objective 1: ERP evidence for pitch selectivity

8.4.1.1 Sensor-based

GA ERP waveforms and latencies

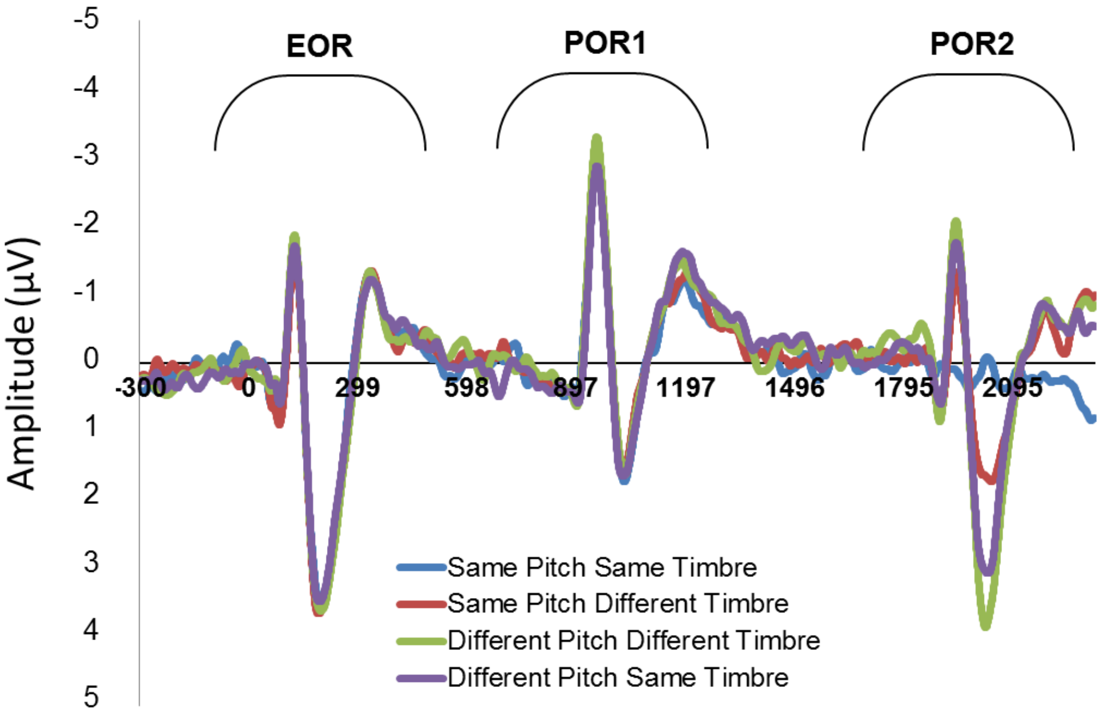

Figure 8.4. GA ERP waveforms measured at the Cz for each condition. Conditions were comprised of 'same pitch, same timbre', 'same pitch, different timbre', 'different pitch, different timbre' and 'different pitch, same timbre'. Note that the horizontal x axis represents time in ms.

The corresponding GA ERP waveforms at the Cz site, across the four sound conditions, are shown in Figure 8.4. Consistent with previous data (Seither-Preisler et al., 2004), the EOR was comprised of two large vertex deflections (N100: Näätänen & Picton, 1987; and P200 Lütkenhöner & Steinsträter, 1998) following the onset of noise. This was followed by two large transitional PORs; POR1 following the onset of the adaptor stimulus (at around 110ms), followed by POR2 (peaking at around 110ms) after the onset of the probe. Peak latencies for EOR, POR1 and POR2 components are shown in Table 8.2. As expected from visual inspection of the GA data, the EORs were not significantly different across conditions for the sensor-based data. Peak-to-peak EORs (µV) from N100-P200 did not reveal any significant main effects for pitch congruence, *F*(1, 17) = 0.38, *p* = .543, timbre congruence, *F*(1, 17) = 0.58, *p* = .458, and no pitch congruence*timbre congruence interaction, *F*(1, 17) = 0.76, *p* = .394, respectively. This was also the case for POR1 adaptor responses observed for the sensor-based data, whereby peak-to-peak POR1s (µV) from N100-P200 revealed no main effects of pitch congruence, *F*(1, 17) = 0.70, *p* = .415, timbre congruence *F*(1, 17) = 0.02, *p* = .900, and no pitch congruence*timbre congruence interaction, *F*(1, 17) = 2.32, *p* = .146, respectively.

ERP peak latencies for N100 and P200 components for the EOR, POR1 adaptor and POR2 probe were obtained. All peak latencies were in accordance with the findings of previous studies (e.g., Bosnyak et al., 2004; Seither-Preisler et al., 2004). Previous studies have found that the POR typically occurs around 150ms, which is markedly later than the EOR (N100; e.g., Chait et al., 2006; Krumbholz et al., 2003). The results reported here also derived PORs which share extremely similar temporal properties. Examining the temporal window of processing for new and repeated stimuli, the facilitation model (see Section 6.3) suggests that the latency and/or duration of the response to repeated items will be shorter for the second presentation than for the first presentation (see Grill-Spector et al., 2006). As shown in Table 8.2, both ‘N100’ and ‘P200’ latencies for POR2 responses were significantly shorter (i.e., 128.14ms, and 201.12ms, respectively) than POR1 (i.e., 143.18ms, and 217.12ms, respectively); *t*(17) = -7.47, *p* < .001, and *t*(17) = -4.25, *p* = .001, respectively. Considering the POR and EOR latencies demonstrated in the current study and the nature of the continuous stimulation paradigm used, I can be confident that the observed cortical responses are attributable to the neural mechanisms underlying pitch and timbre processing specifically, as opposed to sound coding in general.

Table 8.2

Average peak latencies in ms (± SD) derived from N100 and P200 components of interest comprising EOR, POR1 and POR2 at Cz sensor site, collapsed across conditions and listeners

ERP Component	Mean latency in ms (± SD)
EOR (N100)	122.71 (± 10.34)
POR1 (N100)	143.18 (± 5.68)
POR2 (N100)	128.14 (± 9.20)
EOR (P200)	193.96 (± 15.63)
POR1 (P200)	217.12 (± 9.34)
POR2 (P200)	201.12 (± 13.35)

There was no significant difference between the N100 and P200 latencies respectively for the POR2 (probe response) for pitch change only (i.e., different pitch same timbre) and timbre change only conditions (i.e., different timbre, same pitch); *t*(17) = -.57, *p* = .578, and *t*(17) = -.22, *p* = .828. This was contrary to Walker et al. (2011a) observed findings in ferrets that timbre changes are processed earlier than pitch changes. However a possible explanation for this could relate to fact that the current experiment matched pitch and timbre discriminability. This is explored further in Chapter 9.

Visual inspection of the data (see Figure 8.4) revealed there were differences between conditions for the POR2 response. When there was no change in pitch and timbre ('same pitch, same timbre' condition) responses had amplitudes close to zero. When there was a change in pitch, timbre or both, the amplitudes of the N100 and P200 components increased with increasing pitch and timbre differences. ERP response sizes for the probe (POR2) reveal that they increased for increasing pitch and timbre changes between adaptor and probe, and were minimal when both the pitch and timbre was the same. This minimum is likely to reflect noise in the EEG measurements. Overall these results suggest that the POR2 probe responses are pitch and timbre selective. These differences across conditions were further interrogated in the Section 8.4.2 below.

The POR2 probe response did have a similar triphasic morphology as the EOR and probe responses reported in Briley et al. (2013), with a small initial positive peak (referred to as P100; Näätänen and Picton 1987), followed by the N100 and P200. The presence of the P100 component for the probe responses, but not the adaptor responses, prompted further analysis of the probe, to rule out the possibility of any statistically significant differences across conditions. Peak-to-peak amplitudes of the probe were calculated from the onset of the probe stimulus to the peak of the P100 (termed here as 'P100 onset'), for all pitch and/or timbre change conditions (i.e., 'same pitch different timbre', 'different pitch same timbre', and 'different pitch different timbre'). This analysis was not conducted on the 'same pitch same timbre' condition because there was no P100 present, which is characteristic of a stimulus condition where the same sound is repeated. For statistical rigour, this peak-to-peak analysis was repeated from the peak of P100 to the N100 (termed here as 'P100 offset'). Paired sample t-tests, Bonferonni corrected to account for FWE associated with multiple comparisons, were conducted. The mean latency for P100 responses was 79ms, and there was no statistically significant difference in P100 latency across conditions. As anticipated, comparisons between 'different pitch different timbre' conditions (P100 onset: M = 3.02, SD = 1.86; P100 offset: M = 4.61, SD = 2.10) and 'different pitch same timbre' conditions (P100 onset: M = 2.40, SD = 1.65; P100 offset: M = 4.02, SD = 2.24) revealed no statistically significant differences; P100 onset: $t(17)$ = 1.07, p > .05; P100 offset: $t(17)$ = 1.12, p > .05). This was also the case for comparisons between 'different pitch different timbre' conditions and 'same pitch different timbre' conditions (P100 onset: M = 2.49, SD = 1.60, $t(17)$ = 0.96, p > .05 ; P100 offset: M = 3.54, SD = 2.03, $t(17)$ = 2.00, p > .05), and comparisons between 'same pitch different timbre' and 'different pitch same timbre' conditions (P100 onset: $t(17)$ = -0.19, p > .05; P100 offset: $t(17)$ = 1.14, p > .05). These non-significant results for the P100 responses of the probe, confirm that differences in peak-to-peak amplitudes were confined to the N100-P200 components of the probe, and are analysed further in the following section. The P100 could have arisen in the probe responses due to the characteristics of the stimulus sequence, such as differences in the stimuli and duration of sound preceding the adaptor (800ms noise) versus the stimuli and duration of sound preceding the probe (1000ms tone); however, the onset and offset of the P100 was not statistically different across conditions and therefore seems unlikely to have affected the probe adaptation calculated for peak-to-peak measures of N100-P200. Seither-Preisler et al. (2004) found that reducing the time interval between the noise and pitch resulted in a decrease in the amplitude of the POR. In the current experiment, the noise stimulus was shorter than the adaptor that preceded the probe, and therefore might be one explanation as to why the P100 was only evident in the probe response. If the current experiment had used noise and adaptor stimuli with the same lengths, we may have observed the presence of a P100 for the adaptor response also. The possible explanations for the presence of the P100 in the probe response are explored further in the Discussion.

8.4.1.2 Source-based

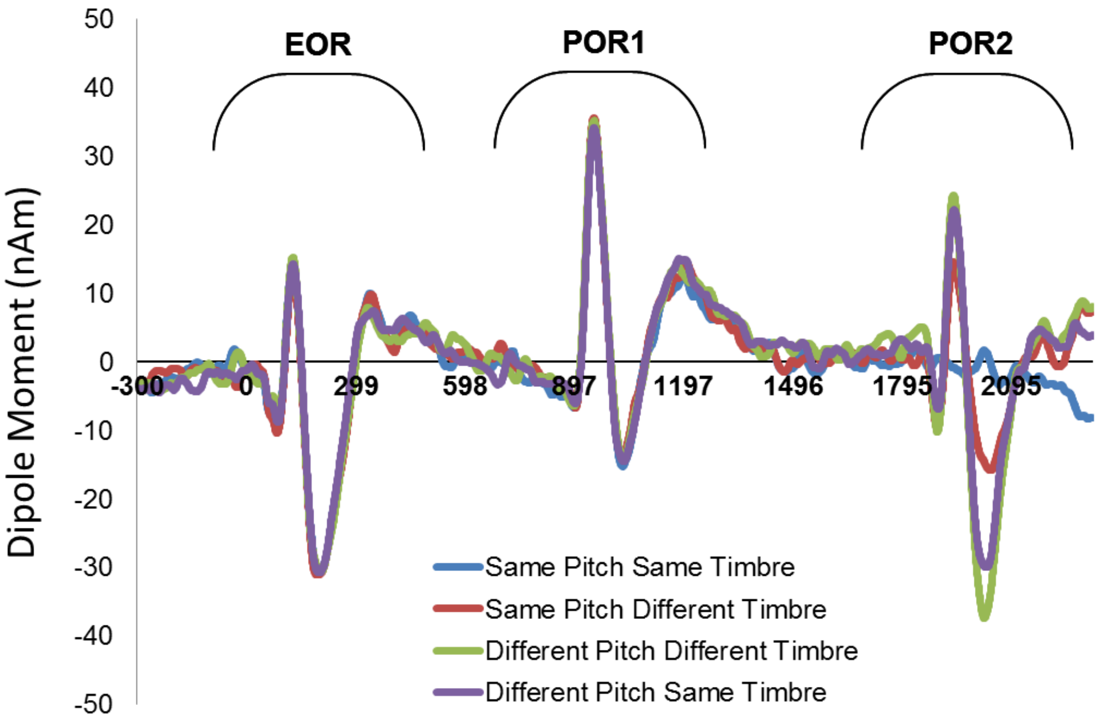

Figure 8.5. Average of right and left auditory dipoles for each condition. Conditions were comprised of 'same pitch, same timbre', 'same pitch, different timbre', 'different pitch, different timbre' and 'different pitch, same timbre'. Note that the horizontal x axis represents time in ms.

As shown in Figure 8.5, source waveforms across conditions were highly comparable with the sensor-based waveforms, whereby the only noticeable difference between conditions related to the POR2 probe responses. EOR and POR1 responses were comparable.

8.4.1.3 Summary

In summary, the findings from both sensor and source-based analyses fulfilled the pitch selectivity criterion for Objective 1. EOR, POR1 as well as P100 responses (for the probe) were not significantly different across sound conditions, and differences between conditions appeared to be confined to the evoked POR2 responses.

8.4.2 Objective 2: ERP evidence for pitch non-invariance to timbre

8.4.2.1 Sensor-based results

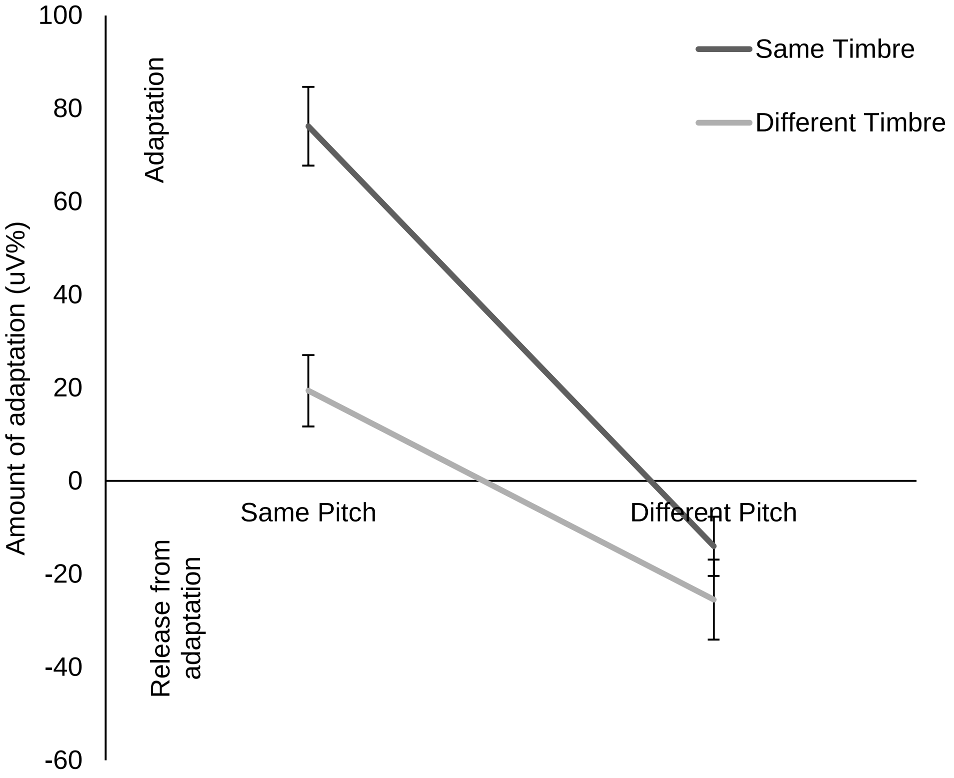

Figure 8.6. Mean amount of adaptation across conditions comprising the pitch congruence and timbre congruence interaction. Cousineau-Morey 95% confidence intervals are plotted. Note: Positive values indicate that the POR2 probe response was reduced relative to the POR1 adaptor response (i.e., adaptation). Negative values indicate the POR2 response was larger than the POR1 response (i.e., release from adaptation). Values approaching 0 indicate little or no adaptation (i.e., smaller values equal reduced adaptation or release from adaptation, values closer to 100% = maximal adaptation).

The amount of adaptation (µV%) from adaptor (POR1) to probe (POR2) did significantly differ across experimental conditions. A highly significant main effect of pitch congruence, *F*(1, 17) = 151.89, p < .001, η_p^2 = .899, indicated that adaptation was maximal for POR2 responses when the same pitch was repeated (*M* = 47.77%, *SD* = 17.14), compared to when the pitch changed (*M* = -19.77%, *SD* = 25.54). Hence a pitch change led to a release from adaptation (i.e., -% = less/no adaptation, greater neuronal response) suggesting that different neuronal populations were recruited to process these changes in the pitch value of the probe. It is therefore possible that observed probe responses do reflect cortical selectivity to the perceived pitch of the stimuli being presented (i.e., more evidence in favour of pitch selectivity). However this argument is based on the assumption that the POR1 and the POR2 are comparable; see the Discussion for potential limitations of this result.

The main effect of timbre also reached significance, *F*(1, 17) = 43.47, p < .001, η_p^2 = .719, which revealed that adaptation effects were greater for POR2 responses when the timbre was repeated (*M* = 31.06%, *SD* = 16.99), relative to when it varied from adaptor to probe (*M* = -3.06%, *SD* = 25.06). Importantly this revealed that changes in timbre also led to a release from adaptation across conditions, suggesting that different neuronal populations were also required to process timbral changes in the probe. This provides some support for the hypothesis that pitch processing neuronal populations are also sensitive to encoding the spectral frequency composition of the pitch-evoking stimuli, and thus pitch and timbre selective (although these neuronal populations may not be the same; see Discussion).

Support for this hypothesis is further evidenced by the significant interaction between pitch and timbre congruence, *F*(1, 17) = 20.09, p < .001, η_p^2 = .542. Figure 8.6 depicts how the most adaptation was observed when the same stimulus was repeated (same pitch, same timbre), and the least amount of adaptation (i.e., the largest ‘release’ from adaptation) when both the pitch and timbre changed from adaptor to probe. Values around zero indicate that the POR2 probe response was equivalent to the POR1 adaptor response, whilst higher values above zero indicate more adaptation (i.e., reduced response for POR2 probe compared to POR1 adaptor), and values below zero indicate release from adaptation (i.e., greater response for POR2 compared to POR1). This highlights how although the greatest adaptation occurred for when both pitch and timbre were repeated, when there was a change in either or both features of the probe, this caused varying degrees of release from adaptation or less adaptation.

Comparing the observed pattern of adaptation in Figure 8.6 to the predictions that were made in Figure 8.1 (see Section 8.1.2), a ‘hybrid’ pattern of results is observed between the pattern of adaptation predicted if pitch is invariant to timbre and if pitch is non-invariant to timbre. Specifically, post hoc t-tests on the amount of adaptation revealed that all conditions were significantly different from one another (p < .001) conforming to the pattern of adaptation predicted if the pitch response is non-invariant, with the exception of the ‘different pitch, different timbre' and the 'different pitch, same timbre' conditions. This pattern indicates that the timbre had no significant influence on the magnitude of the release from adaptation when pitch changed across the sound pair. Importantly, however, there was much less adaptation when the timbre changed but the pitch stayed the same which suggests that pitch processing neuronal populations are non-invariant to timbre, but potentially more sensitive to pitch changes than timbre changes. However, this argument assumes that the neuronal populations responding to both pitch and timbre changes are the same (see Discussion for alternative explanations).

One sample t-tests confirmed that half of the conditions were significantly different from zero baseline ($p < .005$) with the exception of the 'different pitch, same timbre condition' and 'same pitch, different timbre' conditions (after correcting for multiple comparisons).

8.4.2.2 Dipole-based results

The amount of adaptation (nAm) from adaptor to probe averaged over the two dipole sources located approximately in medial HG, were highly comparable and qualitatively similar to the sensor-based results (see Figure 8.7). Adaptation was significantly greater, $F(1, 17) = 49.72$, $p < .001$, $\eta_p^2 = .745$, for probe responses when the same pitch repeated ($M = 24.09$, $SD = 13.54$), compared to when the pitch changed ($M = -9.99$, $SD = 14.98$). Similarly for timbre, there was significantly more adaptation, $F(1, 17) = 18.75$, $p < .001$, $\eta_p^2 = .524$, for probe responses when the same timbre repeated ($M = 14.99$, $SD = 10.90$), compared to when the timbre changed ($M = -0.90$, $SD = 14.13$). There was also a significant interaction between pitch and timbre, $F(1, 17) = 5.34$, $p = .034$, $\eta_p^2 = .239$, with post hoc comparisons revealing all conditions were significantly different from one another ($p < .001$), with the exception of the 'different pitch, different timbre' condition vs. the 'different pitch, same timbre condition', also confirming a 'hybrid' pattern for the predicted adaptation results (see Figure 8.1). One sample t-tests confirmed that the majority of conditions were significantly different from zero baseline ($p < .001$) with the exception of the 'different pitch, same timbre condition' (after correcting for multiple comparisons). In summary these results also provide some evidence in favour of the hypothesis for pitch non-invariance to timbre.

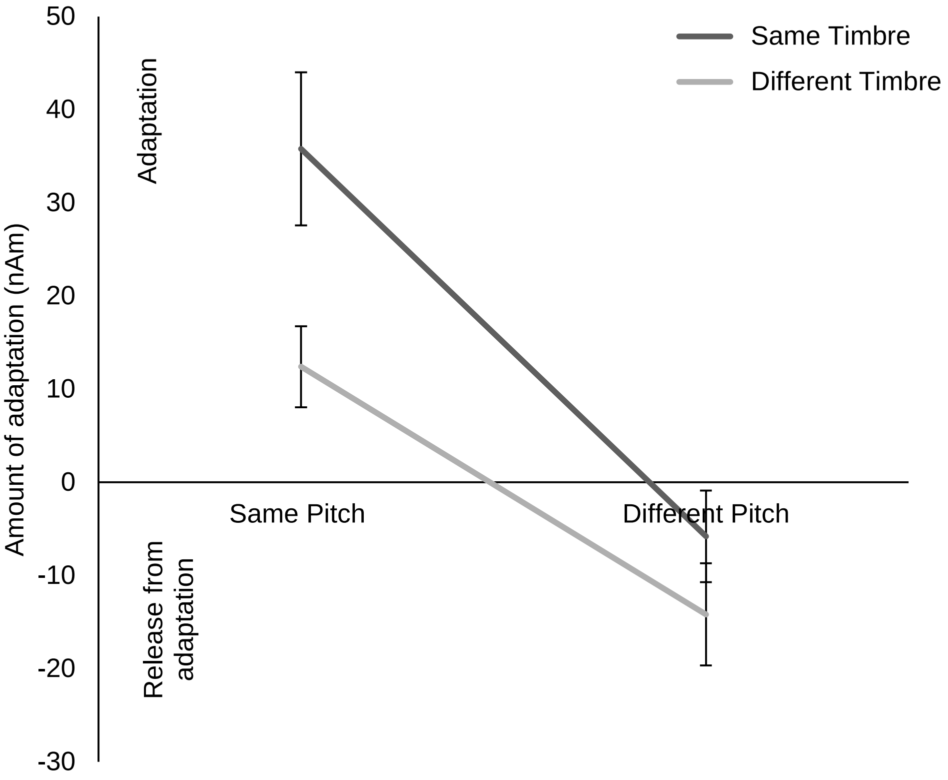

Figure 8.7. *Mean amount of adaptation across conditions comprising the pitch congruence and timbre congruence interaction. Cousineau-Morey 95% confidence intervals are plotted.*

8.4.2.3 Summary

In summary, both the sensor and source-based results showed that adaptation was greater when both pitch, or timbre was repeated, confirming that the adaptation paradigm had been employed successfully and that POR2 probe responses were sensitive to both these changes. There was a release from adaptation or less adaptation when the pitch, timbre or both features changed, which possibly suggests that pitch processing neuronal populations are non-invariant to timbre (and vice versa); however, it is important to note that this argument is based on multiple assumptions, and therefore possible alternative explanations, limitations and future directions are reviewed in the discussion.

8.5. Discussion

Experiment 9 ensured that discriminability for pitch and timbre was well-matched across pitch and timbre tasks and listeners. Experiment 10 described one of the first attempts to investigate pitch invariance to timbre using an ERP adaptation design, and using the same pitch-evoking stimuli that had been previously well-matched for perceptual salience and unconfounded by task requirements. To recap, the main objectives were to explore whether probe responses were selective to pitch (Objective 1, pitch criterion 1), and also to determine the evidence for pitch constancy in auditory cortex, specifically whether pitch is invariant to timbre or not (Objective 2, pitch criterion 2), while controlling for potential confounds (pitch criterion 4).

8.5.1 Probe responses selective to pitch and timbre

Across both sensor and source-based analysis approaches, the current experiment found evidence that suggests that the POR2 probe's responses were selective to pitch and timbre, fulfilling the first criterion surrounding pitch selectively (Objective 1), and supporting previous studies that have used this methodology (i.e., Briley et al., 2013).

8.5.2 Is pitch non-invariant to timbre - no support for pitch constancy?

Experiment 9 found that there was a significant difference within pitch and timbre tasks suggesting an interference effect of timbre on pitch (and vice versa). This supports previous results (Borchert et al., 2011), and studies who claim an interdependence between pitch and other features related to its perception (Bizley et al., 2009).

Across both sensor and source-based analysis approaches, Experiment 10 found evidence in favour of pitch processing being non-invariant to timbre. POR2 probe responses showed release from adaptation or less adaptation when the pitch and timbre changed. These findings provide some support previous human adaptation studies (e.g., Briley et al., 2013; Butler, 1972), and non-human studies that have found multiple areas of primary and non-primary areas of auditory cortex responding to pitch, timbre and spatial location of sounds (e.g., Bizley et al., 2009). These results may also be one explanation as to why previous primate studies have failed to locate neurons in primary auditory cortex that respond to the pitch of a complex tone when the frequency components are different (i.e., Fishman et al., 1998; Schwartz & Tomlinson, 1990, Steinschneider et al., 1998). Although Bendor and Wang (2005, 2010) did find pitch neurons that responded to the pitch of the stimulus despite changes in spectrum, there has been speculation as to whether this positive result is actually more attributable to distortion products (Abel & Kössl, 2009; Briley et al., 2013; McAlpine, 2004).

The current results therefore lend more support towards pitch being co-represented neurally together with timbre. As described earlier, recent studies suggest that other sound features may also be included in this cortical representation (see Bizley et al., 2009; Nelken et al., 2008). Town & Bizley (2013) suggested that there might be a very broadly distributed network of pitch sensitivity, involving multiple brain regions which may spatially overlap with regions sensitive to timbre. Walker et al. (2011a) found that neurons in ferret auditory cortex can represent pitch information in parallel with information about timbre and spatial location, which they referred to as 'multiplexing'. Neurophysiological human evidence (Formisano et al., 2008; Warren et al., 2005; Overath et al., 2008, 2010) suggests that pitch and timbre are processed in similar regions in auditory cortex (i.e., HG, PT, STG and STS), which is consistent with the idea that there are neuronal populations coding pitch that are non-invariant to timbre.

One intriguing finding related to the ‘hybrid’ pattern of adaptation results observed between predicted patterns of adaptation shown in Figure 8.1, Section 8.1.2. Changing pitch always led to a release from adaptation, no matter what the timbre properties of the stimulus pair were, with pitch change conditions not being significantly different from one another. However when the pitch stayed the same and timbre changed, there was still some partial release from ‘maximal’ adaptation. One interpretation of this hybrid result could be that although there was significant release from adaptation for all probe change conditions, neuronal populations were more finely tuned to F0 than to spectral content of the stimuli, and therefore more sensitive to pitch changes than to timbre changes (even though pitch and timbre salience had been equated prior to this experiment). Bizley et al. (2009) did report that neurons in some areas were more sensitive to one feature than to others. For example, primary areas such as A1 were more sensitive to pitch, whilst non-primary areas such as auditory anterior field were more sensitive to timbre. It therefore is possible the generators of responses of the POR2 originated from a region or neuronal populations that were more sensitive to F0 than timbre.

The current experiment shows that novel ERP adaptation methodology can be used to investigate pitch processing. Nevertheless, there are several limitations of the current experiment that need to be addressed in future work. These relate to the presence of the P100, the use of different noise and adaptor durations, the different stimulus presentation contexts for both the adaptor and the probe (i.e., preceded by a noise or preceded by a tone), and the sources of responses to changes in pitch and timbre and whether these are the same or different. The presence of the P100 for the POR2 probe response, although not statistically significant across conditions could have arisen due to differences in the duration of the noise stimulus which preceded the adaptor, and the pitch stimulus adaptor which preceded the probe. One explanation for the presence of the P100 for the probe response could be related to the fact that the noise stimulus was shorter in duration, compared to the pitch stimulus adaptor, and therefore led to smaller amplitudes for the adaptor POR1 response (see Seither-Preisler et al., 2004) which may have diminished the size of the P100, so that it was not visible. It could be argued that if the duration of the noise and adaptor stimuli had been the same, the presence of the P100 may have been visible for both adaptor and probe responses.

Furthermore, the use of noise, adaptor and probe stimuli also meant that the sounds, and therefore the stimulus context, which preceded both the adaptor and probe stimuli, were different; the adaptor was preceded by noise, and the probe was preceded by a tonal sound. Subsequently, we cannot assume that adaptor (POR1) and probe (POR2) responses were strictly comparable, and these two pitch responses might actually be different. For instance, the POR2 response followed similar triphasic morphology as the EOR, with a visible P100, N100, and P200 deflections, whilst the POR1 response's morphology was more representative of a typical POR response, which does not typically have a visible P100 deflection. It might be that the POR1 response which was preceded by a noise stimulus was a conventional POR response, as reported by Krumbholz et al. (2003) using sequences which transitioned from noise to pitch transition. For pitch-evoking probe stimuli (POR2) that were preceded by another pitch stimulus, this is a similar stimulus presentation context as used in Briley et al.'s (2013) ERP adaptation study that investigated pitch processing, and therefore might explain why POR2 responses in the current experiment followed the same triphasic morphology as Briley and colleagues (who also report P1-N1-P2 deflections for the probe). The POR2 responses might therefore actually reflect a pitch-related probe response, which they termed 'PR', rather than a POR2. It seems unlikely that the response to the probe reflects a general response to a change in the auditory stimulus because Krumbholz et al. (2003) found that the transition from an IRN pitch stimulus to noise produced no response whatsoever. Chait et al. (2006) also argued that if the POR was just a detector of change in on-going stimulation, you would expect to see just a single peak rather than two peaks (as observed for PORs); or three peaks, as observed in the current data and Briley et al. (2013). In the present data, the cortical sensitivity to changing pitch stimuli also seemed greater than changes in timbre, which runs counter to the idea that probe responses reflected a change response, as we would have expected adaptation results to be the same regardless of whether the pitch, timbre or both changed. It also seems unlikely that the probe responses reflect auditory deviance detection because only one pitch adaptor preceded probe responses.

The current experiment had three auditory events present in a given sequence (i.e., noise, adaptor and probe); however, future work could undertake the current experiment using just adaptor and probe stimuli, as in Briley et al. (2013), so that the stimulus context is better controlled and pitch responses are therefore more comparable. Alternatively, fMRI in conjunction with EEG measures could be used to explore the responses for POR1 and POR2 and determine whether the sources of these are the same or different, which would give a better indication of whether these responses are comparable. Most studies report slightly different neural generators for POR and N100 or P200 components elicited by any sound stimulus (see Seither-Preisler et al., 2004; Schönwiesner and Zatorre, 2008).

The results showed release from adaptation or less adaption when there was a change in pitch, timbre or both, which suggests that pitch processing neuronal populations may be non-invariant to timbre (and vice versa). However, this argument is difficult to determine given that there is no way that the response to pitch and timbre can be separated in the present data; more specifically, we do not know if the neuronal populations giving rise to POR2 responses for pitch and/or timbre changes, were the same or different. A release from adaptation suggests different neuronal populations have been recruited to process the probe change, but in the present study we were unable to determine if this recruitment of new neurons involved the same or similar neurons for both pitch and timbre. Subsequently, an EEG experiment combined with fMRI methods would enable us to determine whether the responses to pitch and timbre were derived from similar or different neuronal populations. Specifically, we would be able to more definitively rule out whether the responses to timbre changes reflected a timbre onset response (TOR) from separate timbre specific neuronal populations, and whether pitch and timbre changes reflected a summation of different neuronal populations responsible for pitch and timbre perception.

8.5.3 Summary

In summary, the ERP results reported here satisfied the first pitch criterion of pitch selectivity, but failed to find any evidence in favour of the second pitch criterion relating to pitch constancy. Pitch coding neuronal populations might be non-invariant to timbre, but this would need further exploration in follow-up studies. Confirmation of such findings would be counter to the argument that there are dedicated pitch neurons in the auditory cortex.

Despite humans being able to recognise the same note has been played on a piano or a trumpet, it seems reasonable to suggest that that such pitch constant representations might occur at the higher levels beyond auditory cortex. Importantly, however it should be noted that ERP methodology is limited in terms of being able to tell us what specific brain areas responded to given stimuli, and therefore the possibility that observed ERP waveforms did include the summation of cortical activity from two distinct neuronal populations (one responding to the F0, and one responding to the timbre) cannot be completely ruled out in the present data. Therefore further work is required to explore the location of such adaptation effects using a combination other neuroimaging techniques (e.g., fMRI and EEG). This is because fMRI is better at addressing questions surrounding whether (and where) neurons in a particular area are active or not, instead of whether they are tuned to a particular stimulus dimension (Town & Bizley, 2013).

Chapter 9. Searching for pitch invariant representations using behavioural measures

9.1. Introduction

This chapter describes Experiments 11 and 12, where Experiment 12 primarily investigated the criterion of ‘pitch constancy’ using behavioural methods. Adaptation can be measured behaviourally as well as neurophysiologically. The adaptation phenomenon examined in Chapter 8 has been purported as the neural correlate of behavioural performance or perceptual repetition priming (Grill-Spector et al., 2006; Henson & Rugg, 2003; Schacter & Buckner, 1998; Wiggs & Martin, 1998). Behavioural measures are more commonly referred to as ‘perceptual repetition priming’ and believed to represent a learning mechanism that represents repeated or familiar as compared to novel items (Bergerbest et al., 2004). Repetition of stimuli has shown to facilitate discrimination performance (i.e., quicker reaction times and improved accuracy; Dehaene et al., 2001; De Lucia et al., 2010; Henson & Rugg, 2003). Subsequently it is plausible that adaptation results for repeated and different stimuli can generate predictions about discrimination performance measures based on reaction time and accuracy. Like neural adaptation, it may be possible to draw inferences about the neural coding of pitch based on whether or not discrimination performance for listening conditions exhibit pitch constancy or invariance. Discrimination performance for pitch and timbre tasks using reaction time measures and accuracy measures is yet to be fully explored in the human pitch coding literature to date (but see Pitt, 1994 as one example). The experimental approach combined two different behavioural methods. Experiment 11 was a feasibility study which sought to equate behavioural performance for pitch and timbre tasks within subjects, as well as assess between-subject differences in the discriminability between and within pitch and timbre tasks. Experiment 12, the study of interest in this chapter, further investigated the pitch constancy criterion by exploring the reaction time and accuracy effects across the same experimental conditions used in the ERP Experiment 10. It is (to my knowledge) the first of its kind to examine accuracy and reaction time data to pitch and timbre tasks in a fully-crossed design, where stimulus parameters have been well-matched for discriminability.

9.1.1 Behavioural evidence for pitch invariance to timbre

9.1.1.1 Non-human studies

A prerequisite for any study that wishes to make comparisons between animal models and humans is that the non-humans can actually perceive and distinguish between the sounds being tested (Bizley, Shinn-Cunningham, & Lee, 2012). We know that like humans, non-human animals can discriminate pitch and timbre (e.g., Bizley et al., 2012; Walker, Schnupp, Hart-schnupp, King, & Bizley, 2009) across a variety of different listening tasks (e.g., change detection, 2AFC) and stimuli (pure, complex, and naturalistic tones). Walker et al. (2009) used pure tones and artificial vowels to measure pitch discrimination in ferrets. This involved training ferrets to discriminate whether the pitch was higher or lower than the reference sound (ranging from 200-1200 Hz), using a 2AFC task. For comparison, they also measured discrimination performance in naïve human listeners, using a similar paradigm. Ferrets were able to discriminate pitch equally well for both pure tones and vowel stimuli, however human pitch discrimination was substantially better overall. Nevertheless the study confirmed that ferrets can be trained to label complex sounds as high or low. Walker et al. (2011a) also trained ferrets to detect changes in the pitch or timbre of a repeating artificial vowel on a go/no go task. The ferret would initiate the trial by placing its nose in a poke hole situated in the center of the chamber. A sequence of artificial vowel sounds was then presented. The reference sound could change in identity (timbre) or pitch on the 3rd and 7th vowel in the sequence. If the ferret withdrew their nose from the hole when a deviant was presented, they were rewarded with water. If they did not withdraw then they had a 12 s time-out period. Timbre and pitch changes were tested in separate sessions. They found that ferrets accurately responded to changes in timbre more rapidly than they could for pitch. Nevertheless, it should be noted that in none of these animal studies did the authors match conditions for stimulus discriminability like in the present chapter and elsewhere (see Chapters 7 and 8). This might be one explanation for why differences in sensitivity and time courses for timbre have emerged (Bizley et al., 2009; Walker et al., 2011a; but not detected in ERP data here, see Section 8.4.1). It could be argued that timbre changes were more salient, and thus gave rise to the pattern of results reported. Hence the present chapter explored whether timbre discriminations were quicker and more accurate than pitch discrimination

Some evidence for pitch constancy for discrimination tasks has also been reported. A handful of studies have found that non-human animals can universally distinguish between vowel sounds (e.g., timbre), despite changes in the gender of the speaker (i.e., pitch; (Kojima & Kiritani, 1989; Ohms, Gill, Van Heijningen, Beckers, & ten Cate, 2010), or in F0 and spatial location (Bizley et al., 2012). However as previously discussed in Chapter 7, pitch invariance to timbre (and vice versa) has not been specifically addressed by animal studies as they typically observe discrimination thresholds for one target dimension (i.e., pitch or timbre) whilst the other non-target dimension remains fixed, or only focus on exploring one target dimension when the other non-target dimension is varying.

9.1.1.2 Human studies

Psychophysical studies have revealed that humans are able to discriminate pitch (e.g., Jackson & Moore, 2013), and timbre (Grey, 1978) particularly well when all other within- and between-interval non-target acoustic features remain fixed, but not when the non-target dimension varies (see Chapter 7).

9.1.2 Using adaptation evidence to make predictions about behavioural performance

Studies have generally found that when an initial stimulus is repeated, behavioural responses in a given task, such as a reaction time and accuracy tend to be faster and more accurate, and this is known as a priming, or facilitative, effect (DeHaene et al., 2001). For example, Mondor, Hurlburt, and Thorne (2003) conducted two experiments that looked at the effect on pitch categorisation performance when timbre changed from prime to probe. They used a consecutive response paradigm (i.e., two sounds presented in succession) to investigate whether categorisation of the pitch of a probe tone is influenced by the pitch of, and the response made to, a preceding prime tone. The prime and the probe could be drawn from a pool of either low or high frequency pure tones, and participants were required to discriminate whether each of the sounds were high (e.g., 3000 or 3300 Hz) or low (e.g., 500 or 550 Hz) in pitch. The probe and prime were either the same (identical condition; i.e., same pitch, same response), differed in pitch but required the same response as they were drawn from the same frequency pool (equivalent condition), or differed in both the pitch and required different responses (i.e., different condition). By comparing results for identical and different conditions, this allowed the authors to examine the influence of stimulus and response repetition effects, whilst comparing results for identical and equivalent conditions allowed them to assess the importance of pitch repetition over and above that of response repetition. They found that performance was quickest and most accurate when the prime and probe were part of an identical condition, indicating a robust facilitative effect. Performance was intermediate when the two sounds differed in pitch and required different responses (different condition), and slowest when the prime and the probe different in pitch but required the same response (equivalent condition), indicating an inhibitory or interference effect. In a subsequent experiment, they not only found that performance got worse as a function of increased frequency change, but also the magnitude and direction of the pitch change in the ‘equivalent condition’ influenced performance as listeners were quicker and more accurate when the direction of the pitch change was away from the alternative category (i.e., high or low frequency) compared to when it was towards an alternative category. Taken together, the results of the two experiments provided evidence that pitch discrimination performance is influenced by both the pitch of a preceding sound and the response made to it.

Subsequently perceptual priming can be used to investigate the coding or mental representations occurring between stimulus presentation and behavioural responses (Naccache & Dehaene, 2001) and like adaptation, priming can be used to probe the functional characteristics of neuronal populations (Grill-Spector et al., 2006; Henson, 2003). Using these assumptions, it seems reasonable to suggest that given the interference effect that both pitch and timbre have on one another, it might be expected that reactions times as well as accuracy may be impaired for target dimension discrimination, when the non-target dimension is also varying. However, to my knowledge, no prior human (or non-human) study has collectively investigated discrimination performance (reaction times and accuracy) for both pitch and timbre tasks to investigate pitch invariance to timbre in stimuli that have been well-matched.

9.1.3 Objectives and Hypotheses

The current study therefore aims to address this by examining pitch invariance to timbre by comparing behavioural effects to changes in the same pitch-evoking stimuli (Experiment 12), using stimuli that are well-controlled for inferred perceptual salience (Experiment 11). Based on the same logic in Chapter 8, the main objectives were as follows:

Objective 1) Are behavioural responses to pitch invariant to changes in timbre, or non-invariant to changes in timbre? If pitch is non-invariant to timbre changes a 2 x 2 x 2 repeated-measures ANOVA, with target type (pitch or timbre), target dimension (same or different), and non-target dimension (same or different), should reveal a significant main effect of non-target dimension. This would address the second pitch criterion relating to pitch constancy (see Section 1.2).

Objective 2) If timbre discriminations processed quicker than pitch (as per Walker et al., 2011a) it was expected that there would be quicker reaction times and better accuracy for timbre conditions, compared with different pitch conditions (i.e., a significant main effect of target type).

9.2. General methods and materials

9.2.1 Listeners

Twenty-three right-handed (LQ = +87%, Decile R.6) healthy listeners (#46-#68) volunteered to participate in both behavioural Experiments 11 and 12 (5 male, 18 female; M_{age} = 20.52, *SD* = 2.86, age-range 18-30 years). All participants reported normal or corrected-to-normal vision, and had clinically normal hearing (≤25 dB HL between audiometric frequencies 250-8000 Hz). No listeners reported a history of hearing, neurological, and/or psychological impairment, and use of psychiatric medication or substance misuse. Six participants (#49, #52, #55, #57, #62, and #65) were musically trained on the cornet, guitar, piano and violin achieving grades 1-5 across 1-5 years, whilst four listeners (three of whom were not musically trained; #49, #60, #64 and #67) also reported informal musical experience on the guitar, keyboard, saxophone and ukulele gained over 1-8 years. Eighteen subjects (not reported above; #69-#86) did not successfully complete Experiment 11. Two participants withdrew before completion. Sixteen participants were excluded because they did not achieve appropriately matched pitch and timbre discrimination performance during the training phase. Listeners gave written informed consent, and the study was approved and performed in accordance with the College Research Ethics Committee’s guidelines, Nottingham Trent University (ethics no. 2011/46).

9.2.2 Procedure and Design

Experiment 12 was generally completed by listeners on a separate day after Experiment 11, but usually within 14 days. Both experiments took place at Nottingham Trent University. All participants were recruited via Nottingham Trent University’s online Research Participation Scheme or via email. Participants were awarded psychological research credits based on their length of participation (i.e., awarded 1 credit per 10 minutes of participation, where appropriate).

9.3. Experiment 11: Ensuring pitch and timbre discriminability was matched for reaction time Experiment 12

9.3.1 Objective

The following feasibility Experiment 11 was employed to ensure that the salience for pitch and timbre had been equated across listeners in Experiment 12.

9.3.2 Methods

9.3.2.1 Stimuli

The stimuli and trial types in Experiment 11 were the same as those in Experiment 8 (Section 7.7). Also see Section 7.4.2 for details regarding how these were generated and calibrated.

9.3.2.2 Procedure and Design

Experiment 11 followed the same procedure and design as described in Experiment 8 (see Sections 7.7.1 and Section 7.4.3 for more information), and lasted between 2-4 hours, depending on individual training needs. The procedure for training and equating pitch and timbre discrimination is described in section 7.4.3.3. Some participants were invited to complete more than one pitch and timbre discrimination training session. It should be noted that participants could be trained only to discriminate pitch and/or timbre for conditions in which the pitch and/or timbre was ‘different’. Therefore there was no training for the ‘same pitch, same timbre’ condition.

9.3.3 Data Analysis

Data analysis for Experiment 11 followed the same procedure as described in Section 7.4.4.

9.3.4 Results

Average d’ data for all 23 listeners (#46-#68) are plotted in Figure 9.1.

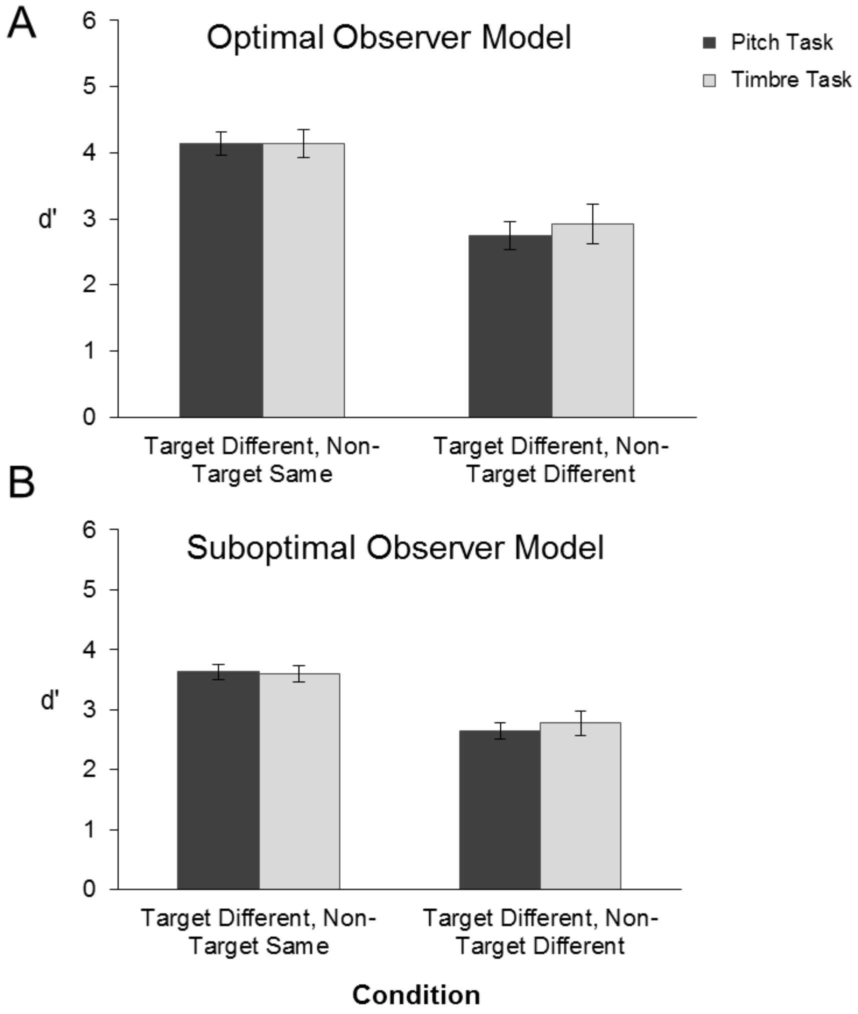

Figure 9.1. Averaged d’ across the 23 listeners for Experiment 11 pitch and timbre tasks for conditions in which the target dimension was different but the non-target remained the same (target different, non-target same), and for conditions in which both the target and non-target were different (target different, non-target different). Panel A illustrates d’ based on the optimal observer model. Panel B illustrates d’ based on the suboptimal observer model. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

Paired sample t-tests conducted on the d’ data for the optimal observer model revealed no statistically significant difference in discrimination performance between pitch and timbre tasks for both ‘target different, non-target same’ conditions (pitch: *M* = 4.14, *SD* = 0.75; timbre: *M* = 4.14, *SD* = 0.91), *t*(22) = 0.02, *p* = .986, and ‘target different, non-target different’ conditions (pitch: *M* = 2.74, *SD* = 0.95; timbre: *M* = 2.92, *SD* = 0.88), *t*(22) = -0.74, *p* = .468), indicating that the discriminability of pitch and timbre was appropriately matched across listeners (see Figure 9.1).

Results indicated that the non-target dimension negatively affected discriminability of the target dimension for both pitch and timbre tasks. As expected, within tasks listeners were significantly worse ($p < .001$) at discriminately the target when the non-target dimension also varied from interval to interval (i.e., ‘target different, non-target different’ conditions) for both the pitch (non-target same: $M = 4.14$, $SD = 0.75$; non-target different: $M = 2.74$, $SD = 0.95$), $t(22) = -6.80$, $p < .001$, and timbre tasks (non-target same: $M = 4.14$, $SD = 0.91$; non-target different: $M = 2.92$, $SD = 0.88$), $t(22) = -4.29$, $p < .001$, compared to when only the target dimension differed and the non-target dimension remained the same (i.e., ‘target different, non-target same’ conditions) for pitch and timbre tasks. These results seem to suggest a detrimental or interference effect of the non-target dimension on discriminating the target (signal), and one that is reciprocal for both dimensions (i.e., pitch affects timbre discrimination and vice versa).

Paired sample t-tests conducted on the data for the suboptimal observer model confirmed the same pattern of results as the optimal model.

9.3.5 Conclusions

Overall, the results showed that discriminability between pitch and timbre tasks were well-matched for the stimulus parameters selected and for this group of participants. The significant differential effect within pitch and timbre tasks was expected given the nature of interference timbre has on pitch (and vice versa), supporting previous results (Borchert et al., 2011), and studies who claim an interdependence between pitch and other features related to its perception (Bizley et al., 2009).

9.4. Experiment 12: Reaction time experiment exploring pitch constancy

9.4.1 Objective

To recap the current behavioural experiment aimed to address pitch criterion 2 relating to pitch constancy, using the same logic and conditions reported in Chapter 8.

9.4.2 Methods

9.4.2.1Stimuli

The stimulus parameters employed in Experiment 12 were identical to those described for Experiments 8-11.

9.4.2.2 Procedure and Design

For both pitch and timbre discrimination tasks, listeners were required to make a judgement across intervals to decide in which interval the target dimension was the same or different. Any non-target within interval changes were the source of interference on performance. To keep the task for both Experiments 11 and 12 as similar as possible and not affect discriminability, each condition was completed in blocks (as per training) but the order of which task (pitch or timbre) they completed first was randomised to eliminate any order effects. Feedback was provided after each trial, in the form of a green “correct” response, or a red “incorrect” response displayed on the computer screen, and enabled listeners to track their performance. Although the procedure and design employed for Experiment 12 was similar to Experiment 11, it differed in a number of important ways which are described in the following three sections below.

Task and Conditions

Given the nature of the present experimental paradigm, I was unable to employ a strictly implicit auditory priming design, as I wanted to ensure the stimuli were as similar as possible to previous behavioural and ERP adaptation experimental conditions, and that the discriminability of pitch and timbre stimuli were effectively matched. Changing the paradigm would have undoubtedly made the task much more difficult and hence affected the results. Subsequently, in the current experiment, listeners not only had to indicate in which interval the pitch or timbre (target dimension) was different, but also indicate in which interval the pitch or timbre was the same (see Table 9.1 for conditions, trial types, and target and non-target intervals). This meant that each of the conditions for pitch and timbre tasks directly mapped onto each of the four conditions presented in the ERP Experiment 10 described in Chapter 8.

Four conditions were created for each task by crossing two levels of pitch congruence (same, different) with 2 levels of timbre congruence (same, different) in a repeated-measures factorial ANOVA design. Pitch and timbre congruence referred to the pairing of stimulus values across the two intervals (see Table 9.1). Only the condition of interest stayed the same or differed between intervals. As shown in Table 9.1, it was the relationship across intervals that gave rise to the condition of interest, where one interval was the target interval and the other was the non-target interval (e.g., for the pitch task condition 2, the only difference between the intervals related to the F0). Unlike Experiments 8, 9 and 11, conditions here were counterbalanced so that each of the four conditions was discriminated for each pitch and timbre task (i.e., totalling 8 conditions).

Table 9.1

Trial types for all conditions in the pitch and timbre discrimination tasks

	Pitch task		Timbre task	
Condition	Target interval	Non-target interval	Target interval	Non-target interval
1: Same pitch, same timbre (i.e., target same, non-target same)	200-1->200-1 200-2->200-2 220-1->220-1 220-1->220-1	200-1->220-1 220-1->200-1 200-2->220-2 220-2->200-2	200-1->200-1 200-2->200-2 220-1->220-1 220-1->220-1	200-1->200-2 200-2->200-1 220-1->220-2 220-2->220-1
2: Same pitch, different timbre	200-1->200-2 200-2->200-1 220-1->220-2 220-2->220-1	200-1->220-2 220-2->200-1 200-2->220-1 220-1->200-2	200-1->200-2 200-2->200-1 220-1->220-2 220-2->220-1	200-1->200-1 200-2->200-2 220-1->220-1 220-1->220-1
3: Different pitch, same timbre	200-1->220-1 220-1->200-1 200-2->220-2 220-2->200-2	200-1->200-1 200-2->200-2 220-1->220-1 220-1->220-1	200-1->220-1 200-2->220-2 220-1->200-1 220-2->200-2	200-1->220-2 200-2->220-1 220-1->200-2 220-2->200-1
4: Different pitch, different timbre (i.e., target different, non-target different)	200-1->220-2 220-2->200-1 200-2->220-1 220-1->200-2	200-1->200-2 200-2->200-1 220-1->220-2 220-2->220-1	200-1->220-2 200-2->220-1 220-1->200-2 220-2->200-1	200-1->220-1 200-2->220-2 220-1->200-1 220-2->200-2

Note: Numbers refer to pitch/F0 types. Suffixes ‘1’ and ‘2’ refer to timbre types. ‘1’ refers to the timbre type created using 3 as the lowest non-overlapping harmonics, whilst ‘2’ refers to the timbre type using 4 as the lowest non-overlapping harmonic (see Section 7.4.2). Solid line separates pitch and timbre tasks, whereas the dashed line separates discrimination intervals on a given trial.

Position of target in the stimulus sequence

For simplicity Experiments 6-10 used a modified version of the dual-pair design, where the target only ever appeared as the second tone in interval pair (e.g., AB-AA or AA-AB). Listeners may have used the position of the target as an additional cue to improve their performance (i.e., optimal observer). Specifically, listeners may have built an internal memory of the reference during training and judged each target as low or high relative to this internalised reference, or compared a given target to internalised low and high target templates (Walker et al., 2009). Humans could be comparing relative pitches of target and reference for each trial. Therefore in Experiment 12, the position of the targets in the stimulus sequence were randomised and occurred either as the first or second tone for each interval an equal number of times. This helped to eliminate the possibility that listeners used this position cue and get a more reliable measure of their reaction time and accuracy across conditions.

Presentation and block order

Presentation and timing of stimulus trials were controlled using Superlab v4.0 (Cedrus Corporation). The timing of stimuli was exactly the same as reported for Experiment 11. Listeners responded ‘interval 1’ or ‘interval 2’ as same or different (depending on the task) using a RB-730 response pad, which accurately records reaction times to about 500 microseconds. The pad is a USB device that requires a USB driver which introduces a constant delay of around 5ms. Reaction times were therefore measured with ms accuracy from the offset of the second interval. The mapping of responding hands to responses was counterbalanced across participants. For a given pitch or timbre task, each condition was presented in individual blocks, where pitch and timbre target types were randomised for each block using the method of constant stimuli. Unlike Experiment 11, the order of completion for the blocks in a given tasks was randomised using a Latin square design to avoid any order effects (i.e., some subjects did easier blocks before harder blocks and vice versa). Participants were given trial examples and practice trials with feedback before each block. A total of 384 trials were presented for each pitch and timbre task. For conditions where the non-target was different, the interval pairings yielded 32 trial types across conditions which were repeated three times. For conditions where the non-target was the same, the interval pairings yielded 16 trial types across conditions which were repeated six times. The presentation of trials was randomised within a given block. Unlike Experiment 11, median reaction times as well as mean accuracy were recorded for each condition per task per listener.

The sound presentation level in Superlab was fixed to ensure stimuli presented binaurally using Sennheiser HD-280 headphones had an overall sound level of 70 dB SPL. Sound level was measured and calibrated prior to testing using the same procedure described in Chapter 7. Experiment 12 took approximately 2 hours per listener to complete.

9.4.3 Data Analysis

For the reaction time analysis, inaccurate trials were discarded, and median reaction times were calculated for the remaining correct trials for each condition, per task and listener. Median reaction times were used instead of mean reaction times; this is because reaction time data is known to be highly variable and prone to outliers (e.g., due to guesses, etc.), where reaction time distributions are often skewed to the right (Ratcliff, 1993). Exploring my own data confirmed that all conditions contained outliers, most conditions had skew and kurtosis values greater than -/+ 2, and histogram plots were positively skewed. Subsequently, median reaction times were used because they are a robust and common way of eliminating outliers in reaction time data (see Ratcliff, 1993).

For the accuracy data, signal detection theory was used to convert PC into corresponding d’ using a similar method as Experiments 6-9 and 11. However given that the design for

Experiment 12 (i.e., the number of possible target combinations/sequences) were now the same or more similar to the conventional a dual-pair design, only the dual-pair design formula for MATLAB described in the Micheyl & Messing's (2006) paper was used (see Section 7.4.4). For dual–pair designs, the formula only calculated d' for PC values of 0.50 to 1.00, otherwise it returned as blank. In these rare cases that the PC did fall below 0.50 but was approaching this value (i.e., Experiment 12) the PC value was set to 0.50 if it was close to this value. Although this meant that some d' values were slightly elevated, it was not anticipated to make a significant difference to the results.

Dependent variables, median reaction time and d' were subjected to a three-way repeated-measures ANOVA (i.e., 2 x 2 x 2). The first factor referred to target type (i.e., pitch or timbre task). The second factor referred to the target dimension (i.e., whether the target dimension was the same or different). The third factor referred to the non-target dimension and whether this was the same or different. Extreme hit rates and false alarm rates were dealt with using the correction procedure described in Section 7.4.4. For all ANOVA testing, an alpha criterion of p < .05 was used to test for statistical significance. All planned and post hoc paired sample t-tests were Bonferonni corrected to account for FWE associated with multiple comparisons. If Mauchly's test of sphericity was violated, a Greenhouse-Geisser correction was used

9.4.4 Results

9.4.4.1 Evaluating the evidence for Objective 1 (Behavioural evidence for pitch non-invariance to timbre) and Objective 2 (No evidence for timbre discriminations being processed quicker than pitch discriminations)

Reaction time data

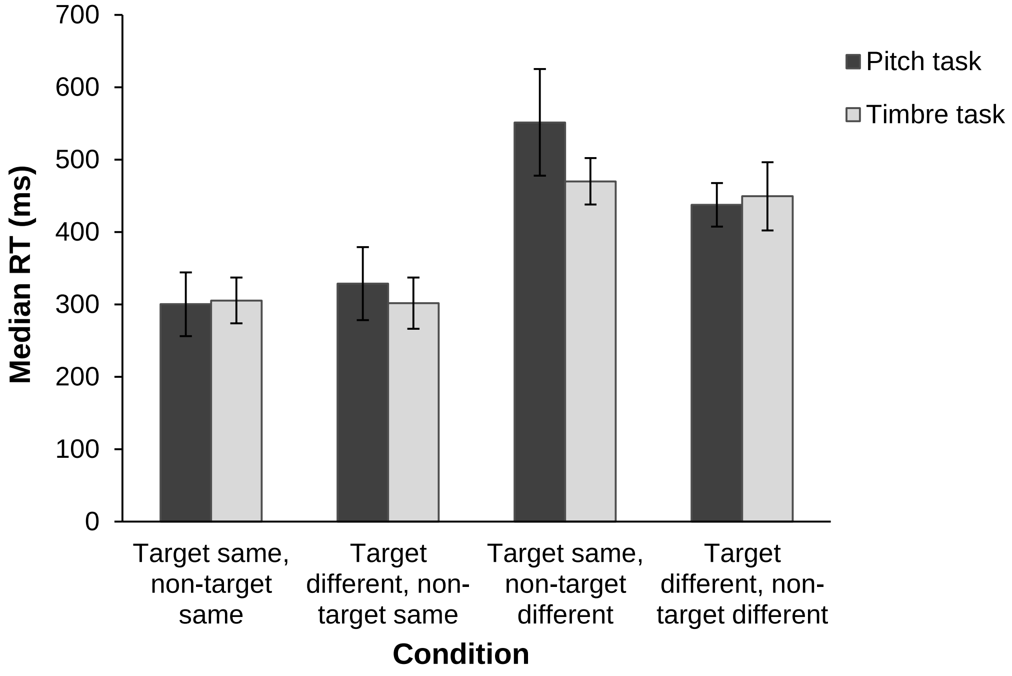

Figure 9.1. Median reaction times for pitch and timbre tasks across conditions. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

The ANOVA revealed a significant main effect of target dimension, *F*(1, 22) = 4.92, *p* = .037, η_p^2 = .183, where across conditions and tasks listeners were 27ms faster at discriminating the target dimension when it was different across intervals (*M* = 379.34ms, *SD* = 135.14) compared to when the target was the same (*M* = 406.76ms, *SD* = 156.59). This result however might be driven by the fact that reaction times for pitch task 'target same, non-target different' condition (i.e., same pitch, different timbre) were longer compared to the timbre task for the same condition (see Figure 9.1).

There was also a significant main effect of the non-target dimension, *F*(1, 22) = 15.26, *p* = .001, η_p^2 = .410, which indicated that listeners were 168 ms slower at discriminating the target when the non-target dimension differed across intervals (*M* = 477.08ms, *SD* = 236.05) than when the non-target dimension stayed the same (*M* = 309.02ms, *SD* = 81.18). This indicates that irrelevant non-target dimension changes interfered with listener's ability to do both pitch and timbre tasks, suggesting pitch is non-invariant to timbre (and vice versa). This provides evidence in support of Objective 1.

Across conditions, there was no significant main effect of target type, *F*(1, 22) = 2.11, *p* = .161, which suggests that discriminability of pitch and timbre tasks was well equated and does not support the directional prediction of Objective 2. Timbre discriminations were not processed quicker than pitch discriminations.

A significant two-way interaction for target dimension*non-target dimension was also found; *F*(1, 22) = 12.39, *p* = .002, η_p^2 = .360. To understand the pattern of this interaction, the data were plotted in Figure 9.2. The null result of the main effect of target type justified collapsing the data across conditions in Figure 9.2.

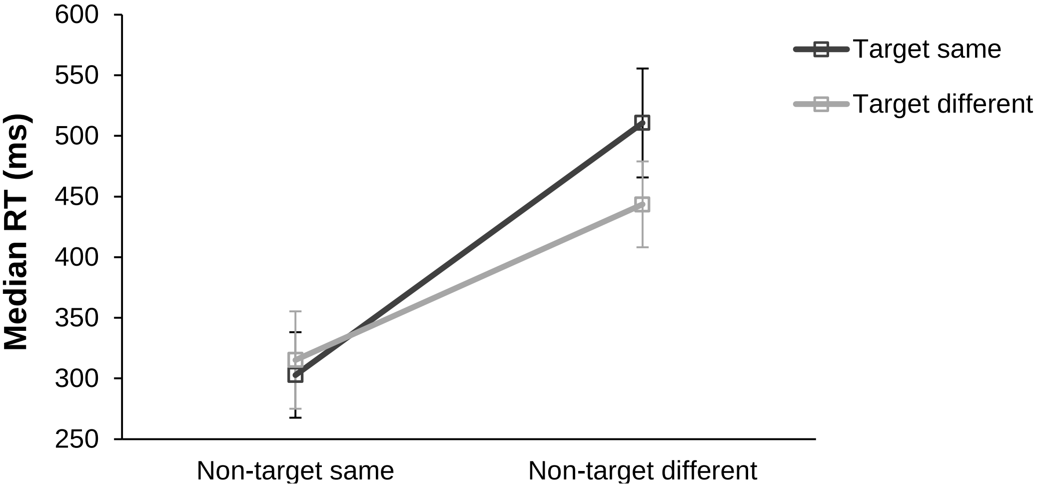

Figure 9.2. Median reaction times for conditions collapsed across pitch and timbre tasks. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

Post hoc t-tests were conducted on the target*non-target interaction reaction time data that had been collapsed across pitch and timbre tasks. As expected, conditions in which both the target and non-target stayed the same across intervals (i.e., ‘target same, non-target same’; same pitch, same timbre condition), were processed significantly faster (*M* = 302.86, *SD* = 95.14) than conditions where the target was the same but the non-target was different (i.e., ‘target same, non-Target different’; *M* = 510.65ms, *SD* = 253.64, *t*(22) = -4.52, *p* < .001), and conditions where the target and non-target dimensions were both different (i.e., ‘target different, non-target different; *M* = 443.50ms, *SD* = 228.15, *t(*22) = -3.51, *p* = .002). These results support neurophysiological and behavioural data that suggest that processing is more efficient for repeated presentations of a stimulus dimension. However there was no significant difference between the ‘target same, non-target same’ conditions (*M* = 302.86, *SD* = 95.14) and the ‘target different, non-target same’ conditions (*M* = 315.19, *SD* = 75.05, *t(*22) = -1.08, *p* = .293). One explanation for this finding could relate to the fact that the target difference was relevant to the task in these conditions, and so it is possible that listeners were looking out for this cue or it was more novel. Furthermore because ‘target different, non-target same’ conditions did not have any interference effects occurring from the non-target this may have also helped to facilitate discrimination performance.

Conditions where the target stayed the same but the non-target was different (i.e., ‘target same, non-target different’) showed significantly slower reactions times (*M* = 510.65, *SD* = 253.64), compared to conditions with only a target difference (*M* = 315.19, *SD* = 75.05), specifically the ‘target different, non-target same’ condition; *t*(22) = -3.99, *p* = .001. This result may indicate an interference effect of 195ms from the non-target dimension, which further supports evidence from previous literature and experiments reported here that pitch is non-invariant to timbre, and vice versa.

As expected, stimulus presentations with two dimension changes (i.e., ‘target different, non-target different) took significantly 128ms longer to discriminate sounds (*M* = 443.50ms, *SD* = 228.15) compared to conditions in which only the task relevant target differed i.e., ‘target different, non-target Same’ conditions (*M* = 315.19, *SD* = 75.05); *t*(22) = -2.99, *p* = .007. However when ‘target different, non-target different’ conditions were compared to ‘target same, non-target different’ conditions (*M* = 510.65ms, *SD* = 253.64), reaction times were actually quicker for the former by 67ms (*M* = 443.50, *SD* = 228.15); *t*(22) = 3.24, *p* = .004). Explanations for this might be because it is more difficult to identify a relevant target dimension as the same when the non-target dimension is varying because there is an aspect of the stimulus that is changing which interferes with processing and listener’s ability to do the task but does not provide a useful cue. Whereas for ‘target different, non-target different’ conditions in which two dimensions are changing, the target of interest is still changing with the irrelevant non-target dimension and thus might facilitate performance to some degree. Other explanations could related to fact that listeners were not trained to identify a target dimension as the same when the irrelevant dimension was also varying, whilst they were trained to discriminate a relevant dimension as different when the irrelevant dimension was also different and thus might have improved performance for two dimension change conditions. Steele & Williams (2006) found that both musicians and non-musicians alike, misjudged identical pitches as being different when they were played on different musical instruments.

There was also a significant three-way interaction between task type, target dimension and non-target dimension, *F*(1, 22) = 4.67, *p* = .042, η_p^2 = .175. However this interaction was not explored further as the comparisons of special interest have already been discussed and deemed to encapsulate any significant effects that may have been present here. No other interactions were significant.

d' sensitivity

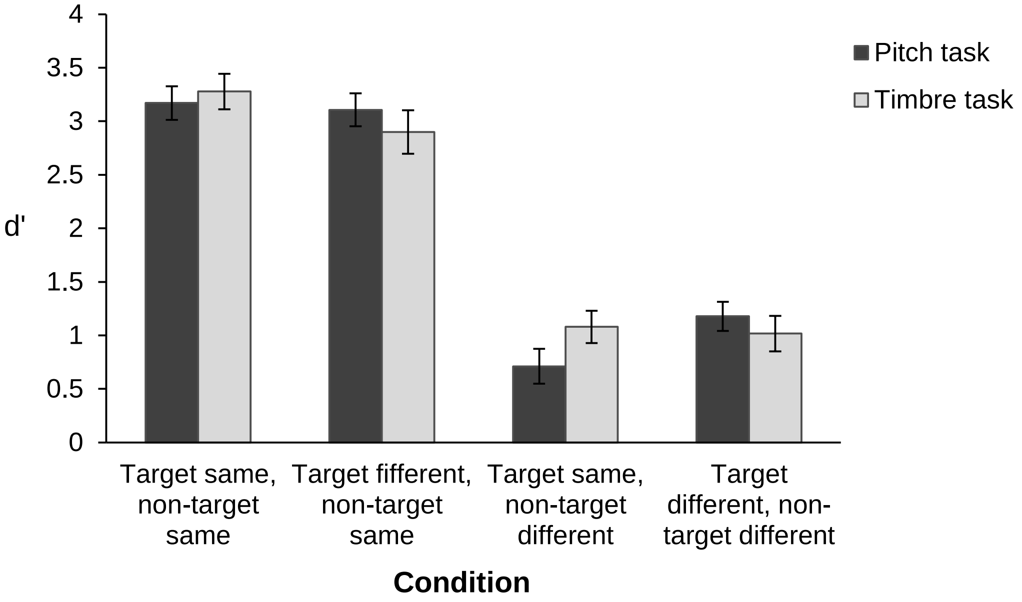

Figure 9.3. Mean d' for pitch and timbre tasks across conditions taken from PC. Within-subjects Cousineau-Morey 95% confidence intervals are plotted. Note that this exactly mirrors the reaction time data.

The same ANOVA model was specified for the accuracy data, where d' was the dependent variable. These data are plotted in Figure 9.3. Despite listeners being trained to discriminate pitch and timbre when the non-target stimulus parameter varied in Experiment 11, performance was impaired (<2 d') when the experimental design was fully counterbalanced in Experiment 12 (see the post hoc t-tests section described below for statistics relating to this result). One explanation might be because some subjects did more difficult blocks before easier blocks, thus increasing errors and decreasing d'.

Contrary to the reaction time data, there was no significant main effect of target dimension, *F*(1, 22) = 0.02, *p* = .904, indicating that discriminability was equivalent, irrespective of whether the target was the same (*M* = 2.06, *SD* = 0.38) or different (*M* = 2.05, *SD* = 0.46).

There was however a significant main effect of the non-target dimension, *F*(1, 22) = 330.46, *p* < .001, η_p^2 = .938, indicating that listeners were much worse at discriminating pitch and timbre targets when the non-target dimension differed across intervals (*M* = 1.00, *SE* = 0.42) than when the non-target dimension stayed the same (*M* = 3.11, *SD* = 0.53). The pattern of results reported for d' were comparable to those reported for the reaction time data (i.e., longer reaction times when non-targets were different), suggesting pitch non-invariance to timbre. This further evidences that irrelevant non-target dimension changes interfered with listener's ability to do pitch and timbre tasks, but also confirmed that there were no-speed accuracy trade-offs. This provides evidence in support of Objective 1.

Across conditions, there was no significant main effect of target type, *F*(1, 22) = 0.05, *p* = .824, which suggests that discriminability of pitch and timbre tasks was well equated and does not support the directional prediction of Objective 2. Timbre discriminations were not processed more accurately than pitch discriminations.

A significant two-way interaction for target dimension and non-target dimension, *F*(1, 22) = 13.45, *p* = .001, η_p^2 = .379, was also confirmed. To explore the nature of this interaction, the data were plotted in Figure 9.4. The null result of the main effect of target type justified collapsing the data across conditions for Figure 9.4.

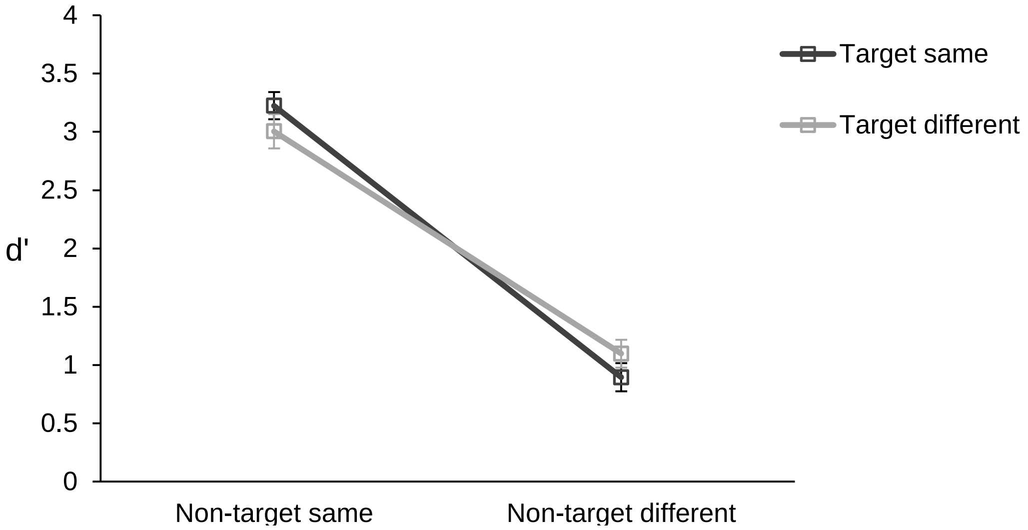

Figure 9.4. Mean d' for conditions collapsed across pitch and timbre tasks. Within-subjects Cousineau-Morey 95% confidence intervals are plotted.

Post hoc t-tests revealed that when the non-target was the same, there was no significant difference in discriminability between target same (*M* = 3.23, *SD* = 0.50) and target different (*M* = 3.00, *SD* = 0.66) conditions; *t*(22) = 2.10, *p* > .05 (after Bonferroni correction for multiple comparisons). This was also the case for when the non-target was different for target same (*M* = 0.90, *SD* = 0.45) and target different (*M* = 1.10, *SD* = 0.47) conditions; *t*(22) = -2.49, *p* > .05 (after Bonferroni correction for multiple comparisons). This further indicated that the discriminability of pitch and timbre had been appropriately matched across listeners and was similar to the pattern of results reported in Experiment 11. However within both target same and target different conditions, discriminability was significantly impaired when the non-target dimension was different (*M* = 0.90, *SD* = 0.45; *M* = 1.10, *SD* = 0.47, respectively) compared to when the non-target was the same (*M* = 3.23, *SD* = 0.50; *M* = 3.00, *SD* = 0.66, *t*(22) = 19.56, *p* < .001 and *t*(22) = 16.68, *p* < .001, respectively). Although this also broadly followed the same patterns of results reported in Experiment 11, as shown in Figure 9.4 and discussed previously, d’ fell considerably below the d’ 2 threshold when the experimental task was fully counterbalanced.

There was also a significant two way interaction between target and task type; *F*(1, 22) = 10.11, $p < .004$, $\eta_p^2 = .315$]. However after correcting for multiple comparisons, none of the post hoc t-tests related to the interaction term were significant. No other interactions were of interest or significant.

Summary

Overall comparisons of interest across reaction time and d’ sensitivity data showed evidence which suggests that:

1) pitch is non-invariant to timbre, providing support for Objective 1.
2) timbre discriminations are not processed any faster, and more accurately, than pitch discriminations, failing to find evidence for Objective 2.

9.5. Discussion

The present experiment described a novel attempt to investigate pitch invariance to timbre using a behavioural approach that explored reaction time and accuracy data for pitch and timbre discrimination tasks in a fully crossed experimental design where the stimuli were well-matched for discriminability. To recap, the main objectives were to explore whether or not the behavioural response to pitch is invariant to changes in timbre (and vice versa; Objective 1), as well as determine whether timbre discrimination performance preceded pitch discrimination performance for stimuli that been matched for discriminability (Objective 2).

9.5.1 Behavioural responses for pitch are non-invariant to timbre (and vice versa) – but is this specific to timbre or related to a more general interference effect?

If pitch is invariant to timbre, then it would be expected that there would be no significant effect for reaction time or accuracy for timbre changes and vice versa (i.e., no significant effect of the non-target dimension). This is because there would be no additional processing resource for conditions in which only the non-target dimension was ‘different’ (i.e., ‘target same, non-target different’ conditions), and therefore reaction times and accuracy would be equivalent to the ‘target same, non-target same’ conditions (i.e., no main effect of non-target and no significant target*nontarget interaction). This is based on the assumption that if pitch processing neuronal populations are only selective to pitch then responses for different timbre conditions should be equally quick and accurate regardless of whether the timbre dimension of the stimulus is the same or different (i.e., repeated or novel per se), and would strongly suggest that the neuronal populations representing these sounds overlap. However during pitch and timbre discrimination tasks discrimination performance was slower and less accurate for conditions in which the non-target dimension was different (i.e., ‘target same, non-target different’ and ‘target different, non-target different’) compared to conditions in which the non-target dimension was the same (i.e., ‘target same, non-target same’ and ‘target different, non-target same’; significant main effect of non-target and a significant target*nontarget interaction). These findings showed that both reaction time and accuracy for the target dimension were impaired when the non-target dimension differed. The interference effect supports evidence from behavioural (Borchert et al., 2011; Moore & Glasberg, 1990; Pitt, 1994; Steele & Williams, 2006) and neurophysiological studies whose results suggest that pitch is non-invariant to timbre (e.g., Bizley et al., 2009; Briley et al., 2013; Walker et al., 2011a) as well as those that indicate pitch and timbre have overlapping neural substrates in auditory cortex (Formisano et al., 2008; Warren et al., 2005; Overath et al., 2008, 2010). However, this argument assumes that this paradigm is tapping very low level feature representation only. An alternative explanation could be that the observed effects stemmed from higher level effects that are responsible for active listening. Nevertheless, future experiments in conjunction with neuroimaging methods could explore this further.

One interesting result related to the finding that there was no significant difference between ‘Target Same, Non-Target Same’ conditions and the ‘Target Different, Non-Target Same’ conditions. This is likely to be related to whether the target needing to be identified as the same or different was relevant to the task. Therefore meaning that there was facilitatory effect on discrimination because target relevant decisions needed to be made in the absence of any non-target interference.

9.5.2 Differences is discriminability across pitch and timbre tasks do not occur when you control for discriminability (Objective 2)

Overall results for reaction time and d’ sensitivity found no evidence for timbre discriminations being processing faster, or more accurately, than pitch discriminations. This finding is contrary to recent ferret evidence by Walker et al. (2011a) who report quicker reaction times by approximately 30ms for timbre discriminations compared to pitch. Walker et al. (2011a) did not match their stimuli for discriminability, so one possible explanation for their contrasting result could be that the timbre salience was greater than the pitch salience, giving rise to their results. Alternatively, it could be that the behavioural methods used here were not sensitive enough to detect such effects, or because reaction time variability between listeners was quite high (see standard deviations reported in Results section above), and thus did not pick up subtle differences.

9.5.3 Evidence against pitch constancy at level of auditory cortex?

There is growing evidence to suggest that although perceptually humans have an ability to invariantly perceive different musical instruments and human speakers, participants find it much more difficult to distinguish timbre when the pitch is varying under controlled psychophysical conditions. This may be because at lower levels of the auditory system acoustic features need to be transformed in a manner that is invariant to other features, but at higher levels in the auditory system neuronal populations responsible for pitch may generalise across other features to recognise a sound, and therefore neuronal populations in auditory cortex represent more than one acoustic feature (e.g., Bizley & Cohen, 2013). However it is perhaps more likely that it may be easier to do in an auditory scene when one has other cues over time to help separate out an auditory object from the background. Whether or not this is the case, is yet to be confirmed by future studies. Nevertheless, the psychophysical evidence does lend some support to the neurophysiological work in favour of an interdependent and non-invariant relationship between pitch and other acoustic features, such as timbre in auditory cortex (Bizley et al., 2009).

9.5.4 Summary

In summary, Experiment 12, failed to find any evidence in favour of the second pitch criterion, pitch constancy. Instead, pitch appears to be non-invariant to timbre. Pitch processing appears to be sensitive to or co-represented with other stimulus features related to its perception (i.e., timbre; evidence against pitch constancy). These findings support previous results from Experiment 10 reported in Chapter 8.

Chapter 10. General discussion

10.1. Thesis overview

The primary aim of this PhD was to evaluate the evidence for a pitch centre within auditory cortex, focusing on the four pitch criteria defined in Hall & Plack (2009):

1) Pitch selectivity
2) Pitch constancy
3) Covariation with pitch salience
4) Accounting for confounding factors

I used a combination of psychophysical, fMRI and EEG approaches. I was especially interested in exploring whether the neural representation of pitch is modular or distributed, and whether there is evidence for pitch constancy at the level of auditory cortex (i.e., are pitch-processing neurons invariant to other attributes of sound, such as timbre). This chapter summarises the main findings in each experimental chapter (Chapters 5, 8 and 9) in relation to the primary aims according to each criteria, as well as discussing the implications for future research.

10.2. Evidence for pitch selectivity

Chapter 5 used an fMRI passive listening experiment (Experiment 5) to evaluate pitch criteria 1, 3 and 4. This experiment was particularly interested in examining the neural representation of pitch and pitch salience using harmonic complex tones that were either resolved (strongly pitch salient) or unresolved (weakly pitch salient). The fMRI study found that the representation of pitch compared to noise was widely distributed across auditory cortex and localised to both primary and non-primary ROIs. This is consistent with findings from a number of fMRI papers (e.g., Garcia et al., 2010; Hall & Plack, 2009)

Chapter 8 used EEG adaptation methodology to evaluate pitch criteria 1, 2 and 4. Experiment 10 primarily aimed at investigating the second pitch criterion relating to pitch constancy, specifically pitch's invariance to timbre, using a novel adaptation design. Harmonic complex sequences were used which transitioned from noise to two pitch tones (adaptor and probe). Sequences either varied in pitch, timbre or both. Both sensor and source-based analyses revealed that the EOR and POR1 were not significantly different across sound conditions, and differences between conditions appeared to be confined to the evoked POR2 responses, thus confirming pitch and timbre selectivity.

The pitch selectivity criterion was therefore satisfied by all neurophysiological experiments reported here.

10.3. Evidence for pitch constancy

Experiment 5 showed some fMRI evidence for pitch constancy in that the incidence map for strong pitch salience > weak pitch salience did report activity in some of the primary and non-primary auditory regions that were found to be selective to pitch. However the distribution of activity still varied widely across multiple auditory areas.

Experiments 9 and 11 both showed behavioural evidence against pitch constancy in that there was a significant interference effect within pitch and timbre tasks when the non-target dimension (pitch or timbre) varied across intervals.

Experiment 10 reported in Chapter 8 revealed that both EEG sensor and source-based analyses revealed significant release in adaptation and a release from maximal adaptation, when the pitch, timbre or both varied for the probe stimuli. These results imply that different neuronal populations were required to process pitch and timbre changes, suggesting pitch may be non-invariant to timbre and thus failed to find evidence of pitch constancy at the level of the auditory cortex using this experimental approach. This lends some support to the findings of Briley et al. (2003) and Bizley et al. (2009), who both suggest pitch and timbre are co-represented in auditory cortex. However the presence of the P100, the use of different noise and adaptor durations, and the different stimulus presentation contexts for both the adaptor and the probe, might have meant that at least two different types of pitch responses were evoked in the present data, a conventional POR response for the adaptor which was preceded by noise (POR1; like Krumbholz et al., 2003), and a probe response relating to the adaptation response of the probe when preceded by a tonal adaptor (POR2; like Briley et al., 2013). The current data also cannot ascertain the source of the probe responses and whether neuronal populations coding pitch and timbre were the same, or derived for partially overlapping or different neuronal populations. Furthermore, we cannot rule out the following: i) that the POR2 responses for pitch and timbre were the same type of response, ii) whether the timbre change reflected a ‘timbre onset response’ (TOR; although the literature has not yet revealed the existence of this), and iii) whether changes in both pitch and timbre gave rise to a summation of these potentially separate onset responses. Given the spatial limitations of EEG methods, further follow-up studies would need to be conducted before we can make definitive conclusions regarding whether pitch is non-invariant to timbre.

Another related problem is associated with using adaptation methodology as the neural mechanisms underpinning this phenomenon are still not understood. Whilst adaptation can be used to probe the neuronal sensitivity to changes in stimulus features, it is difficult to interpret the adaptation effects because different models have different predictions regarding the nature of cortical responses (see Chapter 6.3 for more detail on this). Briley & Krumbholz (2013) found that the pattern of auditory adaptation results fitted the fatigue model best.

However, more work is required to explore which models best reflect auditory-evoked adaptation responses.

EEG has a number of notable strengths, namely its excellent temporal resolution, ability to provide a continuous measure of the overt and covert processing that occurs between a stimulus and a response (e.g., behavioural and/or cognitive), non-invasiveness, and cost effectiveness. However, there are also several limitations which the current experiment has highlighted. This mostly relates to determining the functional significance and location of ERP components, particularly those relating to adaptor and probe responses. These limitations are exacerbated by the limited spatial resolution of EEG because the data recorded may reflect summed contributions from many different neural generators which make it difficult to isolate the source of particular ERPs. However it is becoming increasingly popular to combine ERPs with fMRI, because using both these methods provides a spatiotemporal resolution that neither method alone can provide (see Handy, 2005, Chapter 15, p. 345). To integrate these methods a common spatial reference would be required to establish a more reliable relationship between the ERP effects and the hemodynamic imaging activations. Subsequently, dipole modelling of ERP effects would be needed to facilitate this. Nevertheless a combination of fMRI and EEG might be able to tease apart whether pitch is non-invariant to timbre in human auditory. If sources of adaptation effects for ERPs are the same or similar (overlapping) for pitch and timbre, this would suggest that pitch and timbre processing neurons may be interdependent and therefore non-invariant to one another, as suggested by Bizley et al. (2009).

Although these results should be taken with caution until precise neural representation is confirmed in conjunction with fMRI, adaptation and/or invasive methods, such as intracranial recordings (see Section 10.4), there is some evidence that pitch and timbre have overlapping neural representations in primary and non-primary areas of auditory cortex.

Formisano et al. (2008) conducted an fMRI study to measure spatial patterns of activity in auditory cortex related to speech content (i.e., ‘what’ the person is saying) or speaker identity (i.e., ‘who’ is speaking). They used three Dutch vowels and three native Dutch speakers which evoked different timbres. For both vowel and speaker identity, they found spatially distributed and bilateral patterns of activation in STG, including anterolateral HG, PT and STS, which are also established pitch processing regions of auditory cortex. Interestingly, using a bespoke data-mining algorithm, Formisano and colleagues were able to identify and distinguish between the neuronal fingerprints of listeners in ‘discriminative maps’ created for vowel identity and speaker identity signals. They found that both these maps were significantly different from one another but insensitive to acoustic variations which led them to conclude that timbre information like the low-level ‘what’ (i.e., vowel identity) or the higher-level ‘who’ (i.e., speaker identity) is not only processed at higher-level regions but also in more primary and non-primary auditory areas. This supports other studies that have also found evidence that stimulus features are encoded at multiple levels in the cortical hierarchy (e.g., Staeren et al., 2009). This is something that is often seen in the visual modality, whereby neurons represent a number of different features (e.g., retinal location, orientation and ocular dominance; Hubel and Wiesel, 1977; Briley et al., 2013). Similar to visual processing of objects or faces, at higher levels of auditory processing pitch coding might become more specialised and therefore invariant to other acoustic features. Such a specialised level representation might lie beyond the levels that generate PORs measured in Experiment 10, or alternatively their activation (or invariant presentation) might only occur under active listening conditions (Briley et al., 2013; e.g.,Talja, Alho, & Rinne, 2015).

The behavioural Experiment 12 reported in Chapter 9 used dual-pair (four-interval same-different) discrimination tasks to further investigate pitch constancy (i.e., evidence for pitch invariance to timbre). Findings indicated that timbre changes (non-target stimulus) interfered with listener's ability to discriminate pitch (target stimulus), and vice versa. These results therefore also failed to find evidence for pitch constancy in terms of performance measures, providing some support from neurophysiological studies which suggest pitch and timbre processing are interdependent (e.g., Bizley et al., 2009; Briley et al., 2013). Subsequently, whilst behavioural measures have a number of strengths relating to having a much clearer understanding about what the signal means (i.e., knowing that when a button is pressed, it is understood that any difference in reaction time relates to the amount of time it takes to encode, process and act on that stimulus condition; where longer reaction times indicate longer processing times). It is it is impossible to determine the specific cognitive processes associated with variations in reaction time and accuracy (unlike EEG). Therefore, behavioural results may not necessarily be measuring the same aspects of pitch processing as EEG or fMRI studies because one involves active listening and the other involves passive and/or active listening. Chait et al. (2006) pointed out that behaviour and electrophysiology, studied separately, might lead to different conclusions about the nature of the processing involved, and the simultaneous acquisition of neuroimaging method(s) and behavioural responses provides much more compelling evidence regarding the precise neural mechanisms that are involved.

Overall, all the experiments reported here, across three experimental methods, converged on the same conclusion and failed to find compelling evidence fulfilling the second criterion of pitch constancy, thus increasing confidence in this null result. Pitch processing may be sensitive to or co-represented with other stimulus features related to its perception (i.e., timbre; evidence against pitch constancy). However, as already noted this needs to be determined by follow-up experiments. Nevertheless, pitch constant representations might occur at the higher levels beyond auditory cortex, or that it might be easier to complete such tasks in an auditory scene when one has other cues over time to help separate out an auditory object from the background. Further work is therefore also needed using passive listening (e.g., Hall & Plack., 2009) and active task-dependent methods (e.g., Puschmann et al., 2010; Rinne, Koistinen, Salonen, & Alho, 2009; Talja et al., 2015) to interrogate when and where such pitch constant representations are formed. Alho et al. (2014) conducted a meta-analysis of 115 fMRI studies to investigate the location of pitch and spatial location activation in auditory cortex during passive and active listening conditions for speech and non-speech sounds. Comparing across studies, they found no significant differences in the median loci for pitch processing in passive compared to active listening conditions (i.e., both in middle STG, lateral to HG). Nevertheless, not many (if any) studies have directly compared results from passive vs. active pitch listening conditions within the same study. It is therefore important that future work teases apart the functional aspects of pitch coding by exploring to what extent pitch responses are modulated by task-dependent effects (e.g., a combined ERP and fMRI experiment could investigate pitch invariance to timbre by exploring the nature and location of adaptation effects, both in passive listening and active task-driven designs).

10.4. Evidence for covariation with pitch salience

The evidence for pitch salience reported in Experiment 5 was somewhat questionable given that the weak pitch salience condition was not significantly different from matched noise (did not meet first criterion of pitch selectivity). However activity did appear to be also co-localised to the same regions as the pitch response, and significantly different activation for weak pitch salience vs. noise was reported when lenient uncorrected statistics were used. Subjective ratings did reveal that listeners perceived a significant difference between weak pitch and noise sounds, but this was not detectable in the fMRI for listeners across the group. These findings raise concerns regarding fMRI's sensitivity to pitch salience effects in the context of high individual variability. Experiment 5 therefore failed to find robust evidence meeting criterion 3. Further work is needed in conjunction with other neuroimaging and behavioural methods to confirm whether or not fMRI methods are sensitive enough to detect pitch salience effects. This is because whilst fMRI has excellent spatial resolution, there are a number of limitations to using the methodology, namely relating to the fact the fMRI is not a direct measure of neural activity and the neural mechanisms underlying the hemodynamic response are still not understood. Furthermore, fMRI has poor temporal resolution which means that activation is derived from the average of multiple stimulus events that occur over several seconds. As mentioned previously, combining fMRI with other methods such as EEG provides one way of circumventing such limitations. Specifically, fMRI and EEG methods could be combined to look at whether conditions with differing pitch saliences (confirmed by behavioural measures), give rise to parametric activation in the same or different neuronal populations (i.e., do ERPs with increasing amplitudes with increasing pitch salience map onto the same focal region(s) in auditory cortex?). We would then be able to confirm whether the pitch salience effects that are often reported in MEG and EEG studies (e.g., Krishnan et al., 2012; Krumbholz et al. 2003) actually have the same or different neural locations, thus providing evidence in favour or against the pitch center hypothesis.

Additionally, intracranial recording (electrocortiography or ECoG) is a promising new invasive methodology that is becoming increasingly more popular in the field of auditory neuroscience (e.g., see a recent high-profile tinnitus case study by Sedley et al., 2015), and might be sensitive enough to determine the presence of pitch salience effects in human auditory cortex. This is because it provides a more direct measure of electrical activity (local field potentials, LFPs) of neuronal populations and offers higher spatial and temporal resolution than non-invasive methods, such as fMRI or EEG (Nourski & Howard, 2015). However its application to humans is confined to neurosurgical patients (e.g., epilepsy patients) whose treatment plan requires the placement of electrodes over the cortical surface (surface arrays) or directly into the gray matter (depth electrodes; Nourski & Howard, 2015).

The advantages of using intracranial recordings has only been exemplified in a few auditory studies of pitch (e.g., Griffiths et al., 2010; Schönwiesner & Zatorre, 2008; see Kumar & Schonwiesner, 2012 for a review). For instance, Griffiths et al. (2010) used intracranial recordings to record LFPs in human auditory cortex whilst participants listened to a noise which transitioned into a pitch-evoking stimulus (i.e., IRN). They found that time-locked evoked responses relating to the temporal regularity of the stimulus occurred irrespective of whether the IRN stimulus evoked a pitch percept (e.g., 16 Hz, which is below the lower limit of a pitch percept that is only evoked at ~30 Hz (Krumbholz et al., 2000; Pressnitzer et al., 2001). However time-frequency analysis indicated that sustained oscillatory responses shown in the high gamma range (80-120 Hz) only occurred for stimuli that evoked a pitch percept (e.g., 128 Hz and 256 Hz). Using this method, Griffiths et al. (2010) were able to distinguish between two possible types of pitch-related activity relating to stimulus regularity and perceived pitch, as well as determine that these responses were mapped to similar areas of primary and non-primary auditory cortex (i.e., medial and central HG, but maximal in medial HG). Interestingly, when the authors varied the salience of the IRN pitch stimuli (i.e., higher number of iterations = more salient pitch percept), whilst the pitch value (regularity) was kept fixed, the magnitude of both evoked responses and induced gamma power increased as a function of pitch salience throughout all areas of HG. However given that both Griffiths et al. (2010) and Schönwiesner & Zatorre (2008) used potentially confounded IRN as their pitch-evoking stimulus (see Barker et al., 2012; 2013; Steinmann & Gutschalk, 2012), it would be interestingly to examine the results that arise from intracranial recordings of different types of pitch-evoking stimuli (e.g., resolved and unresolved complex tones), and particularly whether pitch salience effects are more apparent using this methodology.

10.5. Evidence that the study designs accounted for confounding factors

As far as possible, all studies created stimuli in which the potential factors relating to uncontrolled psychoacoustic parameters were considered. All stimuli avoided the use of IRN (see Barker et al., 2012, 2013; Steinman & Gutschalk, 2012). Chapter 5 used well-controlled stimuli that accounted for temporal regularity. A continuous Gaussian wide-band background noise masker was embedded in all of the pitch stimuli at approximately 20 dB above their masked threshold to mask distortion products at lower frequencies caused by nonlinear interaction in the cochlea. Chapters 7, 8 and 9 used stimuli that were matched to activate the same gross spectral region, helping to maintain a stable envelope and prevent cochlea distortions. Stimuli used throughout Chapters 8 and 9 were piloted in Chapter 7 and then equated for pitch and timbre discriminability (and perceptual salience) in Experiments 9 and 11 to ensure differences in the salience of these stimulus parameters did not affect the pattern of results reported. This is an important control that previous pitch studies investigating pitch and timbre representations have failed to address.

Subsequently all experiment chapters fulfilled the fourth pitch criterion related to accounting for confounding factors.

10.6. Evidence for modularity and implications

All four of the pitch criteria described above need to be met in order to establish the existence of a pitch ‘centre’. Throughout all the experiments reported in Chapters 5, 6 and 8, only the first (pitch selectivity) and fourth pitch criteria (controlling for confounding variables) were adequately met. There was no compelling evidence in favour of the second pitch criterion relating to pitch constancy, and questionable evidence in favour of the third pitch criterion (pitch salience).

Overall, across all studies and experimental approaches, the cumulative evidence that has built up across this thesis appears to suggest that there is no modular representation of pitch (pitch centre), but rather pitch processing sites are distributed throughout multiple areas of the primary and non-primary auditory cortex. Evidence for a more distributed representation of pitch is most apparent in the fMRI experiment reported in Experiment 5. It is possible that such a distributed representation of pitch processing may also be sensitive to other acoustic features, as suggested by results from Experiments 10 and 12. However this requires further investigation. Nevertheless these findings might imply that different neuronal populations are required to encode pitch according to each spectral region (Micheyl & Oxenham, 2004; Shamma, 2004). Oxenham et al. (2004) conducted a series of experiments to directly investigate whether pitch processing is dependent upon purely temporal periodic cues, or whether a tonotopic representation of frequency is also required to extract pitch. In an attempt to dissociate temporal from spectral information, they used novel ‘transposed’ stimuli designed to present low-frequency pitch stimuli to high-frequency regions of the cochlea, hence eliminating spectral cues. They hypothesised that if pitch coding is only based on a temporal mechanism, then equivalent behavioural performance would be observed for participants presented with the transposed pitch stimuli compared to the pure tone stimuli, because spectral information is not relevant. However, they found markedly poorer performance for transposed tones, whereby none of their participants could accurately discriminate pitch. They suggested that transposing stimuli to arbitrary spectral locations had a severe impact on the perceived pitch, and hence was responsible for participants’ particularly poor performance. Importantly this led them to conclude that pitch extraction cannot occur unless temporal information is presented to the correct tonotopically organized spectral region. Consistent with the current results and spatio-temporal ‘hybrid’ models of pitch coding, their findings suggested that the neural mechanisms processing pitch require accurate mapping and synthesis of both temporal (pitch; time) and spectral (timbre; place) information to give rise to a perceptible pitch (Oxenham, 2008; Walker et al., 2011).

Furthermore, it might be that different models of pitch processing operate in different neuronal populations in human auditory cortex, and/or under different listening conditions. Further work is required to explore this.

10.7. Additional challenges of interpretation

10.7.1 Comparing across methods

The implementation of differing experimental methodologies has inherently led to a high number of inconsistencies within the literature. In human studies, the problem of identifying specialised pitch regions or addressing questions regarding pitch constancy is particularly pertinent because conventional neuroimaging analyses are typically confined to using subtractive methodology. Nevertheless, even in my own experiments inconsistencies were apparent. For example, the ERP Experiment 10 showed evidence of greater sensitivity to pitch (when pitch and timbre had been equated for salience), but this greater sensitivity to pitch was not observed for the behavioural reaction time Experiment 12 (when the pitch and timbre had been equated). It is important to note that although the current results compared with my previous results, and those reported by others, seem to suggest pitch invariance to timbre, the pitch mechanisms operating under different methods and designs (e.g., neurophysiological passive listening conditions compared to behavioural tasks) might actually be different and involve different brain regions. For instance, the context of stimulus presentation has been shown to modulate the magnitude of pitch-related responses, but subtraction methods do not take this into account (Garcia et al., 2010; Ulanovsky et al., 2004). Hall and Plack (2009) and Puschmann et al. (2010) also report contrasting patterns of pitch activity for multiple pitch-evoking stimuli which might be related to the difference between passive and active listening conditions, respectively.

Discriminability for pitch and timbre was also inconsistent for Experiments 11 and 12 when the task was slightly different (identifying trials that were the same in a given stimulus dimension, as well as different) and the experimental trials were fully counterbalanced. Specifically, d’ values for Experiment 12 conditions where the non-target dimension varied were less than 2. It is therefore important to try and triangulate findings with other methods, designs and species whilst controlling for other variables (e.g., stimuli, see Section 10.7.2). To address this issue and further investigate the evidence for a pitch centre, experiments could aim to confirming the observed pattern of results using another experiment method and/or in another species. Intracranial recordings in particular may allow us to bridge the gap when making comparisons across species, methods or stimuli.

10.7.2 Comparing across pitch-evoking stimuli

Different pitch stimuli often give rise to different results (Hall & Plack., 2009). For example, Experiment 5 used strong and weak pitch salience stimuli, but the consistency and distribution of activation for strong pitch salience conditions was much larger and much more distributed compared to the weak pitch salience condition. Furthermore, Experiments 6-8 illustrate how using different stimulus parameters for pitch and timbre can widely negatively impact the discriminability and perceptual salience of the stimuli. For some parameters, d’ was approaching zero. Subsequently it is important to ensure the stimuli are well controlled and that they can be used to make reliable conclusions regarding pitch-related effects (e.g., see Barker et al., 2012). Unfortunately there is no gold standard pitch stimulus that is used by all studies of pitch, which subsequently makes comparison between studies much more difficult. To address this issue, experts in the field could corroborate and confirm the use of a ‘gold standard’ pitch stimulus to be included by all future investigations of pitch, in an attempt to make the findings across studies and research groups more comparable and reliable. This expands on Hall & Plack’s (2009) recommendation to use multiple pitch-evoking stimuli. However the ‘gold standard’ pitch stimulus would probably exclude IRN given its problematic use for fMRI studies at least (see Steinmann & Gutschalk, 2012), Huggins pitch because it is not universally perceived by listeners (see Hall & Plack., 2009), and unresolved complex tones because they do not produce a salient pitch percept (may not be significantly different from noise when using fMRI; see data from Experiment 5, Section 5.4.5.1). Subsequently, wideband or resolved harmonic complex tones may be a more appropriate choice.

10.8. Future directions

Although used by multiple studies of pitch, subtraction methods are still very restricted in their ability to reliably isolate pitch-related activity. It is imperative that these limitations are now addressed by using more advantageous methodologies which involves triangulation of different methods, contexts and stimuli (as suggested above), as in the approach used throughout this thesis, but in conjunction with one another, rather than separately (i.e., EEG combined with fMRI in active and passive listening paradigms).

Multivariate pattern analysis has been used more recently to overcome problems with conventional analyses and distinguish between cortical fMRI activations to stimuli with differing acoustic features, such as pitch (e.g., Staeren et al., 2009) and timbre (e.g., Formisano et al., 2008). However the results from these studies are still only consistent with some studies and only address whether neurons in a particular area are active or not, instead of whether they are tuned to a particular stimulus dimension (Town & Bizley, 2013). Given that listeners never experience pitch as a perceptual phenomenon in isolation but in combination with varying stimulus characteristics such as timbre and/or spatial location that define an auditory object, it is increasingly important to employ experimental designs that allow for the investigation of such possible interdependent relationships (e.g., adaptation). Subsequently, reconciling invariant pitch perception with distributed and non-invariant sensitivity requires employing different experimental methods that specifically tackle the multidimensional aspects of pitch perception. Combining methods (i.e., EGG, fMRI and behavioural), or intracranial recordings, may be sensitive enough to elucidate the neural representation of pitch, pitch constant and pitch salience effects, and therefore future work should aim to use these methodologies, where possible. Nevertheless, current models of pitch perception do not take into account such multi-dimensional aspects of stimuli. It is paramount that future work aims at addressing the issues highlighted throughout this chapter in order to elucidate the representation of pitch in auditory cortex. Ultimately, this will not only aid academic understanding in the auditory neuroscience community but also prove important in better understanding cortical reorganisation of pitch processing in hearing impaired individuals (e.g. including people with tinnitus; Sereda et al., 2011).

10.9. Summary

This thesis evaluated the evidence for a pitch centre within auditory cortex, focusing on the four pitch criteria defined in Hall & Plack (2009), and using a combination of psychophysical, fMRI and EEG approaches. Whilst evidence of pitch selectivity (pitch criterion 1) was found throughout these experiments (whilst appropriate controlling for confounding factors; pitch criterion 4), the results did not provide compelling evidence for pitch constancy (pitch criterion 2), or covariation with pitch salience (pitch criterion 3). Subsequently, no evidence for a pitch center was found. Instead, the findings imply that pitch processing is distributed throughout auditory cortex, that pitch processing is sensitive to variations in other sound features related to its perception (i.e., timbre), and that this may be reflect a co-representation of pitch and timbre in auditory cortex. Under this assumption, the spatio-temporal model of pitch perception may best describe the neural mechanism underpinning pitch perception. Nevertheless, a number of alternative explanations and limitations to the current experiments are discussed, and it is argued that future work aimed at addressing these challenges to interpretation using a simultaneous combination of fMRI, EEG and psychophysical measures under passive and active listening conditions will prove fruitful in gaining insight into the neural representation of pitch in human auditory cortex.

REFERENCES

Abel, C., & Kössl, M. (2009). Sensitive response to low-frequency cochlear distortion products in the auditory midbrain. *Journal of Neurophysiology*, *101*(3), 1560–74. http://doi.org/10.1152/jn.90805.2008

Akeroyd, M. A., Moore, B. C., & Moore, G. A. (2001). Melody recognition using three types of dichotic-pitch stimulus. *The Journal of the Acoustical Society of America*, *110*(3:1), 1498–504. http://doi.org/10.1121/1.1390336

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, *17*(11), 2601–2608. http://doi.org/10.1093/cercor/bhl167

Amaro, E., & Barker, G. J. (2006). Study design in fMRI: Basic principles. *Brain and Cognition*, *60*(3), 220–232. http://doi.org/10.1016/j.bandc.2005.11.009

American National Standards Institute. (1994). *American National Standard Acoustical Terminology S1.1*. New York: Acoustical Society of America.

American Standards Association. (1960). In *Acoustical Terminology SI. 1*. New York: American Standards Association.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., … Altenmüller, E. (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage*, *30*(3), 917–926. http://doi.org/10.1016/j.neuroimage.2005.10.044

Barker, D., Plack, C. J., & Hall, D. A. (2011). Human auditory cortical responses to pitch and to pitch strength. *Neuroreport*, *22*(3), 111–115. http://doi.org/10.1097/WNR.0b013e328342ba30

Barker, D., Plack, C. J., & Hall, D. A. (2012). Reexamining the evidence for a pitch-sensitive region: A human fMRI study using iterated ripple noise. *Cerebral Cortex*, *22*(4), 745–753. http://doi.org/10.1093/cercor/bhr065

Barker, D., Plack, C. J., & Hall, D. A. (2013). Representations of pitch and slow modulation in auditory cortex. *Frontiers in Systems Neuroscience*, *7*, 62. http://doi.org/10.3389/fnsys.2013.00062

Barrett, D. J. K., & Hall, D. A. (2006). Response preferences for “what” and “where” in human non-primary auditory cortex. *NeuroImage*, *32*(2), 968–977. http://doi.org/10.1016/j.neuroimage.2006.03.050

Behroozmand, R., Korzyukov, O., & Larson, C. R. (2012). ERP correlates of pitch error detection in complex Tone and Voice auditory feedback with missing fundamental. *Brain Research*, *1448*, 89–100. http://doi.org/10.1016/j.brainres.2012.02.012

Bendor, D. (2012). Does a pitch center exist in auditory cortex? *Journal of Neurophysiology*, *107*(3), 743–746. http://doi.org/10.1152/jn.00804.2011

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165. http://doi.org/10.1038/nature03867

Bendor, D., & Wang, X. (2010). Neural coding of periodicity in marmoset auditory cortex. *Journal of Neurophysiology*, *103*(4), 1809–1822. http://doi.org/10.1152/jn.00281.2009

Bennett, C. M., Wolford, G. L., & Miller, M. B. (2009). The principled control of false positives in neuroimaging. *Social Cognitive and Affective Neuroscience*, *4*(4), 417–422. http://doi.org/10.1093/scan/nsp053

Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2004). Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience*, *16*(6), 966–977. http://doi.org/10.1162/0898929041502760

Bernstein, J. G., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *The Journal of the Acoustical Society of America*, *113*(6), 3323–3334. http://doi.org/10.1121/1.1572146

BioSemi EEG reference. (n.d.). Retrieved May 1, 2012, from www.biosemi.com/faq/cms&drl.htm

Bizley, J. K., & Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nature Reviews Neuroscience*, *14*(10), 693–707. http://doi.org/10.1038/nrn3565

Bizley, J. K., Shinn-Cunningham, B. G., & Lee, a. K. C. (2012). Nothing Is Irrelevant in a Noisy World: Sensory Illusions Reveal Obligatory within-and across-Modality Integration. *Journal of Neuroscience*, *32*(39), 13402–13410. http://doi.org/10.1523/JNEUROSCI.2495-12.2012

Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J., & Schnupp, J. W. H. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *The Journal of Neuroscience*, *29*(7), 2064–2075. http://doi.org/10.1523/JNEUROSCI.4755-08.2009

Blackman, G. A., & Hall, D. A. (2011). Reducing the effects of background noise during auditory functional magnetic resonance imaging of speech processing: qualitative and quantitative comparisons between two image acquisition schemes and noise cancellation. *Journal of Speech, Language, and Hearing Research*, *54*(2), 693–704. http://doi.org/10.1044/1092-4388(2010/10-0143)

Blamire, A. M. (2008). The technology of MRI - The next 10 years? *British Journal of Radiology*, *81*(968), 601–617. http://doi.org/10.1259/bjr/96872829

Borchert, E. M. O., Micheyl, C., & Oxenham, A. J. (2011). Perceptual grouping affects pitch judgments across time and frequency. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(1), 257–269. http://doi.org/10.1037/a0020670

Bosnyak, D. J., Eaton, R. A., & Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cerebral Cortex*, *14*(10), 1088–1099. http://doi.org/10.1093/cercor/bhh068

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. http://doi.org/10.1038/nrn756

Briley, P. M., Breakey, C., & Krumbholz, K. (2013). Evidence for Pitch chroma mapping in human auditory cortex. *Cerebral Cortex*, *23*(11), 2601–2610. http://doi.org/10.1093/cercor/bhs242

Briley, P. M., Kitterick, P. T., & Summerfield, A. Q. (2012). Evidence for Opponent Process Analysis of Sound Source Location in Humans. *Journal of the Association for Research in Otolaryngology*, *101*, 83–101. http://doi.org/10.1007/s10162-012-0356-x

Briley, P. M., & Krumbholz, K. (2013). The specificity of stimulus-specific adaptation in human auditory cortex increases with repeated exposure to the adapting stimulus. *Journal of Neurophysiology*, *110*(12), 2679–88. http://doi.org/10.1152/jn.01015.2012

Briley, P. M., & Summerfield, A. Q. (2014). Age-related deterioration of the representation of space in human auditory cortex. *Neurobiology of Aging*, *35*(3), 633–644. http://doi.org/10.1016/j.neurobiolaging.2013.08.033

Büchel, C., Wise, R. J., Mummery, C. J., Poline, J. B., & Friston, K. J. (1996). Nonlinear regression in parametric activation studies. *NeuroImage*, *4*(1), 60–66. http://doi.org/10.1006/nimg.1996.0029

Butler, B. E., & Trainor, L. J. (2012). Sequencing the cortical processing of pitch-evoking stimuli using EEG analysis and source estimation. *Frontiers in Psychology*, *3*(JUN), 1–13. http://doi.org/10.3389/fpsyg.2012.00180

Butler, R. A. (1972). The Auditory Evoked Response to Stimuli Producing Periodicity Pitch. *Psychophysiology*, *9*(2), 233–237. http://doi.org/10.1111/j.1469-8986.1972.tb00758.x

Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, *76*(3), 1698–716. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8890286

Carlyon, R. P., Long, C. J., & Micheyl, C. (2012). Across-channel timing differences as a potential code for the frequency of pure tones. *Journal of the Association for Research in Otolaryngology*, *13*(2), 159–171. http://doi.org/10.1007/s10162-011-0305-0

Cedolin, L., & Delgutte, B. (2005). Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *Journal of Neurophysiology*, *94*(1), 347–362. http://doi.org/10.1152/jn.01114.2004

Cedolin, L., & Delgutte, B. (2010). Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. *The Journal of Neuroscience*, *30*(38), 12712–12724. http://doi.org/10.1523/JNEUROSCI.6365-09.2010

Chait, M., Poeppel, D., & Simon, J. Z. (2006). Neural response correlates of detection of monaurally and binaurally created pitches in humans. *Cerebral Cortex*, *16*(6), 835–848. http://doi.org/10.1093/cercor/bhj027

Chambers, J., Akeroyd, M. A., Summerfield, A. Q., & Palmer, A. R. (2001). Active control of the volume acquisition noise in functional magnetic resonance imaging: Method and psychoacoustical evaluation. *The Journal of the Acoustical Society of America*, *110*(6), 3041. http://doi.org/10.1121/1.1408948

Chuang, C.-K., & Wang, W. S.-Y. (1978). Psychophysical pitch biases related to vowel quality, intensity difference, and sequential order. *The Journal of the Acoustical Society of America*, *64*(4), 1004. http://doi.org/10.1121/1.382083

Clare, S., Francis, S., Morris, P. G., & Bowtell, R. (2001). Single-shotT*2 measurement to establish optimum echo time for fMRI: Studies of the visual, motor, and auditory cortices at 3.0 T. *Magnetic Resonance in Medicine*, *45*(5), 930–933. http://doi.org/10.1002/mrm.1124

Cohen, D., & Erez, A. (1991). Event-Related-Potential Measurements of Cognitive Components in Response to Pitch Patterns. *Music Perception: An Interdisciplinary Journal*, *8*(4), 405–430. http://doi.org/10.2307/40285520

Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *The Journal of the Acoustical Society of America*, *98*(2:1), 862–879. http://doi.org/10.1121/1.413512

Constable, R. T., & Spencer, D. D. (2001). Repetition time in echo planar functional MRI. *Magnetic Resonance in Medicine*, *46*(4), 748–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11590651

Cramer, E. M., & Huggins, W. H. (1958). Creation of Pitch through Binaural Interaction. *The Journal of the Acoustical Society of America*, *30*(5), 413. http://doi.org/10.1121/1.1909628

Creelman, C. D., & Macmillan, N. A. (1979). Auditory phase and frequency discrimination: a comparison of nine procedures. *Journal of Experimental Psychology. Human Perception and Performance*, *5*(1), 146–156. http://doi.org/10.1037/0096-1523.5.1.146

Cusack, R., Brett, M., & Osswald, K. (2003). An Evaluation of the Use of Magnetic Field Maps to Undistort Echo-Planar Images. *NeuroImage*, *18*(1), 127–142. http://doi.org/10.1006/nimg.2002.1281

Da Costa, S., van der Zwaag, W., Marques, J. P., Frackowiak, R. S. J., Clarke, S., & Saenz, M. (2011). Human Primary Auditory Cortex Follows the Shape of Heschl's Gyrus. *Journal of Neuroscience*, *31*(40), 14067–14075. http://doi.org/10.1523/JNEUROSCI.2000-11.2011

Dai, H., & Micheyl, C. (2012). Separating the contributions of primary and unwanted cues in psychophysical studies. *Psychological Review*, *119*(4), 770–88. http://doi.org/10.1037/a0029343

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human Brain Mapping*, *5*(5), 329–40. http://doi.org/10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.0.CO;2-5

Darwin, C., & Carlyon, R. (1995). Auditory Grouping. In *Hearing* (pp. 387–424). Elsevier. http://doi.org/10.1016/B978-012505626-7/50013-3

De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R. a., Clarke, S., & Murray, M. M. (2010). Perceptual and semantic contributions to repetition priming of environmental sounds. *Cerebral Cortex*, *20*(7), 1676–1684. http://doi.org/10.1093/cercor/bhp230

De Martino, F., Schmitter, S., Moerel, M., Tian, J., Ugurbil, K., Formisano, E., … de Moortele, P.-F. (2012). Spin echo functional MRI in bilateral auditory cortices at 7T: An application of B1 shimming. *NeuroImage*, *63*(3), 1313–1320. http://doi.org/10.1016/j.neuroimage.2012.08.029

Degerman, A., Rinne, T., Salmi, J., Salonen, O., & Alho, K. (2006). Selective attention to sound location or pitch studied with fMRI. *Brain Research*, *1077*(1), 123–34. http://doi.org/10.1016/j.brainres.2006.01.025

Dehaene, S., Naccache, L., Cohen, L., Bihan, D. L., Mangin, J. F., Poline, J. B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature Neuroscience*, *4*(7), 752–758. http://doi.org/10.1038/89551

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13494–13499. http://doi.org/10.1073/pnas.93.24.13494

Desmond, J. E., & Glover, G. H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: Statistical power analyses. *Journal of Neuroscience Methods*, *118*(2), 115–128. http://doi.org/10.1016/S0165-0270(02)00121-8

Deutsch, D., & Roll, P. L. (1976). Separate “what” and “where” decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology. Human Perception and Performance*, *2*(1), 23–29. http://doi.org/10.1037/0096-1523.2.1.23

Duyn, J. (2004). Receiver Technology. Second international workshop on parallel MRI. Zurich, Switzerland. Retrieved from www.mr.ethz.ch/parallelmri04/abstracts/pub/Duyn.pdf

Edmister, W. B., Talavage, T. M., Ledden, P. J., & Weisskoff, R. M. (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Human Brain Mapping*, *7*(2), 89–97. http://doi.org/10.1002/(SICI)1097-0193(1999)7:2<89::AID-HBM2>3.0.CO;2-N

Eickhoff, S. B. (2005). The Human Parietal Operculum. I. Cytoarchitectonic Mapping of Subdivisions. *Cerebral Cortex*, *16*(2), 254–267. http://doi.org/10.1093/cercor/bhi105

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325–1335. http://doi.org/10.1016/j.neuroimage.2004.12.034

Ernst, S. M. A., Verhey, J. L., & Uppenkamp, S. (2008). Spatial dissociation of changes of level and signal-to-noise ratio in auditory cortex for tones in noise. *NeuroImage*, *43*(2), 321–328. http://doi.org/10.1016/j.neuroimage.2008.07.046

Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., & Peters, T. M. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. In *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference* (pp. 1813–1817). IEEE. http://doi.org/10.1109/NSSMIC.1993.373602

Faulkner, A. (1985). Pitch discrimination of harmonic complex signals: residue pitch or multiple component discriminations? *The Journal of the Acoustical Society of America*, *78*(6), 1993–2004. http://doi.org/10.1121/1.392656

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Mitchell Steinschneider. (1998). Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. *Brain Research*, *786*(1-2), 18–30. http://doi.org/10.1016/S0006-8993(97)01423-6

Fletcher, H. (1924). The Physical Criterion for Determining the Pitch of a Musical Tone. *Physical Review*, *23*(3), 427–437. http://doi.org/10.1103/PhysRev.23.427

Fodor, J. A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*. MIT Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=e7nrSeibJZYC&pgis=1

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). “Who” Is Saying “What”? Brain-Based Decoding of Human Voice and Speech. *Science*, *322*(5903), 970–973. http://doi.org/10.1126/science.1164318

Foster, J. R., Hall, D. A., Summerfield, A. Q., Palmer, A. R., & Bowtell, R. W. (2000). Sound-Level Measurements and Calculations of Safe Noise Dosage During EPI at 3 T. *Journal of Magnetic Resonance Imaging*, *12*(1), 157–163. http://doi.org/10.1002/1522-2586(200007)12:1<157::AID-JMRI17>3.0.CO;2-M

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J., Heather, J. D., & Frackowiak, R. S. J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, *3*(3), 165–189. http://doi.org/10.1002/hbm.460030303

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage*, *7*(1), 30–40. http://doi.org/10.1006/nimg.1997.0306

Friston, K. J., Holmes, A. P., & Worsley, K. J. (1999). How many subjects constitute a study? *NeuroImage*, *10*(1), 1–5. http://doi.org/10.1006/nimg.1999.0439

Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., & Frith, C. D. (1996b). Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage*, *4*(3:1), 223–35. http://doi.org/10.1006/nimg.1996.0074

Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouquette, S., & Poline, J. (2000). To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage*, *12*(2), 196–208. http://doi.org/10.1006/nimg.2000.0609

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S., & Dolan, R. J. (1996a). The trouble with cognitive subtraction. *NeuroImage*, *4*(2), 97–104. http://doi.org/10.1006/nimg.1996.0033

Garcia, D., Hall, D. A., & Plack, C. J. (2010). The effect of stimulus context on pitch representations in the human auditory cortex. *NeuroImage*, *51*(2), 808–816. http://doi.org/10.1016/j.neuroimage.2010.02.079

Genovese, C. R., Lazar, N. A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, *15*(4), 870–8. http://doi.org/10.1006/nimg.2001.1037

Gizewski, E. R., Maderwald, S., Wanke, I., Goehde, S., Forsting, M., & Ladd, M. E. (2005). Comparison of volume, four- and eight-channel head coils using standard and parallel imaging. *European Radiology*, *15*(8), 1555–1562. http://doi.org/10.1007/s00330-005-2730-0

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, *47*(1-2), 103–138. http://doi.org/10.1016/0378-5955(90)90170-T

Gockel, H. E., Carlyon, R. P., & Plack, C. J. (2011). Combination of spectral and binaurally created harmonics in a common central pitch processor. *Journal of the Association for Research in Otolaryngology*, *12*(2), 253–260. http://doi.org/10.1007/s10162-010-0250-3

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *The Journal of the Acoustical Society of America*, *54*(6), 1496. http://doi.org/10.1121/1.1914448

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. John Wiley and Sons. Retrieved from https://books.google.co.uk/books/about/Signal_Detection_Theory_and_Psychophysic.html?id=yYsQAQAAIAAJ&redir_esc=y

Grey, J. M. (1978). Timbre discrimination in musical patterns. *The Journal of the Acoustical Society of America*, *64*(2), 467–472. http://doi.org/10.1121/1.382018

Griffiths, T. D. (2012). Cortical Mechanisms for Pitch Representation. *Journal of Neuroscience*, *32*(39), 13333–13334. http://doi.org/10.1523/JNEUROSCI.1661-12.2012

Griffiths, T. D., Büchel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nns*, *1*(5), 422–.

Griffiths, T. D., & Hall, D. a. (2012). Mapping Pitch Representation in Neural Ensembles with fMRI. *Journal of Neuroscience*, *32*(39), 13343–13347. http://doi.org/10.1523/JNEUROSCI.3813-12.2012

Griffiths, T. D., Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., … Howard, M. a. (2010). Direct recordings of pitch responses from human auditory cortex. *Current Biology*, *20*(12), 1128–1132. http://doi.org/10.1016/j.cub.2010.04.044

Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., & Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience*, *4*(6), 633–637. http://doi.org/10.1038/88459

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. http://doi.org/10.1016/j.tics.2005.11.006

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187–203. http://doi.org/10.1016/S0896-6273(00)80832-6

Guimaraes, a. R., Melcher, J. R., Talavage, T. M., Baker, J. R., Ledden, P., Rosen, B. R., … Weisskoff, R. M. (1998). Imaging subcortical auditory activity in humans. *Human Brain Mapping*, *6*(1), 33–41. http://doi.org/10.1002/(SICI)1097-0193(1998)6:1<33::AID-HBM3>3.0.CO;2-M

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., & Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *NeuroImage*, *15*(1), 207–216. http://doi.org/10.1006/nimg.2001.0949

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, *22*(2), 755–766. http://doi.org/10.1016/j.neuroimage.2004.01.025

Hall, D. A., & Barker, D. (2012). Coding of Basic Acoustical and Perceptual Components of Sound in Human Auditory Cortex. In D. Poeppel, T. Overath, A. N. Popper, & R. R. Fay (Eds.), *The Human Auditory Cortex* (Vol. 43). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4614-2314-0

Hall, D. A., Barrett, D. J. K., Akeroyd, M. A., & Summerfield, A. Q. (2005). Cortical representations of temporal structure in sound. *Journal of Neurophysiology*, *94*(5), 3181–3191. http://doi.org/10.1152/jn.00271.2005

Hall, D. A., Edmondson-Jones, A. M., & Fridriksson, J. (2006). Periodicity and frequency coding in human auditory cortex. *European Journal of Neuroscience*, *24*(12), 3601–3610. http://doi.org/10.1111/j.1460-9568.2006.05240.x

Hall, D. A., Haggard, M. P., Akeroyd, M. a., Palmer, A. R., Summerfield, a. Q., Elliott, M. R., … Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain*

Mapping, *7*(3), 213–223. http://doi.org/10.1002/(SICI)1097-0193(1999)7:3<213::AID-HBM5>3.0.CO;2-N

Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, *19*(3), 576–585. http://doi.org/10.1093/cercor/bhn108

Hall, D. A., Summerfield, A. Q., Gonçalves, M. S., Foster, J. R., Palmer, A. R., & Bowtell, R. W. (2000). Time-course of the auditory BOLD response to scanner noise. *Magnetic Resonance in Medicine*, *43*(4), 601–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10748437

Hall, D. A., & Susi, K. (2015). Hemodynamic imaging of the auditory cortex. In G. G. Celesia & G. Hickok (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 257–275). http://doi.org/10.1016/B978-0-444-62630-1.00015-9

Hall, D. A., & Wild, D. C. (2012). Fundamental principles underlying MRI and functional MRI. In K. Tremblay & R. Burkard (Eds.), *Translational Perspectives in Auditory Neuroscience: Hearing Across the Lifespan – Assessment and Disorders.* (pp. 283–315). San Diego: Plural Publishing Inc.

Handel, S., & Erickson, M. L. (2001). A Rule of Thumb: The Bandwidth for Timbre Invariance Is One Octave. *Music Perception*, *19*(1), 121–126. http://doi.org/10.1525/mp.2001.19.1.121

Handy, T. C. (2005). *Event-related Potentials: A Methods Handbook*. MIT Press. Retrieved from https://books.google.co.uk/books?id=OQyZEfgEzRUC&printsec=frontcover&dq=Event-related+Potentials:+A+Methods+Handbook&hl=en&sa=X&ei=k9KZVda-GITB7gaUmJDQAg&ved=0CCEQ6AEwAA#v=onepage&q=Event-related Potentials: A Methods Handbook&f=false

Hautus, M. J., van Hout, D., & Lee, H.-S. (2009). Variants of A Not-A and 2AFC tests: Signal Detection Theory models. *Food Quality and Preference*, *20*(3), 222–229. http://doi.org/10.1016/j.foodqual.2008.10.002

Healy, E. W., Moser, D. C., Morrow-Odom, K. L., Hall, D. A., & Fridriksson, J. (2007). Speech perception in MRI scanner noise by persons with aphasia. *Journal of Speech, Language, and Hearing Research*, *50*(2), 323–34. http://doi.org/10.1044/1092-4388(2007/023)

Heeger, D. (n.d.). Lecture Notes. Retrieved from http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/pitch/pitch.html

Henson, R. (2007). Efficient experimental design for fMRI. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, 193–210. http://doi.org/10.1016/B978-012372560-8/50015-2

Henson, R., Büchel, C., Josephs, O., & Friston, K. (1999). The Slice-Timing Problem in Event-related fMRI. *NeuroImage*, *9*, 125. Retrieved from http://discovery.ucl.ac.uk/5679/1/5679.pdf

Henson, R. N. ., & Rugg, M. . (2003). Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia*, *41*(3), 263–270. http://doi.org/10.1016/S0028-3932(02)00159-8

Henson, R. N. A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, *70*(1), 53–81. http://doi.org/10.1016/S0301-0082(03)00086-8

Hoechstetter, K., Berg, P., & Scherg, M. (2010). *BESA Research Tutorial 3: Batch Scripts, Multiple Subjects & Conditions, MATLAB-Interface*. Retrieved from ftp://www.besa.de/be/besa.de/demonstrations_and_tutorials/BESA Research Tutorial 3 - Batch Scripts - Multiple Subjects+Conditions - MATLAB interface.pdf

Hoechstetter, K., Rupp, A., Meinck, H. M., Weckesser, D., Bornfleth, H., Stippich, C., … Scherg, M. (2000). Magnetic source imaging of tactile input shows task-independent attention effects in SII. *Neuroreport*, *11*(11), 2461–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10943704

Holmes, A., & Friston, K. (1998). Generalisability, Random Effects & Population Inference. *Neuroimage*, *7(4(2/3)*, S754. Retrieved from http://ftp.fil.ion.ucl.ac.uk/spm/doc/papers/aph_rfx.pdf

Houtsma, A. J. M., & Goldstein, J. L. (1972). The Central Origin of the Pitch of Complex Tones: Evidence from Musical Interval Recognition. *The Journal of the Acoustical Society of America*, *51*(2B), 520. http://doi.org/10.1121/1.1912873

Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *The Journal of the Acoustical Society of America*, *87*(1), 304. http://doi.org/10.1121/1.399297

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, *160*(1), 106–154.2. http://doi.org/10.1523/JNEUROSCI.1991-09.2009

Huettel, S., Song, A., & McCarthy, G. (2009). *Functional magnetic resonance imaging* (2nd ed.). Sunderland, USA: Sinauer Associates Inc. http://doi.org/978-0-87893-286-3

Humphries, C., Liebenthal, E., & Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *NeuroImage*, *50*(3), 1202–1211. http://doi.org/10.1016/j.neuroimage.2010.01.046

Ille, N., Berg, P., & Scherg, M. (2002). Artifact Correction of the Ongoing EEG Using Spatial Filters Based on Artifact and Brain Signal Topographies. *Journal of Clinical Neurophysiology*, *19*(2), 113–124. http://doi.org/10.1097/00004691-200203000-00002

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., … Belliveau, J. W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6809–6814. http://doi.org/10.1073/pnas.0303760101

Jackson, H. M., & Moore, B. C. J. (2013). The dominant region for the pitch of complex tones with low fundamental frequencies. *The Journal of the Acoustical Society of America*, *134*(2), 1193–204. http://doi.org/10.1121/1.4812754

Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, *34*(1), 65–73. http://doi.org/10.1002/mrm.1910340111

Johnsrude, I. S. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain*, *123*(1), 155–163. http://doi.org/10.1093/brain/123.1.155

Kim, D. Il, Mathalon, D. H., Ford, J. M., Mannell, M., Turner, J. A., Brown, G. G., … Calhoun, V. D. (2009). Auditory oddball deficits in schizophrenia: an independent component analysis of the fMRI multisite function BIRN study. *Schizophrenia Bulletin*, *35*(1), 67–81. http://doi.org/10.1093/schbul/sbn133

Kim, J., Kim, Y., Lee, S., Seo, J.-H., Song, H.-J., Cho, J. H., & Chang, Y. (2012). Alteration of functional connectivity in tinnitus brain revealed by resting-state fMRI?: A pilot study. *International Journal of Audiology*, *51*(5), 413–417. http://doi.org/10.3109/14992027.2011.652677

King, A. J., & Nelken, I. (2009). Unraveling the principles of auditory cortical processing: can we learn from the visual system? *Nature Neuroscience*, *12*(6), 698–701. http://doi.org/10.1038/nn.2308

Kojima, S., & Kiritani, S. (1989). Vocal-auditory functions in the chimpanzee: Vowel perception. *International Journal of Primatology*, *10*(3), 199–213. http://doi.org/10.1007/BF02735200

Kreitewolf, J., Lewald, J., & Getzmann, S. (2011). Effect of attention on cortical processing of sound motion: An EEG study. *NeuroImage*, *54*(3), 2340–2349. http://doi.org/10.1016/j.neuroimage.2010.10.031

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. a. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*(5), 250–256. http://doi.org/10.1016/j.tins.2006.02.008

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. http://doi.org/10.1038/nn.2303

Krishnan, A., Bidelman, G. M., & Gandour, J. T. (2010). Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing Research*, *268*(1-2), 60–66. http://doi.org/10.1016/j.heares.2010.04.016

Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., & Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, *50*(12), 2849–2859. http://doi.org/10.1016/j.neuropsychologia.2012.08.013

Krishnan, A., & Plack, C. J. (2011). Neural encoding in the human brainstem relevant to the pitch of complex tones. *Hearing Research*, *275*(1-2), 110–119. http://doi.org/10.1016/j.heares.2010.12.008

Kropotov, J. D. (2009). *Quantitative EEG, Event-Related Potentials and Neurotherapy*. *Quantitative EEG, Event-Related Potentials and Neurotherapy*. Elsevier. http://doi.org/10.1016/B978-0-12-374512-5.00021-9

Krumbholz, K., Patterson, R. D., & Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *The Journal of the Acoustical Society of America*, *108*(3), 1170. http://doi.org/10.1121/1.1287843

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, *13*(7), 765–772. http://doi.org/10.1093/cercor/13.7.765

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 739–751. http://doi.org/10.1037//0096-1523.18.3.739

Kumar, S., & Schonwiesner, M. (2012). Mapping Human Pitch Representation in a Distributed System Using Depth-Electrode Recordings and Modeling. *Journal of Neuroscience*, *32*(39), 13348–13351. http://doi.org/10.1523/JNEUROSCI.3812-12.2012

Kumar, S., Sedley, W., Nourski, K. V., Kawasaki, H., Oya, H., Patterson, R. D., … Griffiths, T. D. (2011). Predictive Coding and Pitch Processing in the Auditory Cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3084–3094. http://doi.org/10.1162/jocn_a_00021

Labuschagne, I. B., & Hanekom, J. J. (2013). Preparation of stimuli for timbre perception studies. *The Journal of the Acoustical Society of America*, *134*(3), 2256–67. http://doi.org/10.1121/1.4817877

Langner, G., Sams, M., Heil, P., & Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology*, *181*(6), 665–676. http://doi.org/10.1007/s003590050148

Lanting, C. P., Briley, P. M., Sumner, C. J., & Krumbholz, K. (2013). Mechanisms of adaptation in human auditory cortex. *Journal of Neurophysiology*, *110*(4), 973–83. http://doi.org/10.1152/jn.00547.2012

Larsen, E., Cedolin, L., & Delgutte, B. (2008). Pitch representations in the auditory nerve: two concurrent complex tones. *Journal of Neurophysiology*, *100*(3), 1301–19. http://doi.org/10.1152/jn.01361.2007

Levitt, H. (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, *49*(2B), 467. http://doi.org/10.1121/1.1912375

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia*, *7*(4), 128–134. http://doi.org/10.1007/BF02156143

Licklider, J. C. R. (1954). “Periodicity” Pitch and “Place” Pitch. *The Journal of the Acoustical Society of America*, *26*(5), 945. http://doi.org/10.1121/1.1928005

Loeb, G. E., White, M. W., & Merzenich, M. M. (1983). Spatial cross-correlation. A proposed mechanism for acoustic pitch perception. *Biological Cybernetics*, *47*(3), 149–163. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6615914

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–157. http://doi.org/10.1038/35084005

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. MIT Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=y4-uAwAAQBAJ&pgis=1

Lütkenhöner, B., & Steinsträter, O. (1998). High-precision neuromagnetic study of the functional organization of the human auditory cortex. *Audiology & Neuro-Otology*, *3*(2-3), 191–213. http://doi.org/10.1159/000013790

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User’s Guide* (2nd Editio). Mahwah, NJ: Erlbaum. Retrieved from https://books.google.com/books?hl=en&lr=&id=P094AgAAQBAJ&pgis=1

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199. http://doi.org/10.1037//0033-2909.98.1.185

Makeig, S., Jung, T.-P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences*, *94*(20), 10979–10984. http://doi.org/10.1073/pnas.94.20.10979

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–24. http://doi.org/10.3758/s13428-011-0168-7

McAlpine, D. (2004). Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. *Journal of Neurophysiology*, *92*(3), 1295–311. http://doi.org/10.1152/jn.00034.2004

McDermott, H. J. (2004). Music perception with cochlear implants: a review. *Trends in Amplification*, *8*(2), 49–82. http://doi.org/10.1177/108471380400800203

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263–1271. http://doi.org/10.1111/j.1467-9280.2008.02235.x

Meddis, R., & O’Mard, L. (1997). A unitary model of pitch perception. *The Journal of the Acoustical Society of America*, *102*(3), 1811–1820. http://doi.org/10.1121/1.420088

Micheyl, C., & Dai, H. (2009). Likelihood ratio, optimal decision rules, and relationship between proportion correct and d’ in the dual-pair AB-versus-BA identification paradigm. *Attention, Perception & Psychophysics*, *71*(6), 1426–33. http://doi.org/10.3758/APP.71.6.1426

Micheyl, C., & Messing, D. P. (2006). Likelihood ratio, optimal decision rules, and correct response probabilities in a signal detection theoretic, equal-variance Gaussian model of the observer in the 4IAX paradigm. *Perception & Psychophysics*, *68*(5), 725–735. http://doi.org/10.3758/BF03193696

Micheyl, C., & Oxenham, A. J. (2004). Sequential F0 comparisons between resolved and unresolved harmonics: no evidence for translation noise between two pitch mechanisms. *The Journal of the Acoustical Society of America*, *116*(5), 3038–3050. http://doi.org/10.1121/1.1806825

Micheyl, C., & Oxenham, A. J. (2005). Comparing F0 discrimination in sequential and simultaneous conditions. *The Journal of the Acoustical Society of America*, *118*(1), 41–44. http://doi.org/10.1121/1.1929228

Moazami-Goudarzi, M., Michels, L., Weisz, N., & Jeanmonod, D. (2010). Temporo-insular enhancement of EEG low and high frequencies in patients with chronic tinnitus. QEEG study of chronic tinnitus patients. *BMC Neuroscience*, *11*, 40. http://doi.org/10.1186/1471-2202-11-40

Mondor, T. A., Hurlburt, J., & Thorne, L. (2003). Categorizing sounds by pitch: effects of stimulus similarity and response repetition. *Perception & Psychophysics*, *65*(1), 107–114. http://doi.org/10.3758/BF03194787

Moore, B. C. J. (1974). Relation between the critical bandwidth and the frequency-difference limen. *The Journal of the Acoustical Society of America*, *55*(2), 359. http://doi.org/10.1121/1.1914512

Moore, B. C. J. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *The Journal of the Acoustical Society of America*, *77*(5), 1853. http://doi.org/10.1121/1.391936

Moore, B. C. J., & Glasberg, B. R. (1990). Frequency discrimination of complex tones with overlapping and non-overlapping harmonics. *The Journal of the Acoustical Society of America*, *87*(5), 2163–2177. http://doi.org/10.1121/1.399184

Moore, B. C. J., Glasberg, B. R., & Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *The Journal of the Acoustical Society of America*, *75*(2), 550–561. http://doi.org/10.1121/1.390527

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, *13*(4), 684–701. http://doi.org/10.1006/nimg.2000.0715

Näätänen, R., & Picton, T. (1987). The N1 Wave of the Human Electric and Magnetic Response to Sound: A Review and an Analysis of the Component Structure. *Psychophysiology*, *24*(4), 375–425. http://doi.org/10.1111/j.1469-8986.1987.tb00311.x

Naccache, L., & Dehaene, S. (2001). The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex (New York, N.Y. : 1991)*, *11*(10), 966–974. http://doi.org/10.1093/cercor/11.10.966

Nelken, I. (2008). Processing of complex sounds in the auditory system. *Current Opinion in Neurobiology*, *18*(4), 413–417. http://doi.org/10.1016/j.conb.2008.08.014

Nelken, I., & Ulanovsky, N. (2007). Mismatch negativity and stimulus-specific adaptation in animal models. *Journal of Psychophysiology*, *21*(3-4), 214–223. http://doi.org/10.1027/0269-8803.21.34.214

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*(3), 653–660. http://doi.org/10.1016/j.neuroimage.2004.12.005

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*, *12*(5), 419–446. http://doi.org/10.1191/0962280203sm341ra

Noreen, D. (1981). Optimal decision rules for some common psychophysical paradigms. In S. Grossberg (Ed.), *Mathematical psychology and psychophysiology (Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics)* (pp. 237–279). Providence, RI: American Mathematical Society.

Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical Pitch Regions in Humans Respond Primarily to Resolved Harmonics and Are Located in Specific Tonotopic Regions of Anterior Auditory Cortex. *Journal of Neuroscience*, *33*(50), 19451–19469. http://doi.org/10.1523/JNEUROSCI.2880-13.2013

Nourski, K. V, & Howard, M. A. (2015). Invasive recordings in the human auditory cortex. *Handbook of Clinical Neurology*, *129*, 225–44. http://doi.org/10.1016/B978-0-444-62630-1.00013-5

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, *87*(24), 9868–9872. http://doi.org/10.1073/pnas.87.24.9868

Ohms, V. R., Gill, A., Van Heijningen, C. A. A., Beckers, G. J. L., & ten Cate, C. (2010). Zebra finches exhibit speaker-independent phonetic perception of human speech. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1684), 1003–1009. http://doi.org/10.1098/rspb.2009.1788

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. http://doi.org/10.1016/0028-3932(71)90067-4

Olulade, O., Hu, S., Gonzalez-Castillo, J., Tamer, G. G., Luh, W.-M., Ulmer, J. L., & Talavage, T. M. (2011). Assessment of temporal state-dependent interactions between auditory fMRI responses to desired and undesired acoustic sources. *Hearing Research*, *277*(1-2), 67–77. http://doi.org/10.1016/j.heares.2011.03.008

Overath, T., Kumar, S., Stewart, L., von Kriegstein, K., Cusack, R., Rees, A., & Griffiths, T. D. (2010). Cortical Mechanisms for the Segregation and Representation of Acoustic Textures. *Journal of Neuroscience*, *30*(6), 2070–2076. http://doi.org/10.1523/JNEUROSCI.5378-09.2010

Overath, T., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2008). Encoding of Spectral Correlation over Time in Auditory Cortex. *Journal of Neuroscience*, *28*(49), 13268–13273. http://doi.org/10.1523/JNEUROSCI.4596-08.2008

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends in Amplification*, *12*(4), 316–331. http://doi.org/10.1177/1084713808325881

Oxenham, A. J. (2012). Pitch Perception. *Journal of Neuroscience*, *32*(39), 13335–13338. http://doi.org/10.1523/JNEUROSCI.3815-12.2012

Oxenham, A. J., Bernstein, J. G. W., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(5), 1421–1425. http://doi.org/10.1073/pnas.0306958101

Oxenham, A. J., Micheyl, C., & Keebler, M. V. (2009). Can temporal fine structure represent the fundamental frequency of unresolved harmonics? *The Journal of the Acoustical Society of America*, *125*(4), 2189–2199. http://doi.org/10.1121/1.3089220

Oxenham, A. J., Micheyl, C., Keebler, M. V, Loper, A., & Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(18), 7629–7634. http://doi.org/10.1073/pnas.1015291108

Paltoglou, A. E., Sumner, C. J., & Hall, D. A. (2009). Examining the role of frequency specificity in the enhancement and suppression of human cortical activity by auditory selective attention. *Hearing Research*, *257*(1-2), 106–118. http://doi.org/10.1016/j.heares.2009.08.007

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767–776. http://doi.org/10.1016/S0896-6273(02)01060-7

Penagos, H. (2004). A Neural Representation of Pitch Salience in Nonprimary Human Auditory Cortex Revealed with Functional Magnetic Resonance Imaging. *Journal of Neuroscience*, *24*(30), 6810–6815. http://doi.org/10.1523/JNEUROSCI.0383-04.2004

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. a, Johnson, R., … Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, *37*(2), 127–152. http://doi.org/10.1111/1469-8986.3720127

Plack, C. J. (2005). *The Sense of Hearing*. Lawrence Erlbaum Associates. Retrieved from https://books.google.com/books?id=wpYSS8o0PeoC&pgis=1

Plack, C. J., Barker, D., & Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research*, *307*(June), 53–64. http://doi.org/10.1016/j.heares.2013.07.020

Plack, C. J., & Oxenham, A. J. (2005). The Psychophysics of Pitch. In *Pitch: Neural Coding and perception* (pp. 7–55). New York: Springer-Verlag. http://doi.org/10.1007/0-387-28958-5_2

Plack, C., Oxenham, A., & Fay, R. (2005). *Pitch: Neural Coding and Perception*. (C. J. Plack, R. R. Fay, A. J. Oxenham, & A. N. Popper, Eds.) (Vol. 24). New York, NY: Springer New York. http://doi.org/10.1007/0-387-28958-5

Plomp, R. (1967). Pitch of Complex Tones. *The Journal of the Acoustical Society of America*, *41*(6), 1526. http://doi.org/10.1121/1.1910515

Poeppel, D., & Hickok, G. (2015). Electromagnetic recording of the auditory system. *Handbook of Clinical Neurology*, *129*, 245–55. http://doi.org/10.1016/B978-0-444-62630-1.00014-7

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70. http://doi.org/10.1093/scan/nsm006

Poldrack, R. A., Mumford, J., & Nichols, T. (2011). *Handbook of Functional MRI Data Analysis*. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9780511895029

Pressnitzer, D., & Patterson, R. (2001). Distortion products and the perceived pitch of harmonic complex tones. In D. J. Breebart, A. J. M. Houtsma, A. Kohlrausch, V. F. & Prijs, & R. Schoonoven (Eds.), *Physiological and Psychophysical Bases of Auditory Function* (pp. 97–104). Shaker, Maastricht. Retrieved from http://129.199.80.1/Audition/dp/pdfs/pressnitzer-2001-distortion_pitch.pdf

Pressnitzer, D., Patterson, R. D., & Krumbholz, K. (2001). The lower limit of melodic pitch. *The Journal of the Acoustical Society of America*, *109*(5 Pt 1), 2074–2084. http://doi.org/10.1121/1.1359797

Price, C. J., Moore, C. J., & Friston, K. J. (1997). Subtractions, conjunctions, and interactions in experimental design of activation studies. *Human Brain Mapping*, *5*(4), 264–272. http://doi.org/10.1002/(SICI)1097-0193(1997)5:4<264::AID-HBM11>3.0.CO;2-E

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*, *49*(2), 1641–1649. http://doi.org/10.1016/j.neuroimage.2009.09.045

Qin, M. K., & Oxenham, A. J. (2005). Effects of Envelope-Vocoder Processing on F0 Discrimination and Concurrent-Vowel Identification. *Ear and Hearing*, *26*(5), 451–460. http://doi.org/10.1097/01.aud.0000179689.79868.06

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*. http://doi.org/10.1037/0033-2909.114.3.510

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. http://doi.org/10.1016/j.edurev.2010.12.001

Rinne, T., Koistinen, S., Salonen, O., & Alho, K. (2009). Task-dependent activations of human auditory cortex during pitch discrimination and pitch memory tasks. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(42), 13338–43. http://doi.org/10.1523/JNEUROSCI.3012-09.2009

Ritsma, R. J. (1963). On Pitch Discrimination of Residue Tones. *International Journal of Audiology*. http://doi.org/10.3109/05384916309070127

Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *The Journal of the Acoustical Society of America*, *42*(1), 191–198. http://doi.org/10.1121/1.1942972

Ritter, S., Dosch, H. G., Specht, H. J., & Rupp, A. (2005). Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. *NeuroImage*, *27*(3), 533–543. http://doi.org/10.1016/j.neuroimage.2005.05.003

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex*, *15*(8), 1261–1269. http://doi.org/10.1093/cercor/bhi009

Rose, J. E., Brugge, J. F., Anderson, D. J., & Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, *30*(4), 769–793. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4962851

Rousseau, B., & Ennis, D. M. (2001). A Thurstonian model for the dual pair (4IAX) discrimination method. *Perception & Psychophysics*, *63*(6), 1083–1090. http://doi.org/10.3758/BF03194541

Russo, F. A., & Thompson, W. F. (2005). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, *67*(4), 559–568. http://doi.org/10.3758/BF03193514

Sanei, S., & Chambers, J. a. (2007). *EEG Signal Processing*. *Chemistry & biodiversity* (Vol. 1). http://doi.org/10.1002/9780470511923

Sartori, G., & Umiltà, C. (2000). How to avoid the fallacies of cognitive subtraction in brain imaging. *Brain and Language*, *74*(2), 191–212. http://doi.org/10.1006/brln.2000.2334

Scarabino, T., Giannatempo, G. M., Popolizio, T., Tosetti, M., d'Alesio, V., Esposito, F., … Salvolini, U. (2007). 3.0-T functional brain imaging: a 5-year experience. *La Radiologia Medica*, *112*(1), 97–112. http://doi.org/10.1007/s11547-007-0124-x

Schacter, D. L., & Buckner, R. L. (1998). Priming and the brain. *Neuron*, *20*(2), 185–195. http://doi.org/10.1016/S0896-6273(00)80448-1

Schäfer, A., van der Zwaag, W., Francis, S. T., Head, K. E., Gowland, P. A., & Bowtell, R. W. (2008). High resolution SE-fMRI in humans at 3 and 7 T using a motor task. *Magnetic Resonance Materials in Physics, Biology and Medicine*, *21*(1-2), 113–120. http://doi.org/10.1007/s10334-007-0099-6

Scherg, M., Berg, P., & Hoechstetter, K. (2010). *BESA Research Tutorial 5: Artifact Correction*. Retrieved from ftp://www.besa.de/be/besa.de/demonstrations_and_tutorials/BESA Research Tutorial 5 - Artifact Correction.pdf

Schnupp, J. W. H., & Bizley, J. K. (2010). On pitch, the ear and the brain of the beholder. Focus on "neural coding of periodicity in marmoset auditory cortex.". *Journal of Neurophysiology*, *103*(4), 1708–1711. http://doi.org/10.1152/jn.00182.2010

Schönwiesner, M., & Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschl's gyrus. *Experimental Brain Research*, *187*(1), 97–105. http://doi.org/10.1007/s00221-008-1286-z

Schouten, J. F. (1940). The residue and the mechanism of hearing. *Proceedings of the Koninklijke Nederlandse Academie van Wetenschappen*, *43*, 991–999. Retrieved from http://www.dwc.knaw.nl/DL/publications/PU00017508.pdf

Schouten, J. F. (1962). Pitch of the Residue. *The Journal of the Acoustical Society of America*, *34*(9B), 1418. http://doi.org/10.1121/1.1918360

Schroeder, M. R. (1968). Period Histogram and Product Spectrum: New Methods for Fundamental-Frequency Measurement. *The Journal of the Acoustical Society of America*, *43*(4), 829. http://doi.org/10.1121/1.1910902

Schwarz, D. W., & Tomlinson, R. W. (1990). Spectral response patterns of auditory cortex neurons to harmonic complex tones in alert monkey (Macaca mulatta). *Journal of Neurophysiology*, *64*(1), 282–298. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2388072

Sedley, W., Gander, P. E., Kumar, S., Oya, H., Kovach, C. K., Nourski, K. V., … Griffiths, T. D. (2015). Intracranial Mapping of a Cortical Tinnitus System using Residual Inhibition. *Current Biology*, *25*(9), 1208–14. http://doi.org/10.1016/j.cub.2015.02.075

Seither-Preisler, A., Johnson, L., Krumbholz, K., Nobbe, A., Patterson, R., Seither, S., & Lütkenhöner, B. (2007). Tone sequences with conflicting fundamental pitch and timbre changes are heard differently by musicians and nonmusicians. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(3), 743–751. http://doi.org/10.1037/0096-1523.33.3.743

Seither-Preisler, A., Krumbholz, K., Patterson, R., Seither, S., & Lütkenhöner, B. (2004). Interaction between the neuromagnetic responses to sound energy onset and pitch onset suggests common generators. *European Journal of Neuroscience*, *19*(11), 3073–3080. http://doi.org/10.1111/j.0953-816X.2004.03423.x

Seither-Preisler, A., Patterson, R. D., Krumbholz, K., Seither, S., & Lütkenhöner, B. (2006). From noise to pitch: Transient and sustained responses of the auditory evoked field. *Hearing Research*, *218*(1-2), 50–63. http://doi.org/10.1016/j.heares.2006.04.005

Semal, C., & Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *The Journal of the Acoustical Society of America*, *120*(6), 3907. http://doi.org/10.1121/1.2357708

Sereda, M., Hall, D. a, Bosnyak, D. J., Edmondson-Jones, M., Roberts, L. E., Adjamian, P., & Palmer, A. R. (2011). Re-examining the relationship between audiometric profile and tinnitus pitch. *International Journal of Audiology*, *50*(5), 303–312. http://doi.org/10.3109/14992027.2010.551221

Shackleton, T. M., & Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *The Journal of the Acoustical Society of America*, *95*(6), 3529–3540. http://doi.org/10.1121/1.409970

Shahin, A., Bosnyak, D. J., Trainor, L. J., & Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *The Journal of Neuroscience*, *23*(13), 5545–5552. http://doi.org/23/13/5545

Shahin, A. J., Roberts, L. E., Chau, W., Trainor, L. J., & Miller, L. M. (2008). Music training leads to the development of timbre-specific gamma band activity. *NeuroImage*, *41*(1), 113–122. http://doi.org/10.1016/j.neuroimage.2008.01.067

Shahin, A. J., Roberts, L. E., Miller, L. M., McDonald, K. L., & Alain, C. (2007). Sensitivity of EEG and MEG to the N1 and P2 auditory evoked responses modulated by spectral complexity of sounds. *Brain Topography*, *20*(2), 55–61. http://doi.org/10.1007/s10548-007-0031-4

Shamma, S. A. (1985). Speech processing in the auditory system I: The representation of speech sounds in the responses of the auditory nerve. *The Journal of the Acoustical Society of America*, *78*(5), 1612. http://doi.org/10.1121/1.392799

Shamma, S. A. (2004). Topographic organization is essential for pitch perception. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(5), 1114–1115. http://doi.org/10.1073/pnas.0307334101

Shamma, S., & Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *The Journal of the Acoustical Society of America*, *107*(5), 2631. http://doi.org/10.1121/1.428649

Sigalovsky, I. S., & Melcher, J. R. (2006). Effects of sound level on fMRI activation in human brainstem, thalamic and cortical centers. *Hearing Research*, *215*(1-2), 67–76. http://doi.org/10.1016/j.heares.2006.03.002

Singh, P. G., & Hirsh, I. J. (1992). Influence of spectral locus and F0 changes on the pitch and timbre of complex tones. *The Journal of the Acoustical Society of America*, *92*(5), 2650–2661. http://doi.org/10.1121/1.404381

Soeta, Y., Nakagawa, S., & Tonoike, M. (2005). Auditory evoked magnetic fields in relation to iterated rippled noise. *Hearing Research*, *205*(1-2), 256–61. http://doi.org/10.1016/j.heares.2005.03.026

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound Categories Are Represented as Distributed Patterns in the Human Auditory Cortex. *Current Biology*, *19*(6), 498–502. http://doi.org/10.1016/j.cub.2009.01.066

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *31*(1), 137–149. http://doi.org/10.3758/BF03207704

Steele, K. M., & Williams, A. K. (2006). Is the Bandwidth for Timbre Invariance Only One Octave? *Music Perception*, *23*(3), 215–220. http://doi.org/10.1525/mp.2006.23.3.215

Steinmann, I., & Gutschalk, A. (2012). Sustained BOLD and theta activity in auditory cortex are related to slow stimulus fluctuations rather than to pitch. *Journal of Neurophysiology*, *107*(12), 3458–3467. http://doi.org/10.1152/jn.01105.2011

Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., & Arezzo, J. C. (1998). Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. *The Journal of the Acoustical Society of America*, *104*(5), 2935–2955. http://doi.org/10.1121/1.423877

Talavage, T. M., Ledden, P. J., Benson, R. R., Rosen, B. R., & Melcher, J. R. (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hearing Research*, *150*(1-2), 225–244. http://doi.org/10.1016/S0378-5955(00)00203-3

Talja, S., Alho, K., & Rinne, T. (2015). Source Analysis of Event-Related Potentials During Pitch Discrimination and Pitch Memory Tasks. *Brain Topography*, 1–14. http://doi.org/10.1007/s10548-013-0307-9

Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *The Journal of the Acoustical Society of America*, *71*(3), 679. http://doi.org/10.1121/1.387544

Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., & Poline, J.-B. (2007). Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses. *NeuroImage*, *35*(1), 105–20. http://doi.org/10.1016/j.neuroimage.2006.11.054

Town, S. M., & Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Frontiers in Systems Neuroscience*, *7*(November), 88. http://doi.org/10.3389/fnsys.2013.00088

Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., & Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: auditory cortex. *Annals of the New York Academy of Sciences*, *1060*(1), 148–174. http://doi.org/10.1196/annals.1360.011

Triantafyllou, C., Hoge, R. D., Krueger, G., Wiggins, C. J., Potthast, A., Wiggins, G. C., & Wald, L. L. (2005). Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *NeuroImage*, *26*(1), 243–250. http://doi.org/10.1016/j.neuroimage.2005.01.007

Van der Zwaag, W., Francis, S., Head, K., Peters, A., Gowland, P., Morris, P., & Bowtell, R. (2009). fMRI at 1.5, 3 and 7 T: Characterising BOLD signal changes. *NeuroImage*, *47*(4), 1425–1434. http://doi.org/10.1016/j.neuroimage.2009.05.015

Vidyasagar, R., Stancak, A., & Parkes, L. M. (2010). A multimodal brain imaging study of repetition suppression in the human visual cortex. *NeuroImage*, *49*(2), 1612–1621. http://doi.org/10.1016/j.neuroimage.2009.10.020

Von Helmholtz, H. L. F. (1948). *The sensations of tone: As a physiological basis for the theory of music (6th ed.). Notes* (Vol. 12). New York: Peter Smith. http://doi.org/10.1037/12740-000

Vurma, A. (2014). Timbre-induced pitch shift from the perspective of Signal Detection Theory: the impact of musical expertise, silence interval, and pitch region. *Frontiers in Psychology*, *5*, 44. http://doi.org/10.3389/fpsyg.2014.00044

Vurma, A., & Ross, J. (2006). Pitch perception of sounds with different timbre. In *9th International conference on Music Perception and Cognition*. Alma Mater Studiorum University of Bologna. Retrieved from http://www.marcocosta.it/icmpc2006/pdfs/269.pdf

Wald, L. L. (2012). The future of acquisition speed, coverage, sensitivity, and resolution. *NeuroImage*, *62*(2), 1221–1229. http://doi.org/10.1016/j.neuroimage.2012.02.077

Walker, K., Bizley, J., King, A., & Schnupp, J. (2011a). Multiplexed and Robust Representations of Sound Features in Auditory Cortex, *31*(41), 14565–14576. http://doi.org/10.1523/JNEUROSCI.2074-11.2011

Walker, K. M. M., Bizley, J. K., King, A. J., & Schnupp, J. W. H. (2011b). Cortical encoding of pitch: Recent results and open questions. *Hearing Research*, *271*(1-2), 74–87. http://doi.org/10.1016/j.heares.2010.04.015

Walker, K. M. M., Schnupp, J. W. H., Hart-Schnupp, S. M. B., King, A. J., & Bizley, J. K. (2009). Pitch discrimination by ferrets for simple and complex sounds. *The Journal of the Acoustical Society of America*, *126*(3), 1321. http://doi.org/10.1121/1.3179676

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *The Journal of Neuroscience*, *23*(13), 5799–804. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12843284

Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *NeuroImage*, *24*(4), 1052–1057. http://doi.org/10.1016/j.neuroimage.2004.10.031

Warren, J. D., Uppenkamp, S., Patterson, R. D., & Griffiths, T. D. (2003). Analyzing Pitch Chroma and Pitch Height in the Human Brain. *Annals of the New York Academy of Sciences*, *999*(17), 212–214. http://doi.org/10.1196/annals.1284.032

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, *64*(2), 198–207. http://doi.org/10.3758/BF03195786

Westbury, C. F., Zatorre, R. J., & Evans, a. C. (1999). Quantifying variability in the planum temporale: A probability map. *Cerebral Cortex*, *9*(4), 392–405. http://doi.org/10.1093/cercor/9.4.392

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. New York, NY: Oxford University Press. http://doi.org/10.1093/acprof:oso/9780195092509.001.0001

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology*, *8*(2), 227–233. http://doi.org/10.1016/S0959-4388(98)80144-X

Wightman, F. L. (1973). The Pattern-Transformation Model of Pitch. *The Journal of the Acoustical Society of America*, *53*(1), 315. http://doi.org/10.1121/1.1982283

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited--again. *NeuroImage*, *2*(3), 173–81. http://doi.org/10.1006/nimg.1995.1023

Yost, W. A. (1996). Pitch strength discrimination for iterated rippled noise. *The Journal of the Acoustical Society of America*, *99*(4), 2490. http://doi.org/10.1121/1.415613

Zarate, J. M., Ritson, C. R., & Poeppel, D. (2013). The Effect of Instrumental Timbre on Interval Discrimination. *PLoS ONE*, *8*(9). http://doi.org/10.1371/journal.pone.0075410

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. http://doi.org/10.1016/S1364-6613(00)01816-7

Zeki, S. (1983). Colour coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, *9*(4), 741–765. http://doi.org/10.1016/0306-4522(83)90265-8

Zhao, F., Wang, P., & Kim, S.-G. (2004). Cortical depth-dependent gradient-echo and spin-echo BOLD fMRI at 9.4T. *Magnetic Resonance in Medicine*, *51*(3), 518–24. http://doi.org/10.1002/mrm.10720

APPENDIX

Third party copyright material removed. See Hall, D. A., & Susi, K. (2015). Hemodynamic imaging of the auditory cortex. In G. G. Celesia & G. Hickok (Eds.), *Handbook of Clinical Neurology* (Vol. 129, pp. 257–275). http://doi.org/10.1016/B978-0-444-62630-1.00015-9