

REPRESENTATION OF  
STATISTICAL SOUND PROPERTIES IN  
HUMAN AUDITORY CORTEX

A thesis submitted for the degree of Doctor of Philosophy  
University College London  
February 2009

by  
Tobias Overath

Wellcome Trust Centre for Neuroimaging  
Institute of Neurology

## **DECLARATION**

I, Tobias Overath, hereby confirm that the work presented in this thesis is my own. Where information is derived from other sources, I confirm that this has been indicated in the thesis.

Tobias Overath

Date

# ABSTRACT

The work carried out in this doctoral thesis investigated the representation of statistical sound properties in human auditory cortex. It addressed four key aspects in auditory neuroscience: the representation of different analysis time windows in auditory cortex; mechanisms for the analysis and segregation of auditory objects; information-theoretic constraints on pitch sequence processing; and the analysis of local and global pitch patterns. The majority of the studies employed a parametric design in which the statistical properties of a single acoustic parameter were altered along a continuum, while keeping other sound properties fixed.

The thesis is divided into four parts. Part I (Chapter 1) examines principles of anatomical and functional organisation that constrain the problems addressed. Part II (Chapter 2) introduces approaches to digital stimulus design, principles of functional magnetic resonance imaging (fMRI), and the analysis of fMRI data. Part III (Chapters 3-6) reports five experimental studies. Study 1 controlled the spectrotemporal correlation in complex acoustic spectra and showed that activity in auditory association cortex increases as a function of spectrotemporal correlation. Study 2 demonstrated a functional hierarchy of the representation of auditory object boundaries and object salience. Studies 3 and 4 investigated cortical mechanisms for encoding entropy in pitch sequences and showed that the planum temporale acts as a computational hub, requiring more computational resources for sequences with high entropy than for those with high redundancy. Study 5 provided evidence for a hierarchical organisation of local and global pitch pattern processing in neurologically normal participants. Finally, Part IV (Chapter 7) concludes with a general discussion of the results and future perspectives.

---

# TABLE OF CONTENTS

ABBREVIATIONS .....	VIII
LIST OF FIGURES .....	X
LIST OF TABLES .....	XIII
ACKNOWLEDGMENTS.....	XIV
CHAPTER 1. GENERAL INTRODUCTION .....	1
1.1 Functional anatomy of the auditory system .....	3
1.1.1 Macroscopic organisation of the auditory cortex.....	4
1.1.2 Cytoarchitecture.....	8
1.1.3 Electrophysiology.....	11
1.1.4 Tonotopy.....	12
1.1.5 Spectrotemporal receptive fields.....	14
1.1.6 Processing streams .....	15
1.2 Experimental approaches to auditory cortex function.....	17
1.2.1 Auditory object analysis .....	17
1.2.2 Pitch.....	20
1.2.3 Timbre.....	23
1.2.4 Sound and pitch sequences.....	25
1.2.4.1 Mismatch negativity .....	26
1.2.4.2 Auditory streaming.....	28
1.2.4.3 Complex pitch sequences.....	32
1.3 Key problems addressed in this thesis .....	35
1.3.1 Chapter 3 – Study 1 .....	36
1.3.2 Chapter 4 – Study 2.....	37
1.3.3 Chapter 5 – Studies 3 & 4.....	37
1.3.4 Chapter 6 – Study 5.....	38
CHAPTER 2. TECHNIQUES AND METHODS.....	39
2.1 Approaches to stimulus design for auditory neuroscience .....	39
2.1.1 Simple synthetic stimuli.....	40

2.1.1.1 Pure tones and decomposition of complex sounds into pure tone components .....	41
2.1.1.2 Iterated rippled noise (IRN) .....	43
2.1.1.3 Amplitude modulation (AM) .....	44
2.1.1.4 Frequency modulation (FM) .....	45
2.1.2 Sampled natural stimuli.....	45
2.1.3 Complex probabilistic stimuli.....	47
2.1.3.1 Spectrotemporal correlation in complex AM spectra.....	47
2.1.3.2 Spectrotemporal coherence .....	49
2.1.3.3 Information theoretic properties of pitch sequences.....	50
2.2 Magnetic resonance imaging .....	51
2.3 Functional magnetic resonance imaging .....	54
2.3.1 Echo-planar imaging .....	54
2.3.2 Physiological basis of BOLD signal and haemodynamic response function .....	55
2.4 fMRI and auditory stimulus presentation .....	56
2.5 Image pre-processing.....	59
2.5.1 Realignment and unwarping .....	59
2.5.2 Normalisation.....	60
2.5.3 Smoothing.....	61
2.6 Statistical analysis.....	61
2.6.1 General Linear Model (GLM) .....	61
2.6.2 Gaussian Random Field (GRF) theory.....	62
2.6.3 Random-effects analysis.....	63
CHAPTER 3. ENCODING OF SPECTROTEMPORAL CORRELATION IN COMPLEX ACOUSTIC SPECTRA .....	65
Summary .....	65
3.1 Introduction .....	66
3.2 Materials and Methods .....	69
3.2.1 Participants .....	69
3.2.2 Stimuli .....	69
3.2.3 Experimental design .....	70
3.2.4 fMRI protocol and analysis.....	71
3.3 Results .....	73
3.4 Discussion.....	80

---

CHAPTER 4. ENCODING OF SPECTROTEMPORAL COHERENCE IN ‘AUDITORY TEXTURES’ .....	86
Summary .....	86
4.1 Introduction .....	87
4.2 Materials and Methods .....	91
4.2.1 Participants .....	91
4.2.2 Stimuli .....	91
4.2.3 Experimental design .....	91
4.2.4 Behavioural data analysis .....	94
4.2.5 fMRI protocol and analysis.....	94
4.3 Results .....	96
4.4 Discussion.....	102
CHAPTER 5. ENCODING OF THE RATE OF INFORMATION PRODUCTION IN PITCH SEQUENCES .....	105
Summary .....	105
5.1 Introduction .....	106
5.2 Materials and Methods (Study 3) .....	110
5.2.1 Participants .....	110
5.2.2 Stimuli .....	110
5.2.3 Experimental design .....	111
5.2.4 fMRI protocol and analysis.....	112
5.3 Results (Study 3).....	114
5.4 Materials and Methods (Study 4) .....	117
5.4.1 Participants .....	117
5.4.2 Stimuli .....	117
5.4.3 Experimental design .....	118
5.4.4 fMRI protocol and analysis.....	118
5.5 Results (Study 4).....	120
5.6 Further analysis considerations.....	130
5.7 Discussion.....	132
CHAPTER 6. HIERARCHICAL ENCODING OF GLOBAL AND LOCAL INFORMATION IN PITCH PATTERNS .....	136
Summary .....	136

## TABLE OF CONTENTS

---

6.1 Introduction .....	137
6.2 Materials and Methods .....	139
6.2.1 Participants .....	139
6.2.2 Stimuli and Experimental Procedure.....	139
6.2.3 fMRI protocol and analysis.....	140
6.3 Results .....	143
6.3.1 Behavioural results .....	143
6.3.2 Effects of processing contour-preserved and contour-violated differences.....	144
6.3.3 Comparison of Local and Global Processing .....	146
6.4 Discussion.....	148
CHAPTER 7. GENERAL DISCUSSION.....	154
7.1 The length of analysis time windows in auditory cortex increases from primary to association cortex.....	156
7.2 The segregation and representation of auditory objects are hierarchically linked in auditory cortex.....	159
7.3 The planum temporale (PT) acts as a ‘computational hub’ .....	161
7.4 Local and global information in pitch patterns is hierarchically organised .....	163
7.5 Key future problems and implications for auditory neuroscience .....	165
BIBLIOGRAPHY.....	169
APPENDIX I: AUTHOR CONTRIBUTIONS .....	211
APPENDIX II: PUBLICATIONS ARISING FROM THIS THESIS .....	213

---

**ABBREVIATIONS**

<b>2I2AFC</b>	two-interval two-alternatives forced choice
<b>AAC</b>	auditory association cortex
<b>AC</b>	auditory cortex
<b>AL</b>	anterolateral field of non-human primate
<b>AM</b>	amplitude modulation
<b>ANOVA</b>	analysis of variance
<b>AST</b>	asymmetric sampling in time
<b>aSTG</b>	anterior STG
<b>BF, CF</b>	best frequency, centre frequency
<b>BOLD</b>	blood oxygenation level dependent
<b>cHG</b>	central HG
<b>CL</b>	caudolateral field of non-human primate
<b>CM</b>	caudomedial field of non-human primate
<b>CN</b>	cochlear nucleus
<b>CPB</b>	caudal parabelt field of non-human primate
<b>dB</b>	decibels
<b>DCM</b>	dynamic causal modelling
<b>dHb</b>	de-oxyhaemoglobin
<b>DFT</b>	discrete Fourier transform
<b>DLPFC</b>	dorsolateral prefrontal cortex
<b>ECD</b>	equivalent current dipole
<b>EEG</b>	electroencephalogram
<b>EPI</b>	echo planar imaging
<b>ERP</b>	event related potential
<b>FFT</b>	fast Fourier transform
<b>FIR</b>	finite impulse response
<b>FM</b>	frequency modulation
<b>fMRI</b>	functional MRI
<b>FWE</b>	family-wise error
<b>FWHM</b>	full-width-at-half-maximum
<b>GRF</b>	Gaussian random field
<b>GLM</b>	general linear model
<b>HG</b>	Heschl's gyrus
<b>HRF</b>	haemodynamic response function
<b>Hz, kHz</b>	Hertz, kilo Hertz
<b>IC</b>	inferior colliculus
<b>IDFT</b>	inverse DFT
<b>IFFT</b>	inverse FFT
<b>IFG</b>	inferior frontal gyrus
<b>ILD</b>	inter-aural level difference
<b>IPS</b>	inferior parietal sulcus
<b>IRN</b>	iterated rippled noise
<b>ISI</b>	inter stimulus interval
<b>ITD</b>	inter-aural time difference
<b>LB</b>	lateral belt field of non-human primate
<b>LFP</b>	local field potential
<b>LHD</b>	left hemispheric damage

<b>IHG</b>	lateral HG
<b>LL</b>	lateral lemniscus (of the pons)
<b>LS</b>	lateral sulcus
<b>MB</b>	medial belt field of non-human primate
<b>MDS</b>	multidimensional scaling
<b>MEG</b>	magnetoencephalogram
<b>MGB</b>	medial geniculate body
<b>mHG</b>	medial HG
<b>ML</b>	middle lateral field of non-human primate
<b>MM</b>	middle medial field of non-human primate
<b>MMN</b>	mismatch negativity
<b>MNI</b>	Montreal Neurological Institute
<b>MRI</b>	magnetic resonance imaging
<b>mSFG</b>	medial superior frontal gyrus
<b>MUA</b>	multi-unit activity
<b>NC</b>	normal control (participant)
<b>NMR</b>	nuclear magnetic resonance
<b>oHb</b>	oxyhaemoglobin
<b>PAC</b>	primary auditory cortex
<b>PET</b>	positron emission tomography
<b>PP</b>	planum polare
<b>pSTS</b>	posterior STS
<b>PT</b>	planum temporale
<b>RF</b>	radio frequency
<b>RHD</b>	right hemispheric damage
<b>RM</b>	rostromedial field of non-human primate
<b>RPB</b>	rostral parabelt field of non-human primate
<b>RTL</b>	rostrotemporal lateral field of non-human primate
<b>RTM</b>	rostrotemporal medial field of non-human primate
<b>SAC</b>	secondary auditory cortex
<b>SAM</b>	sinusoidal amplitude modulation
<b>SD</b>	standard deviation
<b>SEM</b>	standard error of the mean
<b>SNR</b>	signal to noise ratio
<b>SOC</b>	superior olivary complex
<b>SPL</b>	sound pressure level
<b>SPM</b>	statistical parametric mapping
<b>STG</b>	superior temporal gyrus
<b>STL</b>	superior temporal lobe
<b>STP</b>	superior temporal plane
<b>STRF</b>	spectrotemporal receptive field
<b>TE</b>	time to echo
<b>TPJ</b>	temporo-parietal junction
<b>TR</b>	time to repeat
<b>VOI</b>	volume of interest
<b>VPLC</b>	ventrolateral prefrontal cortex

---

**LIST OF FIGURES**

Figure 1-1	Schematic drawings of the cerebral cortex and the location of auditory cortex in several primates .....	5
Figure 1-2	Schematic diagram of monkey auditory cortex illustrated for the macaque monkey.....	6
Figure 1-3	Parcellation of the human superior temporal cortex by different investigators .....	7
Figure 1-4	Frequency responses at six different positions along the basilar membrane (see von Békésy, 1960) .....	13
Figure 3-1	Auditory stimulus. Spectrograms of representative stimuli from each level of correlation. ....	68
Figure 3-2	Psychometric functions from participants 1-16 .....	74
Figure 3-3	Main results. Areas increasing in activity as a function of spectrotemporal correlation (red) and areas responding to sound in general (blue) .....	75
Figure 3-4	STS lateralisation. Areas showing a significantly stronger increase in activity in right STS than left STS (red) together with areas that show an increase as a function of correlation (blue).....	76
Figure 3-5	Plots of the two local maxima in left aSTG.....	77
Figure 3-6	Plots of the two local maxima in right aSTG.....	78
Figure 3-7	Plots of the local maximum in left PT.....	78
Figure 3-8	Plots of the two local maxima in right PT .....	79
Figure 4-1	Auditory stimulus. Example of a block of sound with four spectrotemporal coherence segments showing absolute coherence values for each segment and the corresponding change in coherence between the segments. ....	90
Figure 4-2	Ramp density .....	93
Figure 4-3	Psychometric functions from all participants .....	97

LIST OF FIGURES

---

Figure 4-4  $d'$  scores for detecting changes in coherence in the MRI scanner .....98

Figure 4-5 Main results. Areas showing an increased haemodynamic response as a function of increasing absolute coherence (blue) and change in coherence (red).....99

Figure 4-6 Changes in increasing vs. decreasing relative coherence ..... 100

Figure 4-7 Pilot study results .....102

Figure 5-1 Auditory stimulus. Examples of fractal waveforms (blue) and the related pitch sequences (red, rounded to the nearest integer) based on inverse Fourier transforms of  $f^{-n}$  power spectra, with exponent  $n = 0$  (top),  $n = 0.9$  (middle),  $n = 1.5$  (bottom) ..... 108

Figure 5-2 Main results I. Areas showing an increase in BOLD signal as a function of increasing entropy (red) and areas that responded to sound in general (blue) ..... 115

Figure 5-3 Experimental design ..... 122

Figure 5-4 Behavioural results. Hits (*hit*) and correct rejections (*cr*) are displayed for the three Delay periods (1, 2, 3) across the five Levels of exponent  $n$ ..... 123

Figure 5-5 Main results II. Areas showing an increase in BOLD signal as a function of increasing entropy (red) and areas that responded to sound in general ..... 124

Figure 5-6 BOLD signal in medial and lateral HG ..... 125

Figure 5-7 Response in IC and MGB. Entropy increase (red) and [Sound – Silence] (blue) contrasts ..... 127

Figure 5-8 Retrieval and comparison results. Areas that show stronger activation for retrieval and comparison than encoding ..... 128

Figure 5-9 Analysis considerations I..... 130

Figure 5-10 Analysis considerations II..... 132

Figure 6-1 Experimental design. Schematic of two consecutive trials ..... 141

---

Figure 6-2	Schematic of the experimental design and modelled haemodynamic response functions (hrf's) .....	143
Figure 6-3	Main effects for <i>Local</i> and <i>Global</i> . Activations for the <i>Local</i> ([Ldiff – Same]) (red) and <i>Global</i> ([Gdiff – Same]) (blue) contrasts .....	144
Figure 6-4	Main effects for <i>Local</i> and <i>Global</i> ( $p < 0.05$ ).....	145
Figure 6-5	Results for the lateralisation test of (left) <i>Local</i> ([Ldiff – Same]) for original – flipped scans, and (right) <i>Global</i> ([Gdiff – Same]) for original – flipped scans.....	146
Figure 6-6	Activations for the <i>Local</i> – <i>Global</i> ([Ldiff – Gdiff]) contrast .....	147
Figure 6-7	Results for the lateralisation test of <i>Local</i> - <i>Global</i> ([Ldiff – Gdiff]) for original – flipped scans.....	148

**LIST OF TABLES**

Table 3-1	MNI coordinates of local maxima.....	76
Table 4-1	Stereotactic MNI-coordinates. ....	100
Table 5-1	Mean sample entropy $H_{\text{SampEn}}$ values (standard error of the mean in parentheses) of the pitch sequences across levels in the two studies...	114
Table 5-2	MNI coordinates of local maxima in PT as a function of increasing entropy in the two studies .....	125
Table 5-3	Local maxima coordinates for the main effect of Retrieval and comparison.....	129
Table 5-4	Descriptive data for the pitch sequences in the two studies with a classification scheme based on the sample entropy estimates. ....	131
Table 6-1	Stereotactic coordinates for the three contrasts <i>Local</i> , <i>Global</i> , and <i>Local – Global</i> . ....	145

## ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to my principal supervisor Tim Griffiths for his encouragement and support throughout. Despite numerous ‘attempts’ of mine to jeopardise the various experiments and their results with my scepticism, Tim always made sure I wasn’t successful and brought me back on the right track. I would also like to thank my secondary supervisor Cathy Price for her help and support.

I am extremely grateful to Katharina von Kriegstein for the many hours helping with scanning, for essentially teaching me SPM and functional imaging data analysis, and for always finding a way out of what until then seemed like a dead end. Sukhbinder Kumar provided invaluable help with Matlab, sacrificed hours of his time discussing stimulus manipulations via Skype and somehow managed to stay friendly throughout, isn’t it?

I would also like to thank Rhodri Cusack for his advice and often ingenious ideas on experimental design and analysis; I am grateful for his selfless help, be it from Cambridge, Paris, San Francisco, or in transit.

Not surprisingly, this thesis would have had to be a theoretical body of work without the numerous participants volunteering to spend what surely sometimes must have seemed like half an eternity in a dark and narrow bore listening to weird sounds. Many thanks!

Peter Aston made me feel welcome from the very first day on. I would like to thank the Functional Imaging Team (Amanda Brennan, Janice Glensman, and David Bradbury) for always being there at the right moment. A special thank you also to Eric Featherstone and Alphonso Reid for helping out with Cogent scripts and uncountable hours fixing the electrostatic headphones. Ric Davis, Rachael Maddock and Chris Freemantle were always very helpful in figuring out PC and networking issues.

A special thank you goes to Marta Garrido, who was a constant presence literally from the first day and with whom I shared many a democratic coffee. I would

also like to thank the many fellows of the FIL for their selfless advice and help with scanning and analysis issues, with special thanks to Alex Leff in particular. Karl Friston and Will Penny also often came to my rescue, for which I am grateful.

Manon Grube seems to have the ability always to be just in the right place so that I can appreciate her tremendous hospitality; I hope that some time the situation reverses so that I can return the favour. Lauren Stewart and Jason Warren were ideal collaborators and I greatly benefited from their advice and knowledge. I also thank Bob Carlyon for candid discussions and playing devil's advocate.

I would like to thank my parents and sister for supporting me in so many ways throughout the many years of my academic education, which is about to culminate in the official closure of student-hood. Last but not least, this thesis would not be what it is now without the 24/7, 4·365 stalwart support of my wife, Lillian; for far too long we needed to overcome some 3,500 miles and gazillions of electrons through iChat, and I am deeply grateful for her commitment and sacrifice spending uncounted hours in airports, on planes and in transit.



## Chapter 1. GENERAL INTRODUCTION

The acoustic environment comprises a multitude of simultaneous and consecutive acoustic events. The human auditory system has evolved remarkable capabilities to decode the acoustic information within this complex heterophony with apparent ease. For example, when engaging in a conversation with a friend at a reception while multiple other people speak simultaneously and music plays in the background, we are able to follow the conversation while ignoring the interfering acoustic information from the other sound sources (the so-called ‘cocktail party effect’; Cherry, 1953).

This thesis addresses fundamental mechanisms of auditory perception in such complex and dynamically changing acoustic environments. Specifically, this thesis investigates cortical principles for segregating and grouping elements within the auditory scene (Bregman, 1990), both at the level of individual auditory objects (Griffiths & Warren, 2004) and at the level of grouped objects or linked streams of objects. Such auditory scene analysis requires the auditory system to assess the statistical properties of the acoustic signal so as to extract those spectrotemporal patterns of the signal that comprise the relevant acoustic features in the signal (e.g. the conversation with the friend), while simultaneously filtering out irrelevant or redundant information (e.g. the background noise emanating from the rest of the room).

From an information-theoretic perspective, the auditory system needs to assess those statistical properties that contribute relevant information comprising an auditory object and optimise its coding of these features. At the same time, the neural code representing auditory objects must be robust enough to allow one and the same auditory object to be recognised irrespective of different instantiations (e.g. recognising the vowel /a/ irrespective of whether it is pronounced with or without background noise or by different speakers). The auditory system must further establish object boundary properties, which allow it to detect transitions between auditory objects, and thus segregate between auditory objects at points in time when those properties change.

The information-theoretic approach regards the brain as a Bayesian inference generator, which forms predictions from the statistical properties of sensory input and evaluates these predictions based on stored, experience-dependent templates or priors (Friston, 2003a, 2005). Within this framework, the auditory system is constantly evaluating the incoming signal with respect to its statistical properties, from which it forms predictions that are the basis for detecting transitions in the auditory scene when the signal properties change.

There is a considerable body of work investigating the principles for processing basic and relatively deterministic acoustic signal properties, yet the knowledge obtained from these studies, while valuable, is necessarily limited by the simplicity of the experimental stimuli used. What is needed is an approach that offers a bridge between the representation of highly controlled, deterministic, low-level acoustic features, and the highly complex nature of the real acoustic environment, while still maintaining control over the experimental stimulus manipulations. This thesis aims to provide such an approach by taking advantage of the computational power available for signal processing and digital sound synthesis in order to create generic sounds that approximate a level of complexity that is comparable to many naturally occurring sounds. The use of synthetic sounds allows the precise manipulation and control of sophisticated higher-order statistical signal properties that are characteristic of complex ethological sounds. At the same time, the generic nature of the experimental stimuli allows inferences with respect to general principles of auditory processing that likely apply to a variety of sound classes such as speech or music.

Such higher-order signal properties are likely represented at advanced levels of the auditory system such as auditory cortex. The methodology used in all experimental studies of this work, functional magnetic resonance imaging (fMRI), allows the investigation of acoustic information processing at the neural network level across multiple levels of the auditory system, including auditory cortex, with high spatial (on the order of a few cubic millimetres) and considerable temporal (on the order of a couple seconds) resolution (Logothetis, 2008). It is therefore well suited to investigating the neural substrates underlying the processing of higher-order statistical properties in acoustic signals.

The aim of the General Introduction is to provide a conceptual framework for the studies carried out in this thesis. In particular, following a brief overview of the functional anatomy of the auditory cortex (Section 1.1), the General Introduction will review experimental approaches towards elucidating the encoding of statistical properties in acoustic signals (Section 1.2), before introducing the key problems addressed in the experimental work of this thesis (Section 1.3).

## **1.1 Functional anatomy of the auditory system**

The acoustic information that reaches the cochlea is processed via a series of brain structures that form the ascending auditory pathway. Briefly, the mechano-electrical transduction of the travelling sound wave into neuronal signals is accomplished by the hair cells on the basilar membrane of the inner ear (Hudspeth, 1989). Once the incoming mechanical signal has been transduced into an electro-chemical signal, it is then projected along the auditory nerve, which terminates in the cochlear nucleus (CN). Most fibres leaving the CN cross the midline and convey acoustic information to auditory structures in the contralateral hemisphere, while a small number continue ipsilaterally. The major subcortical structures of the auditory system are the superior olivary complex (SOC) and the nucleus of the lateral lemniscus (LL) of the pons, the inferior colliculus (IC) of the midbrain, and the medial geniculate body (MGB) of the thalamus. Neurons of the MGB then project to primary and secondary auditory areas within the temporal lobe of the cortex cerebrum.

The peripheral and subcortical auditory structures set up important constraints or principles for subsequent stages of auditory information processing, such as the orderly series of frequency bandpass filters instantiated on the basilar membrane (von Békésy, 1960), or binaural analysis mechanisms in SOC and IC that allow sound localisation via interaural time and level differences (ITD and ILD, respectively; McAlpine, 2005). Furthermore, far from being mere relay stations in the ascending auditory pathway, subcortical structures already perform a significant amount of information processing of complex acoustic features (Miller *et al.*, 2001; Rauschecker, 1998; Winer & Lee, 2007).

Nevertheless, the processing of higher-order statistical properties in acoustic signals, which is the focus of the present body of work, depends on the involvement of auditory cortex for accurate and efficient representation. The following sections therefore provide in more detail a review of the anatomical organisation and parcellation of auditory cortex (Sections 1.1.1 and 1.1.2), before discussing some of its basic functional properties (Sections 1.1.3 to 1.1.6). Specifically, the following overview takes an approach that focuses on certain characteristic features of auditory cortex (e.g. macroanatomy, cytoarchitecture, or functional organisation) that are preserved across a variety of species, instead of a species-centred approach. As an organisational principle, within each section these properties will first be presented for non-human primates and then for humans; further, the general organisation proceeds from the description of primary areas to higher-level areas. This approach enables a direct comparison between species and highlights the phylogenetic emergence of generic principles of auditory cortex organisation (Hackett, 2007).

### 1.1.1 Macroscopic organisation of the auditory cortex

Auditory cortex covers much of the superior temporal plane (STP) of the temporal lobe and is conventionally defined as the region that receives its primary afferents from ventral or dorsal MGB (Hackett, 2007). Allometric measurements (surface area and volume) of the superior temporal gyrus (STG) reveal that it increases about threefold each time in the progression from squirrel monkey, macaque monkey, and chimpanzee, to human (Rilling & Seligman, 2002) (Figure 1-1). In fact, most of the increase in humans can be attributed to a relative expansion of auditory association cortex in the temporal lobe. In all primates, a significant portion of auditory cortex is ‘hidden’ beneath the Sylvian fissure (or lateral sulcus, LS, in non-human primates), separating the parietal and temporal lobes, on the dorso-medial surface of the STP.

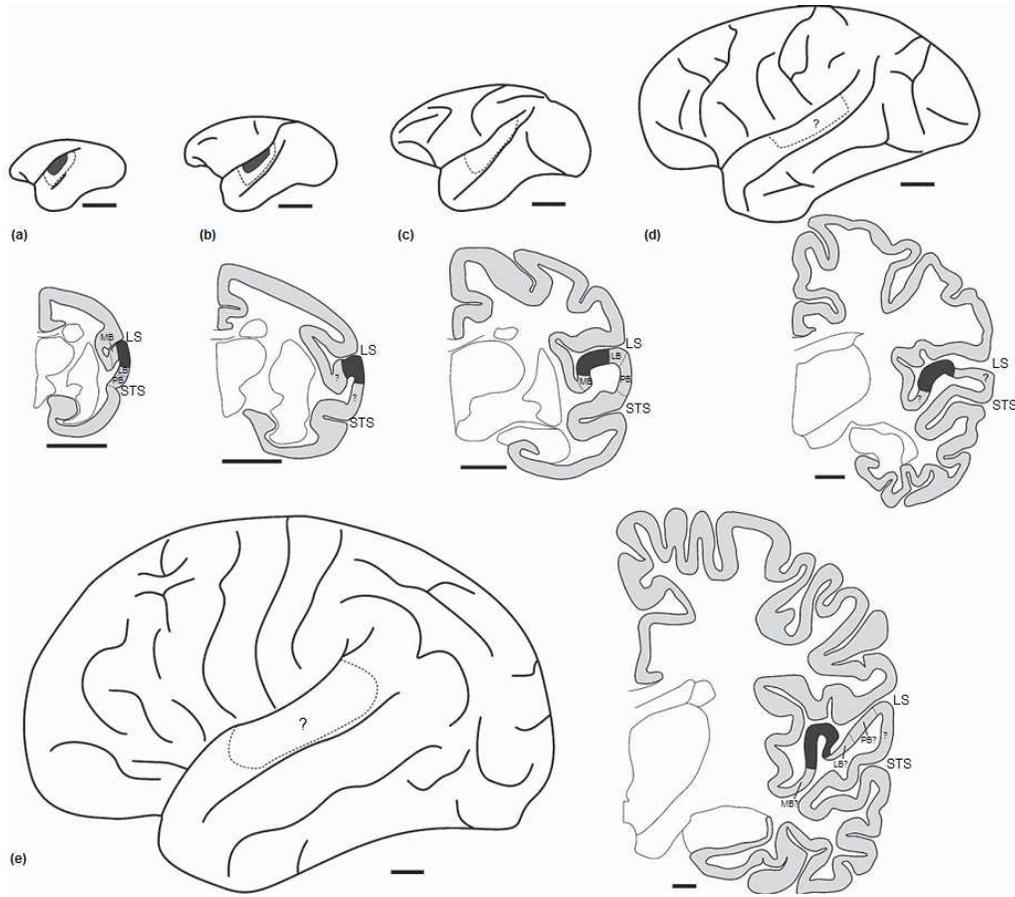


Figure 1-1 Schematic drawings of the cerebral cortex and the location of auditory cortex in several primates. Each panel depicts a lateral view of the left hemisphere and a coronal section through right auditory cortex. (a) Marmoset monkey (*Callithrix jacchus jacchus*); (b) squirrel monkey (*Saimiri squireus*); (c) macaque monkey (*Macaca mulatta*); (d) chimpanzee (*Pan troglodytes*); (e) human (*Homo sapiens*). The core area is shaded in dark grey. MB, medial belt region; LB, lateral belt region; PB, parabelt region; ?, region not defined; LS, lateral sulcus; STS, superior temporal sulcus. Scale bars: 5 mm (coronal sections); 10 mm (lateral views). Modified from Hackett (2007) with permission from Elsevier.

The general organisation of the auditory cortex of non-human primates is commonly divided into core, belt, and parabelt regions. Three primary subfields form the core of auditory cortex, while about seven to eight surrounding belt subfields and some two to three parabelt subfields have been identified (Hackett *et al.*, 2001; Hackett *et al.*, 1998a) (Figure 1-2). The core area consists of a primary area (A1), and more anterior

rostral (R) and rostrotemporal (RT) areas. The belt and parabelt regions are labelled according to their anatomical location (e.g. anterolateral, AL, or caudomedial, CM; see Figure 1-2).

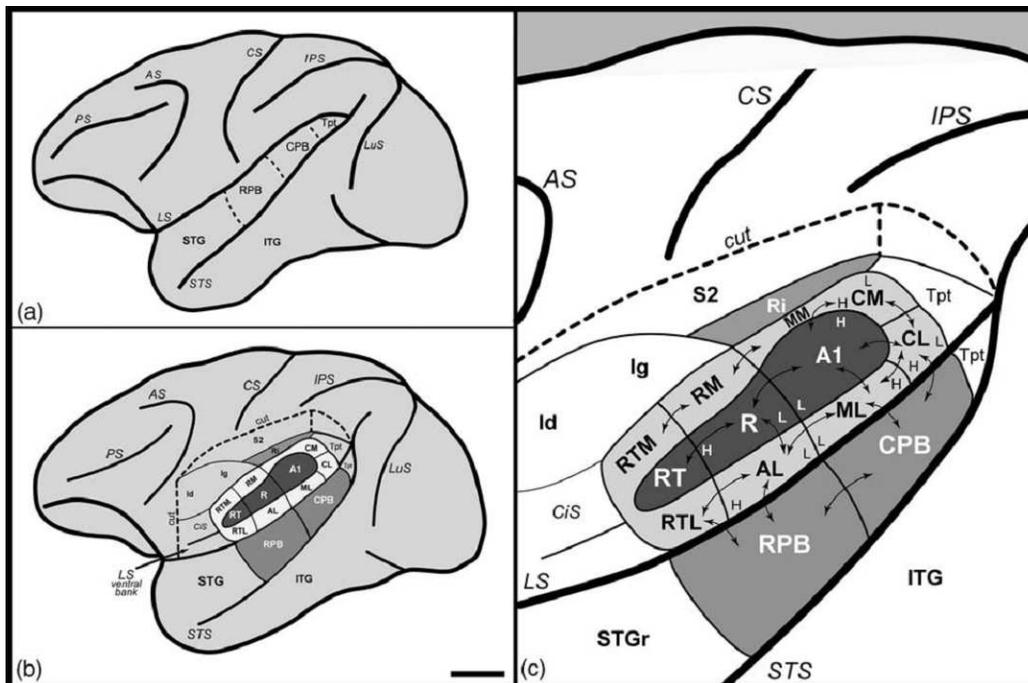


Figure 1-2 Schematic diagram of monkey auditory cortex illustrated for the macaque monkey. (a) Lateral view of the superior temporal gyrus (STG) with rostral (RPB) and caudal (CPB) parabelt areas. (b) Virtual cut to reveal auditory areas lying on the lower bank of the lateral sulcus (LS). (c) Close-up of (b), also revealing the flow of information (arrows) and tonotopic gradients (white letters H, high frequency, and L, low frequency). The three areas with dark shading represent the core of auditory cortex (auditory area 1, A1; rostral, R; rostrotemporal, RT); surrounding it are the eight belt regions in light shading (caudomedial, CM; caudolateral, CL; middle medial, MM; middle lateral, ML; rostromedial, RM; anterolateral, AL; rostrotemporal medial, RTM; rostrotemporal lateral, RTL). The rostral (RPB) and caudal (CBP) parabelt areas on the STG are shown in medium shading. Scale bar: 10 mm. From Hackett (2007) with permission from Elsevier.

In humans, core, belt, and parabelt regions likely correspond to primary, secondary, and association cortex, respectively, comprising some 30 functionally distinct subfields (Hackett, 2007); however, the precise delineation of the subfields varies considerably between researchers (Figure 1-3). Anatomically, these regions encompass the first transverse gyrus of Heschl, or Heschl's gyrus (HG), the posterior lying planum temporale (PT), and the STG. These correspond to Brodmann areas (BA) 41, 42, 52, and 22 (Brodmann, 1909). Areas in the superior temporal sulcus (STS) and, more rostrally towards the planum polare (PP), at the temporal pole, are considered auditory related cortex (Hackett, 2007).

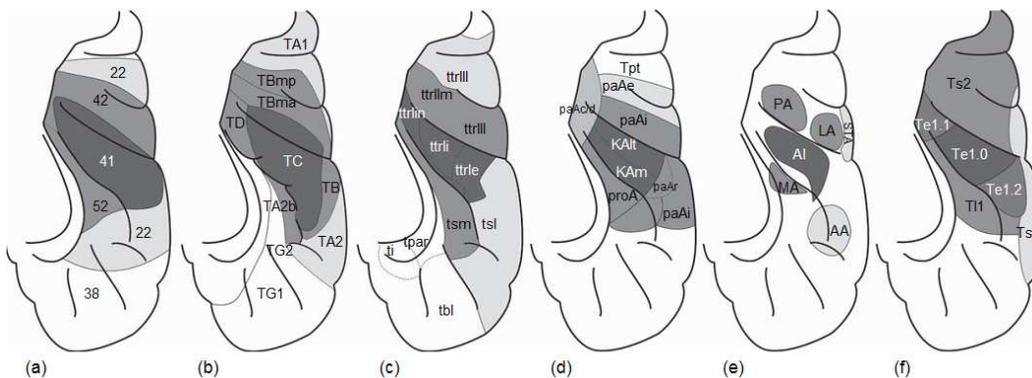


Figure 1-3 Parcellation of the human superior temporal cortex by different investigators. Schematic figures are standardised and normalised (the STG is not visible). Dark shading indicates the core region, medium shading indicates belt regions, and light shading indicates parabelt and possibly other regions. Posterior is up, lateral is right. For detailed description of the abbreviations see the original publications: (a) Brodmann (1909); (b) von Economo & Koskinas (1925) and von Economo & Horn (1930); (c) Beck (1928); (d) Galaburda & Sanides (1980); (e) Rivier & Clarke (1997) and Wallace *et al.* (2002); (f) Morosan *et al.* (2001; 2005). Modified from Hackett (2007) with permission from Elsevier.

In humans and chimpanzees, some individuals display a duplicate or bifid HG in one or both hemispheres (this feature is absent in all other non-human primates; Hackett *et al.*, 2001; Leonard *et al.*, 1998; Rademacher *et al.*, 1993; Sweet *et al.*, 2005). In the

case of a single HG, primary auditory cortex covers about two thirds of HG and does not extend past its anterior and sulcal boundaries. Where there is a duplicate HG, primary auditory cortex usually spans parts of both gyri and the intermediate transverse sulcus.

Perhaps the best described human auditory association area is the PT, lying posterolateral to HG on the STP; it is thought to contain multiple auditory subfields (Rivier & Clarke, 1997; Wallace *et al.*, 2002), which might explain its participation in a multitude of perceptual processes (Griffiths & Warren, 2002). The gross morphology of PT varies considerably (Westbury *et al.*, 1999), and often shows a left-hemispheric asymmetry (Dorsaint-Pierre *et al.*, 2006; Eckert *et al.*, 2006; Foundas *et al.*, 1994; Steinmetz *et al.*, 1989); however, such demonstration might in part reflect the stereotactic method used (Westbury *et al.*, 1999). This has traditionally led to the conclusion that PT forms a special role in language processing (Foundas *et al.*, 1994); more recent findings suggest, however, that there is no direct correspondence between leftward asymmetry and language function (Dorsaint-Pierre *et al.*, 2006; Eckert *et al.*, 2006).

Thus, both in non-human and human primates, auditory cortex displays a parcellation into three anatomically distinct regions (core, belt and parabelt in non-human primates; primary, secondary, and association cortex in humans). However, while homologies between different species are often assumed (especially between non-human primates and humans), this has only been shown convincingly for primary auditory cortex; homologies of higher-order auditory cortex are problematic, and certain areas of human association cortex may not have homologues in primates (Hackett *et al.*, 2001).

### 1.1.2 Cytoarchitecture

The core area of monkey auditory cortex and its three subfields display a distinct cytoarchitecture. The core region as a whole exhibits typical features of primary- or koniocortex, as it contains a dense layer IV, indicating prominent thalamocortical

connections (Galaburda & Pandya, 1983). Furthermore, the core region stains prominently for the calcium binding protein parvalbumin, is highly granular and myelinated, and displays high metabolic activity (Jones *et al.*, 1995; Kaas & Hackett, 2000; Morel *et al.*, 1993; Pandya, 1995). These features are most pronounced for area A1 and least present for area RT (Hackett *et al.*, 1998a). The narrow band of belt subfields surrounding the core displays reduced cell density and columnar spacing, larger pyramidal cells and a less dense myelination than the core areas (Hackett *et al.*, 1998a). The parabelt region stains less darkly than core and belt regions and displays other features that distinguish parabelt from core and belt, such as a stronger tendency to be arranged in vertical columns and an even lower cell density (Hackett *et al.*, 1998a). A subdivision of the parabelt region into two subfields is not obvious from cytoarchitectonic markers, but rather is based on cortico-cortical connectivity: the rostral parabelt (RPB) subfield shares connections within the anterior temporal lobe and ventrolateral frontal cortex, while the caudal parabelt (CPB) subfield projects caudally to the temporo-parietal junction and dorsolateral frontal cortex (Hackett *et al.*, 1999; Kaas & Hackett, 2000; Romanski *et al.*, 1999).

The human homologue of the core in non-human primates can be similarly divided into three subfields: one primary area in central HG, and two secondary areas in medial and anterolateral HG (Morosan *et al.*, 2001; Rademacher *et al.*, 2001); these are sometimes denoted as areas Te1.0, Te1.1, and Te1.2, respectively (Morosan *et al.*, 2001; Morosan *et al.*, 2005) (see also Figure 1-3). The cytoarchitectonic characteristics of these three subfields are similar to those in the core of monkey auditory cortex: all areas display prominent cytochrome oxidase, parvalbumin and acetylcholinesterase staining in cortical layers IIIc and IV (Clarke & Rivier, 1998; Hackett *et al.*, 2001; Hackett *et al.*, 1998a; Rivier & Clarke, 1997; Wallace *et al.*, 2002); a high density of small cells is particularly marked in layers II and IV; and layers V and VI are relatively thick (Galaburda & Sanides, 1980; Wallace *et al.*, 2002). The secondary subfields in medial and lateral HG show lower metabolic activity in layer IV than the primary subfield (Wallace *et al.*, 2002). The cytoarchitectonic characteristics of human auditory association regions are less well defined, rendering comparisons to belt and parabelt regions in non-human primates problematic (Hackett, 2007).

While cytoarchitectonic studies of post-mortem brains have significantly advanced our understanding of auditory cortex, the results can only provide limited functional information, for example on whether or not a given neuron is an ‘auditory neuron’; this caveat is sometimes circumvented by exposing animals to certain sounds or behavioural tasks before sacrifice and subsequent mapping of histochemical markers (Bajo *et al.*, 2007; Kaczmarek & Robertson, 2002; Overath, 2004b; Rauschecker *et al.*, 1997). Immunohistochemistry using antero- and retrograde tracers can inform about the macroscopic organisational connectivity between cortical subfields. These studies highlight both hierarchical and parallel processing streams that show a considerable degree of preservation between species.

In monkeys, the core region receives its main afferents from ventral MGB, while projecting to ipsilateral and contralateral core regions as well as to adjacent belt regions (Hackett *et al.*, 1998b; Kaas & Hackett, 2000; Morel *et al.*, 1993). The ipsilateral connections between core and belt are related to their spatial positions, in that anatomically neighbouring areas share stronger connections than non-adjacent areas (Galaburda & Pandya, 1983; Hackett *et al.*, 1998a; Morel *et al.*, 1993). The dorsal nucleus of the MGB provides the main afferents to belt subfields, which share multiple interconnections with each other as well as with the core and parabelt regions. Parabelt subfields receive subcortical input from the dorsal MGB (as well as strong afferents from medial pulvinar, supragenicolate and limitans nuclei, see Hackett *et al.*, 1998b) and are mainly indirectly connected to the core via belt areas. Direct connections from the core to parabelt areas are minimal (Hackett *et al.*, 1998a). Parabelt areas project to destinations in frontal, parietal, and temporal cortex for higher-level stages of processing.

In humans, similar processing schemes have been suggested by imaging techniques that track the diffusion in white matter axon bundles to elucidate intrinsic functional connectivity between brain areas (Behrens & Johansen-Berg, 2005; Behrens *et al.*, 2003). These studies confirm the general functional architecture within human auditory cortex (Upadhyay *et al.*, 2007; Upadhyay *et al.*, 2008).

### 1.1.3 Electrophysiology

A large part of our understanding of the functional characteristics of the subfields of auditory cortex is derived from invasive techniques applied in non-human primates, such as electrophysiology and lesion studies. These techniques provide excellent spatial and temporal (in the case of electrophysiology) resolution and have been invaluable in advancing auditory neuroscience. The most common measures are peristimulus response histograms (reflecting the firing rate of a certain neuron or neural ensemble to a specific stimulus), and response synchronisation (reflecting temporal distributions of neural discharges). However, caution is needed when interpreting electrophysiological responses obtained in anaesthetised animals, since these can differ from those of awake and behaving animals, while recordings in non-anaesthetised animals can also address effects of attentional modulation (Wang, 2000). In rare cases, depth-electrode and surface grid measurements can be acquired in humans during the pre-surgical evaluation of patients with intractable epilepsy (e.g. Howard *et al.*, 1996; Howard *et al.*, 2000; Liégeois-Chauvel *et al.*, 2001; Liégeois-Chauvel *et al.*, 1994; Liégeois-Chauvel *et al.*, 1991).

A general processing scheme in auditory cortex based on electrophysiological data across species including humans suggests that core and possibly belt areas encode basic spectral and temporal acoustic features, before more complex signal attributes are processed in parabelt regions (Rauschecker, 1998). For example, core areas respond more vigorously to pure sounds (or sinusoids), while belt and parabelt areas respond more strongly to complex sounds (rather than to sinusoids), and also to species-specific vocalisations (Rauschecker, 1998; Rauschecker & Tian, 2000; Rauschecker *et al.*, 1995; Rauschecker *et al.*, 1997; Tian *et al.*, 2001). Neurons in the core regions show narrow frequency tuning to pure tones, while the tuning properties of belt and parabelt neurons are increasingly broad (Bendor & Wang, 2008; Rauschecker *et al.*, 1995; Rauschecker *et al.*, 1997). However, ablation of primary auditory cortex in rhesus monkeys does not abolish neuronal pure tone sensitivity (Heffner & Heffner, 1986), suggesting an additional parallel processing stream that likely relies on the preservation of an orderly frequency encoding (see Section 1.1.4 below) in lower structures of the auditory system.

The hierarchical flow of information in auditory cortex is also confirmed by the latency of neuronal responses. In humans, responses in medial HG can already be recorded approximately 20 ms post stimulus onset, while central and lateral HG show slightly longer latencies at 50 ms and 60-75 ms, respectively, and responses in PT peak at around 100 ms (Brugge *et al.*, 2008; Liégeois-Chauvel *et al.*, 1994; 1991). Similarly, in some cases direct intracortical stimulation of primary auditory cortex in posteromedial HG can produce propagated responses towards lateral HG and association areas (Brugge *et al.*, 2003; Howard *et al.*, 2000; Liégeois-Chauvel *et al.*, 1994; 1991). Response latencies increase and become more variable as one progresses to structures in the STG and parietal operculum (Celesia, 1976; Liégeois-Chauvel *et al.*, 1991); there is also evidence for back-projections from STG to HG (Brugge *et al.*, 2003), likely facilitating neuronal modulation of afferent signals. Responses in the hemisphere contralateral to acoustic stimulation are stronger than those to ipsilateral stimulation, while binaural stimulation leads to the strongest response amplitudes (Liégeois-Chauvel *et al.*, 1991).

#### 1.1.4 Tonotopy

One prominent principle of auditory functional organisation across species is the conservation of an orderly frequency representation, or tonotopy, throughout the ascending auditory pathway. This is similar to comparable anatomical organising schemes in other senses (e.g. retinotopy in vision). The mechanical characteristics of the basilar membrane set up a series of frequency bandpass filters, in which high frequencies lead to maximum movement excursions at the base, while lower frequencies are represented towards the apex of the basilar membrane (von Békésy, 1960) (Figure 1-4). This mapping of frequency to spatial position is maintained in the auditory nerve leading to the cochlear nucleus, and is subsequently conserved in subcortical structures including IC (e.g. Schreiner & Langner, 1988, 1997) as well as the ventral MGB as part of the lemniscal ascending auditory pathway (the non-lemniscal dorsal MGB does not show a tonotopic organisation; for a review, see Jones, 2003). This is usually demonstrated via best or centre frequencies (BF or CF, respectively) using electrophysiological recordings. Several subfields of auditory

cortex each show tonotopic frequency gradients (see below); in fact, the demonstration of frequency gradient reversals often functions as evidence for boundaries between functionally different cortical fields.

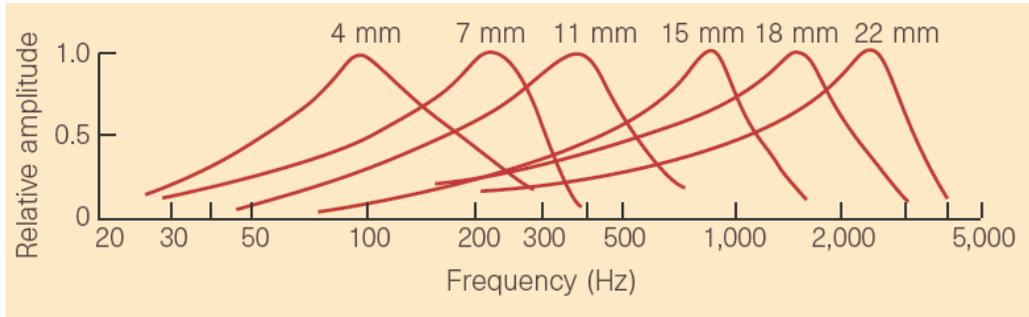


Figure 1-4 Frequency responses at six different positions along the basilar membrane (see von Békésy, 1960). Distance is indicated with respect to the apex of the basilar membrane. Reprinted from Moore (1999) with permission from Macmillan Publishers Ltd.

Tonotopic gradients in primary auditory cortex have been shown in a variety of species with electrophysiological and immunohistochemical tracing studies, as well as a post hoc combination of the two (Joris *et al.*, 2004; Kosaki *et al.*, 1997; Morel *et al.*, 1993). In monkeys, the primary field A1 displays a tonotopic gradient from high to low as one moves from rostral to caudal; subfield R shows a reverse tonotopic gradient, while the tonotopic gradient in RT is similar to A1 (Figure 1-2). At least four of the approximately seven belt areas (namely AL, CL, CM, ML) also display a tonotopic organisation (Kosaki *et al.*, 1997; Morel *et al.*, 1993). However, the frequency tuning in belt areas is generally broader than those of neurons in the core areas (Kosaki *et al.*, 1997; Recanzone *et al.*, 2000).

Recently, tonotopic frequency gradient reversals have been confirmed in monkey auditory cortex using fMRI (Petkov *et al.*, 2006). In humans, Talavage and colleagues (2000) demonstrated eight distinct tonotopic fields in both HG and STG.

However, an alternative account posits that areas with preferred high or low frequencies represent distinct fields instead of representing the border between fields (Schönwiesner *et al.*, 2002). One potentially significant caveat of many functional imaging studies investigating tonotopic organisation in the auditory system is, however, that stimuli are generally presented at a level far above their response threshold (Kim & Molnar, 1979), thereby blurring the frequency selectivity and spatial resolution of the results.

For frequencies up to a certain frequency cut-off (which varies between species, see Joris *et al.*, 2004; Middlebrooks, 2008), responses in the auditory nerve are phase-locked to the instantaneous phase of the motion produced on the basilar membrane. Frequencies above this threshold are likely represented via an average firing pattern across different neurons, where each neuron might only fire at a multiple of the frequency. Both midbrain and cortical structures retain these principles of frequency representation, although the upper frequency limit for phase-locking decreases with each step along the ascending auditory pathway (Joris *et al.*, 2004).

### 1.1.5 Spectrotemporal receptive fields

A further approach to elucidating functional characteristics of auditory cortex is the description of spectrotemporal receptive field (STRF) properties in auditory neurons (Aertsen & Johannesma, 1981a, b; Eggermont *et al.*, 1981). The STRF of a neuron is represented by a kernel (in the spectral and temporal domain) that describes its dynamic spectrotemporal response properties (Coath *et al.*, 2008). STRFs can therefore be regarded as representing characteristic computational properties in single neurons or small neuronal ensembles (Elhilali *et al.*, 2007). Importantly, the calculation of a neuron's STRF reveals both excitatory and inhibitory response properties in both spectral and temporal domains and thus provides important information beyond other descriptions, for example a neuron's BF as introduced in the previous Section 1.1.4.

Studies investigating STRF properties in the primary auditory cortex of awake ferrets demonstrated a co-existence of stability and plasticity that allows optimal coding of acoustic features in a dynamically changing world (Elhilali *et al.*, 2007). For example, while neurons display stable STRFs to certain acoustic features when these are not behaviourally relevant, STRF properties can undergo plastic modulation within hours or minutes if these acoustic features become task relevant (Fritz *et al.*, 2003; Fritz *et al.*, 2007; Fritz *et al.*, 2005).

In the primary auditory cortex of non-anaesthetised monkeys, STRF responses display on-excitation as well as off-excitation or on-inhibition characteristics, providing a sophisticated and flexible neural code to integrate complex spectrotemporal information, such as in natural sounds and species-specific vocalisations (deCharms *et al.*, 1998; Pelleg-Toiba & Wollberg, 1989; Shamma & Symmes, 1985). Furthermore, inhibitory responses slightly lag excitatory responses, which opens a short time window during which neurons can initially respond, thus establishing an equilibrium of excitatory and inhibitory responses that allows fine-tuning of firing precision and optimal information processing (Wehr & Zador, 2003; Zhang *et al.*, 2003).

### 1.1.6 Processing streams

The above Section 1.1.3 on electrophysiological recordings in auditory cortex discussed a processing hierarchy in which the encoded acoustic features increase in complexity as one ascends in the auditory hierarchy. There is also evidence of two parallel processing streams in the auditory cortex that represent two types of auditory information; the identity ('what') and spatial location ('where') of acoustic signals (Kaas & Hackett, 1999; Rauschecker & Tian, 2000; Romanski *et al.*, 1999; Tian *et al.*, 2001). According to this model, auditory 'what' information is processed in a rostral or anterior pathway along the temporal lobe and prefrontal cortex, while auditory 'where' information is processed along a dorsal route via posterior temporal cortex, posterior parietal lobe and frontal cortex. This organisational scheme is similar to the 'what' and 'where' pathways in vision (Ungerleider & Haxby, 1994).

Anterior and caudal belt areas in the auditory cortex of rhesus monkeys display differential sensitivity to the type or spatial position of con-specific vocalisations, respectively (Tian *et al.*, 2001). In humans, the processing scheme is now relatively well established for the rostral ‘what’ pathway. For example, Zatorre and colleagues (2004) demonstrated increased activation within the upper bank of the right anterior STS as well as the right inferior frontal gyrus as a function of the distinctness of auditory object identities (see also Section 1.2.1). However, it might be more appropriate to conceive of relative gradients for processing ‘what’ (and ‘where’) information instead of exclusive processing streams (Kikuchi *et al.*, 2007).

Evidence for the dorsal ‘where’ pathway has proven to be more controversial, mainly because posterior temporal cortex also responds to con-specific vocalisations in addition to anterior temporal cortex (Poremba *et al.*, 2004; Tian *et al.*, 2001), suggesting that a distinction between ‘what’ and ‘where’ information may be relative rather than absolute. This has led some researchers to emphasise the importance of procedural ‘how’ information in the acoustic signal to explain the functional significance of a dorsal processing stream (Belin & Zatorre, 2000; Middlebrooks, 2002; Zatorre *et al.*, 2002b). This view highlights articulatory aspects of the stimuli used (i.e. conspecific vocalisations and speech) to explain the role of the dorsal pathway (Hickok & Poeppel, 2007; Warren *et al.*, 2005b). This is in agreement with later conceptualisations of the original ‘where’ pathway in vision that emphasise its functional role in visuomotor integration (Milner & Goodale, 1995; Rizzolatti *et al.*, 1997).

The previous sections provided a general framework of the functional anatomy of auditory cortex, highlighting basic principles of auditory processing, such as the co-existence of serial and parallel processing streams and the representation of increasingly complex acoustic features along the ascending auditory pathway. These provide a foundation for investigating more complex functional properties reviewed next.

## **1.2 Experimental approaches to auditory cortex function**

The following sections review several experimental approaches to elucidating the representation of statistical signal properties within subdivisions of auditory cortex. Following a brief introduction to auditory objects (Section 1.2.1), regarded as the ‘building blocks’ of auditory scene analysis, the two subsequent sections will discuss pitch (Section 1.2.2) and timbre (Section 1.2.3) as critical object features that require the integration of higher-order statistical properties. Subsequently, processes underlying the grouping of linked objects will be reviewed (Section 1.2.4), with a special emphasis on the extraction of statistical regularities in auditory streams.

### **1.2.1 Auditory object analysis**

The concept and definition of auditory objects and auditory object analysis are controversial (Griffiths & Warren, 2004; Kubovy & Van Valkenburg, 2001; Nelken, 2004; Scholl, 2001; Scott, 2005). Griffiths and Warren (2004) propose four principles of auditory object analysis. First, auditory object analysis comprises the processing of information that corresponds to entities in the physical or sensory acoustic world. Second, auditory object analysis requires perceptual mechanisms that segregate the object itself from other objects and from the rest of the acoustic environment. Third, auditory object analysis must abstract characteristic object properties in order to enable a stable representation or object identity even when these properties undergo minor stochastic variations (e.g. the characteristic features of a speaker’s voice must be maintained irrespective of the speaker’s spatial position). Fourth, in a multimodal world, object information should generalise across senses, such as when associating a face with a voice. From an information theoretic perspective (Shannon, 1948), auditory object analysis requires computational mechanisms that are both robust (allowing the maintenance of object identity) as well as sensitive to critical acoustic changes (allowing the detection of transitions and the segregation between objects).

A slightly different conceptualisation highlights the importance of perceptual segregation mechanisms for auditory object analysis (Kubovy & Van Valkenburg,

2001). The authors propose that the defining process of auditory objects is figure-ground segregation. This posits that auditory object analysis is based on the detection of boundaries in frequency-time space within which statistical signal regularities apply. This allows the integration and disambiguation of auditory objects in complex auditory scenes. Further, Kubovy & Van Valkenburg (2001) conceive of auditory object analysis as a hierarchical process: the output of early, subcortical processing stages represents basic acoustic features in the auditory scene that require grouping; grouping in turn produces putative perceptual objects according to Gestalt principles (Wertheimer, 1922, 1923); in a complex auditory scene, attentional processes subsequently select which of the perceptual objects become figure, and which become (back-) ground.

In the present context, auditory objects are conceptualised in information theoretic terms, such that a given auditory object is characterised through its probabilistic higher-order statistical properties; in turn, boundaries between auditory objects are indicated by transitions in these statistical regularities (Kubovy & Van Valkenburg, 2001). That is, at a generic descriptive level, auditory objects are defined in terms of their distinct statistical signal characteristics, which simultaneously distinguish them from other auditory objects (and possibly other object classes). Statistical regularities thus provide important information for auditory scene analysis, as they allow the perceptual organisation of the acoustic environment (e.g. figure-ground segregation). The focus here is on the analysis of statistical characteristics that are inherent in the acoustic signal; in the case of natural sounds, these are represented as abstract templates and thereby provide distinct information on auditory objects.

The four principles proposed by Griffiths & Warren (2004) and the emphasis on figure-ground segregation (Kubovy & Van Valkenburg, 2001) together with the conceptualisation of auditory objects as characterised by statistical properties provide a general framework for experimental investigations of auditory object analysis. However, the devil is in the detail (just as in the visual system: Feldman, 2003). Consider a speaker producing the vowel /a/: should the speaker (or sound source) be regarded as an auditory object, or the vowel /a/ itself, or the speaker's position, or all of them together? Griffiths & Warren (2004) argue that any of these object characteristics can be regarded as an auditory object; any of these can define an

auditory object based on the four basic principles. For example, the auditory object could be the speaker or the vowel; at the same time, the combination of the constituent objects could also comprise an auditory object, similar to visual objects and perceptual Gestalt phenomena (Wertheimer, 1922, 1923). Here, it is argued that each of the constituent objects has characteristic statistical properties that can be extracted, maintained and disambiguated according to the principles introduced above (Griffiths & Warren, 2004).

Bregman (1990) emphasised the temporal aspect of audition and suggested that auditory streams are the equivalent of visual objects (see Chapter 1.2.4.2); in this regard, auditory events that are grouped together or perceived as a distinct auditory stream form an object of audition. A further view suggests yet a slightly different conceptualisation by equating auditory objects with auditory streams (Shamma, 2008). However, a direct comparison between visual and auditory objects is necessarily limited by their differential reliance on space and frequency: whereas visual objects exist in space-time (and can be static), auditory objects exist over frequency-time space (and are rarely static). For example, two different pitches arising in the same location can be heard as distinct, while two identical pitches in two different locations will likely be perceived as a single sound (Kubovy & Van Valkenburg, 2001). Conversely, two colours emerging from the same spatial location will likely blend into one colour percept, while they will be perceived as two light sources if they emerge from two spatial location (Kubovy & Van Valkenburg, 2001).

These conceptualisations of auditory object analysis address two fundamental and complementary requirements of the auditory system, often referred to as simultaneous and sequential grouping (Bregman, 1990; Carlyon, 2004; Darwin, 1997; Darwin & Carlyon, 1995; Griffiths & Warren, 2004). First, acoustic events must be grouped together, for example with respect to their identity or source. Second, acoustic events must be parsed or segregated, for example to distinguish between different sound sources. At relatively short time scales, these fundamental processes are relevant for binding together acoustic information and perceiving as an entity individual auditory objects (Griffiths & Warren, 2004); at the level of longer time scales, they are relevant for the analysis of linked or grouped streams of objects (Bregman, 1990).

The studies in this thesis explored mechanisms for both the analysis of individual objects characterised by higher-order statistical signal properties (Studies 1-2, Chapters 3-4) as well as for processing higher-order statistical signal properties in linked streams of objects such as in pitch sequences (Studies 3-5, Chapters 5-6). The following two sections briefly review two features of auditory objects, pitch (Section 1.2.2) and timbre (Section 1.2.3), which are relevant for the experimental work of this thesis.

### 1.2.2 Pitch

Pitch is a fundamental feature of the auditory scene (Helmholtz, 1875) and a universal element in music across human cultures (McDermott & Hauser, 2005). The recent American National Standards definition states: “Pitch [is] that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends primarily on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus” (ANSI, 1994). Nevertheless, after decades of research, many aspects of the neuronal and neural mechanisms underlying the perception of pitch are still not completely understood (Plack *et al.*, 2005), and the current section will therefore attempt only a brief overview.

Pitch and frequency are not identical. Pitch refers to the (psychological) percept and can differ from (physical) frequency. In the case of a pure tone with a fixed frequency, pitch is identical to frequency. However, most naturally occurring sounds are complex sounds consisting of multiple frequencies that are harmonics of the lowest or fundamental frequency  $f_0$  that is present in the sound. As a general approximation, the perceived pitch of a complex periodic tone corresponds to that of the  $f_0$  (however, see Fastl & Zwicker, 2007). Therefore, instead of perceiving several distinct frequencies making up the signal, the spectral information is integrated across frequency bands, leading to a single, coherent pitch percept that may be different from its physical stimulus properties. This is likely achieved by higher-order pattern

matching mechanisms that operate over multiple frequency bands (de Cheveigné, 2005).

A prominent example highlighting the pattern matching mechanisms underlying pitch perception is the ‘missing fundamental’ phenomenon (de Boer, 1976; Terhardt, 1974; Winkler *et al.*, 1997). In this auditory illusion, a complex periodic tone of several harmonics is presented without its fundamental frequency  $f_0$ . The percept of this complex tone is that of a pitch an octave lower than its lowest physically present frequency, leading to an ‘illusory’ percept (corresponding to the missing  $f_0$ ). As indicated by the missing fundamental phenomenon, pitch perception can diverge from the spectral information present in the sound, and is best described via the repetition rate of a sound. For natural sounds consisting of a fundamental frequency and several harmonics that are integer multiples of  $f_0$ , the slowest repetition rate of the complex sound is equal to the period of the fundamental frequency. When the fundamental frequency is removed, the repetition rate of the complex sound remains unchanged. This also holds when removing more than one frequency (e.g. the first five frequencies of a complex periodic sound consisting of ten harmonics). Therefore, sounds with the same repetition rate, but very different spectra, can evoke the same pitch (Plack & Oxenham, 2005). Furthermore, pitch can be evoked by unresolved harmonics alone; however, the strength of the pitch percept is an order of magnitude stronger in the presence of resolved harmonics. In fact, the third to fifth harmonics contribute most strongly to a pitch percept (Moore, 2003; Plack & Oxenham, 2005).

There are two prominent theories of pitch perception mechanisms (Plack *et al.*, 2005). One posits that the place on the basilar membrane with maximum movement excursion determines the pitch percept (place pitch theory); the other highlights the temporal aspect of pitch and posits that the repetition rate of the periodic sound determines its pitch (temporal pitch theory). Neither theory can explain all experimental data. In fact, since at least some spatial or place information is necessary for pitch perception (Oxenham *et al.*, 2004), it has been suggested that pitch is coded by a place-specific temporal code (Shannon, 2005). It is further conceivable that there is more than one pitch encoding mechanism, since resolved and unresolved harmonics are represented via different neural codes (Plack *et al.*, 2005). Thus, pitch perception is situated at the interface between sensation and perception and offers a unique

window into how the auditory system makes inferences based on its statistical evaluation of (sometimes ambiguous) acoustic information.

Given the complexity of pitch perception and its reliance on pattern matching mechanisms, it seems reasonable to assume that pitch extracting mechanisms emerge relatively late in the auditory hierarchy. Early evidence from patients with temporal lobectomies suggested that the right temporal lobe including HG is important for detecting the direction (up or down) of a pitch pair with missing fundamentals, but not with present fundamental frequencies (Zatorre, 1988). Subsequent studies have refined these findings and suggest that the lateral aspect of Heschl's gyrus as part of secondary auditory cortex plays a crucial role in processing pitch information.

Griffiths and colleagues (2001) created a repetition pitch by systematically varying the temporal regularity of sounds via a delay-and-add algorithm (Yost *et al.*, 1996); generally, pitch salience increases with the temporal regularity created by the number of delay-and-add iterations. They demonstrated activity in HG that was stronger for pitch evoking sounds than for noise (see also Patterson *et al.*, 2002). However, the manipulation of repetition pitch via the number of iterations cannot directly address whether activation increases due to (physical) temporal regularity or (perceptual) pitch salience. Making use of the fact that pitch salience of sounds with only resolved or only unresolved harmonics differs, while temporal regularity remains constant, Penagos and colleagues (2004) were able to localise pitch salience to lateral Heschl's gyrus as part of secondary auditory cortex. Similarly, activity in lateral HG is specific to the percept of temporal pitch and not generalised to analysis of temporal structure as such, as in spatial width perception introduced by interaural delay (Hall *et al.*, 2005). Electrophysiological recordings in the primary auditory cortex of the marmoset monkey confirmed these findings for various types of complex sounds that all shared a pitch percept within a low-frequency region at the anterolateral border of areas A1 and R (Bendor & Wang, 2006; Bendor & Wang, 2005).

Warren and colleagues (2003) expanded these findings for two further aspects of pitch, namely pitch chroma and pitch height. Pitch can be modelled along a helix, where pitch chroma is represented by angular position along the helix, and pitch height by elevation (Bachem, 1950; Krumhansl, 1990). In this helical model, pitches an octave apart (frequency ratio of 2:1) share the same pitch chroma, but are of

different pitch height. The authors showed that pitch chroma and pitch height are encoded in distinct cortical areas anterior and posterior to primary auditory cortex, respectively. These results suggest a processing hierarchy in which the basic features of pitch are encoded first in lateral HG, and are subsequently rendered to a more refined analysis with respect to pitch chroma and pitch height in auditory association cortex.

### 1.2.3 Timbre

Timbre is a further perceptual attribute that reflects higher-order auditory object perception and segregation. Timbre, known evocatively in German as *Klangfarbe* ('sound colour'), commonly refers to those aspects of a sound that distinguish it from another sound with identical pitch, duration and intensity (ANSI, 1994). Rather than being a unitary entity, however, timbre is best conceptualised within a perceptual 'timbre space' consisting of multiple temporal and spectral dimensions (Caclin *et al.*, 2005; Hajda *et al.*, 1997; McAdams *et al.*, 1995). Studies using multidimensional scaling (MDS) techniques (Caclin *et al.*, 2005; McAdams *et al.*, 1995) suggest three to four principal dimensions: the attack or log-rise time, spectral centroid, spectral fine structure, and to a lesser extent spectral variation or spectral flux (for a slightly different weighting of timbral dimensions see Kendall *et al.*, 1999). Thus, timbre relies on the statistical evaluation and integration of multiple stimulus dimensions.

The principles underlying timbre perception provide important information for auditory object analysis, such as object recognition and segregation (Handel, 1995; McAdams, 1993). For example, for short tones produced by impulsive instruments (e.g. piano or drums), the information in the log-rise time is vital for auditory object or instrument identification, while the spectral information contained in the sustained part of longer instrument sounds is sufficient for recognition (Hajda, 1999). Furthermore, auditory streaming (Section 1.2.4.2) is strongly influenced by timbre (Bregman, 1990).

Until about 1850, in Western tonal music, timbre was mainly a means to distinguish the instrument carrying the melodic line from those playing the accompaniment, or to signal boundaries by changes in instrumentation. Subsequently, the richness of the symphonic orchestra of the late Romantic era enabled composers to make timbre itself a distinctive attribute via sophisticated orchestration techniques, e.g. by combining different instruments to form different timbres. The aesthetic epitome of this timbre feature was the concept of a *Klangfarbenmelodie* ('tone colour melody') based on timbral sequences, introduced in the 20<sup>th</sup> century by Arnold Schönberg and Anton Webern in their compositions. Electronically generated synthetic sounds later opened the door to an unlimited variety of musical timbres (McAdams, 1996). The complexity of timbre perception is underscored by the fact that even today's sophisticated computational pattern matching algorithms are not capable of disambiguating the various instruments in an orchestra; a feat that the human auditory system performs with apparent ease. In speech, vowels are the prime example of timbre. We distinguish an /a/ from a /u/ by virtue of their different spectral shape, or the different profile of their harmonics, commonly referred to as formants (Rogers, 2000).

On a more abstract scale, affine transpositions of timbral relations are perceived as more similar than non-affine timbral transpositions, underscoring the higher-order perceptual properties of timbre (McAdams & Cunibile, 1992). Similarly, it is possible, though generally quite difficult, to disambiguate via timbre alone two sound sources or auditory objects that differ in pitch; for example, distinguishing a violin playing an A4 from a flute playing a C5 (Handel & Erickson, 2004). As a general rule of thumb (Handel & Erickson, 2001), this generalisation of timbral attributes across instrumental pitch is only possible for sounds that are within an octave of each other (for similar findings concerning speech vowels, see Erickson *et al.*, 2001; Erickson & Perry, 2003).

Given the complex interplay of multiple acoustic features or dimensions within the perceptual attribute timbre, it is reasonable to hypothesise that timbre is processed or assembled in auditory association areas. At the same time, possibly because of its complexity, the precise neural correlates of timbre and its dimensions are still relatively little understood. A study in which multiple timbral attributes were varied

simultaneously found bilateral posterior superior temporal activity (Menon *et al.*, 2002), while changes in the spectral envelope activated a network of PT and posterior STS (Warren *et al.*, 2005a). Furthermore, the latter network seems to be serially organised, so that the extraction of the spectral envelope could be attained between HG and PT, where PT functions as a computational hub, forwarding the information to posterior STS (Kumar *et al.*, 2007). These results are supported by a patient with a right-lateralised lesion confined to areas within this network (right lateral HG, PT, posterior STG, and posterior STS), who had impaired differentiation of various timbres (dystimbria) (Griffiths *et al.*, 2007). However, there is also evidence for an at least partially parallel representation of the different dimensions of timbre (Caclin *et al.*, 2006), with potential hierarchies only arising at later processing stages (Caclin *et al.*, 2007; Caclin *et al.*, 2008).

Despite important recent advances with respect to elucidating the neural representation or representations of the multidimensional auditory attribute timbre, these likely reveal only the tip of the iceberg and many questions still remain. The precise neural representations of the different timbral dimensions that have been established via MDS techniques remain unclear, as does the question at which stages in the auditory system these dimensions interact. Further, it is unresolved to what extent the processing of the different timbral dimensions is hierarchical versus parallel (Caclin *et al.*, 2006; Caclin *et al.*, 2007).

#### 1.2.4 Sound and pitch sequences

The acoustic scene does not only consist of single auditory objects (e.g. pitch or timbre), of course, but is often a pattern or sequence of sound objects that conveys acoustic information, such as in musical melodies or the prosody of spoken sentences ('t Hart *et al.*, 1990; Patel, 2008). These auditory streams can themselves be regarded as forming auditory objects (Bregman, 1990). Several experimental paradigms have investigated the underlying neural mechanisms for organising the auditory scene, such as tracking and grouping sound sequences. The following sections review three

approaches to elucidating how the auditory system evaluates the statistical signal properties of sound and pitch sequences.

#### 1.2.4.1 *Mismatch negativity*

A classical experimental paradigm that provides a window into how the auditory system makes inferences based on statistical stimulus characteristics in the auditory scene is the mismatch negativity (MMN) paradigm (Näätänen, 1995; Näätänen *et al.*, 2007). The MMN is traditionally recorded non-invasively with electrophysiological techniques, such as electro- and magneto-encephalography (EEG and MEG, respectively). Subtracting the event-related response (ERP) to frequent ‘standard’ stimuli (e.g. a sinusoid of 1000 Hz) from the response to infrequent ‘deviant’ or oddball stimuli (e.g. a sinusoid of 1100 Hz) reveals a negative component, the MMN, at 150-250 ms post stimulus onset, with a maximal deflection at fronto-central electrodes and a typical sign reversal at the mastoids when using a nose reference (Näätänen *et al.*, 1978; Sams *et al.*, 1985a; Sams *et al.*, 1985b).

The MMN is thought to reflect the pre-attentive processing of any discernable violation of a previously established context. The larger the difference between standard and deviant stimuli, the earlier and larger the MMN (Sams *et al.*, 1985a; Tiitinen *et al.*, 1994). Furthermore, equivalent current dipole (ECD) modelling suggests that there might be feature-specific MMN generators, since some studies show spatially distinct ECD sources for frequency, intensity, inter-stimulus-interval, and duration oddballs (Deouell & Bentin, 1998; Deouell *et al.*, 1998; Frodl-Bauch *et al.*, 1997; Giard *et al.*, 1995; Levänen *et al.*, 1993; Rosburg, 2003) (but see Sams *et al.*, 1991). The MMN is also elicited by more complex acoustic changes such as deviating phonemes (for a review see Näätänen, 2001; see also Näätänen *et al.*, 1997; Winkler *et al.*, 1999), rhythms (Vuust *et al.*, 2005), timbre (Caclin *et al.*, 2006, 2007, 2008), and musical pitch sequences (van Zuijen *et al.*, 2004; Winkler *et al.*, 2006). Its pre-attentive nature makes it a promising tool in infants and clinical populations (for reviews see Kujala *et al.*, 2007; Näätänen, 2000, 2003).

The cortical generators of the mismatch response are likely localised in secondary or association auditory cortex on the STP, with a potential additional source in right inferior frontal cortex, as has been demonstrated using ECD modelling in EEG (Giard *et al.*, 1995; Jemel *et al.*, 2002; Scherg *et al.*, 1989) and MEG studies (Alho *et al.*, 1998a; Alho *et al.*, 1998b; Hari *et al.*, 1984; Levänen *et al.*, 1996; Sams *et al.*, 1991), as well as with fMRI (Opitz *et al.*, 1999; Opitz *et al.*, 2002; Rinne *et al.*, 2005) and intracortical recordings (Halgren *et al.*, 1995; Halgren *et al.*, 1998; Kropotov *et al.*, 2000; Kropotov *et al.*, 1995; Rosburg *et al.*, 2005).

Whether the MMN reflects a distinct neural mismatch source in STP or is due to sensory adaptation processes within a single neuronal population is currently a matter of debate (Jääskeläinen *et al.*, 2004; Näätänen *et al.*, 2005). The classical, echoic memory-trace hypothesis argues that the mismatch response reflects the activity of a specific subpopulation that compares the current sensory input with a previously established context and signals upon detection of a violation (for a review see Näätänen *et al.*, 2005). In this framework, the MMN would reflect activity that is distinct from other, temporally close evoked responses, such as the N100 or N1 (or its magnetic equivalent the N100m or N1m), which is the main evoked negative deflection to sound onset and is composed of at least two subcomponents (Näätänen & Picton, 1987). The alternative adaptation hypothesis of the MMN mechanisms highlights the existence of these different N100 subcomponents, in particular the more posterior N1p (which peaks at around 85 ms) and more anterior N1a (which peaks at around 150 ms) components. It argues that the MMN is a subtraction artefact attributable to different features of N1p and N1a: the narrow frequency tuning of the N1a generators would show sensory adaptation to the standard stimulus and a subsequently larger response to a frequency deviant; conversely, the broad frequency tuning of the N1p generators would also adapt to the standard, but would show a larger response only to wide frequency deviations. Thus, subtracting evoked responses to standards from deviants would produce a negative deflection at the latency of the N1a, leading to a misattribution of the MMN.

There are several results that suggest that sensory adaptation cannot explain all results in the vast MMN literature (reviewed in Näätänen *et al.*, 2005). For example, an MMN is also present when the oddball is not physically different from the standard

or reflects a complex rule change that is unlikely to be confounded by sensory adaptation (van Zuijen *et al.*, 2004). However, recent studies suggest that the MMN might reflect a temporal combination of sensory (adaptation) and cognitive (memory-trace) processes (Maess *et al.*, 2007). Similarly, a scheme based on predictive coding (Friston, 2003a, 2005) of sensory input processing that encompasses both hypotheses has been proposed to underlie cortical MMN mechanisms (Garrido *et al.*, 2008).

#### 1.2.4.2 *Auditory streaming*

One fundamental requirement of the auditory system is to be able to follow a sound source irrespective of distraction from other sound sources, often described as auditory scene analysis (Bregman, 1990). One prominent example of such figure-ground segregation is the ‘cocktail party effect’ mentioned above (Cherry, 1953); the brain far surpasses the ability of sophisticated computational algorithms in its ability to segregate sound sources (Haykin & Chen, 2005). Classical studies investigating the underlying perceptual mechanisms have employed an auditory streaming paradigm, first introduced by Bregman and Campbell (1971; further explored in detail in the doctoral thesis by van Noorden, 1975), which highlights the neural mechanisms underlying the extraction of statistical signal properties relevant for figure-ground segregation.

In the classical auditory streaming paradigm, two pure tones A and B of different frequencies are presented in an alternating pattern, the most common patterns being ABAB and ABA\_ sequences (where ‘\_’ represents a silent gap before the repetition of the triplet). Whether the pattern is heard as a single stream of alternating A and B frequencies (stream integration or fusion) or as two separate streams where each stream is formed by A or B tones only (stream segregation or fission), depends on the presentation rate and the difference in frequency ( $\Delta f$ ) between A and B. As a general principle, the faster the presentation rate and the larger  $\Delta f$ , the more likely stream segregation occurs.

The stimulus paradigm may seem overly simplistic and far removed from the complex acoustic environment. However, its generality and relevance for auditory perception is underscored by the demonstration of behavioural streaming phenomena in songbirds (Hulse *et al.*, 1997; MacDougall-Shackleton *et al.*, 1998) and non-human primates (Izumi, 2002). Furthermore, as will be discussed below, rather than being a stimulus-driven bottom-up phenomenon, it can be modulated by attention, providing evidence for statistical inference in the brain.

One prominent theory – the peripheral channelling hypothesis (Hartmann & Johnson, 1991) – that attempts to describe the neural mechanisms underlying stream segregation, highlights the role of auditory filters in the sensory periphery (see also Beauvois & Meddis, 1996). According to the peripheral channelling hypothesis, if A and B are close in frequency and therefore pass through the same auditory filter on the basilar membrane, they will excite the same neural population in the ascending auditory system, leading to a percept of a single stream of alternating A and B tones. Conversely, if the difference in frequency  $\Delta f$  between A and B is large enough so that A and B will pass through different filters and excite separate neural populations, this leads to a two-stream percept or stream segregation. While this theory is able to explain much of the experimental data of stream segregation (as reviewed in Darwin, 1997; Micheyl *et al.*, 2007; Moore & Gockel, 2002; Snyder & Alain, 2007), there are two aspects in particular that cannot be explained by peripheral mechanisms alone.

First, stream segregation can also occur for sounds in which the peripheral coding does not differ: for example, when two complex tones A and B differ only with respect to their unresolved harmonics, while the resolved spectrum of the two tones is kept identical (Grimault *et al.*, 2000; Gutschalk *et al.*, 2007; Vliegen *et al.*, 1999; Vliegen & Oxenham, 1999). Similarly, streaming can occur with temporal cues only, but otherwise identical long-term power spectra (Grimault *et al.*, 2002; Roberts *et al.*, 2002). Second, stream segregation is sensitive to attention. This was already acknowledged by van Noorden (1975), who discovered that attention can bias perception towards segregation or integration in ambiguous ABA\_ sequences with intermediate  $\Delta f$ . A related aspect is that stream segregation has a temporal component that can be influenced by attention, in the sense that stream segregation often requires

several seconds to build up (Carlyon *et al.*, 2001; Carlyon *et al.*, 2003), and this build-up seems to be ‘reset’ when switching attention (Cusack *et al.*, 2004).

The physiological bases of streaming have only been studied relatively recently. A general principle underlying stream segregation seems to be ‘forward suppression’. For example, intracortical recordings (multiunit activity and current source density) in A1 of awake monkeys, where the A tone in ABAB sequences was centred on the best frequency (BF) of the recording site, showed decreased responses to B tones as a function of presentation rate and  $\Delta f$  (Fishman *et al.*, 2001). For slow presentation rates, both A and B frequencies elicited large responses; however, for fast presentation rates, only A tones elicited clear responses, while the magnitude of responses to B tones decreased. This effect increased with  $\Delta f$ . Fishman and colleagues (2001) interpreted these findings in terms of a forward masking effect that A tones exert on B tones. While these effects could also be explained due to the fact that with increasing  $\Delta f$ , B tones moved farther away from the best frequency region of the recording site, Micheyl and colleagues (2005), recording in macaque A1, quantified that forward suppression is at least a contributing factor by comparing responses to B tones in an ABA\_ paradigm, to responses to B tones alone. A subsequent study found that inter-tone-interval, i.e. the time interval between tones, is more crucial for stream segregation than the presentation rate (Fishman *et al.*, 2004); this is consistent with psychoacoustic results (Bregman *et al.*, 2000).

As mentioned earlier, stream segregation often has a build-up time of several seconds (Carlyon *et al.*, 2001; Cusack *et al.*, 2004). Generally, the faster the presentation rate, the faster the build-up of stream segregation. Micheyl and colleagues (2005) showed that this perceptual switch (as measured in humans) has a similar time-course representation to the corresponding neural response (as measured in monkey A1). This is similar to a reduced N1m/P1m complex, reflecting the stimulus onset response to the B tone in ABA sequences as a function of  $\Delta f$  (Gutschalk *et al.*, 2005; see also Snyder *et al.*, 2006). For ambiguous  $\Delta f$ , the N1m/P1m complex was similarly reduced whenever participants indicated a two-stream percept as opposed to a one-stream percept (Gutschalk *et al.*, 2005). Somewhat at odds with the neurophysiological data in non-human primates (Fishman *et al.*, 2004; Fishman *et al.*, 2001; Micheyl *et al.*, 2005), the source reconstruction indicated

an origin in secondary and association cortex (Gutschalk *et al.*, 2005; Snyder *et al.*, 2006). However, a more recent study (Gutschalk *et al.*, 2007) using both MEG and fMRI reported a source origin and haemodynamic activity in auditory cortex, including primary cortex, that correlated with a two-stream percept (see also Deike *et al.*, 2004; Wilson *et al.*, 2007). However, there is also evidence for extra-auditory areas participating in streaming: Cusack (Cusack, 2005) found no correlate of stream segregation in auditory cortex, but instead activity in the intraparietal sulcus (IPS) that covaried with a two-stream percept in ambiguous  $\Delta f$  sequences. An intriguing recent result suggests that stream segregation is not specific to auditory cortex, but is already present in the cochlear nucleus of the anaesthetised guinea pig (Pressnitzer *et al.*, 2008).

While these studies provide a critical insight into the cortical areas relevant for stream segregation, they nevertheless say relatively little about the neuronal mechanism. For example, some studies also report a general effect of adaptation for both A and B responses as a function of presentation rate (e.g. Fishman *et al.*, 2004; Fishman *et al.*, 2001; Gutschalk *et al.*, 2005; Micheyl *et al.*, 2005). How forward suppression and adaptation mechanisms interact in the paradigm is currently unclear; for example, as stream segregation builds up (forward suppression), the strength of the neuronal responses recedes (adaptation). Furthermore, the synaptic mechanisms underlying forward suppression have yet to be elucidated (Brosch & Schreiner, 1997; Calford & Semple, 1995; Denham, 2001; Eggermont, 1999). The multiplicity of factors determining auditory streaming (Moore & Gockel, 2002) suggests a distributed network contributing to auditory streaming instead of one single substrate. Nevertheless, theories emphasising the degree of spectral or tonotopic separation (Hartmann & Johnson, 1991) likely explain most, while certainly not all, of the stream segregation phenomenon. At the same time, there still remain outstanding questions regarding (i) the role of peripheral vs. central mechanisms, (ii) the role of attention on stream formation, and (iii) the contribution of different auditory areas as well as that of non-auditory areas.

### 1.2.4.3 *Complex pitch sequences*

While the two previous paradigms (MMN and auditory streaming) highlighted basic properties of auditory perception, one significant restriction is their repetitive and thus deterministic nature. The MMN reflects sensory and comparator mechanisms over relatively short time scales and under arguably simple or overly deterministic acoustic stimulation conditions (except for the occasional rule violation exemplified by the oddball stimulus). This is similar in the auditory streaming paradigm, in which a set pattern generally consisting of only two frequencies is repeated over and over.

However, real-life acoustic events generally have a complexity that is magnitudes greater. For example, pitch sequences such as those in musical melodies extend over several seconds and often span multiple pitches combined in a rhythmic and harmonic structure, and thus are likely to require perceptual processes with a complexity that far surpasses those necessary for simple mismatch detection in a repetitive sequence. In fact, a significant aspect contributing to music appreciation is its complexity, exemplified by the balance between expected and unexpected musical events (Huron, 2006). Similarly, speech perception involves the processing of complex consonant-vowel transitions over different time-scales (Rosen, 1992), as well as the tracking of prosodic structure.

Section 1.1.4 noted that initial perception of single frequencies engages primary and secondary auditory areas, while a more integrated pitch percept likely arises in lateral HG (Section 1.2.2) (Bendor & Wang, 2005; Griffiths, 2001, 2005; Patterson *et al.*, 2002). In contrast, the perception of a complex series of pitches requires perceptual grouping mechanisms and is thus likely to engage areas beyond primary cortex (see also the two preceding Sections 1.2.4.1 and 1.2.4.2). Lesions to parabelt areas in macaques (Cowey & Weiskrantz, 1976) and association cortex in humans (Patel *et al.*, 1998; Peretz *et al.*, 1994), sparing primary cortex, lead to impairments of auditory pattern perception such as pitch sequence processing.

Both the direction (up or down) and the precise size of intervals between successive pitches provide important information for pitch sequence perception (Peretz, 1990). The simplest process required is the detection of the direction of pitch change between two successive pitches; patient studies indicate that this computation depends on right lateral HG (Johnsrude *et al.*, 2000). In one of the first studies

investigating the perception of longer musical pitch sequences, Zatorre and colleagues (1994) used positron emission tomography (PET) to compare the response in auditory cortex to simple melodies against the response to acoustically matched noise sequences; this revealed increased activity in right STG for melodies. Lateral HG and PT participate in the perception of lively pitch sequences compared to fixed-pitch sequences (Griffiths *et al.*, 2001), while anterior and posterior temporal cortices show a parametric increase with temporal structure for musical pitch sequences (Griffiths *et al.*, 1998). Furthermore, there is evidence that duration and pitch sequences are processed in overlapping areas in the temporal lobes (Griffiths *et al.*, 1999).

Patterson and colleagues (2002) suggest a hierarchical organisation of pitch sequence processing, in which precise temporal information is represented in subcortical structures, before the emergence of a pitch percept in lateral HG and the representation or integration of increasingly complex pitch sequences (as in a melody) in association areas. The authors presented four types of sound sequences: noise, fixed-pitch, random, and diatonic pitch sequences (similar to melodies). The stimuli used were repetition pitch (iterated rippled noise, IRN) to control for low-level acoustic complexity. All four types of sound sequences activated primary, secondary and association cortex. Furthermore, primary auditory cortex responded strongest to noise, while activation for fixed-pitch sequences extended into lateral Heschl's gyrus. Random and diatonic pitch sequences resulted in activation beyond primary and secondary auditory cortices along the anterior superior temporal gyrus, notably the planum polare (PP). Interestingly, the authors found no consistent differences across participants for the random and diatonic pitch sequences (similar to the results by Griffiths *et al.*, 2001). Nevertheless, the results demonstrate a hierarchy in processing temporal pitch information at the level of the cortex, extending from an initial analysis in primary and secondary areas to auditory association areas on the superior temporal plane.

Further, studies of pitch sequence processing have highlighted the perceptual cues of global and local contour information (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990). The global contour of a pitch sequence describes its pattern of 'ups' and 'downs', while the local contour denotes the exact interval size. Study 5 of this thesis

(Chapter 6) investigated the underlying neural correlates for global and local processing of pitch patterns in healthy volunteers.

In summary, some general principles of auditory cortex function across species can be highlighted. First, the information flow in auditory cortex is both hierarchical and parallel, enabling dynamic and highly adaptable processing of complex acoustic information, in which bottom-up, top-down and lateral signalling co-exist. Second, the complexity of acoustic information represented in each region increases from core via belt to parabelt cortex (and their likely human homologues). However, there is currently no consensus as to whether structures in the ascending auditory pathway process similar acoustic features (e.g. tonotopic maps at various stages in the auditory pathway) at different levels of complexity and generality, or whether these structures share a division of labour where each processes different sound attributes (Griffiths *et al.*, 2004; Nelken, 2004; Pressnitzer *et al.*, 2008; Scott, 2005). A related outstanding problem is whether auditory cortex represents complex auditory objects irrespective of local stochastic variations (e.g. represents a voice irrespective of its location or background noise) or, rather, according to invariant acoustic features that only lead to the emergence of auditory objects via the synchronisation of larger neural populations, which each code a particular invariant acoustic feature. Nelken (2004) argues that these levels co-exist and that it is indeed possible for auditory cortex to represent complex auditory objects as such. It has been hypothesised that such redundancy reduction may in fact be a general principle in the ascending auditory system: while neuronal responses in auditory cortex may indeed be noisier than those in subcortical structures, at the same time they can be argued to be less redundant than subcortical structures, which often faithfully represent a single physical stimulus attribute multiple times (Chechik *et al.*, 2006). The increasing complexity in auditory cortex implicates computational processes at the level of single neurons, and quite probably neuronal populations, which provide a less faithful representation of precise physical stimulus attributes than that provided by previous structures in the auditory hierarchy. Such a computational scheme suggests that redundancy is reduced as one ascends along the auditory hierarchy and that redundant signals require fewer computational resources.

### 1.3 Key problems addressed in this thesis

Previous approaches to auditory processing, such as the MMN and auditory streaming paradigms, have typically used deterministic stimulus designs in which the statistical properties of the experimental stimulus were relatively constrained. While the mechanisms and topography of the underlying processes are now relatively well understood, inferences from these data are necessarily limited by the scope of the experimental stimuli. Thus, current outstanding questions concern the representation of less deterministic and higher-order stimulus properties. This doctoral thesis addresses the representation of higher-order statistical properties of acoustic signals in human auditory cortex; in particular, it investigates the participating structures and their organisation (e.g. hierarchical or parallel) for representing various higher-order statistical properties in acoustic signals.

The majority of the studies herein employ a parametric design in which the statistical properties of a single acoustic parameter – such as entropy, correlation over time, or spectrotemporal coherence – are altered along a continuum (Friston, 2003b). The advantage of a parametric approach is that it provides a sharper understanding of the representation of a particular acoustic stimulus parameter across different instantiations; in particular, parametric designs enable the investigator to probe precisely the brain's response to a particular parameter across various instantiations, while keeping other lower-level acoustic properties, e.g. spectral power or bandwidth, constant. This is in contrast to classical categorical or factorial designs, which contrast the effect of an experimental variable with a control condition, i.e. the presence versus the absence of the experimental variable, on the response in cortex, and thus cannot inform on the precise effect of the experimental variable across different instantiations. Furthermore, while categorical designs hinge on the adequacy of a control condition and are prone to problems associated with cognitive subtraction or pure insertion (Friston *et al.*, 1996a), the different levels of parametric designs function as their own internal control stimulus and allow the detection of non-linear responses across levels (Friston, 2003b). This thesis specifically takes an information-theoretic approach in the sense that statistical signal properties are systematically varied, thereby tracking cortical areas that encode the signal properties over different instantiations.

The detailed investigation of complex sound properties is only made possible by modern signal processing techniques, whose computational power has increased significantly over the last decade. Current signal processing software allows the sophisticated manipulation of complex higher-order statistical properties in generic, synthetic sounds that share many of the characteristic features of naturally occurring sounds, while avoiding their semantic connotations. As a consequence, this work addresses neural processing mechanisms for the abstraction of generic higher-order statistical acoustic properties at the level of neural populations in neurologically normal participants. The following sections provide a brief description of the motivation for each of the five studies comprising the thesis.

### 1.3.1 Chapter 3 – Study 1

*Are there distinct time scales over which the auditory cortex addresses statistical signal properties, and can these operations be assigned to distinct areas within auditory cortex?*

Acoustic information evolves over several time scales, from microseconds (relevant for sound localisation and spectral pitch resolution), to tens of milliseconds (e.g. phonemes) and hundreds of milliseconds (e.g. syllables), to several seconds (e.g. musical melodies or spoken sentences) (Rosen, 1992). Accordingly, the auditory system needs to assess acoustic information over a range of time scales or time windows. One way to achieve this is to vary the correlation in the acoustic signal over different time scales to assess brain activation that tracks acoustic information evolving over one of those time windows (Luo & Poeppel, 2007). There is evidence that processing of information encapsulated in time windows of tens and hundreds of milliseconds is lateralised towards the left and right auditory cortices, respectively (Boemio *et al.*, 2005; Poeppel, 2003). However, the precise conceptualisation and representation suggested by different studies is not consistent (c.f. Obleser *et al.*, 2008; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001). Further, the division of labour between primary, secondary, and association areas for these different time windows of analysis has so far not been clearly established. This study investigated

the notion of distinct time windows by representing the different spectrotemporal time windows via correlation in the acoustic spectrum.

### 1.3.2 Chapter 4 – Study 2

*How are auditory objects represented and segregated in the auditory cortex?*

The auditory system has developed remarkable precision in identifying, as well as segregating, distinct auditory objects (Griffiths & Warren, 2004; Nelken, 2004; Scott, 2005). In order to achieve this, it needs to assess the statistical object properties to detect boundaries or transitions between objects as well as maintain object constancy. However, little is known about the underlying cortical mechanisms for these two fundamental perceptual mechanisms. This study introduces a novel stimulus, in which auditory objects were identified by the percentage of randomly distributed frequency ramps with identical direction and trajectory. Thus, auditory object perception depended on the detection of higher-order spectrotemporal coherence; similarly, object segregation depended on the detection of a change in coherence over frequency-time space. This study investigated the cortical representation for segregating and integrating auditory objects based on higher-order statistical properties such as spectrotemporal coherence.

### 1.3.3 Chapter 5 – Studies 3 & 4

*Can the planum temporale be described as a neural engine that requires fewer computational resources for redundant signals than for those with high information content?*

Within the auditory system, the planum temporale of human auditory association cortex is thought to represent a ‘computational hub’ that compares the neuronal pattern of incoming information to pre-existing templates and subsequently

gates the information along the auditory hierarchy for further processing (Griffiths & Warren, 2002). The ‘computational hub’ model of PT function integrates results from a variety of studies investigating a multitude of auditory functions. However, the model has not been tested explicitly, and studies 3 and 4 investigated one specific prediction that arises from the ‘computational hub’ model: activity in PT should increase as a function of the entropy, or information production, in pitch sequences.

#### 1.3.4 Chapter 6 – Study 5

*Is there a cortical hierarchy and lateralisation scheme for processing local and global information in pitch patterns?*

Pitch sequences consist of two structural levels; the global level comprises the pattern of ‘ups’ and ‘downs’ that forms the contour of the pitch sequence, while the local level denotes the precise interval size between pitches (Dowling, 1978; Dowling & Fujitani, 1971; Dowling *et al.*, 1987). Behavioural and patient studies have demonstrated a hierarchical organisation of pitch pattern perception, such that global processing precedes local processing. However, the notion that global processing is right-lateralised, while local processing is left-lateralised, has been less consistently supported between studies (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990; Schuppert *et al.*, 2000). Further, accounts of the underlying brain structures have so far only come from patients with cerebral damage. This study tested the hierarchy and lateralisation accounts in neurologically normal participants.

## Chapter 2. TECHNIQUES AND METHODS

Signal processing and digital stimulus synthesis form the theoretical core of this work. The advances over the past decade in computing power and software sophistication have made it possible to generate complex acoustic stimuli that allow researchers to generate and manipulate sounds seemingly at their will. The core experimental methodology employed in this thesis is functional magnetic resonance imaging (fMRI). Here, too, increased magnetic field strengths and innovations in the design of acquisition sequences and data analysis have made large strides and now provide a powerful tool to investigate brain function. This chapter introduces applications of signal processing and digital stimulus design in auditory neuroscience (Section 2.1), followed by a brief review of the basis of MRI and fMRI (Sections 2.2 and 2.3), specific considerations for studying auditory perception with fMRI (Section 2.4), and an outline of fMRI data analysis (Section 2.5).

### 2.1 Approaches to stimulus design for auditory neuroscience

Digital stimulus design has become an invaluable tool in auditory neuroscience. With respect to this thesis, digital stimulus synthesis provides a bridge between bottom-up and top-down approaches to acoustic information processing. The bottom-up approach allows individual acoustic characteristics to be systematically created and manipulated from first principles, thereby investigating basic rules determining auditory scene analysis (Section 2.1.1). The top-down approach allows the systematic manipulation of natural acoustic signals (Section 2.1.2). A complementary approach using digital stimulus synthesis enables the generation of sophisticated acoustic signals that can be as complex as natural sounds, but in which characteristic stimulus features can be tightly controlled (Section 2.1.3). Furthermore, individual sound features (such as spectrotemporal correlation in complex acoustic spectra or information production in pitch sequences) can be designed to obey probabilistic rather than deterministic principles, and thus approximate properties of ethological sounds.

### 2.1.1 Simple synthetic stimuli

Acoustic signals in the real world are continuous, or analogue. However, most contemporary computing devices operate in the digital domain; therefore acoustic signals are discretised or sampled at a certain sampling rate. According to Shannon's sampling theorem, the sampling rate must be at least twice as high as the highest frequency in the original analogue signal in order to faithfully capture the signal (Shannon, 1949). This is also often called the Nyquist rate. A common sampling rate discretises the signal at 44100 Hz; according to the sampling theorem, this faithfully captures frequencies in the signal of up to 22050 Hz, which is above the normal hearing threshold of humans. The number of bits per sample that are used to encode the signal provides a second dimension of fidelity; the higher the number of bits, the more precise the correspondence between analogue and digital signal. A common bit rate is 16 bits/sample.

The following sections give brief mathematical descriptions of fundamental types of stimuli employed to elucidate generic mechanisms of auditory perception (frequency and pitch, Sections 2.1.1.1 and 2.1.1.2; amplitude modulation, Section 2.1.1.3; frequency modulation, Section 2.1.1.4). These have been used successfully in numerous experimental studies employing electrophysiology, functional imaging and psychophysics in humans and animals (e.g. Fastl & Zwicker, 2007; Joris *et al.*, 2004; Laureys *et al.*, 2003; Moore, 2003; Patterson *et al.*, 2002; Rees & Malmierca, 2005; Warren, 2008). Their experimental power lies in the tight control that the experimenter has over the acoustic feature that is manipulated; this ensures that any observable effect (behavioural or neural) can be attributed precisely to the stimulus manipulation.

However, a notable caveat is their relative simplicity and deterministic nature, limiting the validity of direct comparisons with natural sounds. Deducing the neural representation of complex natural sounds from that of pure tones would need to assume that the auditory system behaves in a linear way, which is not the case (Hart *et al.*, 2003; Malone *et al.*, 2007; Rauschecker *et al.*, 1995).

### 2.1.1.1 *Pure tones and decomposition of complex sounds into pure tone components*

In mathematical terms, a pure tone can be described as

$$x(t) = A \sin\left(2\pi \frac{t}{T} + \phi\right), \quad (\text{Eq. 2-1})$$

which, as a function of time  $t$ , has amplitude  $A$ , period (or time to repeat)  $T$ , and initial phase  $\phi$ . A commonly used measure of the period is in fact its inverse  $1/T$ , called the frequency  $f$ , which denotes the number of periods per second. The angular frequency  $\omega$ , in radians, is then defined as  $\omega = 2\pi f$ . Note that a pure tone can also be expressed in mathematical terms when substituting cosine for sine, where the relation  $\cos(x) = \sin(x + \pi/2)$  holds. For brevity, a pure tone wave  $x(t)$  will sometimes also be referred to as a frequency.

According to the principles of Fourier analysis, all sounds can be described as a composition of one or more frequencies. Most naturally occurring sounds originating from animate sources are periodic sounds and consist of multiple frequencies, where the constituent frequencies are integer multiples of the lowest frequency ( $f_0$ ) present in the sound. The principle of Fourier analysis is that every periodic sound can be written as the sum of its constituent frequencies, and the Fourier series provides a mathematical description of this principle.

The Fourier series of a signal  $x(t)$  with period  $T$  may be defined as

$$x(t) = A_0 + \sum_{n=1}^{\infty} (A_n \cos(\omega_n t) + B_n \sin(\omega_n t)). \quad (\text{Eq. 2-2})$$

where  $\omega_n = 2\pi n f_0$  (Hartmann, 2000). The Fourier coefficients  $A_n$  and  $B_n$  are defined for  $n > 0$  as

$$A_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \cos(\omega_n t) dt, \quad \text{and} \quad B_n = \frac{2}{T} \int_{-T/2}^{T/2} x(t) \sin(\omega_n t) dt,$$

and for  $n = 0$  by

$$A_0 = \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt.$$

The equations defining the Fourier coefficients define the Fourier transform in the context of the Fourier series. In general, the Fourier transform is an operator that allows one to pass freely between two types of representations for a signal  $x(t)$ : the time and the frequency domains. In the time domain, the signal is defined in terms of sound pressure present at each point in time, while in the frequency domain the signal is defined via the amplitude (and phase) of each constituent frequency.

For digitally sampled signals, the discrete Fourier transform (DFT) must be used (Hartmann, 2000). For a time series

$$x_1, x_2, \dots, x_N$$

with  $N$  sampling points, the DFT is defined by computing for each  $1 \leq n \leq N$

$$\hat{x}_n = \sum_{k=1}^N x_k e^{-2\pi i kn / N}. \quad (\text{Eq. 2-3})$$

This allows the time series to be reconstructed via the inverse DFT (IDFT)

$$x_n = \frac{1}{N} \sum_{k=1}^N \hat{x}_k e^{-2\pi i kn / N}. \quad (\text{Eq. 2-4})$$

A naïve algorithm for computing the DFT would require approximately  $N^2$  operations, which for large  $N$ , would require substantial computational resources. For example, with a sampling rate of 44.1 kHz (as for the stimuli in this thesis), sounds of several seconds duration would have hundreds of thousands of sampling points. Therefore, the fast Fourier transform (FFT), which performs the same computation in only approximately  $N \log N$  operations, is commonly used in digital signal processing (the same holds for the inverse FFT, IFFT).

### 2.1.1.2 Iterated rippled noise (IRN)

A large body of work emphasises the temporal structure of sounds, rather than the spectral structure considered above, as relevant to pitch (de Cheveigné, 2005) (see also Section 1.2.2). Work by Yost and Patterson (Patterson *et al.*, 1996; Yost *et al.*, 1996) introduced noise stimuli associated with pitch in which the temporal regularity of the sound and associated pitch can be manipulated, whilst controlling the spectral structure that is resolved by the auditory system. Such stimuli are used in this thesis as a way of controlling the resolved spectrum in pitch sequences, to allow clearer interpretation of the data in terms of complex pitch sequence properties rather than lower level frequency representations (see Chapter 5). These stimuli are referred to as iterated rippled noise (IRN) or regular interval sounds (Patterson *et al.*, 1996; Yost *et al.*, 1996). In IRN sounds, a noise sample is iteratively added to itself with a delay (in ms), where the delay determines the period of the sound that is produced. There are two basic methods to generate IRN sounds: one can either add the original noise sample, or the running noise sample. The temporal structure, or periodic quality, introduced by iteratively adding the samples with a fixed delay increases with the number of iterations. Thus, in the spectral domain, IRN sounds contain all frequencies within a certain passband (a property inherent to noise).

In mathematical terms, the original noise sample method can be described as follows. If  $x(t)$  is the original noise sample, then

$$X(t) = \sum_{k=0}^K x(t - k\Delta t) \quad (\text{Eq. 2-5})$$

gives the IRN with  $K$  iterations and delay  $\Delta t$ . Technically, this equation is restricted to  $t \geq K\Delta t$ , but this simply amounts to cropping the sound at the beginning.

The running noise sampling method is defined iteratively. If  $x_0(t) = x(t)$ , then the first iteration is defined as  $x_1(t) = x_0(t) + x_0(t - \Delta t)$ , the second as  $x_2(t) = x_1(t) + x_1(t - \Delta t)$ , and so on. The  $K$ -th iteration

$$X(t) = x_K(t) = x_{K-1}(t) + x_{K-1}(t - \Delta t)$$

then gives a running noise IRN with  $K$  iterations and delay time  $\Delta t$ .

### 2.1.1.3 Amplitude modulation (AM)

Amplitude modulation (AM) of natural sounds can be considered at the level of the whole waveform (when it is determined by the ‘envelope’) and at the level of changes in intensity within particular frequency regions, as used to simulate natural sounds in ‘vocoding’ (see Section 2.1.2). Most naturally occurring sounds vary constantly not only because of their sinusoidal nature, but also because of varying amplitude or frequency. AM occurs when the amplitude  $A$  of the signal is also a function of time,  $A(t)$ . The two most common forms of AM are sinusoidal and linear AM. For sinusoidal AM (SAM), the modulating amplitude  $A(t)$  is itself given by a wave,

$$A(t) = 1 + m \sin(2\pi f_m t + \phi_m).$$

The modulation depth of  $A(t)$  is defined as its fixed amplitude  $m$  and can be any value between  $m = 0$  (0%, no modulation) and  $m = 1$  (100%, maximum modulation). Furthermore, the function  $A(t)$ , as a scaling factor, must always be non-negative, that is  $0 \leq m \leq 1$ . The frequency of the original signal  $x(t)$  is now called the carrier frequency  $f_c$ , while the frequency  $f_m$  of  $A(t)$  is called the modulation frequency. The combined amplitude modulated signal is then given by

$$x(t) = A(t) \sin(2\pi f_c t + \phi_c). \quad (\text{Eq. 2-6})$$

In the frequency domain, an AM sound with carrier frequency  $f_c$  and modulation frequency  $f_m$  is represented by a central peak at  $f_c$  with two subsidiary sideband peaks at  $f_c \pm f_m$ . The amplitudes of the two side bands are always identical, and are precisely half that of  $f_c$  at 100% modulation depth.

SAM is the simplest form of AM, consisting of a single modulation frequency. Different modulation waveforms can be constructed using Fourier series. For example, a square wave has multiple harmonics at odd integer multiples of the fundamental that decrease in magnitude with increasing frequency. A square wave can therefore be constructed from the sum of a number of sinusoidal modulations. The same applies to other modulation waveforms such as sawtooth or ramp changes in intensity.

#### 2.1.1.4 Frequency modulation (FM)

Frequency modulation (FM) occurs when the frequency  $f$  of the signal  $x(t)$  is a function of time,  $f(t)$ . Again, the most common forms of FM are sinusoidal and linear. When the instantaneous frequency of  $x(t)$  is sinusoidal, the FM signal can be written as

$$x(t) = A \sin(\omega_c t + \Delta f \sin \omega_m t), \quad (\text{Eq. 2-7})$$

where  $A$  is a fixed amplitude,  $\omega_c = 2\pi f_c$  denotes the carrier frequency in radians, and  $\omega_m = 2\pi f_m$  denotes the modulation frequency in radians.  $\Delta f$  specifies the maximum frequency modulation; for  $\Delta f \neq 0$ , the frequency spectrum of the resulting FM sound gains multiple sidebands around  $f_c$ , forming a complex spectrum. Depending on the exact value of  $\Delta f$ , the carrier frequency  $f_c$  can even disappear from the spectrum. Vibrato is one example of naturally occurring sinusoidal FM.

When the instantaneous frequency  $f(t)$  is linear, this is given in general by a linear equation of the form  $mt+b$ . An explicit linear equation of the form

$$f(t) = \left( \frac{f_2 - f_1}{T} \right) t + f_1$$

produces a linear sweep from the starting frequency  $f_1$  to the ending frequency  $f_2$  over a time period  $T$ , with a slope  $\frac{f_2 - f_1}{T}$ .

In the linear case, the total FM signal is described by

$$x(t) = \cos(2\pi f(t)t + \phi). \quad (\text{Eq. 2-8})$$

#### 2.1.2 Sampled natural stimuli

A second approach to auditory stimulus design is the use and specific manipulation of natural sounds. This has been particularly useful for understanding the perception of ethological sounds such as speech (or species-specific vocalisations in general) and

music. The advantage of this approach is its ethological validity, since it most closely approaches real world listening situations. For example, the STS of both human and non-human primates, respectively, is sensitive to specific timbral cues such as those present in voices or con-specific vocalisations (Belin, 2006; von Kriegstein & Giraud, 2004). However, this approach faces the difficulty of controlling stimulus parameters such that the observed responses are not confounded by low-level acoustic confounds (such as spectral or temporal complexity). A further caveat is that the use of natural stimuli constrains the possible inferences to the stimulus material used; that is, natural stimuli automatically evoke semantic associations, in which case the degree to which behavioural or neural responses are attributable to the acoustic parameters or stimulus semantics is difficult to assess, or needs to be specifically addressed in the experimental design (von Kriegstein *et al.*, 2003).

Two approaches can be distinguished: one uses natural sounds as such, or natural sounds that are minimally manipulated to fit certain secondary criteria (e.g. sound duration), so as to most closely emulate perception in a complex world (Nelken *et al.*, 1999; Schnupp *et al.*, 2006; Wang, 2000; Wang & Kadia, 2001). This approach has yielded important insights into higher cognitive processes such as language (Hickok & Poeppel, 2007; Price, 2000) or music perception (Stewart *et al.*, 2006).

The second approach directly manipulates specific characteristics of the auditory signal so as to isolate critical determinants of the signal. A prominent experimental paradigm is the use of vocoding techniques and its variations, which allow the control of spectral and temporal information in the acoustic signal (Davis & Johnsruide, 2003; Narain *et al.*, 2003; Obleser *et al.*, 2008; Overath, 2004a; Scott *et al.*, 2000; Scott *et al.*, 2006; Shannon *et al.*, 1995; Smith *et al.*, 2002). For example, in noise vocoding (Shannon *et al.*, 1995), the speech signal is divided into a fixed number of spectral bands (from just one band to multiple contiguous bands), and the spectral information of the speech signal in all the bands is replaced with white noise; this procedure retains the overall temporal structure of the speech signal in the spectral bands, while altering the spectral information. Speech intelligibility is retained with the presence of only a few spectral bands; thus, much of the information for speech intelligibility is carried in the envelope (Shannon *et al.*, 1995; Smith *et al.*, 2002). At the level of the cortex, Scott and colleagues (2000; Narain, 2003) showed

that both normal speech and noise-vocoded speech that was still intelligible led to activations in a left-lateralised temporal lobe network when compared with acoustically matched, but unintelligible, speech (see also Giraud *et al.*, 2004).

Conversely, spectral resolution is critical for music perception (Smith *et al.*, 2002); spectral information in speech operates at a coarser spectral resolution (e.g. in the case of formants). The importance of spectral resolution for music perception is also evident from patients with cochlear implants who generally are unable to appreciate music, since the physical constraints of cochlear implants limit the number of spectral channels (Shannon, 2005). The spectrotemporal trade-off theory (Zatorre & Belin, 2001; Zatorre *et al.*, 2002a) draws on this dissociation by highlighting the differential importance of spectral and temporal information for music and speech perception, respectively (see also Study 1, Chapter 3).

### 2.1.3 Complex probabilistic stimuli

#### 2.1.3.1 *Spectrotemporal correlation in complex AM spectra*

As described in Section 2.1.1.3 above, the amplitude of natural sounds fluctuates over time. A common approach for investigating principles of AM processing in the auditory system utilises SAM. The systematic independent or interactive manipulation of modulation rate and modulation depth in SAM sounds enables the investigation of basic principles of AM processing in the auditory system. However, SAM is a rather deterministic stimulus manipulation, since a given sound commonly has a fixed modulation rate  $f_m$  and a fixed modulation depth  $m$ . In contrast, the envelope in ethological sounds varies over different time scales and to different degrees in a complex and often non-deterministic manner. The auditory system needs to both track and integrate the information in the signal over different time scales and frequency regions so as to perceive auditory objects and disambiguate between auditory objects.

One approach to conceive of the probabilistic nature of ethological sounds is in terms of spectrotemporal correlation. In this framework, correlation can be regarded as a probabilistic principle, such that the individual instantiations of the signal obey

certain global constraints over a given time frame, while their precise instantiations are non-deterministic and can take any form within the constraints. This approach allows the synthesis of sounds with specific global constraints, set by a given degree of correlation, while keeping other stimulus parameters constant.

In general, many ethological sounds contain various AM rates or modulation rates simultaneously; for example, phonemes and syllables in speech operate over different modulation rates, the former at the order of tens of milliseconds, the latter over hundreds of milliseconds (Poeppel, 2003; Rosen, 1992). These rates or time scales can be described in terms of spectrotemporal correlation and implemented in synthetic sounds (see Study 1 in Chapter 3). Given a unit frame length of 20 ms within a longer sound, the degree of correlation between any two consecutive frames introduces time windows of different lengths: low correlation values result in short time windows within which correlation exists (i.e. some 20 ms), while high correlation values result in longer time windows spanning hundreds of milliseconds. The probabilistic nature of such synthetic sounds, given by the degree of correlation, closely resembles that of naturally occurring speech sounds such as phonemes or syllables.

In mathematical terms, the degree of AM in complex sounds can be operationalized in terms of the Pearson product moment correlation  $r$ . For a complex spectrum with a fixed set of  $n$  constituent frequencies, the degree of correlation between two frames  $x$  and  $y$  can be described as

$$r(x, y) = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y}. \quad (\text{Eq. 2-8})$$

Here,  $x$  and  $y$  are vectors representing the instantaneous amplitude values (in dB) of the frequencies in the spectrum,  $\bar{x}$  and  $\bar{y}$  are the arithmetic means of  $x$  and  $y$ , and  $s_x$  and  $s_y$  represent the standard deviations of  $x$  and  $y$ .

In this implementation, at the local level, each frequency has an independent or stochastic AM profile that is constrained at a global level by the overall degree of correlation between the two spectra as a whole. From an information theoretic perspective, the processing of such signals requires the assessment of higher-order

statistical properties, while ignoring local fluctuations. Essentially, the auditory system must track the properties of the acoustic signal with respect to different time windows and detect higher-order statistical properties such as spectrotemporal correlation in complex acoustic spectra consisting of multiple frequencies. Such sounds approach the complexity of natural sounds, while their statistical properties can be controlled.

### 2.1.3.2 *Spectrotemporal coherence*

Many naturally occurring sounds are characterised by FM, and mechanisms for processing single FM ramps of different slope and range, at the level of subcortical and cortical structures, have been studied in detail (Rees & Malmierca, 2005). However, single FM sweeps are still far removed from the complexity of FM in ethological sounds. For example, many monkey vocalisations and speech sounds comprise coherent FM across multiple frequencies, in which the fundamental frequency  $f_0$  and its harmonics move coherently up or down in frequency (Rees & Malmierca, 2005).

In a complex acoustic world, coherently moving frequencies are likely to emanate from the same source and are thus interpreted as being part of or forming an auditory object. To achieve this, the auditory system needs to assess the acoustic signal over various frequency bands and time scales simultaneously. Again, this is a higher-order process that requires the detection of global statistical properties irrespective of local fluctuations. Furthermore, changes in the global statistical properties are likely to signal transitions between objects, and the auditory system thus needs to detect these changes at a higher-order level of integration that allows it to disambiguate them from mere (stochastic) local fluctuations.

The stimulus used in Study 2 (Chapter 4) assessed the neural correlates underlying the higher-order integration and segregation of auditory objects in complex sounds, in which distinct objects were identified by higher-order spectrotemporal coherence. Specifically, sounds consisted of multiple linear FM

ramps that were randomly distributed in frequency-time space, forming an ‘auditory texture’. The degree of coherence was defined as the percentage of FM ramps with identical frequency trajectory (slope) and direction (up or down) (while the remaining FM ramps had random trajectories and directions). Thus, since the FM ramps were randomly distributed, a mechanism detecting spectrotemporal coherence needed to assess the acoustic signal over multiple frequency bands and time windows covering hundreds of milliseconds. Similarly, mechanisms detecting spectrotemporal coherence transitions needed to assess stimulus transitions at a higher-order statistical level covering multiple FM ramps, rather than at the level of individual FM ramps.

### 2.1.3.3 *Information theoretic properties of pitch sequences*

As described in the General Introduction, experimental paradigms such as the MMN or auditory streaming paradigms, which investigate generic principles of auditory scene analysis, are limited by their deterministic nature. While it is inherent to the MMN paradigm to have ‘non-deterministic’ oddball stimuli, which violate the seemingly established statistical rules represented by the standard stimulus, the overall complexity still does not compare to the complexity encountered in natural sounds.

An elegant way to define global information theoretic properties of pitch sequences without explicitly defining the precise intervals or local fluctuations is to derive pitch sequences from exponential power spectra. In mathematical terms, exponential power spectra are defined by

$$I_n = kf^{-n}, \quad (\text{Eq. 2-9})$$

where  $k$  is a constant and the exponent  $n$  determines the slope of the amplitude spectrum across frequencies  $f$ . For  $n = 0$ ,  $I_0$  is constant across all frequencies (white noise); as  $n$  increases, the exponential slope increases and acts as a low-pass filter. Performing an inverse FFT (IFFT) on a power spectrum with a given exponent  $n$  gives a time series; the time points in this series can then be treated as representing pitches to form a pitch series. ‘Fractal’ pitch sequences based on inverse Fourier

transforms of  $f^{-n}$  power spectra (Patel & Balaban, 2000; Schmuckler & Gilden, 1993) provide a means to control directly the entropy of the sequence via the exponent  $n$  (see Chapter 5).

## 2.2 Magnetic resonance imaging

Magnetic resonance imaging (MRI) has its basis in nuclear magnetic resonance (NMR) of the nucleus of individual atoms. NMR can be traced back to the 1940s (Bloch et al., 1946; Purcell et al., 1946). The word ‘nuclear’ was later dropped in the clinical environment in favour of MRI, to avoid a false connection with nuclear radioactivity.

If an atomic nucleus has an odd number of protons and nucleons, this imbalance causes the atomic nucleus to spin around its axis, thus creating a magnetic momentum. Atomic nuclei with an even number of protons and nucleons, or an even atomic mass number, do not have a net spin or angular momentum and thus do not emit NMR signals. In particular, hydrogen ( $H$ ) atoms not only have a pronounced nuclear momentum, but are also abundant in natural tissues. In the absence of a strong magnetic field, the orientation of the atomic nuclei in normal tissue is random, and no net magnetic field can be detected. However, if a strong external magnetic field  $B_0$  is applied, the magnetic moments align either parallel to  $B_0$  (low-energy state) or anti-parallel to  $B_0$  (high-energy-state). The magnetic field strength is denoted in Tesla ( $T$ ), where 1 Tesla = 10,000 Gauss. For comparison, the magnetic field strength of the earth is approximately 0.5 Gauss. Typical magnetic field strengths for human-compatible MRI scanners are 1.5, 3, or 4.7  $T$ , while some high-field MRI scanners reach 7  $T$ .

NMR makes use of two properties of atomic nuclei: their alignment and their precession around themselves. A small majority of the atomic nuclei align in the low-energy state parallel to  $B_0$ , leading to a small net magnetization effect (in a 1.5 T scanner, this ratio is about 1/10,000 000). The nuclei do not align precisely parallel or

anti-parallel to  $B0$ , but precess around  $B0$ , much like a spinning top. The speed of the precession is proportional to  $B0$  and is defined by the Larmor equation:

$$\omega = \gamma B0 \quad (\text{Eq. 2-10})$$

Here,  $\omega$  is the (Larmor) frequency in MHz ( $\omega/2\pi$ ),  $\gamma$  is the gyromagnetic ratio in MHz/Tesla for the spin under consideration, and  $B0$  is the external magnetic field strength. The Larmor frequency of hydrogen atoms is 42.578 MHz/Tesla. In a three-dimensional reference frame or coordinate system x-y-z, there are two magnetisations at work due to the alignment and precession of the nuclei. The longitudinal magnetisation denotes the magnetic moment due to the orientation of the nuclei (the z-axis is defined to be aligned with  $B0$ ); the transverse magnetisation denotes the magnetisation in the x-y plane due to the precession of the nuclei around the z-axis. Since all nuclei precess in random phase, there is no net detectable transverse magnetisation. At this stage, the longitudinal magnetisation is the only detectable magnetisation due to the slight excess of nuclei that are aligned in the low-energy state parallel to  $B0$ .

The application of a brief ( $\sim 1$  ms) radio frequency (RF) excitation pulse at time  $t0$  with the same frequency as the precessing  $H$  atoms introduces an additional magnetic field  $B1$  perpendicular to  $B0$ . This has two effects: firstly, the longitudinal magnetisation of the nuclei is tilted towards  $B1$  and the magnetisation in the z-direction is reduced. The time for the longitudinal magnetisation to *increase* to within 63% of its magnetisation at time  $t0$  is denoted as  $T1$ . Secondly, the RF pulse with Larmor frequency  $\omega$  causes the nuclei to precess in phase, thus producing, for the first time, a transverse magnetisation in the x-y plane (hence the term magnetic resonance). After cessation of the RF signal, the precession of the individual nuclei begins to de-phase again (mainly through mutual interference due to spatial proximity), thereby releasing a small amount of radiation at the Larmor frequency and consequently reducing the transverse magnetisation. The time for the transverse magnetisation to *decrease* to within 63% of its magnetisation at time  $t0$  is denoted as  $T2$ .  $T2$  assumes an ideal tissue where magnetisation is homogenous throughout. However, the macroscopic geometry and composition of the imaged sample, e.g. the head, vary greatly and influence the magnetic susceptibility. The effective time in normal tissue for the transverse magnetisation to decrease to within 63% of its magnetisation at time

$t_0$  is denoted as  $T2^*$ . For example, the transition from tissue to air (which have different magnetic susceptibilities) at the sinuses is particularly pronounced. Importantly for fMRI, magnetic susceptibility variations are also present around blood vessels, where the de-oxyhaemoglobin (dHb) level affects  $T2^*$  in the surrounding tissue. An RF receiver coil can then detect and amplify the signal related to  $T1$  and  $T2^*$ .

Importantly, the NMR signal is proportional to the density of the protons in each tissue and  $T1$  and  $T2^*$  differ for different tissues. The most common technique in NMR is a spin-echo technique. Generally, a  $90^\circ$  RF pulse perturbs the tissue at time  $t_0$  and the time at which the decay signal is read out with an RF receiver coil is the time to echo (TE). The calibration of TE and the time to repeat (TR) determine the contrast and quality of an MR image. Due to inhomogeneities in the magnetic field and those introduced by the tissue, the transverse magnetisation decay will vary across different spatial locations ( $T2^*$ ). A second ‘echo’ RF pulse of  $180^\circ$  is therefore applied to ‘refocus’ the transverse magnetisation decay at time TE/2 and essentially neutralise the effects of  $T2^*$  dephasing (i.e. the spin-echo acquisition is less susceptible to  $T2^*$  effects). The NMR signal received at the RF receiver coil at time TE is then decomposed via a Fourier transform. An alternative technique in NMR is a gradient-echo technique. This acquisition technique records the signal after the initial  $90^\circ$  RF pulse without phase refocusing and is thus more susceptible to  $T2^*$  effects; for this reason, it is commonly used in fMRI.

As described thus far, NMR has been restricted to tissue classification without any spatial information. The birth of MRI can be traced to the 1970s, when it was realised that NMR could also reveal spatial properties of tissues by spatially varying the magnetic field, and consequently the Larmor frequency, along a gradient (Damadian *et al.*, 1977; Lauterbur, 1973; Mansfield & Grannel, 1973). By inducing a gradient field along the three main coordinates ( $x$ - $y$ - $z$ ), the received signal can then be decomposed using Fourier transforms and it is possible to spatially reproduce anatomical properties of the tissue. That is, the received or resonant frequency is now a function of spatial position within the imaged tissue.

Generally, the steeper the slope of the magnetic gradient and the longer its application, the higher the spatial frequency resolution, and vice versa (DeLaPaz,

1994). Thus, the combination of amplitude and duration of the gradients determine the spatial frequency encoding. The spatial frequency information is represented in planar k-space, where the x-axis represents the ‘read-out’ gradient ( $G_x$ ) that encodes the spatial frequency, and the y-axis represents the ‘phase-encode’ gradient ( $G_y$ ). High spatial frequencies are represented towards the periphery and low frequencies at the centre of k-space. The k-space trajectory along which the signal is encoded traverses the different phases (via a series of appropriate RF pulses) of a spatial frequency before advancing to and repeating the same procedure for the next spatial frequency. The application of a slice-selection gradient ( $G_z$ ) that is perpendicular to the x-y plane enables the acquisition of multiple planes of the imaged tissue. This gradient ensures that only protons in a selected slice (x-y plane) are ‘resonant’ to the applied RF pulses and emit a signal. Finally, an inverse Fourier transform of the frequency-phase information within each plane can reveal the spatial properties of the imaged tissue.

## **2.3 Functional magnetic resonance imaging**

### **2.3.1 Echo-planar imaging**

The image acquisition techniques described so far are relatively slow (on the order of several minutes for one volume, e.g. to cover the head of a person), since essentially each row of the k-space is preceded by an RF excitation pulse. However, in order to track physiological changes, e.g. changes due to different oxygenation levels in blood vessels (see Chapter 2.3.2), scan volumes need to be acquired faster. Echo-planar imaging (EPI; Mansfield, 1977) allows ultra-fast acquisition of the x-y plane with a single RF excitation pulse (‘single shot’), which is on the order of tens of milliseconds per volume. This is achieved by rapid ( $\sim 1$  kHz) switching of the frequency ( $G_x$ ) and phase ( $G_y$ ) gradients to cover the entire plane. Commonly, EPI is performed with a gradient-echo acquisition sequence instead of a spin-echo sequence, since the former is more sensitive to  $T2^*$  changes (see Chapters 2.2 and 2.3.2). Such gradient echoes are generated via an oscillating gradient along the read-out ( $G_x$ ) direction, following a ‘zig-zag’ trajectory in k-space. In EPI, the TE is defined as the time from the RF excitation pulse to the centre of k-space, which is approximately equal to  $T2^*$

(Logothetis, 2002). While EPI is extremely powerful, it requires dedicated hardware that can withstand the taxing physical stress of magnetic gradient switching, and it is prone to artefacts (DeLaPaz, 1994).

### 2.3.2 Physiological basis of BOLD signal and haemodynamic response function

In their seminal studies, Ogawa and colleagues (Ogawa & Lee, 1990; Ogawa *et al.*, 1990a; Ogawa *et al.*, 1990b) discovered that the NMR signal in blood vessels in the rat brain varied with changes in blood oxygenation demand, or blood flow. Specifically, paramagnetic dHb (Pauling & Coryell, 1936) causes a susceptibility difference between the vessel and its surrounding tissue, which in turn leads to an increased dephasing of the protons and a decrease in the associated T2\* signal. The diamagnetic oxyhaemoglobin (oHb) does not produce such an effect. If a brain area is activated, cells consume oxygen from nearby blood vessels, leading to a temporary increase of the dHb/oHb ratio. However, soon after, in an overshoot mechanism, blood is directed towards the active site, leading to a net increase in oHb and a decrease in the dHb/oHb ratio; this increase in oHb causes an increase in the previously disturbed spin coherence (T2\*) and consequently an NMR signal intensity increase. The resulting blood oxygen level dependent (BOLD) signal is thus an indirect index of neural activity.

The BOLD signal has its physiological basis in the so-called haemodynamic response function (hrf) and can be divided into several characteristic phases (Logothetis, 2002). An initial undershoot (Malonek & Grinvald, 1996) is followed after about 2 seconds by an increase in the BOLD signal that is mainly due to increase in blood flow directed towards the active region (Fox & Raichle, 1986). The hrf peaks approximately 4-6 seconds after onset of stimulation and decreases fairly rapidly after cessation of the activating stimulus. This is followed by an undershoot phase in which the BOLD signal decreases to below its initial magnitude, which can be explained by vasodilatation and an increase in local venous blood volume (Buxton *et al.*, 1998). The BOLD signal returns to normal after approximately 32 seconds. The associated

BOLD signal changes are minute, on the order of 1-1.5% in the auditory cortex (Talavage *et al.*, 1999).

From first principles, the BOLD signal is considered an ‘indirect’ measure of neural activity and its precise neuronal underpinnings are still under investigation (e.g. Logothetis, 2002, 2003, 2004, 2008). However, pioneering work by Logothetis and colleagues, combining intracortical electrophysiological recordings and fMRI in anaesthetised (Logothetis *et al.*, 2001) and unanaesthetised macaques (Goense & Logothetis, 2008), showed that the fMRI BOLD signal correlates better with local field potentials (LFPs), an index of pre-synaptic integration, than with post-synaptic action potentials, as assessed via multi-unit activity (MUA). Furthermore, when MUA responses adapt while LFPs remain unaffected, the BOLD signal remains unaltered. Thus, the best predictor of the BOLD signal is LFP activity in the gamma frequency range (20-60 Hz), implicating the importance of neuromodulatory processes (Goense & Logothetis, 2008).

## **2.4 fMRI and auditory stimulus presentation**

Acoustic noise due to the mechanical switching of the magnetic gradient coils is a serious constraint for fMRI studies investigating auditory processing. In conventional continuous 1.5 T and 3 T EPI, sound pressure levels in the bore of the scanner exceed 120 dB (Price *et al.*, 2001; Ravicz *et al.*, 2000). The main source of acoustic noise is the readout phase during the imaging protocol, with other ambient noise factors such as the helium cooling pump and air conditioning system only contributing relatively little to the total noise level (Ravicz *et al.*, 2000). Furthermore, the noise produced by the scanner has a broad spectrum from 250 Hz to 4 kHz with a typical peak at around 1-1.5 kHz (Chambers *et al.*, 2001; Hall *et al.*, 1999; Ravicz *et al.*, 2000), which covers a crucial frequency range in human auditory perception. While ear protection such as ear defenders can achieve some 20-40 dB noise reduction, a significant part of the scanner noise is still conducted via the ear canal (< 500 Hz) and bones (> 500Hz) (Ravicz & Melcher, 2001). Active noise cancellation techniques (Amaro *et al.*, 2002; Chambers *et al.*, 2001; Moelker & Pattynama, 2003) can further reduce the effect of

the scanner noise, but their effectiveness is similarly limited by the contribution of bone conduction (Ravicz & Melcher, 2001).

Apart from the generally unpleasant experience for the participant, the noise also considerably reduces the signal to noise ratio (SNR). That is, when presenting an auditory stimulus while acquiring scan volumes, the haemodynamic response (and consequently the BOLD signal) due to the stimulus and that due to the scanner noise are confounded and difficult to disambiguate. The constant background noise of the scanner furthermore introduces a continuous stimulus for the auditory system, resulting in adaptation or habituation, as well as inhibitory processes, particularly in subcortical and primary auditory structures. Since the haemodynamic response does not behave linearly across sound levels, the effects of loud background noise levels are often unpredictable (Belin *et al.*, 1999; Edmister *et al.*, 1999; Talavage & Edmister, 2004), rendering the notion of a ‘silent’ baseline condition problematic. For example, subtracting the haemodynamic response to a ‘silent’ baseline during the presence of scanner noise from the haemodynamic response to an experimental stimulus during the presence of scanner noise ([stimulus + scanner noise] – [‘silence’ + scanner noise]) is not identical to subtracting silence from an experimental stimulus ([stimulus – silence]) (Gaab *et al.*, 2007). A further, more cognitive, constraint is that the acoustic stimuli are difficult for participants to hear due to the background noise, and the experiment essentially becomes a figure-ground task rather than a true sensory or perceptual representation of the experimental stimulus attributes per se (Scheich *et al.*, 1998).

A variety of imaging protocols have been introduced to circumvent or avoid these effects for studies investigating auditory perception. Principally, there are three variants of ‘silent’ imaging designs that offer different temporal resolutions. Eden and colleagues (1999) used an imaging protocol in which a silent period (~ 2 seconds) is inserted between scan volume acquisitions that was long enough to present a short acoustic stimulus. The temporal resolution of such behaviour interleaved gradients (Eden *et al.*, 1999) or compressed (Amaro *et al.*, 2002) protocols is not as precise as conventional continuous imaging, but still on the order of a few seconds. Belin and colleagues (1999) introduced an event-related design by lengthening the silent period between volume acquisitions to some 9 seconds while jittering the presentation of the

acoustic stimulus within that silent period; the haemodynamic response to the acoustic stimulus is thus captured at different time points. Hall and colleagues (1999) acquired a volume only at the predicted peak of the haemodynamic response. This ‘sparse’ imaging protocol carries minimal information about the shape of the haemodynamic response to different experimental stimuli, i.e. its temporal resolution is very limited. However, it maximises the dissociation between the haemodynamic response of the experimental stimulus and that of the scanner noise and thereby significantly improves the SNR (Hall *et al.*, 1999).

Apart from the trade-off with respect to time resolution, a further constraint of these designs is the considerable length of the scanning procedure required to obtain reasonable SNR, which makes its use difficult in certain subject populations (e.g. clinical patients). In general, when the experimental question is weighted towards the sensory representation of certain acoustic attributes that are likely to be influenced by adaptation and background noise, ‘silent’ or ‘sparse’ imaging protocols should be preferred. However, if temporal resolution is of the essence or the experimental question addresses higher level auditory processes in non-primary cortex that are less likely to be significantly affected by the scanner noise, continuous imaging protocols should be preferred. In the work presented in this thesis, both continuous and sparse acquisition protocols were used. Continuous acquisition was the protocol of choice for Studies 1 and 2, since in these studies the temporal dynamics of the data were of particular interest; furthermore, the experimental question for Study 2 required an experimental design whose effects could only be captured with the temporal resolution of a continuous imaging protocol. In contrast, Studies 3-5 used a sparse acquisition paradigm, since the pitch sequences employed in these studies extended over multiple seconds, making it possible for the BOLD signal to be captured at the end of each stimulus block or pitch sequence.

## 2.5 Image pre-processing

The analysis of functional imaging data requires elaborate pre-processing algorithms before any statistical analysis with respect to the experimental effects can be performed. Specifically, the successive scans need to be realigned to account for movement of the participants, normalised to a standard stereotactic reference frame that allows between-subject comparisons, and finally smoothed to increase the SNR. Each of these steps, executed within Statistical Parametric Mapping (SPM) software (<http://www.fil.ion.ucl.ac.uk/spm>), and encompassing a sophisticated theoretical and mathematical background, is described briefly below.

### 2.5.1 Realignment and unwarping

The successive scan volumes in fMRI are treated as a time series. In an ideal world, a given image volume element (voxel) would represent the same cortical area across scans. However, the spatial resolution of fMRI ( $\sim 3 \times 3 \times 3$  mm) means that even tiny movements lead to misalignment across successive scans; in fact, movement on the scale of micro-millimetres can significantly affect the data (Friston *et al.*, 1995a) and can contribute as much as 90% of the variance of the data (Friston *et al.*, 1996b). This leads to signal changes in a given voxel that might then be misattributed as ‘activation’. While the most serious movement artefacts are due to participants’ head movements, even small movements due to the cardiac cycle movement are a source of scan misalignment, particularly in brainstem structures. Motion that is uncorrelated with the experimental conditions generally introduces external noise and consequently decreases the detection of true activation; conversely, motion that is correlated with an experimental task can lead to misattribution of signal changes as ‘activation’. Thus, it is essential to remove movement artefacts.

Typically, the first image of the time series is treated as a global reference for the first scans of subsequent sessions, which are then in turn used as references for the remaining scans within their session. The realignment routine uses a least squares approach and a 6 parameter (three translations and three rotations) affine ‘rigid-body’

spatial transformation to calculate the motion associated with each scan (Andersson *et al.*, 2001; Friston *et al.*, 1995a). These parameters are then used to ‘reslice’ the scan to the new grid coordinates, usually via sinc interpolation (Grootenck *et al.*, 2000). There are, however, additional non-linear movement-related artefacts; specifically, magnetic inhomogeneities particularly in regions with an air-tissue interface, such as the orbitofrontal cortex or the anterior inferior temporal lobes, cause deformations in the sampling matrix (Andersson *et al.*, 2001) and are further distorted by movement. The unwarping routine takes account of such susceptibility-by-movement interactions. It can be further informed by the acquisition of  $B_0$  magnetic fieldmaps for each participant, which provide an explicit measure of  $B_0$  inhomogeneities and associated geometric distortions (Cusack *et al.*, 2003; Hutton *et al.*, 2002).

### 2.5.2 Normalisation

Since the anatomy of individual brains differs, it is necessary to transform the scans into a stereotactical reference space that allows comparisons across participants. The realign and unwarping routine implemented within SPM creates a mean functional image of all functional scans; this image is subsequently used to estimate warping parameters that map it onto a common stereotactic space (Talairach & Tournoux, 1988; Toga *et al.*, 1994) or an average brain derived from large sets of previous imaging data (Evans *et al.*, 1993; Mazziotta *et al.*, 1995; Roland & Zilles, 1994). The estimation is commonly achieved via a 12-parameter affine transformation (translations, rotations, zooms, and shears), where the parameters constitute a spatial transformation matrix (Friston, 2003b). This is then followed by an iterative non-linear estimation of spatial deformation patterns.

### 2.5.3 Smoothing

The final stage of the preprocessing routines convolutes the data with an isotropic Gaussian smoothing kernel. The primary motivation for smoothing the data is to increase the SNR ratio (Friston, 2003b), albeit at the cost of spatial resolution. Confounding effects of noise at the level of individual voxels can be reduced by convolution with a smoothing kernel whose support is about 2-3 times the voxel size. Generally, a smoothing kernel of 6 mm full-width-at-half-maximum (FWHM) is appropriate at the single-subject level, while an 8 mm kernel at the group level is able to take into consideration the morphological differences between participants. However, small structures, such as nuclei in the brainstem, require smaller smoothing kernels of about 4 mm FWHM.

## 2.6 Statistical analysis

The pre-processing routines described above enable the examination of regionally specific effects of the experimentally manipulated variable(s) within a statistical framework. SPM software combines the General Linear Model (GLM), described in Section 2.6.1, to estimate the effects of interest due to the experimental variable(s) with Gaussian Random Field (GRF) theory to model spatially extended processes (Section 2.6.2). The result is a statistical parametric map (SPM) that represents the regionally-specific effects of the experimental variable(s).

### 2.6.1 General Linear Model (GLM)

The GLM provides a framework for the statistical analysis of functional imaging data. It incorporates common statistical tests such as Student's *t*-test or analyses of variance (ANOVAs). Essentially, the signal intensity of each voxel within a scan is treated as a time series across scans and approximated via the general equation

$$Y = X\beta + \varepsilon \quad (\text{Eq. 2-11})$$

Here,  $Y$  is the data matrix with cells  $y_{ij}$  for  $j$  voxels (rows) across  $i$  scans (columns);  $X$  is the design matrix with cells  $x_{ik}$ , with one column for every effect of interest in each row;  $\beta$  is a parameter matrix where  $\beta_j$  is a column vector of parameter estimates for each row;  $\varepsilon$  is a matrix of normally distributed error terms (Friston *et al.*, 1995b). The effects of interest in  $\beta$  are modelled via convolution with a canonical hrf (see Section 2.3.2). The experimental conditions and their corresponding parameter estimates are contrasted against each other by appropriately weighting the columns  $\beta_j$ . A normal  $t$  statistic can then be obtained for each voxel via the ratio of contrast-weighted parameter estimates to the estimated standard error term.

### 2.6.2 Gaussian Random Field (GRF) theory

GRF theory assumes that, under the null hypothesis, SPMs of the voxel parameter estimates for a given condition are distributed according to a known probability density function, normally Student's  $t$  or  $F$  distributions. Any deviations of this distribution that surpass a set (significance) threshold can be attributed to the experimental variable with a certainty  $1-\alpha$ , where  $\alpha$  is the Type I error of falsely rejecting the null hypothesis.

In a typical fMRI study with whole head coverage, a single scan volume comprises tens of thousands of voxels. The statistical comparison of each voxel with all other voxels so as to estimate the effect of the experimental variable(s) introduces a considerable likelihood of false positives (Hochberg & Tamhane, 1987). For example, when testing 200,000 voxels at a significance threshold of  $p < 0.001$  (a common statistical threshold in fMRI), about 200 voxels can be expected to show chance, and therefore potentially false, 'activation'. A conservative, straightforward approach would be to control for multiple comparisons via the Bonferroni correction by dividing the statistical threshold by the number of independent comparisons (Logan & Rowe, 2004). However, in this case, the statistical threshold would be very stringent ( $p = 0.05/200,000 = 0.0000025$ ). Furthermore, the voxel time series are not

strictly independent, since neighbouring voxels are likely correlated due to macroscopic anatomy (e.g. blood vessels) and the spatial extent of the hrf.

Using a conservative threshold also reduces the power to detect activation (or it increases the likelihood of falsely rejecting truly active voxels). Thus, a common convention sets a significance threshold of  $p < 0.001$  for brain areas where the researcher has an *a priori* hypothesis (e.g. auditory cortex). In more exploratory studies or for activations in brain areas where the researcher did not have an *a priori* hypothesis, a more conservative threshold that accounts for the problem of multiple comparisons is prudent so as to avoid false rejections of the null hypothesis. A common approach is the family-wise error (FWE) rate (Logan & Rowe, 2004; Nandy & Cordes, 2007). Here, the tens of thousands of individual voxels  $\omega$  are collected in a family  $\Omega$ , with an associated family of null hypotheses  $\{H_\omega : \omega \in \Omega\}$ . The omnibus null hypothesis  $H_Q$  is then rejected if at least one  $H_\omega$  is rejected for a set threshold  $u$ ; the individual threshold  $u$  can be chosen so that the threshold for the full family gives the desired level of certainty (commonly  $\alpha = 0.05$ ).

### 2.6.3 Random-effects analysis

There are two main types of analyses, which differ with respect to their scope of inference (Friston *et al.*, 1999). Generally, fixed-effects analyses allow inferences concerning the typical behaviour of the group of participants tested in the study, while random-effects analyses allow inferences to be drawn about the average behaviour of the general population. Specifically, fixed-effects analyses disregard inter-subject variability and essentially treat each participant as a session within a longer time series; thus, the only error source that is modelled and accounted for is the error variance between scans. Fixed-effects analyses have a high number of degrees of freedom (slightly less than the number of scans total).

Conversely, random-effects analyses also account for inter-subject variability as an additional source of variance, resulting in  $n - 1$  degrees of freedom for  $n$  participants, reducing the effect of subject outliers. A random-effects analysis requires

a sufficient number of participants (at least eight) so as to reliably estimate inter-subject variability and obtain adequate power to detect effects of interest. Random-effects analyses incorporate a two stage procedure, where the contrast of interest is computed at the single-subject level before it is evaluated at the group level. All studies in this thesis used random-effects analyses.

## **Chapter 3. ENCODING OF SPECTROTEMPORAL CORRELATION IN COMPLEX ACOUSTIC SPECTRA**

### **Summary**

Acoustic information in natural sounds evolves over a range of time scales. In speech, for example, phonemes and syllables unfold over two distinct time windows, the former on the order of tens of milliseconds, the latter over hundreds of milliseconds. The auditory system needs to track the acoustic information over these different analysis windows. Recent studies suggest differences in the encoding of short (tens of ms) and longer (hundreds of ms) time windows in left and right auditory cortex, respectively. Study 1 assessed brain activation in response to the systematic variation of the time window over which complex spectra change. The different time windows were realised by controlling the degree of correlation between successive time frames of the spectrum: the greater the correlation of the spectrum between successive time frames, the longer the time window for a given change. The parameters were chosen such that stimuli corresponded to time windows between 20-300 ms. The data show bilateral activity in the planum temporale (PT) and anterior superior temporal gyrus (aSTG) as a function of increasing time window, as well as activity in the superior temporal sulcus (STS) that was significantly lateralised to the right. No cortical areas increased their activity as a function of decreasing time windows. The network revealed as a function of increasing time windows represents a generic mechanism for the analysis of temporal structure in natural sounds. Furthermore, the data suggest a complex lateralisation model where different levels of analysis occur within different subareas of auditory cortex.

### 3.1 Introduction

This study considers the neural bases of different temporal analysis windows for natural sounds. There is accumulating evidence that auditory perception extracts acoustic information over different time scales. In speech, for example, phonemic and syllabic rates operate over two distinct time scales, the former on the order of tens of milliseconds, the latter over hundreds of milliseconds (Rosen, 1992). One model of speech perception, the ‘asymmetric sampling in time’ (AST) hypothesis (Poehpel, 2003), draws on this dissociation. It posits a lateralisation scheme in auditory cortex (AC) by which slower modulations (~3-6 Hz or ~150-300 ms) preferentially engage right AC, whereas fast modulations (~20-40 Hz or ~25-50 ms) are preferentially processed in left AC. The present study considers generic mechanisms for the analysis of the temporal structure of novel sounds with a similar level of complexity to that of speech sounds.

Previous investigations (Boemio *et al.*, 2005; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001) have manipulated the acoustic segment length within multiple-segment sounds to probe for distinct processing of different temporal modulations or time windows. However, results in these studies differed with respect to specialisations of different subareas in auditory cortex for temporal modulations, either within or between hemispheres. For example, Boemio and colleagues (2005) demonstrated sensitivity to decreasing temporal modulation rates in auditory association cortex (AAC), with a right-hemispheric bias that was most pronounced in right superior temporal sulcus (STS). The authors did not find evidence for differential temporal sensitivity in primary or secondary auditory cortices (PAC and SAC) as part of Heschl’s gyrus (HG).

However, others (Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001) have shown sensitivity to increasing temporal modulation rates in HG, which was more marked on the left. These studies (Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001) did not find evidence for differential temporal sensitivity in AAC. Obleser and colleagues (2008) used a more natural stimulus and instead manipulated the spectral and temporal resolution of speech signals, demonstrating slight lateralisation preferences in right and left AAC (specifically STS) for spectral and temporal resolution, respectively. Thus, critical yet unresolved questions relate to (i) the extent

to which the analysis of different levels of temporal structure depends on primary and secondary auditory cortex as opposed to ‘higher-order’ auditory association cortex, and (ii) the lateralisation of temporal analysis within these different areas (Hickok & Poeppel, 2007; Zatorre *et al.*, 2002a; Zatorre & Gandour, 2008).

Furthermore, natural sounds including speech comprise modulations of a complex spectrum over time, where the spectrum changes dynamically and where changes in the spectrum convey information about sound events (relevant to communication) and sources (relevant to identification). The temporal variation of the spectrum of naturally occurring sounds generally conforms to statistical distributions; the spectrum at any given point cannot be precisely predicted, but it will be within a range that can be defined by a statistical distribution. None of the previous studies captured this complexity of natural sounds in their experimental manipulations.

In this study, a novel stimulus (Figure 3-1) is introduced that is based on the systematic manipulation of the degree of statistical fluctuation over time in complex acoustic spectra. The rate of fluctuation is operationalized as the mean Pearson product-moment correlation ( $r$ ) between amplitude spectra in adjacent time frames, as used in previous behavioural studies of timbre (Caclin *et al.*, 2005; Krimphoff, 1993; Krimphoff *et al.*, 1994; Krumhansl, 1989; McAdams *et al.*, 1995). Rapid modulation of the spectrum (at the phonemic rate in speech sounds) corresponds to short time windows within which a given degree of correlation is always present between any two time frames of the spectrum, even if these are not adjacent. Slow modulation of the spectrum (at the syllabic rate in speech sounds) corresponds to long time windows within which a given degree of correlation is always present between any two time frames.

For a sound composed of 20 randomly chosen frequencies, the intensity of each frequency was allowed to vary between adjacent time frames such that the Pearson correlation ( $r$ ) between the adjacent time frames as a whole corresponded to a fixed value  $r$ . For a sound with high correlation between adjacent time frames (e.g.  $r = 0.9$ ), the correlation between non-adjacent time frames decays exponentially with the number of time frames (or lag) between the non-adjacent time frames. The window length of this decay process is defined as the duration over which the correlation between any two non-adjacent time frames reaches a minimum value ( $r = 0.2$ , in the

present case) from its initial value  $r$  between adjacent time frames. It is calculated by the following Equation 3-1:

$$window\_length(0.2) = frame\_duration \cdot \frac{\ln(0.2)}{\ln(|r|)} \quad (\text{Eq. 3-1})$$

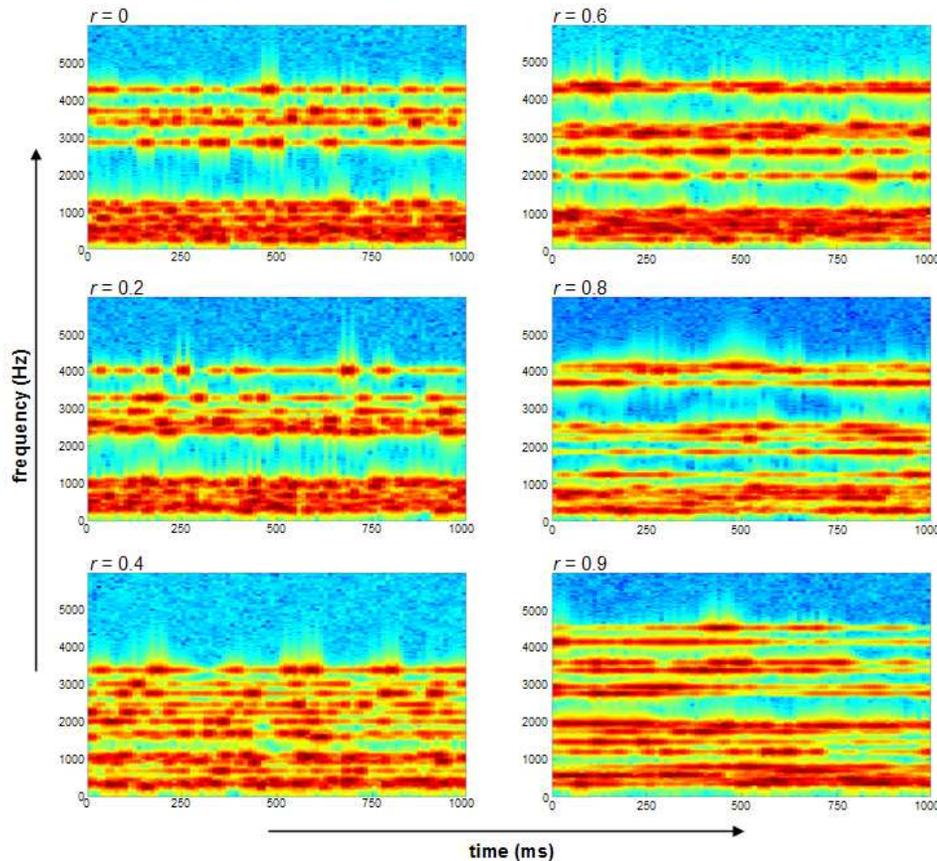


Figure 3-1 Auditory stimulus. Spectrograms of representative stimuli from each level of correlation.

In the case of small initial values of  $r$  (e.g.  $r = 0$  or  $r = 0.2$ ), the correlation between non-adjacent time frames is not appreciably different from that between adjacent time frames. Figure 3-3 (inset) shows the relationship between the correlation ( $r$ ) and the window length when the frame duration is 20ms: the window length corresponding to values of  $r$  between 0.2 and 0.9 varies between 20ms and 305ms, encompassing windows relevant to phonemic and syllabic processing, respectively (Rosen, 1992).

The stimuli are more similar to the acoustic complexity of speech and other naturally occurring sounds than the stimuli used in previous studies (Boemio *et al.*, 2005; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001). Unlike speech, however, the stimuli allow systematic manipulation of the time windows over which correlation-controlled change in the spectrum occurs without any semantic confound, enabling the investigation of fundamental mechanisms for timing analysis.

Using fMRI, haemodynamic activity was measured while participants listened to stimuli with multiple components where the correlation ( $r$ ) across the spectrum was varied in six steps between a value producing no correlation between adjacent time frames ( $r = 0$ ) to one producing strong correlation ( $r = 0.9$ ). Specifically, differences in activation were sought as a function of increasing and decreasing spectrotemporal correlation and the associated window length between (i) primary and secondary cortices in HG, and AAC, and (ii) between the two hemispheres of the brain.

## 3.2 Materials and Methods

### 3.2.1 Participants

17 right-handed participants (aged 18-31, mean age = 25.35, 9 females) with normal hearing and no history of audiological or neurological disorders provided written consent prior to the study. The study was approved by the National NHS Research Ethics Committee.

### 3.2.2 Stimuli

All stimuli were created digitally in the frequency domain using Matlab (<http://www.mathworks.com>) at a sampling rate of 44.1 kHz and 16 bit resolution. Each sound consisted of 20 sinusoids randomly chosen from a pool of 101 logarithmically spaced frequencies between 246 – 4435 Hz. The particular parameters were chosen so as to approximate respective features in naturally occurring sounds, which typically have complex spectra with multiple frequencies present. The

bandpass (246 – 4435 Hz) covers the acoustic range that is most important for human auditory perception, and the number of frequencies within this pool (101 frequencies) are a result of this range. The amplitude spectrum was defined in 20 ms frames such that the correlation from one frame to the next was operationalized as the Pearson product moment correlation  $r$ :

$$r(x, y) = \frac{1}{n} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x s_y} \quad (\text{Eq. 3-2})$$

where  $x$  and  $y$  are the amplitude (in dB) vectors over the 20 frequency components of two consecutive frames,  $n$  is the number of frequencies,  $s_x$  and  $s_y$  represent the standard deviations of  $x$  and  $y$ , and  $\bar{x}$  and  $\bar{y}$  are the arithmetic means of  $x$  and  $y$ , respectively. Thus, the amplitude spectrum of a given sound varied with a specified correlation ( $r = 0, 0.2, 0.4, 0.6, 0.8, 0.9$ ) between the 20 ms segments. Linear spline interpolation amplitude transitions were applied between frames, so that sounds were continuous and did not have any sudden amplitude jumps. This was applied in order to render the sounds more similar to most ethological sounds; however, some speech sounds like plosives or consonant-vowel do display sudden amplitude jumps (Rogers, 2000). Importantly, the mean amplitude (65 dB) and standard deviation (SD = 15) were identical for each frequency component in a given sound and across correlation levels. Each sound had a rise and fall time of 20 ms.

### 3.2.3 Experimental design

Prior to the experiment in the MRI scanner, participants were familiarised with the stimuli and then performed 2I2AFC psychophysics with  $r = 0$  as reference sounds. Stimuli were 2 sec. long and were different exemplars from the ones subsequently used in the scanner. Psychophysics ensured that participants were able to distinguish a highly correlated sound from the reference sound, and they needed to reach at least 90% correct performance for the strongest correlation ( $r = 0.9$ ) to be included in the fMRI study. Psychometric functions and 95% correct perceptual thresholds were estimated via a Weibull bootstrapping procedure (Wichmann & Hill, 2001).

Stimuli in the scanner were of different durations (1, 2, 3, or 4 sec.) and separated by a mean inter-stimulus interval (ISI) of 2 sec. (range: 1.5-2.5 sec.) as well as occasional silence trials of 6 sec. duration (20 per session). Stimuli were presented in a pseudorandom order, with 20 exemplars for each level per session (80 stimuli per level in total, amounting to a total presentation time of 200 sec. per level). Participants performed a stimulus-irrelevant task by pressing a button after each sound.

Stimuli were presented via NordicNeuroLab (<http://www.nordicneurolab.com>) electrostatic headphones at 80 dB sound pressure level (SPL) using Cogent software (<http://www.vislab.ucl.ac.uk/Cogent>).

### 3.2.4 fMRI protocol and analysis

Gradient weighted echo planar images (EPI) (see Section 2.3.1) were acquired on a 3 Tesla Siemens Allegra system (Erlangen, Germany), using a continuous imaging protocol with 42 contiguous slices per volume (time to repeat/time to echo, 2730/30 ms). Continuous imaging was chosen to ensure a better temporal resolution than that offered by sparse imaging protocols; this would then facilitate an additional examination of the data with respect to its network dynamics using analysis techniques such as dynamic causal modelling (DCM Friston *et al.*, 2003; Penny *et al.*, 2004). The volume was tilted forward such that slices were parallel to the superior temporal plane. Participants completed four sessions of 250 volumes each, resulting in a total of 1000 volumes. To correct for geometric distortions in the EPI due to B0 field variations, Siemens fieldmaps were acquired for each subject, usually after the second session (Cusack *et al.*, 2003; Hutton *et al.*, 2002). A structural T1 weighted scan was acquired for each participant (Deichmann *et al.*, 2004).

Imaging data were processed and analysed using Statistical Parametric Mapping software (SPM5, <http://www.fil.ion.ucl.ac.uk/spm>) (see also Sections 2.5-2.6). The first four volumes in each session were discarded to control for saturation effects. The resulting 984 volumes were realigned to the first volume and unwarped using the fieldmap parameters, spatially normalised to stereotactic space (Friston *et al.*, 1995a) and smoothed with an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum (FWHM). The standard exponential decay function (Eq. 3-1) for the six levels of

correlation yielded time windows of [0 20.00 35.13 63.01 144.25 305.51] milliseconds or [0 1 1.76 3.15 7.21 15.28] lags (as depicted on the right and left y-axis y-axis in the top inset of Figure 3-3, respectively). The corresponding contrast values at the single subject level probing for an effect of time window length over the six levels were then mean centred to yield [-2.94, -2.32, -1.85, -0.98, 1.54, 6.55]. Statistical analysis at the group level used a random-effects model within the context of the general linear model (Friston *et al.*, 1995b), and data were thresholded at  $p < 0.001$  (uncorrected for multiple comparisons across the brain) for areas with an *a priori* hypothesis, i.e. auditory cortex. Where the results survived a more conservative threshold of  $p < 0.05$  (family-wise error corrected for multiple comparisons across the brain), results are reported at this threshold.

For the test of lateralisation, two sets of images were created: both a set of ‘flipped’ left-right unwarped images as well as the original unwarped images were normalised to a symmetrical template so as to enable a direct comparison between the activations in the left and right AC. Note that the resulting symmetrical stereotactic space will differ slightly from MNI stereotactic space. These original and flipped normalised scans were smoothed with an 8 mm FWHM smoothing kernel, as above. Both original and flipped scans were then combined in one design to enable a direct comparison. Statistical analysis at the group level was thresholded at  $p < 0.001$  (uncorrected for multiple comparisons across the brain).

To compare in detail the response in subareas of auditory cortex as a function of spectrotemporal correlation, local maxima coordinates were identified based on the main contrast of spectrotemporal correlation (for PT, aSTG, and STS), and based on a [sound – silence] contrast for left and right HG that are most similar to central HG or SAC (Morosan *et al.*, 2001; Patterson *et al.*, 2002; Rademacher *et al.*, 2001). Finally, the parameter estimates of the BOLD signal were extracted at these coordinates.

### 3.3 Results

The psychometric functions obtained from the psychophysics prior to scanning are displayed in Figure 3-2. For all participants, the perceptual threshold lay between the second and third levels, corresponding to  $r = 0.2$  and  $r = 0.4$ .

An analysis was carried out to seek areas where the activity increased or decreased as a function of correlation and the associated window length (see Materials and Methods). The results show bilateral activity in AAC as a function of increasing correlation or temporal window, in particular in planum temporale (PT) and anterior superior temporal gyrus (aSTG), while also extending into right STS (Figure 3-3, see also Table 1 for coordinates of local maxima). It was formally tested whether this effect arises in and is specific to AAC in PT and aSTG and is not already present in HG (see also Figure 3-3) by extracting the BOLD signal (see Materials and Methods) in central HG, which is most similar to SAC (Morosan *et al.*, 2001; Patterson *et al.*, 2002; Rademacher *et al.*, 2001), and the association areas that showed an increase in activity as a function of correlation. Two separate (for PT and aSTG) repeated measures ANOVAs with factors 2 Hemisphere (left, right)  $\times$  2 Area (HG, [PT or aSTG])  $\times$  6 Correlation level (1-6) demonstrated an Area  $\times$  Correlation level interaction ( $F_{(5,80)} = 8.28$ ,  $p < 0.001$  for PT; and  $F_{(5,80)} = 5.19$ ,  $p < 0.01$  for aSTG).

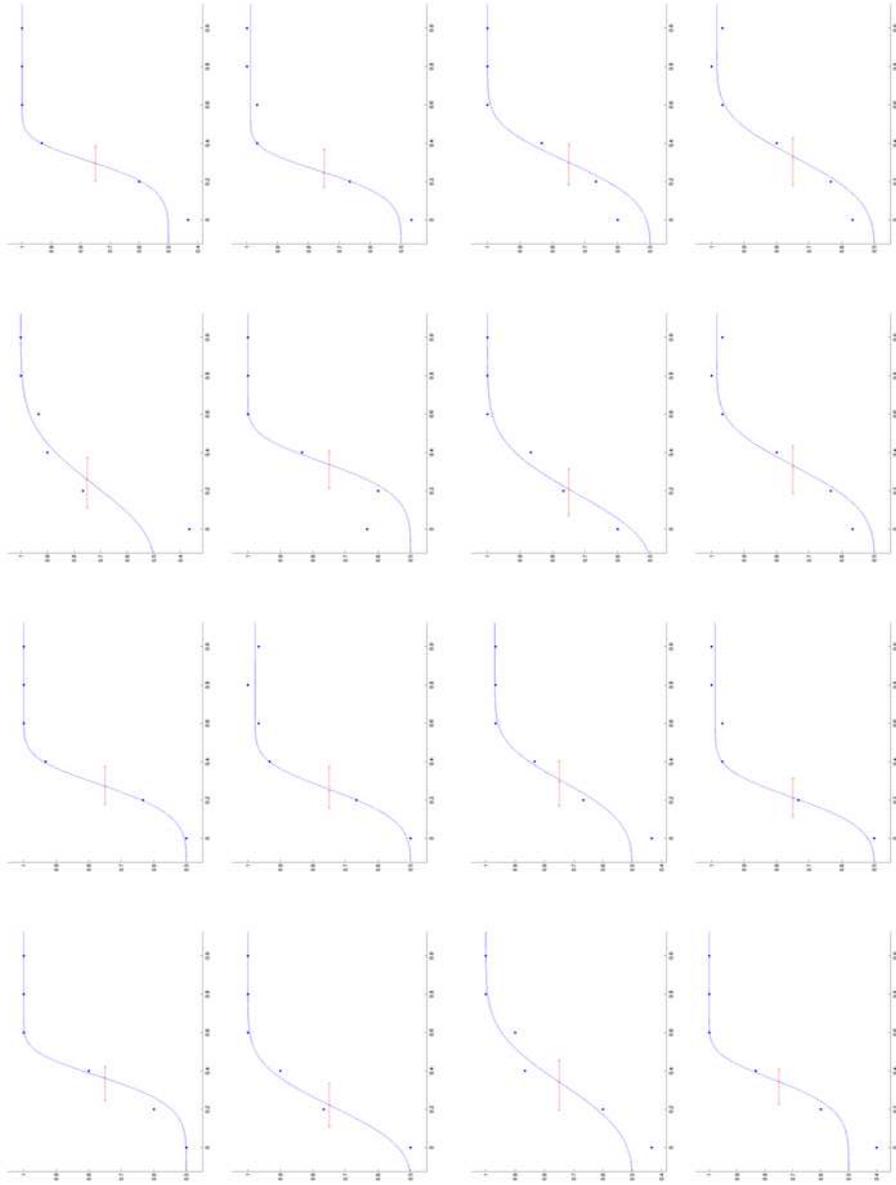


Figure 3-2 Psychometric functions from participants 1-16 (participant 17 is not displayed for lack of space). The x-axis denotes the six levels of correlation, the y-axis denotes performance. The red bar indicates the 95% confidence limits for the perceptual threshold.

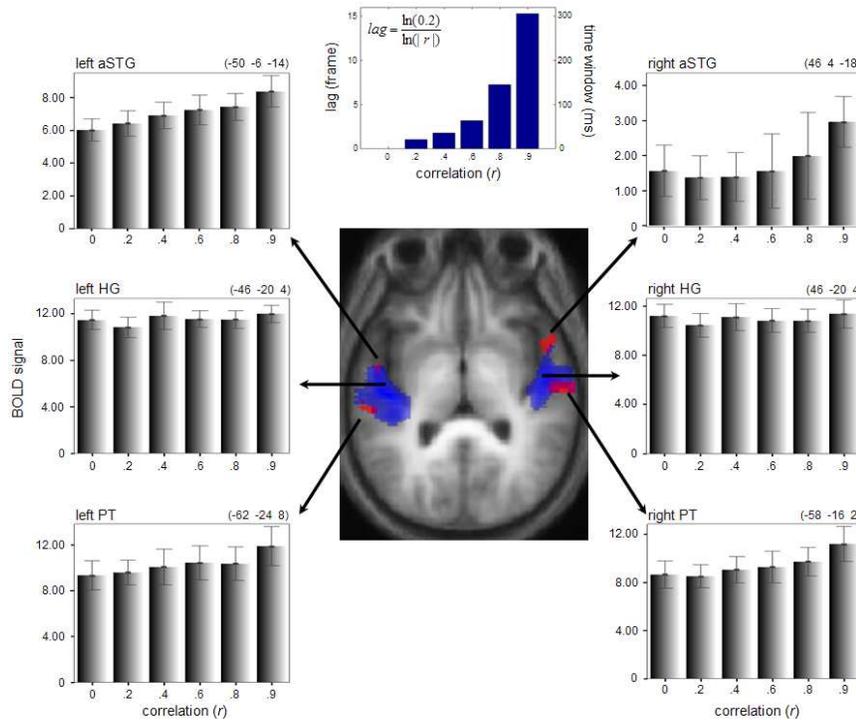


Figure 3-3 Main results. Areas increasing in activity as a function of spectrotemporal correlation (red) and areas responding to sound in general (blue). Results are rendered on a tilted (pitch =  $-0.5$  radians) section of the normalised average structural along STG and thresholded at  $p < 0.05$  (FWE corrected). The bar plots at the sides show the signal at the respective coordinates for the six levels of correlation ( $\pm 95\%$  confidence interval). The top figure displays the average lag (in 20 ms frames) and associated time window length (in ms) for which there exists a correlation  $r > 0.2$  for the six levels of correlation, as determined by the exponential decay function (inset formula).

To compare directly the response in left and right auditory cortices, a formal test of lateralisation was performed by ‘flipping’ and normalising the functional scans to a symmetrical template (see Materials and Methods). Activity in PT and aSTG did not differ between left and right hemispheres. However, right STS showed significantly stronger activation as a function of increasing correlation than its left hemisphere homologue (Figure 3-4 and Table 1). A repeated-measures ANOVA with 2 Area (left STS, right STS)  $\times$  6 Correlation level (1-6) as factors revealed a significant interaction ( $F_{(5,80)} = 2.33$ ,  $p = 0.05$ ). That is, while PT and aSTG in both hemispheres

are equally involved in processing longer spectrotemporal correlation over time, the data suggest a right-lateralised preference in STS.

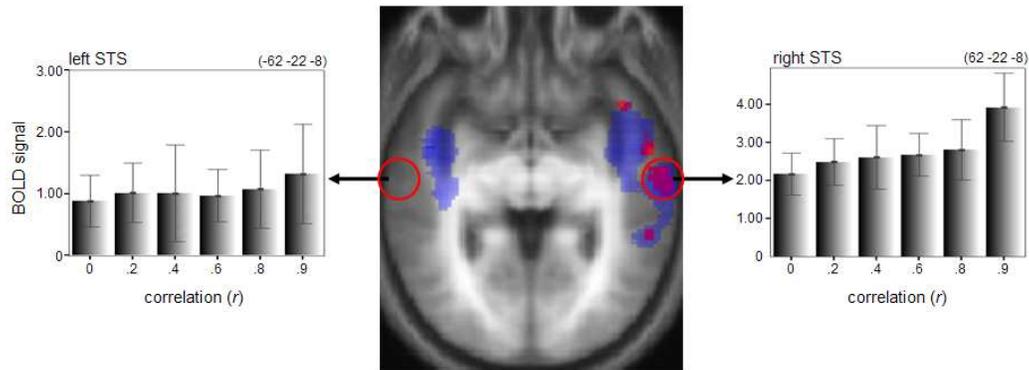


Figure 3-4 STS lateralisation. Areas showing a significantly stronger increase in activity in right STS than left STS (red) together with areas that show an increase as a function of correlation (blue). Results are rendered on a tilted (pitch = -0.5 radians) section of the symmetrical normalised average structural along STS and thresholded at  $p < 0.001$  (uncorrected). The bar plots at the sides show the signal at the respective coordinates for the six levels of correlation ( $\pm 95\%$  confidence interval).

Table 3-1 MNI coordinates of local maxima. The table displays MNI coordinates of local maxima ( $p < 0.05$ , FWE) in PT and aSTG as a function of increasing time window correlation and coordinates of local maxima ( $p < 0.001$ , uncorrected) in right STS for the lateralisation test. Note that the coordinates for STS are only approximations, since they were normalised to a symmetrical template.

Contrast	Area	left hemisphere				right hemisphere			
		x	y	z	t-value	x	y	z	t-value
window increase	PT	-62	-24	8	10.52	58	-16	2	13.49
						66	-16	0	12.83
	aSTG	-50	-6	-14	11.10	46	4	-18	9.91
		-54	-6	-2	8.71	56	10	-12	9.54
lateralisation	STS					66	-22	-8	5.10
						56	-44	6	4.56
						54	-10	-16	5.34
						48	12	-24	3.90

The relationship between the BOLD signal and correlation  $r$  appears non-linear in the right hemisphere and linear in the left (Figure 3-3). It was therefore specifically tested i) whether the relationship between BOLD and  $r$  is better fitted by a linear or exponential function in these areas, and ii) whether carrying out a linear contrast based on  $r$  rather than a contrast based on window length better fits the data. Curve-fitting algorithms were performed on areas in higher-order auditory cortex that were revealed by the analysis of the imaging data to test whether the relationship between BOLD signal and correlation  $r$  better fits a linear versus exponential function. The curve-fitting algorithms were part of the Ezyfit Toolbox for Matlab ([www.fast.u-psud.fr/ezyfit](http://www.fast.u-psud.fr/ezyfit)). The linear function was of the form:  $y(x) = ax + b$ , where  $a$  represents the slope and  $b$  the y-intercept; the exponential function was of the form  $y(x) = a^{bx} + c$ , where  $a$  is the base,  $b$  is the slope and  $c$  represents the y-intercept.

Areas in left aSTG show a better fit for a linear function (no exponential function can be fitted to the data) (Figure 3-5). Conversely, the response in right aSTG is better described by exponential functions (Figure 3-6).

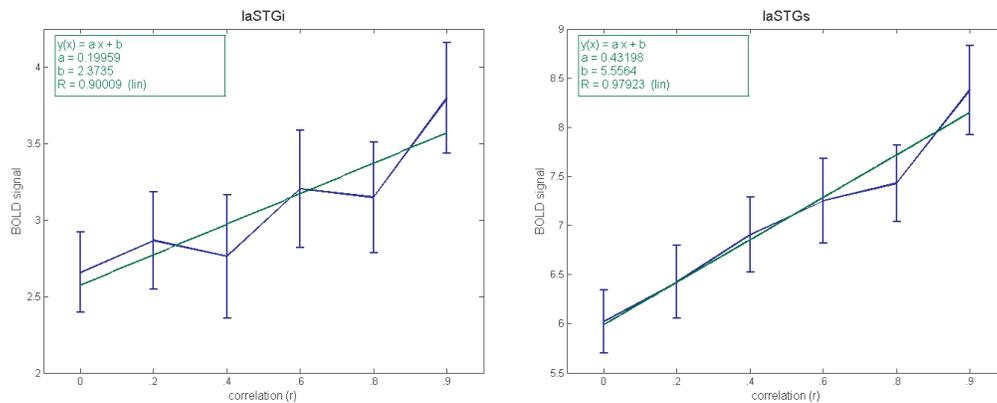


Figure 3-5 Plots of the two local maxima in left aSTG (laSTGi refers to the more inferior [-50 -6 -14], and laSTGs refers to the more superior [-54 -6 -2]). The data are plotted in blue along with the linear (green) fitted function and its parameters. An exponential function could not be fitted to these data.

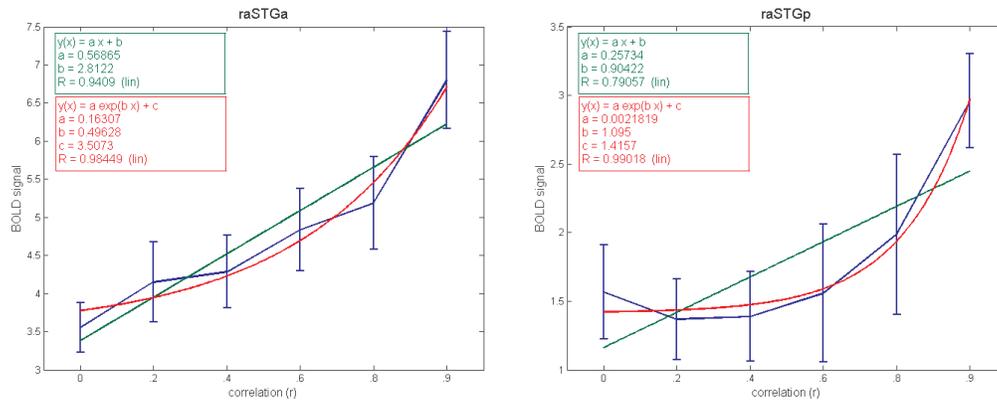


Figure 3-6 Plots of the two local maxima in right aSTG (raSTGa refers to more anterior [56 10 -12], and raSTGp refers to the more posterior [46 4 -18]). The data are plotted in blue along with the linear (green) and exponential (red) fitted functions and their parameters.

With respect to PT, left PT (IPT) shows a slightly better fit to an exponential function than linear (Figure 3-7); the more medial local maximum in right PT (rPTm) also shows a better fit to an exponential function, while no exponential function could be fitted to the more lateral local maximum in right PT (rPTl) (Figure 3-8).

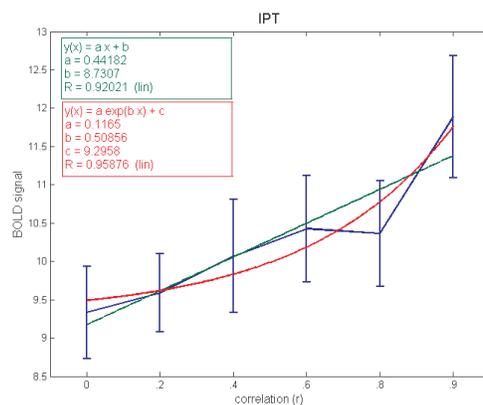


Figure 3-7 Plots of the local maximum in left PT (IPT refers to [-64 -24 8]). The data are plotted in blue along with the linear (green) and exponential (red) fitted functions and their parameters.

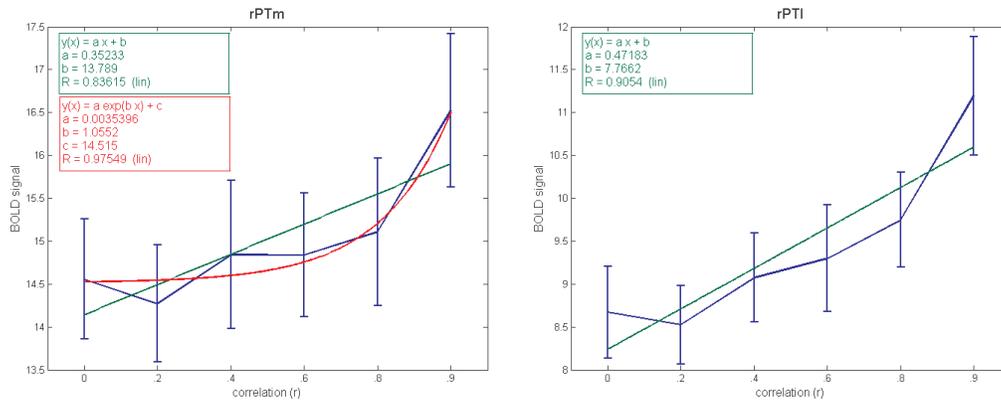


Figure 3-8 Plots of the two local maxima in right PT (rPTm refers to the more medial [58 -16 2], and rPTl refers to the more lateral [66 -16 0]). The data are plotted in blue along with the linear (green) and exponential (red) fitted functions and their parameters. An exponential function could not be fitted to these data for rPTl.

A linear contrast [-2.5 -1.5 -0.5 0.5 1.5 2.5] was performed to assess whether this form of increasing function showed a different network of areas from the (exponential) contrast based on the window length. As can be seen in Table 3-2, the coordinates in both PT and aSTG are generally very similar and in some cases identical. This is the same when using a contrast based on the individual psychometric function obtained for each participant, highlighting the fact that the statistical routines within SPM are not very sensitive to different shapes of responses across levels.

Table 3-2 MNI coordinates of local maxima (FWE,  $p < 0.05$ ) in PT and aSTG as a function of a linear increase and based on participants psychometric functions (see Table 3-1 for comparison).

Contrast	Area	left			t-value	right			t-value
		x	y	z		x	y	z	
window increase (lin)	PT	-62	-24	8	9.44	60	-16	0	13.64
	aSTG	-54	-6	-2	9.70	58	10	-12	10.00
		54	-2	-6	9.30				
window increase (psy)	PT	-62	-22	8	8.66	60	-14	0	9.79
	aSTG	-52	-6	-4	11.13	58	10	-10	8.65
		54	-2	-6	9.11				

In summary, the data show that in the right hemispheric areas the relationship between BOLD and  $r$  is generally better fitted by an exponential function and in the left areas by a linear function. Nevertheless, a very similar network of areas is demonstrated by contrasts based on either  $r$  (linear) or window length (exponential).

There was no evidence for an effect of decreasing correlation; that is, no area showed a signal increase as the time window associated with each level of correlation decreased. Even lowering the statistical threshold to a very lenient  $p = 0.1$  (uncorrected for multiple comparisons) did not yield any activation in the auditory system. Consequently, there was no detectable lateralisation as a function of decreasing correlation.

### **3.4 Discussion**

In this study, the spectrotemporal correlation in complex sounds was systematically varied, demonstrating an increase in activation in AAC as a function of spectral correlation over time (or equivalently as a function of time window length). PT and aSTG showed a bilateral increase in activity with increasing correlation and it was shown that this relationship arises in AAC and is not already present in Heschl's gyrus (i.e. in PAC or SAC). Furthermore, activity along the upper bank of right STS increased to a greater extent than left STS as a function of increasing correlation. There were no areas that showed an increase in activity as a function of decreasing correlation over time (shorter time windows).

The stimuli in the current study were based on complex spectra with multiple frequencies which varied over time in statistically controlled ways that are similar to ethological sounds including speech. In contrast, previous neurophysiological studies of temporal analysis in animal cortex have generally used more deterministic stimuli including sinusoidal amplitude modulation of narrow-band stimuli or noise (Joris *et al.*, 2004; but see Malone *et al.*, 2007). Neurophysiological studies of amplitude modulation in (mainly primary) auditory cortex in humans (Liégeois-Chauvel *et al.*,

2004) and mammals (Joris *et al.*, 2004) show preferred responses to rates of less than 20 Hz, corresponding to temporal windows at the level of tens to hundreds of milliseconds, as used in the present study. Several human imaging studies have used more complex types of temporal modulation (Boemio *et al.*, 2005; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001), but none have controlled the change in complex spectra from one moment to the next as in the present study.

The contribution of different areas of auditory cortex was explicitly tested and this revealed different response profiles across the six levels of correlation between HG on the one hand and AAC on the other. HG did not differentiate between the experimental levels, while AAC – with maxima in PT and aSTG – displayed a systematic BOLD signal increase as a function of spectral correlation. Previous models have tended to emphasise differences in temporal analysis between hemispheres, rather than differences between the specific areas of auditory cortex within hemispheres or differences between lateralisation in different areas. Human anatomical (Morosan *et al.*, 2001; Rademacher *et al.*, 2001) and functional imaging studies (e.g. Patterson *et al.*, 2002) have demonstrated one primary area and two secondary areas in Heschl’s Gyrus that might correspond to ‘core’ areas in macaque, as opposed to human homologues of belt areas of AAC in the planum temporale (PT) and superior temporal gyrus STG (Hackett, 2007). Areas of AAC in the superior temporal sulcus may correspond to parabelt in the macaque. Whilst the homology with macaque schemes is still being explored, it is clear that there is an extensive functional architecture for auditory analysis that might allow different subspecialisations for various types of temporal analysis between areas and between the hemispheres. The present study demonstrates subspecialisations for auditory analysis between different areas, and does not support any simple model based on similar temporal analysis in all the auditory cortical areas on either side.

Using BOLD as a measure of local ensemble activity, the data did not show a preference for short time windows (at the level of tens of milliseconds) in any area of auditory cortex, even when the statistical threshold was lowered substantially. A potential explanation for this might be the existence of different neural coding schemes for slow versus fast temporal modulations to which the BOLD signal might not be as sensitive. For example, Lu and colleagues (2001; see also Wang *et al.*, 2003)

have demonstrated that slow temporal modulations are encoded explicitly via synchronised discharge rates, whereas fast modulations are encoded implicitly via non-synchronised discharges. There is further evidence that neural synchronisation in the gamma frequency range is tightly coupled to the haemodynamic response in cortex (Niessing *et al.*, 2005).

However, some studies have reported signal increases as a function of increasing rates of temporal modulation. Specifically, Zatorre and Belin (2001) and Schönwiesner and colleagues (2005) demonstrated increased activity in primary and secondary auditory cortex with increasing rate of sound fluctuation (see also Jamison *et al.*, 2006). Zatorre and Belin (2001) altered the fluctuation rate of two sinusoidal components (500 and 1000 Hz) and thus arguably used a substantially different stimulus compared to the present study; however, the broadband stimuli in Schönwiesner and colleagues (2005) are similar in acoustic complexity to those used in the present study, although not controlling changes in the spectral shape from one moment to the next as here.

Thus, so far, both the hypothesised preference for short temporal windows in left auditory cortex (Poeppl, 2003; Zatorre & Belin, 2001) as well as the hypothesised preference for longer temporal windows in right auditory cortex (Poeppl, 2003) have been demonstrated; however, each of them has only been demonstrated between studies, but not within studies. In other words, studies that found evidence for a leftwards asymmetry for processing shorter temporal windows did not find evidence for a rightward asymmetry for processing longer temporal windows (Jamison *et al.*, 2006; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001); conversely, those studies that found a rightward asymmetry for processing longer temporal windows did not find evidence for a leftward asymmetry for processing shorter temporal windows (Belin *et al.*, 1998; Boemio *et al.*, 2005; as well as the current study).

Spectrotemporal trade-off (Zatorre *et al.*, 2002a) and AST (Poeppl, 2003) theories describe similar phenomena from slightly different viewpoints. Both theories posit an increased temporal resolution in left auditory cortex, while their view of the sensitivity of the right auditory cortex is complementary: according to the spectrotemporal trade-off theory, increased spectral resolution can only be achieved at

the cost of temporal resolution, i.e. longer temporal windows, which is a slightly different but convergent formulation of what the AST posits for right auditory cortex. In this context, the results for a right-lateralised sensitivity for increasing spectral resolution (Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001) and increasing temporal windows or decreasing temporal resolution (Boemio *et al.*, 2005; current study) seem to converge. However, this does not explain why one set of studies does not find evidence for a right-lateralised sensitivity for increasing temporal windows (Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001), while the other set of studies does not find evidence for a left-lateralised sensitivity of increasing temporal windows (Boemio *et al.*, 2005; current study), despite their use of a similar experimental manipulation (sound segment length) and similar temporal window lengths to test their complementary hypotheses.

A promising recent approach (Giraud *et al.*, 2007) combined fMRI and EEG recordings of spontaneous spectral power (in the absence of any experimental acoustic stimulation, but in the presence of scanner noise) to test the AST hypothesis and found activity in left and right HG (but not AAC) that correlated with fast (~28-40 Hz) and slow (~3-6 Hz) neural oscillations, respectively. While these findings somewhat contradict the precise anatomical locations of the earlier findings of Boemio and colleagues (2005) for longer temporal windows, they might nevertheless offer a bridge in that they show a left-lateralisation for fast temporal modulations (as posited by both AST and spectrotemporal trade-off theories) and a right-lateralisation for slower temporal modulations (as posited by AST). However, comparisons between studies of spontaneous activity in the absence of experimental acoustic input and the temporal structure of stimuli producing the greatest regional activity need to be made with caution. A convincing explanation for the divergence of results between the previous studies despite similar experimental manipulations has yet to be provided (Hickok & Poeppel, 2007; Zatorre & Gandour, 2008).

The psychometric functions show a clear perceptual threshold that is situated between levels 2 and 3 for most participants. This is in contrast to the haemodynamic response to increasing temporal windows in AAC, which shows a more exponential function, especially in right AAC (see Figures 3-5 to 3-8). However, as demonstrated by probing the imaging data with different response functions across the six levels of

correlation (Table 3-2), the statistics employed within SPM are rather insensitive to different response functions. Further investigations are required to elucidate whether, and to what degree, left and right AAC indeed show different response functions (e.g. linear vs. exponential). For example, introducing a stimulus-relevant task might yield response functions that are more similar to the behavioural data; in the current Study 1, the psychophysics prior to scanning employed an explicit threshold detection task, while in the functional imaging paradigm participants were not asked to evaluate the spectrotemporal statistics of the stimuli. Although occurring in ‘higher’ association cortex, the main effect of increasing window length is present even though participants in the studies (Boemio *et al.*, 2005; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001; Study 1) performed no task or a stimulus-irrelevant task, and can thus be argued to be an obligatory correlate of perception. Introducing a stimulus-relevant task might also reveal the engagement of prefrontal areas. For example, Johnsrude and colleagues (1997) found no effect in auditory cortex for differential temporal rates in either hemisphere with an explicit stimulus-relevant task; however, left prefrontal cortex showed a short temporal window preference (see also Temple *et al.*, 2000). In the current Study 1, no areas in prefrontal cortex showed a parametric preference for shorter time windows, even at very lenient statistical thresholds.

The present study has demonstrated the analysis of longer temporal windows in the syllabic range that is bilateral in AAC in STG and right lateralised in STS. In Boemio and colleagues (2005), the analysis of longer time segments was similarly right lateralised and involved STS. An open question remains as to why longer temporal windows, which are important for syllabic information and speech intelligibility (Greenberg *et al.*, 2003; Luo & Poeppel, 2007) and which have been shown to engage a left-lateralised network (Narain *et al.*, 2003; Scott *et al.*, 2000), should be lateralised towards right AC, as posited by the AST hypothesis (Poeppel, 2003). In contrast, the present data, as well as those of Boemio and colleagues (2005), revealed a significant right-lateralisation in STS for increasing time windows (see also Belin *et al.*, 1998). It should be pointed out that this level of temporal analysis is relevant to a variety of sounds including speech, but is not specific to speech.

The present study highlights the power of parametrically varying statistical properties of complex acoustic stimuli to investigate systematically principles of

processing in auditory cortex (Nelken & Chechik, 2007; Overath *et al.*, 2007). This study introduces a novel stimulus with statistical stimulus characteristics that vary in a similar fashion to naturally occurring sounds including speech and demonstrates a network comprising auditory association cortex that plays a crucial role in tracking spectral correlation over different time scales.

## **Chapter 4. ENCODING OF SPECTROTEMPORAL COHERENCE IN ‘AUDITORY TEXTURES’**

### **Summary**

In a complex and dynamic acoustic environment, we must constantly identify and segregate the many often rapidly changing sound elements or auditory objects that constitute the auditory scene. This auditory object analysis requires two fundamental perceptual mechanisms. Firstly, it must define boundaries between two adjacent objects. Secondly, it requires abstraction processes that allow defining features of an object to be recognised, irrespective of local stochastic variation. Study 2 considered the cortical bases for these two processes by creating a novel stimulus (an ‘auditory texture’) in which auditory objects are defined by their spectrotemporal coherence. Auditory objects were identified by the percentage of randomly distributed frequency-modulated (FM) ramps in frequency-time space that had identical direction and trajectory (spectrotemporal coherence), while boundaries were introduced by juxtaposing auditory objects of different spectrotemporal coherence levels. Using fMRI, Study 2 sought areas that signal the detection of a boundary between auditory objects of different coherence levels (change in spectrotemporal coherence) from areas that encode the salience of the object (absolute coherence). The data show that mechanisms defining object boundaries (changes in coherence) are represented in primary and association auditory cortex. In contrast, the representation of the salience of the object (percentage of coherence) occurs only in auditory association cortex. Furthermore, participants’ superior detection of boundaries across which coherence increased was reflected in a greater neural response at these boundaries. The anatomical organisation revealed by these results suggests a hierarchical mechanism for the analysis of auditory objects: boundaries between objects are first detected as a change in statistical coherence over frequency-time space, before a representation that corresponds to the salience of the perceived object is formed.

## 4.1 Introduction

The analysis of auditory objects requires two fundamental perceptual processes (Griffiths & Warren, 2004). The boundaries between two adjacent objects must be defined (Chait *et al.*, 2007; Chait *et al.*, 2008; Kubovy & Van Valkenburg, 2001), whilst characteristic features of an object must be abstracted, irrespective of stochastic variations (Griffiths & Warren, 2004; Nelken, 2004). Such object boundaries often do not ‘exist’ as low-level physical sound features at one particular point, especially in the presence of acoustic noise (Gutschalk *et al.*, 2008; Nahum *et al.*, 2008); rather, detection of boundaries requires a mechanism that identifies a change in the statistical rules governing areas of frequency-time space corresponding to different objects. Further, auditory object recognition requires abstraction processes that allow characteristic features of objects in frequency-time space to be recognised, while ignoring local stochastic variation within one object region (Griffiths & Warren, 2004; Nelken, 2004).

While it is intuitive to assume that the detection of a statistical change at object boundaries precedes the subsequent precise representation of the object (Chait *et al.*, 2007; Chait *et al.*, 2008; Ohl *et al.*, 2001; Scholte *et al.*, 2008), the specific underlying cortical mechanisms for segregating and encoding auditory objects within the auditory scene have not been addressed directly. For example, Zatorre and colleagues (2004) parametrically varied the distinctiveness or identity of auditory object features by combining auditory objects to create a new object (distinctiveness decreased with the number of auditory objects that were combined). The authors presented several sounds with a fixed level of distinctiveness within one 60 s trial (each sound was 500 ms in duration) and demonstrated activity within right STS and right inferior frontal gyrus (IFG) that increased with object distinctness. These results support an anterior processing stream for object identity or auditory ‘what’ information (Kaas & Hackett, 1999; Rauschecker & Tian, 2000; Romanski *et al.*, 1999; Tian *et al.*, 2001) (however, see Belin & Zatorre, 2000; Middlebrooks, 2002; Zatorre *et al.*, 2002b).

However, strictly speaking, this design cannot differentiate whether parametric increase in activation in STS and IFG was due to object distinctness or due to a change percept between objects. That is, as the distinctness between objects increased, participants would also increasingly hear an object change at object boundaries.

Furthermore, the auditory objects used were ethological sounds and are likely confounded by semantic associations instead of highlighting which acoustic features were perceptually relevant (or which acoustic features were important for object distinctness). Other experimental approaches have focussed on the neural correlates of boundary or ‘auditory edge’ detection without investigating in detail processes necessary for object formation (Chait *et al.*, 2007; Chait *et al.*, 2008). One notable study (Schönwiesner *et al.*, 2007) investigated the perception of different levels of acoustic duration changes in the context of an MMN paradigm. The authors found a cortical hierarchy as indicated by three distinct stages for processing duration changes: an initial automatic change detection mechanism in primary auditory cortex, followed by a more detailed analysis in association cortex and attentional mechanisms originating in frontal cortex.

The present study used a form of FM to create object regions and object boundaries in frequency-time space. The stimulus, an ‘auditory texture’, was based on randomly distributed linear FM ramps with varying trajectories (Figure 4-1, see also Materials and Methods). The percentage of coherent modulation, i.e. the proportion of ramps with identical direction (slope-sign) and trajectory (slope-value), was systematically varied, creating different auditory objects, the salience of which increases with coherence. The analysis of such auditory objects comprising different spectrotemporal coherence requires perceptual mechanisms that can assess common statistical properties of the stimulus irrespective of local stochastic variation within an object, and detect transitions when these properties change. Such generic mechanisms are fundamental for auditory object formation and object segregation in ethological sounds, where statistical properties of the acoustic signal need to be evaluated with respect to pre-existing templates (Griffiths & Warren, 2002). It should be noted that while this manipulation is merely one way to define auditory objects and is not intended to speak for all possible auditory objects, it nevertheless addresses generic processes underlying complex auditory object perception. While coherent FM is arguably a relatively weak grouping cue (Carlyon, 1991; Darwin & Carlyon, 1995; Summerfield & Culling, 1992), coherent FM nevertheless is one basis upon which figure-ground selection can occur (McAdams, 1989).

In Figure 4-1, a 3.5-s segment with 100% coherence (all ramps move upwards and with the same trajectory) is followed by a 4.5-s segment of 0% coherence (ramps with random direction and trajectories) and so forth. The associated change in coherence at the boundary between segments is also shown. Note that a +40% change in coherence can be obtained in a number of ways by arranging successive pairs of stimuli with certain absolute coherence levels (0%-40% and 40%-80%, in Figure 4-1). Thus, this stimulus enables a direct assessment of (i) the mechanisms detecting boundaries between auditory objects, represented by the change in coherence between sound segments, and (ii) the representation of the salience of complex auditory objects, determined by the absolute coherence of a sound segment.

These two factors were orthogonalised in the experimental design so as to dissociate neural processes signalling object boundaries from those representing absolute object properties (see Section 4.2). Within the framework of auditory object analysis, it was hypothesised that the detection of a change in coherence would engage auditory areas including primary cortex (Schönwiesner *et al.*, 2007), while auditory object salience would be encoded in higher-level auditory areas only (Zatorre *et al.*, 2004).

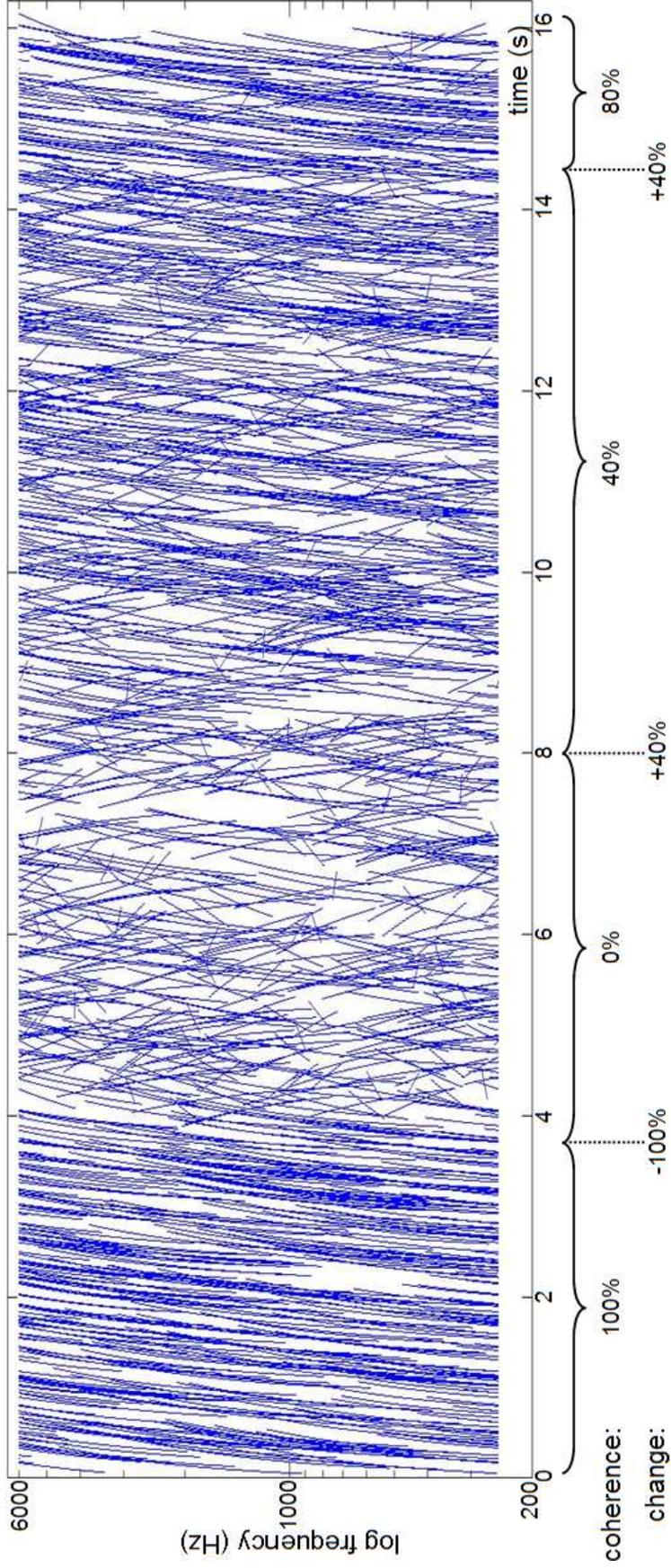


Figure 4-1 Auditory stimulus. Example of a block of sound with four spectrotemporal coherence segments showing absolute coherence values for each segment and the corresponding change in coherence between the segments.

## 4.2 Materials and Methods

### 4.2.1 Participants

23 right-handed participants (aged 18-31, mean age = 25.04, 12 females) with normal hearing and no history of audiological or neurological disorders provided written consent prior to the study. The study was approved by the National NHS Research Ethics Committee.

### 4.2.2 Stimuli

All stimuli were created digitally in the frequency domain using Matlab (<http://www.mathworks.com>) at a sampling frequency of 44.1 kHz and 16 bit resolution. Stimuli consisted of a dense texture of linear FM ramps; each ramp had a duration of 300 ms and started at a random time and frequency (passband 250-6000 Hz), with a density of 80 glides per second, roughly equalling one ramp per critical band (see Figure 4-2). For ramps that extended beyond the passband, i.e. went below 250 Hz or beyond 6000 Hz, a wraparound was implemented such that the ramps ‘continued’ at the other extreme of the frequency band, i.e. at 6000 Hz or 250 Hz, respectively. Stimuli differed in terms of the coherent movement of the ramps: six different coherence conditions were created, where the percentage of ramps moving in the same direction for a given sound segment was systematically varied from 0% coherence to 100% coherence in 20% increments. Thus, for a given sound segment with 40% coherence, 40% of the ramps increased (or decreased) in frequency with an excursion traversing 2.5 octaves / 300 ms; the direction and excursion of the remaining 60% of the FM ramps were randomised. Crucially, the only difference between the six levels is the degree of coherence or ‘common fate’ of the ramps; the total number of ramps, the number of ramps in a critical band as well as the passband of each stimulus did not differ systematically across the levels (Figure 4-2).

### 4.2.3 Experimental design

Prior to scanning, participants were familiarised with and trained on the stimuli and then performed 2I2AFC psychophysics distinguishing the non-random against a

random reference (0% coherence) sound. Stimuli were two seconds long and the direction of the FM glides (up versus down) was counterbalanced. There were 30 pairs for each of the six levels (0%-100% coherence in 20% steps). Participants had to reach at least 90% correct performance for the last level (100% coherence) to be included in the fMRI study. Psychometric functions and 95% correct perceptual thresholds were estimated via a Weibull bootstrapping procedure (Wichmann & Hill, 2001).

Stimuli in the scanner were presented in blocks of sound with an average duration of 16 seconds (range: 11 to 18 seconds). The blocks contained four contiguous segments with a given absolute spectrotemporal coherence (0%, 20%, 40%, 60%, 80%, or 100%). Within a block, the direction (up versus down) of the coherent ramps was maintained. The length of the segments varied (1.5, 3, 3.5, 4.5, 5, or 6.5 seconds) and was randomised within a block. Thus, a given block might have [20% 100% 60% 60%] contiguous coherence segments with durations [1.5 6.5 3.5 4.5] seconds. The associated change in coherence between segments within this block of sound is [+80 -40 0], between segments two through four of the block. Stimuli were presented in one of six pseudorandom permutations which orthogonalised absolute coherence and change in coherence (average correlation between absolute coherence and change in coherence across the six permutations:  $r = 0.06$ ,  $p > 0.1$ ).

The task of participants was to detect a change in coherence within the block, regardless of whether that change was from less coherent to more coherent or vice versa. Participants were required to press a button whenever they heard such a change and were instructed that the frequency of perceptual changes within one block likely ranged from no perceptual change (e.g. a block consisting of [0% 20% 40% 20%] coherence segments, since here the changes are likely to be too small to be detected) to a few changes (e.g. a block consisting of [0% 100% 20% 80%] segments). Sound blocks were separated by a silence of 6 seconds, in which participants were told to relax.

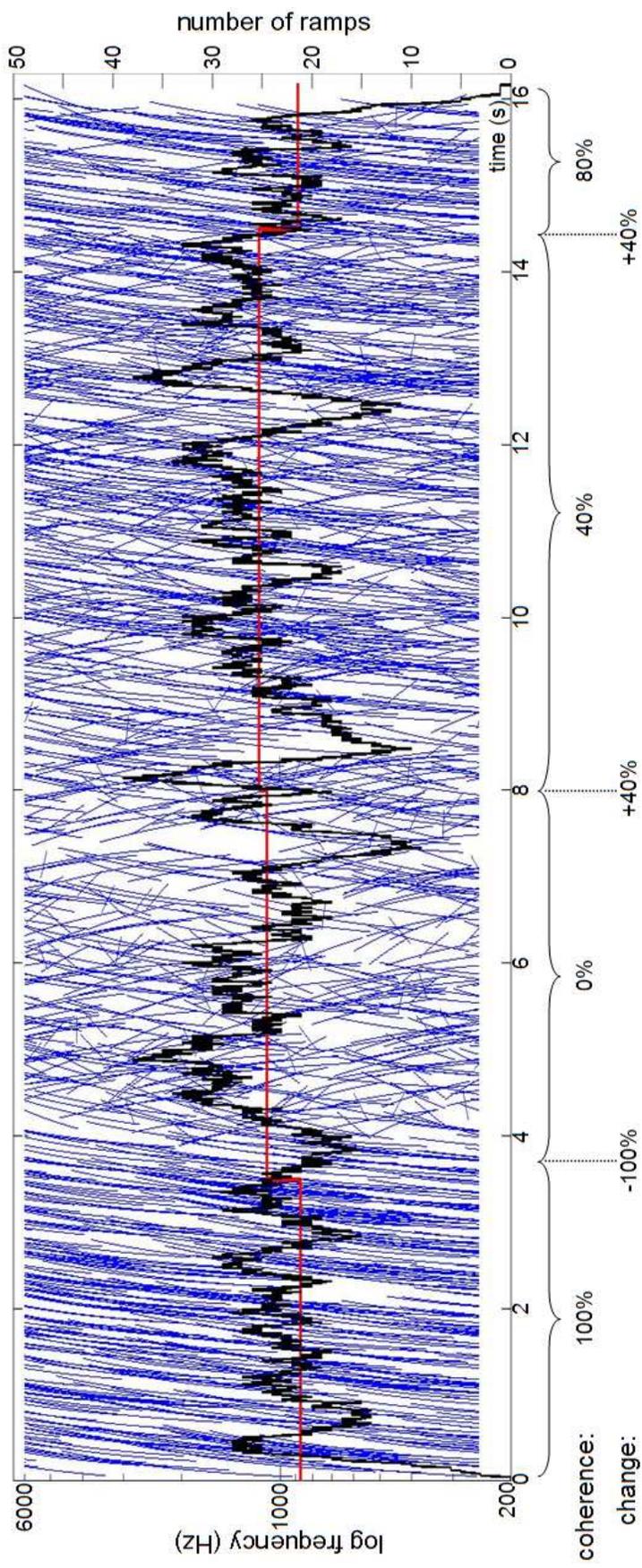


Figure 4-2 Ramp density. The figure displays the same stimulus as in Figure 4-1, together with the number of ramps at every 100th sampling point (black; sampling rate: 44100 Hz) and the mean number of ramps for each coherence segment (red).

In each of three experimental sessions, each coherence level was presented 30 times, amounting to a total of 7.2 minutes presentation time per coherence level. The number of times each of the six different levels of change in coherence (regardless of their direction) occurred can, consequently, not be perfectly balanced; however, permutations were created such that the change that occurred most often occurred less than three times as often as the change that occurred least frequently.

Stimuli were presented via electrostatic headphones (NordicNeuroLab, <http://www.nordicneurolab.com>) at a sound pressure level of 85 dB. Participants saw a cross at the centre of the screen and were asked to look at this cross during the experiment.

#### 4.2.4 Behavioural data analysis

Participants' button presses were recorded and analysed with respect to the onset of each segment within a sound block. Responses were only counted if they occurred within three seconds after the onset of a segment (and within 1.5 seconds after the onset of the shortest segments). The average percentage correct response was then computed by comparing the number of responses to a given change in spectrotemporal coherence to the actual number of those changes. 'Responses' to 0% changes served as a chance baseline.

#### 4.2.5 fMRI protocol and analysis

Gradient weighted echo planar images (EPI) (see Section 2.3.1) were acquired on a 3 Tesla Siemens Allegra system (Erlangen, Germany), using a continuous imaging design with 42 contiguous slices per volume (time to repeat/time to echo, 2730/30 ms). A continuous instead of a sparse imaging protocol was used, since the experimental question required a design whose effects could only be captured with the superior temporal resolution of a continuous imaging protocol. The volume was tilted forward such that slices were parallel to and centred on the superior temporal gyrus. Participants completed three sessions of 372 volumes each, resulting in a total of 1116 volumes. To correct for geometric distortions in the EPI due to B0 field variations,

Siemens fieldmaps were acquired for each subject, usually after the second session (Cusack *et al.*, 2003; Hutton *et al.*, 2002). A structural T1 weighted scan was acquired for each participant (Deichmann *et al.*, 2004).

Imaging data were processed and analysed using Statistical Parametric Mapping software (SPM5, <http://www.fil.ion.ucl.ac.uk/spm>) (see also Sections 2.5-2.6). The first four volumes in each session were discarded to control for saturation effects. The resulting 1104 volumes were realigned to the first volume and unwarped using the fieldmap parameters, spatially normalised to stereotactic space (Friston *et al.*, 1995a) and smoothed with an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum (FWHM). Statistical analysis used a random-effects model within the context of the general linear model (Friston *et al.*, 1995b), and data were thresholded at  $p < 0.001$  for areas with an *a priori* hypothesis, i.e. auditory cortex.

Each design matrix consisted of 18 regressors. All regressors collapsed across the direction of the coherent ramps, i.e. 100% coherent segments in which the ramps moved up were collapsed with 100% coherent segments in which the ramps moved down. The first regressor modelled the haemodynamic response to the onset of each block as a stick function. Regressors 2-7 modelled the onset and duration of the segments within a block corresponding to one of the six levels of spectrotemporal coherence (0%, 20%, 40%, 60%, 80%, 100%). Regressors 8-18 modelled the response to changes in coherence as stick functions, with the eighth regressor modelling 0% changes (i.e. all consecutive coherence pairs of 0-0, 20-20, 40-40, 60-60, 80-80, 100-100), while the subsequent pairs of regressors modelled positive and negative changes of a given magnitude (+20%, -20%, +40%, -40%, +60%, -60%, +80%, -80%, +100%, -100%).

The following planned contrasts were performed. To probe for an effect of increase in activity with increasing absolute coherence, regressors 2-7 were weighted

$$[0 \ -2.5 \ -1.5 \ -0.5 \ 0.5 \ 1.5 \ 2.5 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0].$$

To probe for an effect of increasing change in coherence, regressors 8-18 were weighted

$$[0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -2.73 \ -1.73 \ -1.73 \ -0.73 \ -0.73 \ 0.27 \ 0.27 \ 1.27 \ 1.27 \ 2.27 \ 2.27].$$

These values are all mean centred on zero.

To probe for an effect of relative object salience, changes across which coherence increased ('positive' changes or regressors 9, 11, 13, 15, 17) were exclusively masked with changes across which coherence decreased ('negative' changes or regressors 10, 12, 14, 16, 18).

### 4.3 Results

The psychometric functions obtained from the psychophysics prior to scanning are displayed in Figure 4-3. For the majority of participants, the perceptual threshold lay between the second and fourth levels, corresponding to 20% and 60% coherence.

Behavioural results ( $d'$  scores) for detecting a change in coherence during scanning are shown in Figure 4-4. Performance increased with the magnitude of change (both for changes across which coherence increased or decreased) and was significantly better than chance performance corresponding to 0% change: two separate repeated-measures ANOVAs with factor ChangeLevel (0% - 100%) for either changes across which coherence increased or decreased revealed main effects of ChangeLevel(increase),  $F_{(5,110)} = 58.0$ ,  $p < 0.001$ , and ChangeLevel(decrease),  $F_{(5,110)} = 23.04$ ,  $p < 0.001$ . Pairwise comparisons (two-tailed t-tests) with 0% performance were all significant ( $p < 0.05$ ) for change levels greater than 0% (increase) and 40% (decrease). Furthermore, performance was better for changes across which coherence increased: a repeated-measures ANOVA with factors ChangeLevel (0% - 100%) and ChangeType (increase vs. decrease) revealed main effects for ChangeLevel ( $F_{(5,110)} = 52.05$ ,  $p < 0.001$ ) and ChangeType ( $F_{(1,22)} = 52.32$ ,  $p < 0.001$ ), as well as a significant interaction ( $F_{(5,110)} = 7.87$ ,  $p < 0.001$ ).

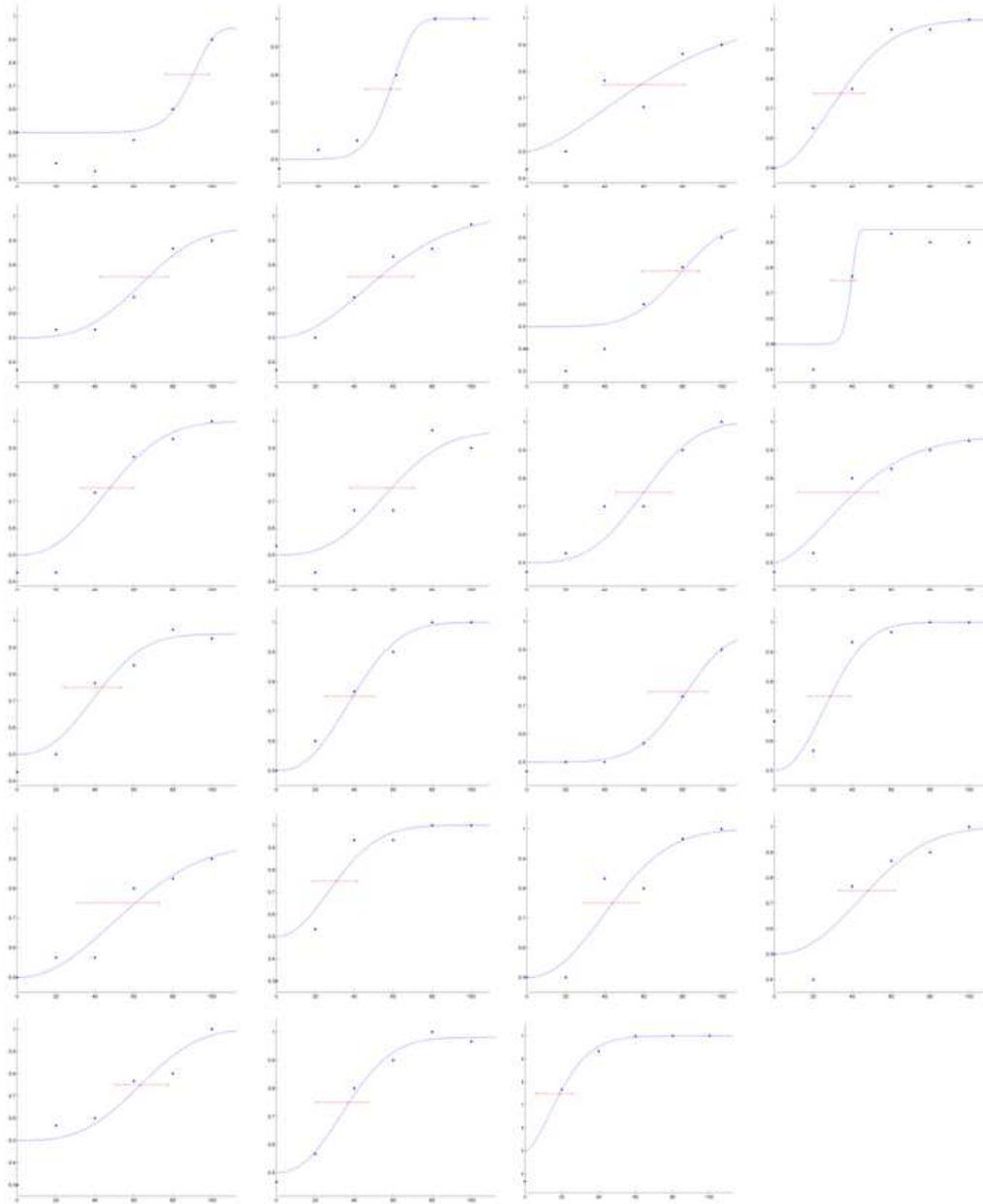


Figure 4-3 Psychometric functions from all participants. The x-axis denotes the six levels of coherence, the y-axis denotes performance. The red bar indicates the 95% confidence limits for the perceptual threshold.

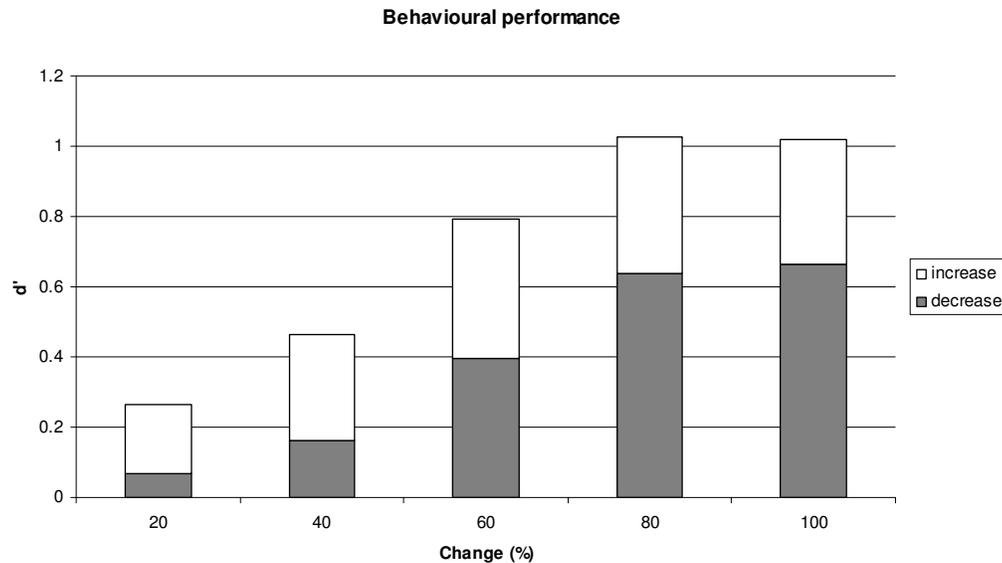


Figure 4-4  $d'$  scores for detecting changes in coherence in the MRI scanner. White bars indicate performance for changes with increasing coherence, and grey bars indicate performance for changes with decreasing coherence.

An analysis was carried out to seek areas in auditory cortex that parametrically varied in activity as a function of increasing change in coherence at the boundaries between adjacent segments. The analysis revealed strong activation increases in HG, PT, TPJ and superior temporal sulcus (STS) as a function of change magnitude (see Table 4-1). The bar charts in Figure 4-5 show (in red) the BOLD signal across the different degrees of change in coherence in all of these areas of auditory cortex.

Next, an analysis was carried out that sought activity within areas of auditory cortex that varied as a function of increasing auditory object salience or spectrotemporal coherence. Bilateral areas in auditory association cortex, including PT and extending into TPJ (Figure 4-5, in blue, and Table 4-1), showed a BOLD signal increase with increasing absolute spectrotemporal coherence. Crucially, activity in HG and STS did not differ across the six levels of spectrotemporal coherence and thus was significantly different from the responses in these areas to increasing change in coherence: two separate 2 Hemisphere (left, right)  $\times$  2 Condition (absolute coherence, change in coherence)  $\times$  6 Level (1-6) repeated-measures ANOVAs

revealed Condition  $\times$  Level interactions for both HG ( $F_{(5,110)} = 3.63, p < 0.01$ ) and STS ( $F_{(5,110)} = 3.98, p < 0.01$ ).

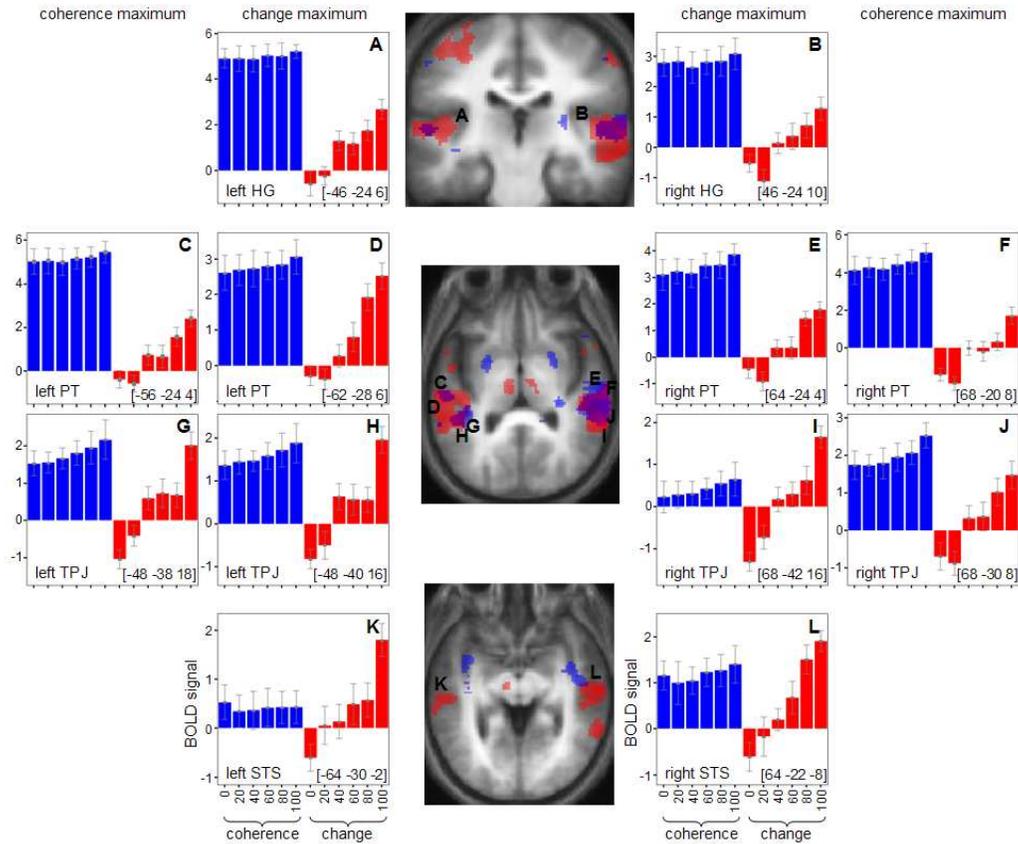


Figure 4-5 Main results. Areas showing an increased haemodynamic response as a function of increasing absolute coherence (blue) and change in coherence (red). Results are rendered on coronal ( $y = -24$ , top) and tilted (pitch =  $-0.5$  radians, middle (superior temporal plane) and bottom (STS)) sections of participants' normalised average structural scans. The bar charts show the BOLD signal ( $\pm$  SEM) corresponding to the six levels of absolute coherence (blue) and the six levels of change in coherence (red). The charts nearest the brain show the response at the local maxima for increasing change in coherence; those at the sides show the local maxima for increasing absolute coherence. Note that the placement of the identifying letter in the brain sections only approximate the precise stereotactic  $[x\ y\ z]$  coordinates at the bottom corner of each chart, since no single planar section can contain all the local maxima simultaneously.

Table 4-1 MNI coordinates of local maxima. Local maxima for the effects of increasing change in coherence and increasing absolute coherence.

Contrast	left hemisphere				right hemisphere				
		x	y	z	t-value	x	y	z	t-value
change in coherence	HG	-46	-24	6	4.86	46	-24	10	3.80
	PT	-62	-28	6	5.45	64	-24	4	6.94
	TPJ	-48	-40	16	6.34	68	-42	16	7.63
	STS	-64	-30	-2	4.55	64	-22	-8	6.16
absolute coherence	PT	-56	-24	4	4.58	68	-20	8	6.71
	TPJ	-48	-38	18	4.91	64	-30	8	7.39

The experimental design also enabled a more detailed investigation of an effect of object salience by way of changes in relative coherence. Behavioural results in Figure 4-4 showed that changes across which coherence increased are generally more salient than changes across which coherence decreased, supporting the notion of increasing object salience with increasing spectrotemporal coherence. It was tested whether this perceptual asymmetry (Cusack & Carlyon, 2003) was also reflected at the neural level (see Section 4.2). Figure 4-6 shows this was the case in PT and STS, which showed stronger responses to changes with increasing relative coherence than vice versa.

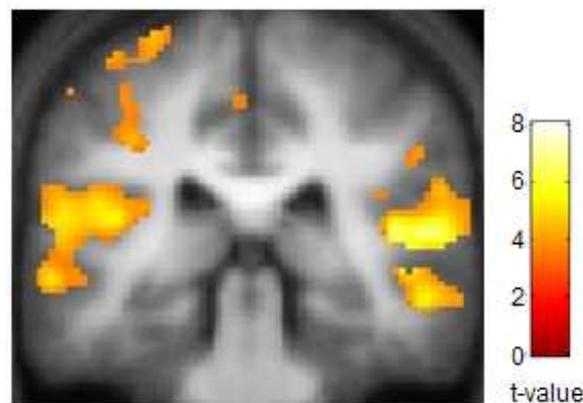


Figure 4-6 Changes in increasing vs. decreasing relative coherence. Coronal ( $y = -32$ ) section showing areas that display a stronger increase for changes across which coherence increased than for changes across which coherence decreased.

### 4.3.1 Pilot study results

It could be argued that the results are confounded by the behavioural task, which required participants to detect changes in coherence. At this point it is helpful to point out a pilot study in which participants were required to detect the overall coherence of the stimuli, but which nevertheless yielded very similar results to the main study. In this pilot study, 6 s long sounds of set coherence (0%, 20%, 40%, 60%, 80%, or 100%) were presented, in which the direction of the coherent ramps changed every 1.5 seconds (i.e. either up-down-up-down or vice versa). In every other respect, for example bandwidth or number of ramps per second, the stimulus of the pilot study was identical to the main study. In a sparse imaging protocol (TR = 8.8 s, Belin *et al.*, 1999; Hall *et al.*, 1999), four participants categorically evaluated the coherence of the stimuli by indicating whether the stimulus had been 'random' or 'coherent' (participants pressed on of two buttons during the acquisition of a scan volume following the presentation of the sound). Importantly, participants in this pilot study evaluated the overall spectrotemporal coherence of the sound and not a change in coherence as in the paradigm of the main study. As can be seen when comparing the results of the pilot study (Figure 4-7) with those of the main study (Figure 4-5 and Figure 4-6), the results in both studies are very similar despite the different tasks for participants, making it unlikely that the results reported in the main text are due to a task confound.

However, the stimulus design in the pilot study, which had a change in direction (up/down) every 1.5 seconds, confounded absolute coherence and change percept, since increasing absolute coherence was accompanied by an increasing change percept at the boundaries of the 1.5 segments making up the 6 s long sound. That is, for 6 s long sounds with alternating up/down 100% coherent ramps, the encoding of the absolute coherence would be accompanied by a perceived change in direction every 1.5 seconds; in contrast, for a 20% coherent sound, the change in direction of 20% of the ramps every 1.5 seconds would not be noticeable (and completely absent for 0% coherent sounds). For this reason, the main study employed an experimental design that allowed the disambiguation of processing absolute coherence versus change in coherence.

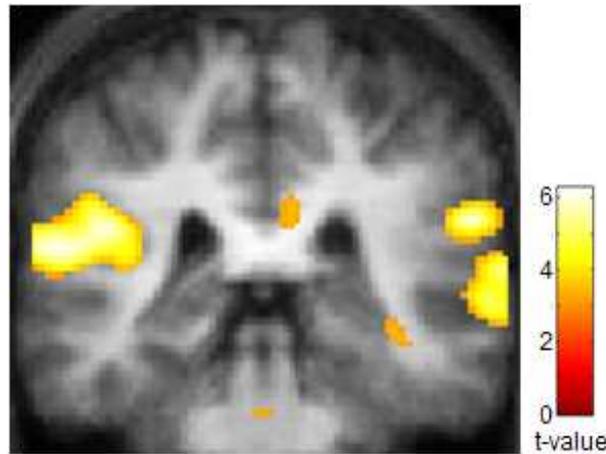


Figure 4-7 Pilot study results. The figure shows areas in PT, TPJ and STS that increased with increasing coherence of the sound (or increasing change in direction, since these are not separable effects in this study). Results are superimposed on participants' mean structural scan ( $y = -36$ ) and are based on a fixed-effects analysis within the context of the general linear model (thresholded at  $p < 0.001$ , uncorrected for multiple comparisons across the brain).

#### 4.4 Discussion

The results demonstrate a specific mapping of object boundaries and object salience to distinct regions of auditory cortex. Activity in auditory cortex including HG, PT, TPJ and STS increased as a function of the change in spectrotemporal coherence at the object boundaries. Further, activity as a function of the absolute spectrotemporal coherence and object salience increased in auditory association cortex in PT and in TPJ. Finally, increases in spectrotemporal coherence at segment boundaries were more perceptually salient than decreases in spectrotemporal coherence at segment boundaries, and this was reflected by stronger neural activity for these changes.

While the observed parametric responses to absolute coherence and change in coherence show some overlap in cortical resources (in PT and TPJ), they are likely separable processes, since the experimental design orthogonalised absolute coherence and change in coherence. This indicates that the overlapping representations of

change in coherence and absolute coherence in the non-primary auditory areas in PT and TPJ represent a distinct mapping of these two processes in similar cortical areas; it is hypothesised that these mappings are subserved by activity within distinct units or networks in those areas. Furthermore, the results are unlikely to be confounded by the behavioural change detection task, since a pilot study with a different task that asked participants to detect the absolute coherence in the sounds yielded very similar results (Figure 4-7). Nevertheless, the perceptual task was quite demanding, as indicated by the psychophysical thresholds (Figure 4-3) and the relatively low  $d'$  scores, particularly for changes across which coherence decreased (Figure 4-4). This can be attributed to the relative weakness of FM as grouping cues (Carlyon, 1991; Darwin & Carlyon, 1995; Summerfield & Culling, 1992), and will need to be addressed further in subsequent studies.

The response to increasing change in spectrotemporal coherence requires mechanisms that integrate statistical features across spectrotemporal regions and assess statistical changes across the integrated whole. That is, boundary detection must depend on the assessment of statistical properties (i.e. the percentage of coherent FM ramps), since other low-level acoustic features such as the density of FM ramps and overall frequency-time space were kept constant (Figure 4-2). This response occurs as early as primary auditory cortex, in which FM direction sensitive neurons have been demonstrated in rats (Ricketts *et al.*, 1998), cats (Heil *et al.*, 1992; Mendelson & Cynader, 1985), and rhesus monkeys (Tian & Rauschecker, 2004) (for a review see Rees & Malmierca, 2005). Single-unit studies of coherent FM have generally not investigated coherent FM across different spectrotemporal regions, but it is hypothesised that this property is encoded at the level of neural ensembles rather than the single neuron level, since the tuning of most single units would be too narrow to encompass the spectral range of the stimulus.

The present study provides a contrasting approach to change detection mechanisms from the classical mismatch negativity (MMN) paradigm, which is thought to reflect the violation of a previously established regularity (Näätänen & Winkler, 1999). The results suggest that, in the current stimulus paradigm, the emergence of regularity or coherence has a different representation to its disappearance. Recently, Chait and colleagues (2007; 2008) demonstrated distinct

cortical mechanisms for the detection of auditory ‘edges’ based on statistical signal properties, where the detection of a statistical regularity (in violation of a previous irregularity) had a different cortical signature than the discovery of a violation of a statistical regularity. The current results support the existence of such a perceptual asymmetry (Figure 4-4). It is proposed that the degree of spectrotemporal coherence is encoded in a continuous manner, with neurons tuned to sounds that are equal or greater in coherence than the neurons’ thresholds. Such a cumulative neural code contains an inherent asymmetry (Cusack & Carlyon, 2003; Treisman & Gelade, 1980): transitions to more coherent sounds excite a larger neural population, rendering them more perceptually salient. This is then also reflected in the haemodynamic response (Figure 4-6).

The data reported here move beyond the analysis of simple FM sounds to the analysis of auditory object patterns within stochastic stimuli which is dependent on mechanisms that are fundamental for the analysis of ethological sounds in a dynamic acoustic environment. This study demonstrates a mechanism for the assessment of auditory object boundaries that is already present in primary cortex, based on integrating dynamic statistical properties governing the object region within a spectrotemporal field. Such a mechanism precedes the encoding in higher-level auditory association cortex of the absolute properties of the object region.

## **Chapter 5. ENCODING OF THE RATE OF INFORMATION PRODUCTION IN PITCH SEQUENCES**

### **Summary**

The entropy metric derived from information theory provides a means to quantify the amount of information transmitted in acoustic streams like speech or music. Brain areas in which neural activity and energetic demands increase as a function of entropy or the rate of information production can be investigated by systematically varying the entropy of pitch sequences. Such a relationship between acoustic information content and neural activity is predicted to occur via an efficient encoding mechanism that uses fewer computational resources when less information is present in the signal. Specifically, it was hypothesised that such a relationship is present in the planum temporale (PT), which has been described as a 'computational hub' within auditory cortex. In two convergent fMRI studies (Studies 3 and 4), this relationship is demonstrated in PT for encoding of pitch sequences: activity in PT increased as a function of the amount of information in the pitch sequences. In contrast, a distributed fronto-parietal network for retrieval of acoustic information operated independently of entropy. The results establish PT as an efficient neural engine that demands fewer computational resources to encode redundant signals than those with high information content.

## 5.1 Introduction

We are constantly required to perceive, distinguish and identify signals in our acoustic environment. A critical first stage of these processes is the encoding of the information into a robust neural code that allows efficient subsequent processing in the auditory system (Lewicki, 2002). In the current study, the properties of such a robust neural code at the level of the cortex were investigated by varying the amount of information, or entropy, in the acoustic signal.

In the context of information theory (Attneave, 1959; Shannon, 1948), entropy ( $H$ ) denotes the uncertainty associated with an event and thus provides a metric to quantify information content: a rare, or uncertain, event carries more information than a common, or predictable, event. The properties of many information transmitting systems can be characterised in terms of entropy. Indeed, Shannon originally applied information entropy to describe transitional probabilities in language (Shannon, 1948): in English, less common letters (e.g. ‘k’) have a lower probability (or higher uncertainty) than more common letters (e.g. ‘e’) and therefore carry higher information and entropy. Similarly, entropy can be used to characterise pitch transition probabilities in simple musical melodies (Pearce & Wiggins, 2004; Pearce & Wiggins, 2006). In the present context, entropy is applied to quantify the information content of pitch sequences.

‘Fractal’ pitch sequences based on inverse Fourier transforms of  $f^{-n}$  power spectra (Patel & Balaban, 2000; Schmuckler & Gilden, 1993) provide a means to control directly the entropy of the sequence via the exponent  $n$  (Figure 5-1). For  $n = 0$ , the excursion of the pitch sequence is equivalent to fixed-amplitude-random-phase noise and thus is completely random (high entropy). In the context of information theory, the high degree of randomness in this signal does not correspond to noise that must be removed by the system, but to a low predictability of the stimulus that results in each individual element of the sequence making a high degree of contribution to the information in the sequence. As  $n$  increases, a single stream gradually dominates the local pitch fluctuations and successive pitches become increasingly predictable (low entropy). Such stimuli are more predictable so that each element of the sequence makes little contribution to the overall information in the stimulus. These families of pitch sequences with different values of  $n$  are statistical ‘fractals’ (Eke *et al.*, 2002) in

the sense that their statistical properties are scale-independent (Schmuckler & Gilden, 1993). For present purposes, the critical property of these pitch sequences that is exploited here is not their ‘fractal’ behaviour, but the variation of entropy that is produced as  $n$  varies, whilst pitch range, tempo and pitch probability remain largely constant (however, it is inherent to the system that for large exponents  $n > 4$  the pitch distribution approaches a sinusoid and consequently the corresponding probability density function is tilted towards the extremes of the pitch range and also that the average interval size between successive pitches decreases for increasing exponents  $n$ ).

Entropy for pitch sequences generated with a given value of exponent  $n$  can be determined by computing the sample entropy ( $H_{Samplen}$ ) (Richman & Moorman, 2000). Intuitively,  $H_{Samplen}$  is based on the conditional probability that two subsequences of length  $m$  that match within a tolerance of  $r$  standard deviations remain within a tolerance  $r$  of each other at the next point  $m+1$ . Explicitly, for a signal or time series of length  $N$ ,  $H_{Samplen}$  is defined as:

$$H_{Samplen}(m, r, N) = -\ln\left(\frac{A_r(m+1)}{A_r(m)}\right), \quad (\text{Eq. 5-1})$$

where  $A_r(m)$  (or  $A_r(m+1)$ ) denotes the probability that two subsequences of length  $m$  (or  $m+1$ ) match within a tolerance  $r$ . Two sequences ‘match’ if their maximum absolute point-by-point difference is within a tolerance of  $r$  standard deviations. That is, sample entropy is essentially a measure of self-similarity, where highly self-similar time series signify high redundancy and therefore low entropy, while time series with low self-similarity represent a high degree of uncertainty and therefore high entropy. Furthermore, sample entropy is a non-parametric measure in the sense that it does not require a priori knowledge of the true probability density function of the underlying time series. In the present case, the parameters were chosen as  $m = 2$ ,  $r = 0.5$ , while  $N$  represents the number of tones of the pitch sequence.

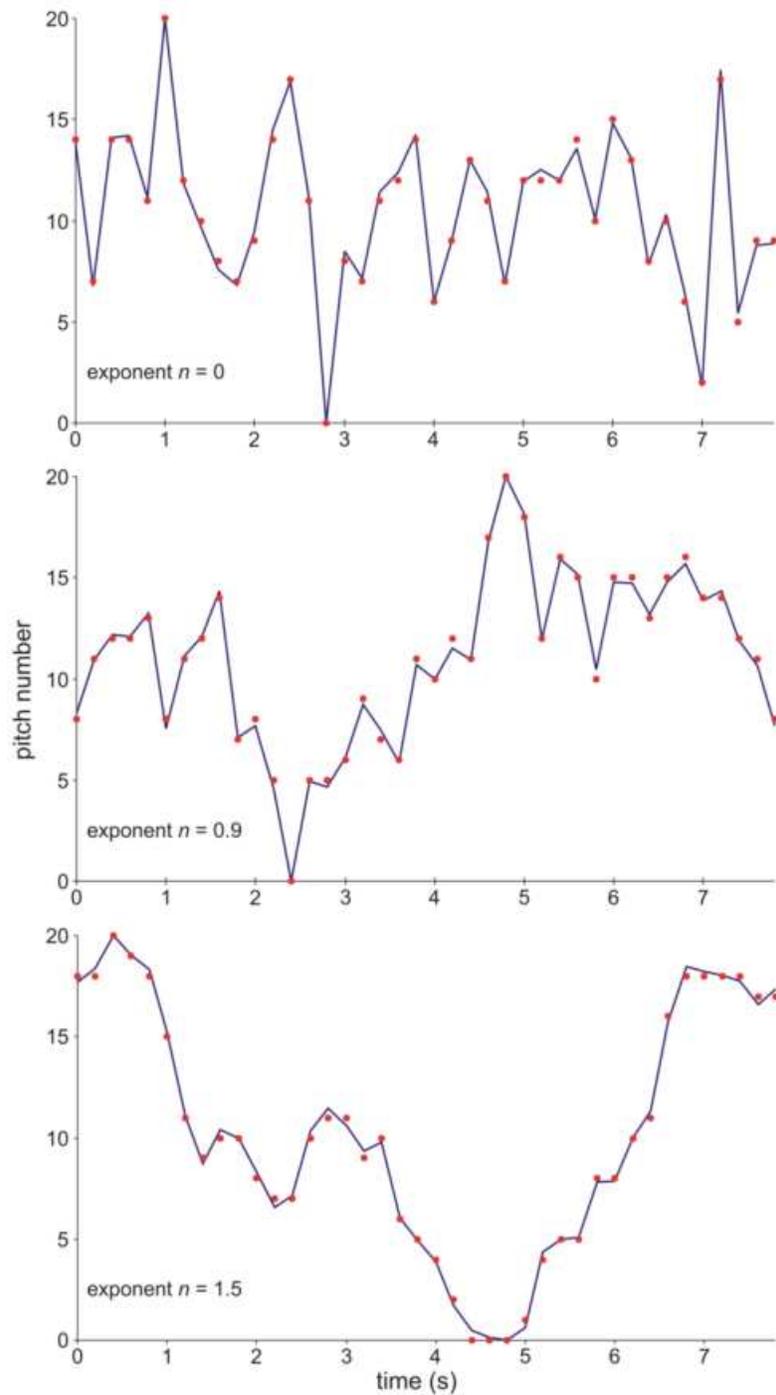


Figure 5-1 Auditory stimulus. Examples of fractal waveforms (blue) and the related pitch sequences (red, rounded to the nearest integer) based on inverse Fourier transforms of  $f^{-n}$  power spectra, with exponent  $n = 0$  (top),  $n = 0.9$  (middle),  $n = 1.5$  (bottom). Equitempered pitch (10-note octave, ranging over two octaves, resulting in 21 possible pitches, with ordinal indices 0 to 21 corresponding to 300 Hz to 1200 Hz) is denoted on the y-axis, time (in seconds) on the x-axis. Entropy is largest for the top pitch sequence and decreases as exponent  $n$  increases.

By varying information theoretic properties of pitch sequences, this addresses encoding mechanisms applied to sounds at a level of generic processing that is not specific to any semantic category. Even before such encoding mechanisms are engaged, the auditory system must represent spectrotemporal features of the stimulus in sufficient detail such that a number of different aspects of the stimulus can be encoded, in order to allow different types of subsequent categorical and semantic processing. In the current context, encoding constitutes the stage of analysis between the detailed representation of the spectrotemporal structure of the stimulus and the subsequent categorical analysis of abstracted acoustic forms. A single sound may be associated with more than one abstracted form: for example, one might obtain vowel, speaker and position from a single sound, where each feature can undergo subsequent categorical and semantic processing. Here, information theory is used to demonstrate encoding mechanisms in the brain that result in the abstraction of a form of the stimulus.

It was hypothesised that, if such encoding mechanisms are efficient, they will use less computational resource for stimuli that have low information content compared to stimuli that have high information content. This hypothesis is tested by measuring the fMRI BOLD signal as an estimate of neural activity and computational resource during encoding of auditory stimuli in which the information content is systematically varied. It was further hypothesised that processing in primary auditory cortex in Heschl's Gyrus (HG) corresponds to a stage at which the detailed spectrotemporal structure of sounds is represented (deCharms *et al.*, 1998; Nelken *et al.*, 1999; Schnupp, 2001), and where such a relationship will *not* be observed. Instead, such a relationship is expected to be observed in distinct auditory association cortex in planum temporale (PT), which has previously been characterised as a 'computational hub' (Griffiths & Warren, 2002) that is required to convert spectrotemporal representations into 'templates': sparse symbolic neural representations that are the basis for categorical, semantic and spatial processing. For example, the spectral envelope of a sound would represent such a 'template' for vowel processing (Warren *et al.*, 2005a). The model was developed to account for the involvement of PT in the analysis of a variety of complex sounds that can be processed categorically (speech, music, and environmental sounds) as well as different spatial attributes (for a review, see Griffiths & Warren, 2002).

Study 3 investigated the encoding of pitch sequences that can be like melodies in their structure, but in which the structure and information content is determined by statistical rules. It was predicted that brain areas will display a positive relationship between the information content or entropy of pitch sequences and neural activity as assessed by the BOLD signal during encoding. Specifically, it was hypothesised that such a relationship exists in PT but not in earlier auditory areas.

## **5.2 Materials and Methods (Study 3)**

### **5.2.1 Participants**

30 right-handed participants (aged 18-43 years, mean age = 24.9; 19 females) with normal hearing and no history of audiological or neurological disorders provided written consent prior to the experiment. None of the participants were professional musicians. The study was approved by the National NHS Research Ethics Committee. Eight participants had to be excluded due to excessive head movements (more than 5 mm translation or 5 degree rotation within one session) or not meeting the psychophysical assessment criteria (see below), leaving a total of 22 participants (aged 18-40, mean age = 24.2 years; 12 females).

### **5.2.2 Stimuli**

All stimuli were created digitally in the frequency domain using Matlab ([www.mathworks.com](http://www.mathworks.com)) at a sampling frequency of 44.1 kHz and 16 bit resolution. Stimuli were 'fractal' sine tone sequences based on inverse Fourier transforms of  $f^{-n}$  power spectra (Patel & Balaban; Schmuckler & Gilden, 1993) for six levels of  $n$  (0, 0.3, 0.6, 0.9, 1.2, 1.5), where pitch sequences ranged from totally random ( $n = 0$ ; high entropy) to more coherent or predictable ( $n = 1.5$ ; low entropy). By randomising the phase spectrum, each exemplar is unique, while at the same time displaying the same characteristic correlational properties of a given level. The pitch range spanned two octaves from 300-1200 Hz, with each octave split into 10 discrete equidistant pitches.

Pitch sequences were presented at a tempo of five notes per second, with a total duration of 7.6 seconds for each pitch sequence (38 notes per sequence). There were 60 exemplars for  $n = 0$  and 30 exemplars for the remaining five levels of  $n$ .

The mean entropy for each level of exponent  $n$  was calculated using the sample entropy  $H_{Samplen}$  (Richman & Moorman, 2000) measure, as described in the Introduction:

$$H_{Samplen}(m, r, N) = -\ln\left(\frac{A_r(m+1)}{A_r(m)}\right).$$

$A_r(m)$  denotes the probability that two subsequences of length  $m$  match within a tolerance  $r$ , i.e.  $A_r(m)$  is the ratio of [all pairs of subsequences of length  $m$  that match] divided by [all possible pairs of subsequences of length  $m$ ]; the same applies to  $A_r(m+1)$ . Guided by Lake and colleagues (2002), a tolerance  $r = 0.5$  and length of subsequence  $m = 2$  as parameter values were chosen. As Eke and colleagues (2002) point out, taking a subset of data points from a fractal time series essentially introduces noise into the resulting time series, leading to lower  $n$  and consequently higher entropy estimates relative to the original values. Table 5-1 therefore lists the mean sample entropy values for the time series of the 38 notes in each pitch sequence.

### 5.2.3 Experimental design

In a behavioural experiment prior to scanning, full psychometric functions were acquired from participants discriminating the non-random pitch sequence against a random reference ( $n = 0$ ) in a 2I2AFC paradigm. Participants were not given feedback. Stimuli were not the same as in the subsequent imaging paradigm and there were 72 trials (12 trials per level). Psychometric functions and 75% correct thresholds were estimated via a Weibull boot-strapping procedure (Wichmann & Hill, 2001). Participants who did not reach at least 80% performance for levels 5 or 6 were not included in the fMRI analysis. In the functional imaging paradigm, participants were asked to categorise whether or not the pitch sequence was random by pressing the corresponding button at the end of each pitch sequence, bearing in mind that pitch sequences of intermediate levels ( $n = 0.6$  to  $0.9$ ) are neither completely random nor

completely coherent (in these cases, participants should nevertheless indicate their predominant percept). Stimuli were presented via custom-built electrostatic headphones at 70 dB SPL using Cogent software (<http://www.vislab.ucl.ac.uk/Cogent/>).

#### 5.2.4 fMRI protocol and analysis

Gradient weighted echo planar images (EPI) (see Section 2.3.1) were acquired with a 3 Tesla Siemens Allegra MRI system (Erlangen, Germany), using a sparse temporal sampling technique (Belin *et al.*, 1999; Hall *et al.*, 1999) (time to repeat/time to echo, TR/TE = 10,530/30 ms), where each volume was cardiac gated. A total of 246 volumes (42 slices, 3x3x3 mm voxel resolution) were acquired over three sessions (82 per session), including 60 volumes for  $n = 0$  and 30 volumes for the other levels of  $n$ , as well as 30 silent control trials (the first two volumes of each session were discarded to allow for saturation effects). To correct for geometric distortions in the EPI images due to B0 field variations, Siemens fieldmaps were acquired for each participant (Cusack *et al.*, 2003; Hutton *et al.*, 2002). A structural T1 weighted scan was acquired for each participant (Deichmann *et al.*, 2004).

Imaging data were processed and analysed using Statistical Parametric Mapping software (SPM2, <http://www.fil.ion.ucl.ac.uk/spm>) (see also Sections 2.5-2.6). Volumes were realigned and unwarped using the fieldmap parameters, spatially normalised (Friston *et al.*, 1995a) to standard stereotactic space and smoothed with an isotropic Gaussian kernel of 8 mm full-width-at-half-maximum. Statistical parametric maps were generated using a Finite Impulse Response (FIR) box-car function in the context of the general linear model (Friston *et al.*, 1995b). The six conditions were parametrically modulated based on the average sample entropy (Richman & Moorman, 2000) value for each level of  $n$  (see Table 5-1), statistically evaluated using a random-effects model and thresholded at  $p < 0.001$  (uncorrected for multiple comparisons across the brain) for areas with an *a priori* hypothesis, i.e. in auditory cortex and specifically PT. In addition, a volume-of-interest analysis was carried out controlling for multiple comparisons within PT by centring a 1 cm sphere around the centroid of the triangular anterior part of PT situated within the superior temporal plane as opposed to the more posterior part that abuts the parietal lobe ([-56 -28 6]

and [58 -24 8] for left and right PT, respectively). The choice of volume was based on the identification of the anterior part of PT in the studies that suggested the ‘computational hub’ model (Griffiths & Warren, 2002). For areas that were not predicted *a priori*, a statistical threshold of  $p < 0.05$  after family-wise error (FWE) correction was adopted.

A potential effect of adaptation in frequency bands at an earlier sensory level was investigated in detail, since pitch sequences with low entropy tend to spend more time within each critical band. Study 3 did not allow disambiguation of the three cytoarchitecturally (Morosan *et al.*, 2001) and functionally (Patterson *et al.*, 2002) distinct areas in HG, namely medial, central, and lateral HG (see Study 4 below for further discussion). Therefore, single coordinates were identified based on local maxima of a sound minus silence contrast for left [-46 -24 6] and right [50 -24 8] HG that are most similar to central HG (Morosan *et al.*, 2001; Patterson *et al.*, 2002), and the first eigenvariate of the BOLD signal at these coordinates was extracted (see Figure 5-2).

The BOLD signal was extracted using a standard procedure in SPM: the time series of a given voxel (e.g. the peak activation voxel for the entropy effect) is provided by SPM via a volume-of-interest (VOI) routine. At the second level statistical analysis, this results in a time series for each contrast where each data point corresponds to a participant. The routine is executed for each contrast, in the current case either six (Study 3) or five (Study 4) [Level – Silence] contrasts, resulting in a 22X6 or 24X5 matrix (22 or 24 participants, respectively), where each row corresponds to a participant and each column to a contrast. The threshold at which the BOLD signal was extracted was  $p < 0.05$  (uncorrected for multiple comparisons). The values are then normalised to the maximum value.

Note that the interaction described here between the BOLD signal in HG and PT across levels assumes that the coupling between neuronal response and the haemodynamic BOLD signal is identical in the two brain regions. While there is no reason to assume the contrary, it has also not been proven that this is indeed the case.

### 5.3 Results (Study 3)

Participants were presented with pure tone pitch sequences which were based on  $f^{-n}$  power spectra with  $n$  ranging from  $n = 0$  to 1.5 in five steps of 0.3. In a behavioural experiment prior to scanning, full psychometric functions were acquired demonstrating that all of the 22 participants could reliably distinguish a non-random pitch sequence from a random ( $n = 0$ ) reference in a 2I2AFC paradigm (see Materials and Methods). Perceptual thresholds for discriminating non-random from a random pitch sequence lay between  $n = 0.6$  and  $n = 0.9$  for the majority of participants.

In a sparse fMRI paradigm (Belin *et al.*, 1999; Hall *et al.*, 1999), participants listened to pitch sequences of a given value for  $n$  and indicated whether it was random or not. A parametric regressor based on the mean sample entropy (Richman & Moorman, 2000) value at each of the six levels of  $n$  (Table 5-1) was used to probe for cortical areas that increased their activity with increasing entropy. The fMRI analysis revealed a BOLD signal increase in PT as a function of increasing entropy at a significance level of  $p < 0.001$  (uncorrected for multiple comparisons, see Figure 5-2 and Table 5-2) and using a small volume correction for the anterior part of PT at a significance level of  $p < 0.05$  (see Section 5.2). No area increased its activity as a function of decreasing entropy, i.e. increasing predictability or redundancy.

Table 5-1 Mean sample entropy  $H_{\text{SampEn}}$  values (standard error of the mean in parentheses) of the pitch sequences across levels in the two studies. The values for each level differ slightly between the studies because pitch sequences in Study 3 consisted of 38 notes, while those in Study 4 consisted of 24 notes.

	Level1	Level2	Level3	Level4	Level5	Level6
Study 3	1.38 (.25)	1.39 (.24)	1.32 (.21)	1.05 (.19)	0.75 (.19)	0.48 (.16)
Study 4	1.49 (.33)	1.54 (.35)	1.39 (.31)	1.18 (.27)	0.87 (.23)	

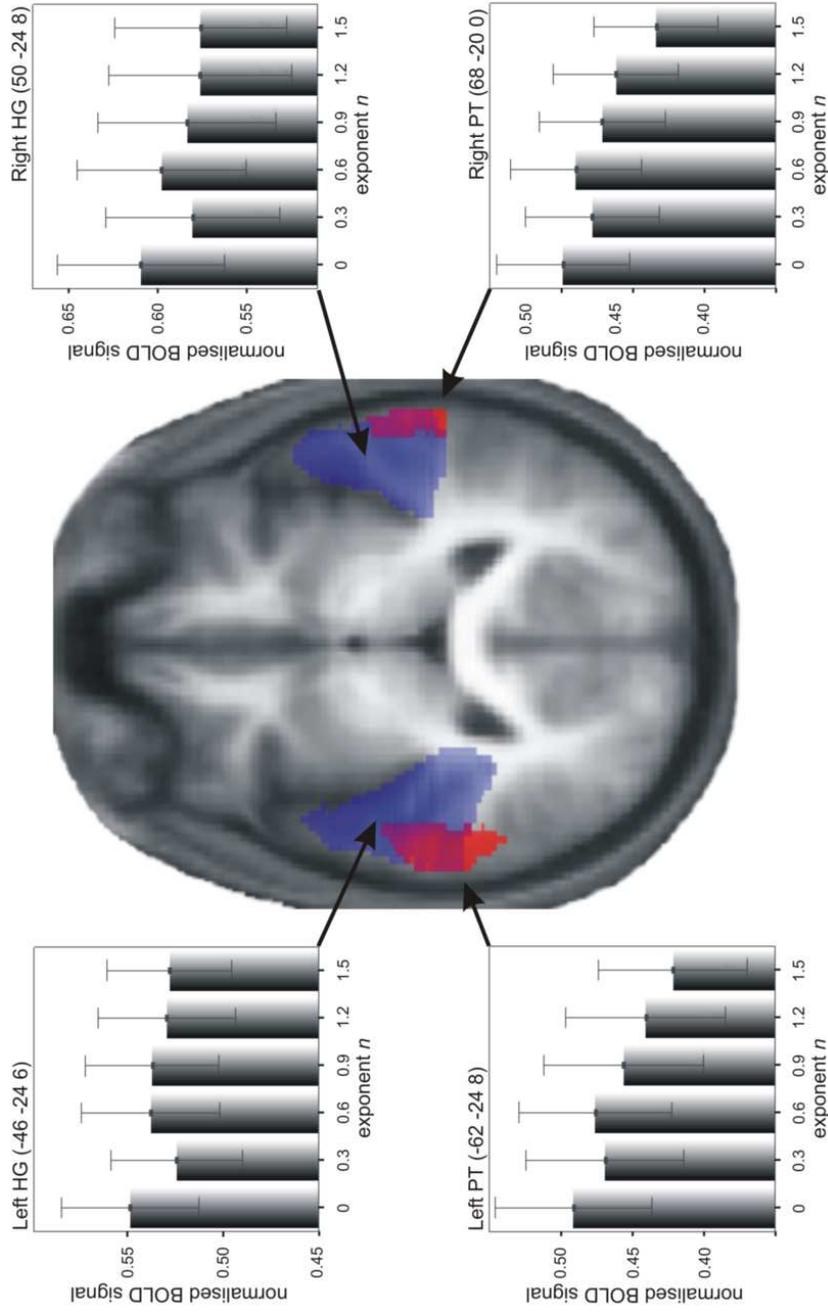


Figure 5-2 Main results I. Areas showing an increase in BOLD signal ( $p < 0.001$ , uncorrected) for multiple comparisons across the brain) as a function of increasing entropy (red) and areas that responded to sound in general ([Sound - Silence] contrast,  $p < 0.05$ , FWE corrected) (blue) rendered on a tilted (pitch = -0.4 radians) axial section of participants' normalised average structural scan. Normalised mean percent BOLD signal change ( $\pm$ SEM) at the local maxima in left and right PT (bottom) and HG (top) is plotted for the six levels of exponent  $n$ .

These results suggest a greater computational and energetic demand for encoding in PT as the information content of acoustic sequences (as assessed by entropy) increases. However, the third study has three potential confounds, which were addressed in a fourth study. Firstly, the effect of entropy in PT might reflect adaptation of the sensory cortical representation of frequency, as the pitch sequences were based on pure tones: for low values of exponent  $n$ , the frequency excursions or intervals are greater on average, so that the signal moves more between specific frequency representations and PT might adapt less and thus produce a greater local activity. Such a mechanism would also be expected to occur in primary and secondary auditory cortex within HG. Therefore, the specific relationship between fractal exponent and local activity in HG and PT was explored by extracting the first eigenvariate of the BOLD signal in left and right HG as well as the local maxima in PT (see Materials and Methods). No significant difference across entropy levels was demonstrated in HG (2 (Hemisphere)  $\times$  6 (Entropy level) repeated measures ANOVA: no main effect of Entropy level ( $F_{(5,17)} = 1.11$ ,  $p > 0.1$ ); Figure 5-2). Furthermore, a 2 (Area [PT vs. HG])  $\times$  6 (Entropy level)  $\times$  2 (Hemisphere) repeated measures ANOVA demonstrated a significant difference in the relationship between BOLD signal across Entropy levels in PT versus HG: Area  $\times$  Entropy level interaction ( $F_{(5,17)} = 4.86$ ,  $p < 0.001$ ).

The existence of the effect in auditory association cortex in PT, the absence of an effect in HG, and a significant interaction between effects in the two areas are indirect evidence against an explanation of the results based on sensory adaptation. Nevertheless, a putative sensory explanation was addressed in a fourth study by using regular-interval noise where sounds have identical passband regardless of their pitch (Griffiths *et al.*, 1998; Patterson *et al.*, 1996; Yost *et al.*, 1996).

Secondly, the effect of entropy might reflect perceptual adaptation at the level of the representation of pitch. Again, such an effect would not be expected in association cortex, but in a proposed ‘pitch centre’ in secondary cortex (Bendor & Wang, 2005; Patterson *et al.*, 2002; Penagos *et al.*, 2004). The fourth study therefore incorporated a more suitable design to detect a potential differential response to the entropy of the acoustic stimuli in cytoarchitectonic (Morosan *et al.*, 2001) and

functional (Patterson *et al.*, 2002) subdivisions of HG in medial, central and lateral HG.

Finally, participants in the third study were explicitly required to assess whether the sequences were random or not. This made it possible that the results reflected a category judgment rather than a fundamental encoding mechanism. To test this, the fourth study differentially examined encoding and retrieval components as a function of entropy, but independent of any other stimulus-related classification task.

## **5.4 Materials and Methods (Study 4)**

### **5.4.1 Participants**

30 right-handed participants (aged 20-44 years, mean age = 28.0 years; 16 females) with normal hearing and no history of audiological or neurological disorders provided written consent prior to the experiment. The study was approved by the National NHS Research Ethics Committee. Six participants had to be excluded because of excessive head movements (more than 5 mm translation or 5 degree rotation within one session), leaving a total of 24 participants (aged 20-44, mean age = 28.58 years; 12 females).

### **5.4.2 Stimuli**

As in Study 3, pitch sequences were again based on  $f^{-n}$  power spectra for five levels of  $n$  (0, 0.3, 0.6, 0.9, 1.2). Each pitch was based on regular-interval noise created using a delay-and-add algorithm (Griffiths *et al.*, 1998; Patterson *et al.*, 1996; Yost *et al.*, 1996) with 16 iterations. The pitch range spanned two octaves from 150-600 Hz, with each octave split into 10 discrete equidistant pitches. Pitch sequences were presented at a tempo of four notes per second, with a total duration of 6 seconds for each pitch sequence (24 notes per sequence). The mean entropy values for each level of  $n$  are depicted in Table 5-1 and are slightly different from Study 3 because each pitch sequence had 24 notes instead of 38. There were 30 exemplars for each level of

$n$ , and stimuli were presented via custom-built electrostatic headphones at 70 dB SPL using Cogent software (<http://www.vislab.ucl.ac.uk/Cogent/>).

### 5.4.3 Experimental design

In a sparse-imaging paradigm (Belin *et al.*, 1999; Hall *et al.*, 1999), participants were scanned 1) after being required to encode a pitch sequence with a particular entropy value and 2) after listening to a second pitch sequence that was either the same sequence or a different sequence from the same entropy level and indicating whether this was the same pitch sequence or different (see also Figure 5-3). To de-correlate (Henson, 2006) activations due to the first and second pitch sequence, the second pitch sequence followed the first pitch sequence either immediately in the next TR, or with two or three TR's delay (within-trial delay). Similarly, the first pitch sequence of the next pair could follow the second pitch sequence of the previous pair immediately, or with one or two TR's delay (between-trial delay). There were 20 pitch sequence pairs for each level, amounting to 100 encoding and 100 retrieval stimuli across the five levels of exponent  $n$ . In addition, there were a total of 100 within-trial volumes and 100 between-trial rest volumes. For each level of exponent  $n$ , 10 out of 20 pairs were identical, and 10 were different. Stimuli were counterbalanced between participants.

To guide participants, a '1' was displayed at the centre of the screen from the start of the first pitch sequence until the start of the second pitch sequence, when a '2' was displayed. At the end of the second pitch sequence, participants briefly saw a '?' to indicate they should now give their response as to whether they thought the second pitch sequence was the same as or different from the first pitch sequence. Participants received immediate feedback. During the rest period between trials, participants saw a fixation cross '+' at the centre of the screen and were instructed to relax.

### 5.4.4 fMRI protocol and analysis

Gradient weighted echo planar images (EPI) were acquired with a 3 Tesla Siemens Allegra MRI system (Erlangen, Germany), using a sparse temporal sampling

technique (Belin *et al.*, 1999; Hall *et al.*), where each volume was cardiac gated to reduce motion artefacts (TR/TE = ca. 8,800/30 ms). A total of 404 volumes (42 slices, 3x3x3 mm voxel resolution) were acquired over two sessions (the first two volumes of each session were discarded to allow for saturation effects). Subsequent to the functional paradigm, a structural T1 weighted scan was acquired for each participant (Deichmann *et al.*, 2004).

Imaging data were processed and analysed using Statistical Parametric Mapping software (SPM5, <http://www.fil.ion.ucl.ac.uk/spm>) (see also Sections 2.5-2.6). Volumes were realigned and unwrapped, spatially normalised (Friston *et al.*, 1995a) to MNI standard stereotactic space and smoothed with an isotropic Gaussian kernel of 8 mm FWHM. Statistical parametric maps were generated by modelling the evoked haemodynamic response to the stimuli and the delay period in the context of the general linear model (Friston *et al.*, 1995b).

To probe for an effect of entropy on encoding, a contrast was carried out to identify areas in which the BOLD signal in the first and second scans increased as a function of a parametric regressor based on the mean sample entropy value at each level (Table 5-1). A second contrast investigated the effect of retrieval and comparison independent of encoding by subtracting the effect of encoding of the first stimulus only (corresponding to the first scan) from that to encoding of the second stimulus, retrieval of the first, and comparison of the two (corresponding to the second scan). A third contrast examined the effect of entropy on retrieval by subtracting [first scan entropy increase] from [second scan entropy increase]. Statistical results are based on a random-effects model and thresholded at  $p < 0.001$  (uncorrected for multiple comparisons across the brain) for areas with an *a priori* prediction, i.e. PT, in addition to the same small volume correction ( $p < 0.05$  corrected for multiple comparisons) as in Study 3. For areas that were not predicted *a priori*, a more conservative statistical threshold of  $p < 0.05$  after FWE correction was adopted.

The fourth study was better suited to identify the three cytoarchitectonically (Morosan *et al.*, 2001) and functionally (Patterson *et al.*, 2002) distinct areas within HG based on the sound minus silence contrast because of 1) the greater number of silent trials and 2) the use of broadband stimuli. Three activations were identified in

HG in either hemisphere, primarily to locate the lateral area previously implicated in perceptual pitch analysis (Patterson *et al.*, 2002; Penagos *et al.*, 2004) and to allow a comparison of the effect of entropy on activity here with that in PT (for individual coordinates see Table 5-2 for PT, Figure 5-5 for central and Figure 5-6 for medial and lateral HG).

Cardiac gating in Study 4 produced a reliable signal in subcortical structures IC and MGB (Figure 5-7). The data were therefore reanalysed with a 4 mm FWHM smoothing kernel that is appropriate to these structures. Local maxima based on a sound minus silence contrast were identified in left IC ([-6 -34 -12]) and right IC ([6 -34 -10]) and left MGB ([-14 -26 -8]) and right MGB ([12 -24 -8]).

For further analysis considerations see Section also 5.6.

## 5.5 Results (Study 4)

In a sparse fMRI paradigm (Belin *et al.*, 1999; Hall *et al.*, 1999), participants were presented with fractal pitch sequences based on  $f^{-n}$  power spectra, with  $n$  ranging from  $n = 0$  to 1.2 in four steps of 0.3. The separate pitches corresponded to regular-interval noise (Griffiths *et al.*, 1998; Patterson *et al.*, 1996; Yost *et al.*, 1996) (see Materials and Methods). By using broadband stimuli and an increased number of silent trials, the fourth study employed a more suitable design to allow disambiguation of the medial functional area in HG that corresponds to primary auditory cortex and areas in lateral HG that correspond to secondary cortices, including the area within which activity corresponds to pitch salience (Patterson *et al.*, 2002; Penagos *et al.*, 2004). The second paradigm also enabled the disambiguation of encoding and retrieval mechanisms. Participants were scanned 1) after being required to encode a pitch sequence with a particular entropy value and 2) after listening to a second pitch sequence that was either identical to the first sequence or different from the first sequence but with the same entropy value. Activity during the first scan reflects the energetic demands of encoding the first sequence, whilst activity during the second scan reflects encoding of the second sequence, retrieval of the first, and comparison of

the two. In order to decorrelate the two scans (Henson, 2006), a delay of one, two, or three scans was introduced between the pitch sequences (see Section 5.4.3 and Figure 5-3). In contrast to the third study, participants were not informed about the nature of the pitch sequences and instead were only told that they would hear pairs of pitch sequences and that their task would be to say whether the second was same or different.

Participants' behavioural performance in the scanner was assessed via hits (*hit*) and correct rejections (*cr*) percent scores (see also Figure 5-4). Both mean *hit* ( $74.25\% \pm 3.14$  SEM) and mean *cr* ( $73.42\% \pm 3.31$  SEM) scores were significantly above chance (50%) (one-sample t-test, *hit*:  $t_{23} = 7.73$ ; *cr*:  $t_{23} = 7.08$ , both  $p < 0.001$ ). Furthermore, a 2 (Response [*hit* vs. *cr*])  $\times$  5 (Entropy Level)  $\times$  3 (Delay) repeated-measures ANOVA showed no main effect in any of the three factors ( $F_{(23,1)} = 0.33$ ;  $F_{(20,4)} = 1.1$ ;  $F_{(22,2)} = 0.53$ ; all  $p > 0.05$ , for Response, Entropy Level and Delay, respectively). There was no Response  $\times$  Entropy Level interaction ( $F_{(20,4)} = 1.01$ ,  $p > 0.05$ ), indicating that participants' performance was not influenced by the Entropy Level of the pitch sequences. Participants had higher *cr* than *hit* scores for Delay 3, while there were more *hits* than *cr* for Delays 1 and 2 (Response  $\times$  Delay interaction;  $F_{(22,2)} = 7.91$ ,  $p = 0.001$ ). An Entropy Level  $\times$  Delay interaction ( $F_{(16,8)} = 2.14$ ,  $p < 0.05$ ) showed a performance increase for Delay 1 from Entropy Level 1 to Entropy Level 5, while there was no such systematic effect for Delay 2 or Delay 3. There was no Response  $\times$  Entropy Level  $\times$  Delay interaction ( $F_{(16,8)} = 0.45$ ,  $p > 0.1$ ).

The imaging results replicate the findings of the third study, demonstrating that activity in PT for encoding (as assessed by both the first and second scan of each pair) increased significantly as a function of entropy for the same significance thresholds as in the third study (Figure 5-5 and Table 5-2). The effect at the level of primary and secondary auditory cortex was examined in detail by extracting the BOLD signal in medial, central and lateral HG (Morosan *et al.*, 2001; Patterson *et al.*, 2002) (Figure 5-5 and Figure 5-6): three separate 5 (Entropy level)  $\times$  2 (Hemisphere) repeated measures ANOVAs showed no main effect of Entropy Level ( $F_{(4,20)} = 0.85$ ,  $F_{(4,20)} = 0.77$ ,  $F_{(4,20)} = 1.83$ , all  $p > 0.1$ , for medial, central and lateral HG, respectively).

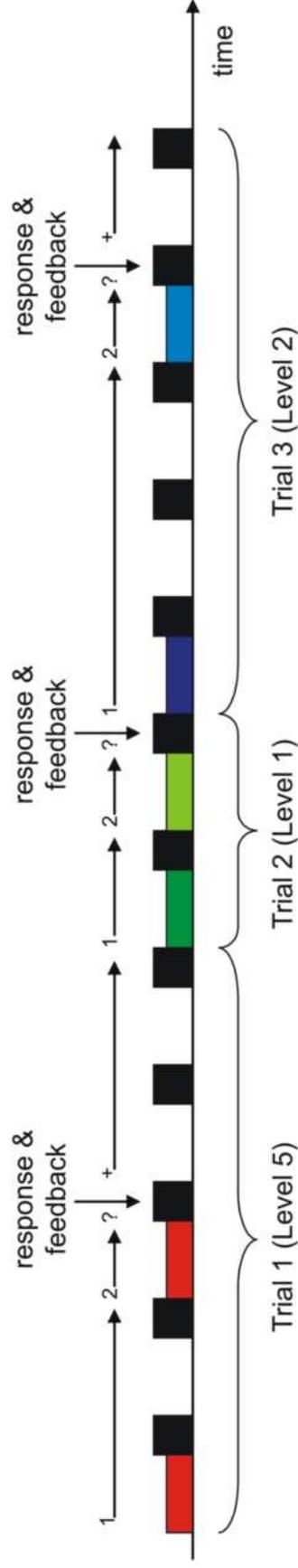


Figure 5-3 Experimental design. Depicted are three consecutive trials with pairs of pitch sequences drawn from entropy Levels 5, 1 and 2, respectively. Coloured boxes indicate the presentation of a pitch sequence, black boxes indicate the acquisition of a scan volume, and the white gaps between scans denote silent periods. Identical colours within one trial indicate that the two pitch sequences of a pair are identical (Trial 1), while slightly different hues indicate trials where the second pitch sequence was different from the first one, but drawn from the same entropy Level (Trials 2 & 3). There were three possible delays within and between trials; for example, Trial 1 has a within-trial delay of 2 scans before the presentation of the second pitch sequences, and a between-trial delay of 3 scans before the beginning of the subsequent trial. Trial 2 has within-trial and between-trial delays of 1 scan, etc. Visual cues as depicted above the schematic of the design presented to guide participants through the experiment. Participants received immediate feedback (correct/incorrect) after giving their response.

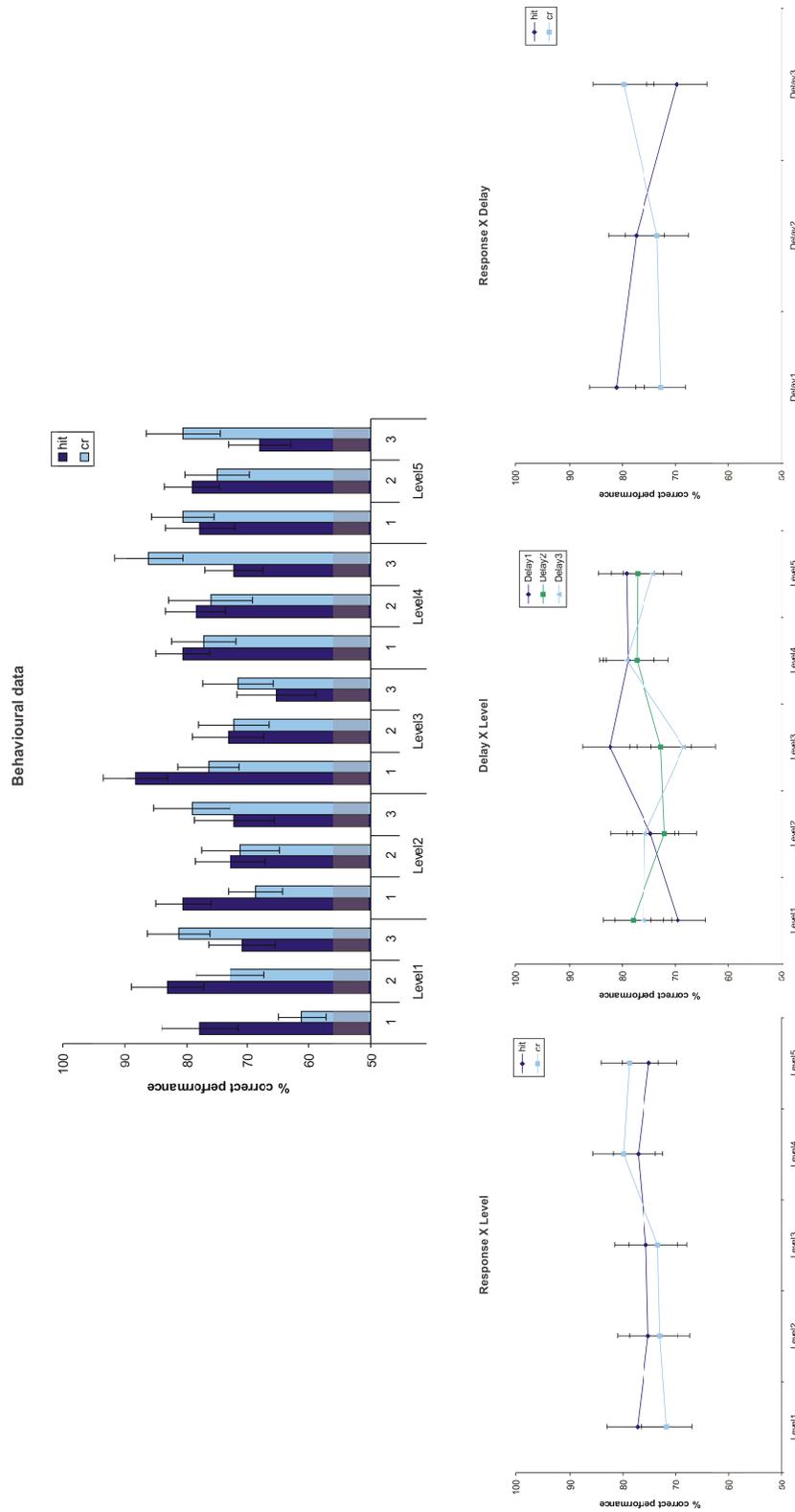


Figure 5-4 Behavioural results. Hits (*hit*) and correct rejections (*cr*) are displayed for the three Delay periods (1, 2, 3) across the five Levels of exponent  $n$ . A repeated measure ANOVA showed no main effect of either Response (*hit* vs. *cr*), Delay (1, 2, 3) or Level (Levels 1-5). There was no Response  $\times$  Level interaction, but a Response  $\times$  Delay and Delay  $\times$  Level interaction (for detailed statistical data see Results section for Study 4).

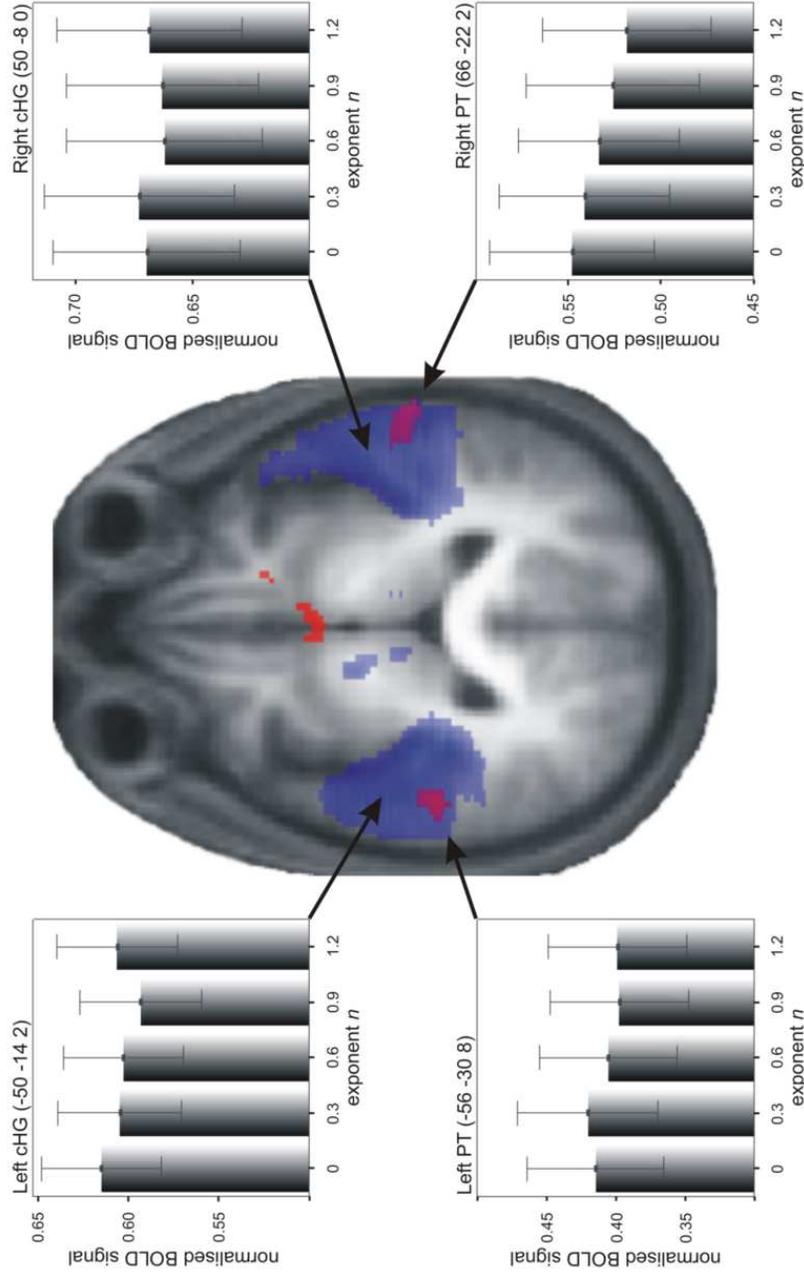


Figure 5-5 Main results II. Areas showing an increase in BOLD signal ( $p < 0.005$ , uncorrected) for multiple comparisons across the brain) as a function of increasing entropy (red) and areas that responded to sound in general ([Sound - Silence] contrast, blue) rendered on a tilted (pitch = -0.5 radians) axial section of participants' normalised average structural scan. Normalised mean BOLD signal change ( $\pm$ SEM) at the local maxima in left and right PT (bottom) and central HG (cHG, top) is plotted for the five levels of exponent  $n$ . See Figure 5-6 for corresponding plots of BOLD signal in medial and lateral HG.

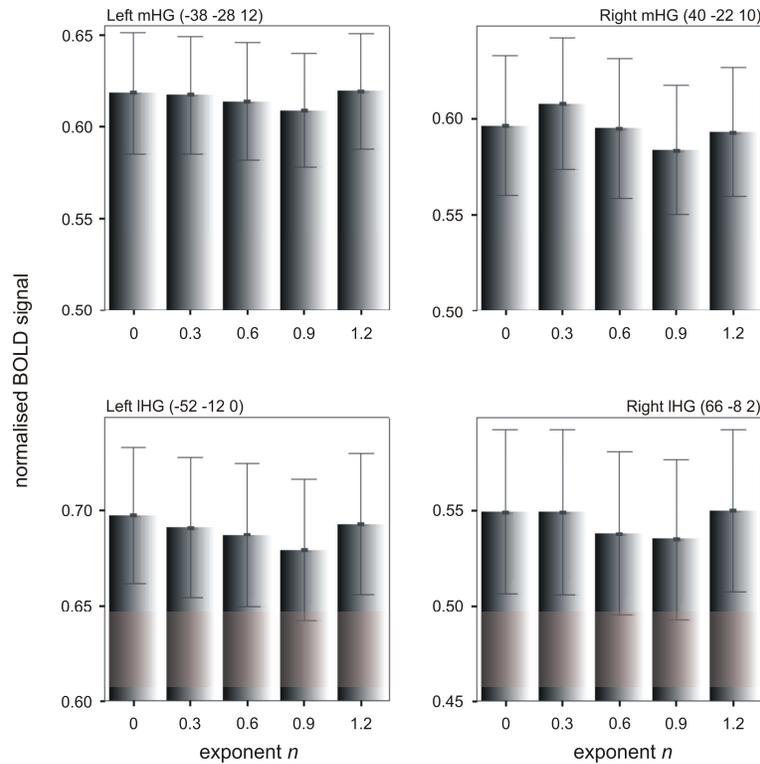


Figure 5-6 BOLD signal in medial and lateral HG. Normalised BOLD signal change (y-axis) in left and right medial (top) and lateral HG (bottom) (mHG and IHG, respectively) plotted against the five levels of exponent  $n$  (x-axis) for Study 4. See Figure 5-5 for corresponding plots of BOLD signal in central HG.

Table 5-2 MNI coordinates of local maxima in PT as a function of increasing entropy in the two studies. Coordinates in italics depict the arithmetic mean of the geometric MNI coordinates for left and right PT in the two studies. The last column lists the geometric distance between the arithmetic means; note that this is smaller than the smoothing kernel (8mm) applied to the data.

Hemisphere	Study 3				Study 4				geometric distance Study 3 vs. Study 4
	entropy increase				entropy increase				
	x	y	z	t-value	x	y	z	t-value	
left PT	-62	-24	8	5.70	-56	-30	8	3.83	
	-60	-38	16	5.02					
mean	-61	-31	12		-56	-30	8		6.48
right PT	68	-20	0	4.74	66	-22	2	3.93	
	66	-30	4	4.39					
	66	-12	0	3.72					
mean	67	-21	1		66	-22	2		1.73

Furthermore, the relationship between entropy and BOLD signal was significantly different between PT and all three subdivisions of HG: 2 (Area [PT vs. medial, central or lateral HG])  $\times$  5 (Entropy level)  $\times$  2 (Hemisphere) repeated measures ANOVAs carried out separately for medial, central or lateral HG showed an Area  $\times$  Entropy level interaction ( $F_{(4,20)} = 2.61, p < 0.05$ ;  $F_{(4,20)} = 3.31, p < 0.05$ ;  $F_{(4,20)} = 5.55, p < 0.001$ ), for medial, central and lateral HG, respectively).

The cardiac gated image acquisition in Study 4 furthermore allowed an examination of a potential effect of stimulus entropy in subcortical auditory structures. The relationship between entropy and the activity in the medial geniculate body (MGB) and inferior colliculus (IC) was examined using a smaller smoothing kernel (4mm FWHM) appropriate for these subcortical structures (Figure 5-7). This analysis showed no main effect of entropy on the BOLD response in these areas (two separate 5 (Entropy Level)  $\times$  2 (Hemisphere) repeated measures ANOVAs:  $F_{(4,20)} = 0.35, p > 0.1$ , for IC;  $F_{(4,20)} = 1.32, p > 0.1$ , for MGB). Due to the different spatial smoothing, no meaningful interaction with the response in cortical structures can be computed.

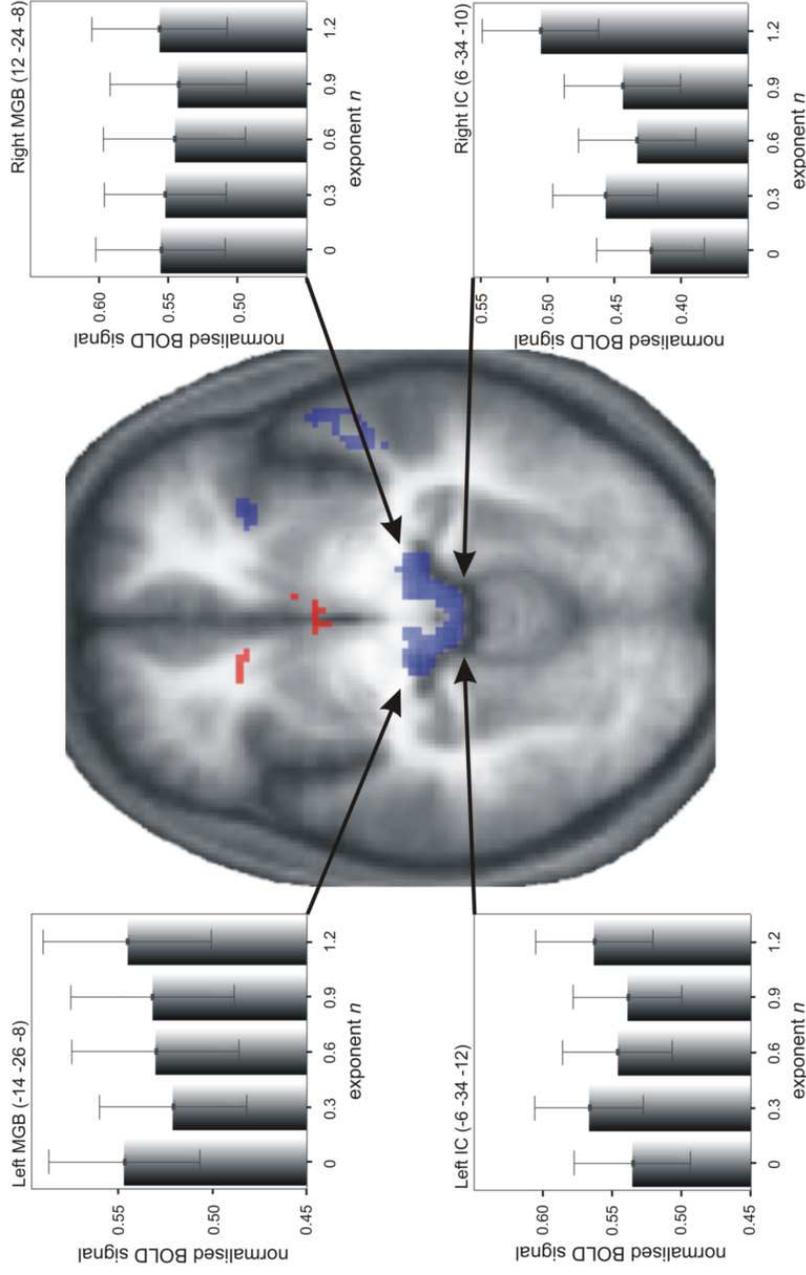


Figure 5-7 Response in IC and MGB. Entropy increase ( $p < 0.001$ , uncorrected) (red) and [Sound – Silence] ( $p < 0.05$ , FWE corrected) (blue) contrasts superimposed on a horizontal section ( $z = -10$ ) of participants' normalised average structural scan that covers IC and MGB (note that  $z$  coordinates vary slightly between maxima and arrows therefore are only indicative of the exact location). Normalised mean percent BOLD signal change ( $\pm$ SEM) at the local maxima in left and right IC (bottom) and MGB (top) is plotted for the five levels of exponent  $n$ .

A second analysis based on the contrast between the second and first scans sought areas involved in retrieval and comparison, but not encoding. This contrast highlighted activity within a bilateral fronto-parietal network, including the mid-ventrolateral prefrontal cortex (mid-VLPFC) area 47/12, inferior parietal sulci, medial superior frontal gyri and dorsolateral prefrontal cortex (DLPFC) ( $p < 0.05$ , family-wise error (FWE) corrected for multiple comparisons; Figure 5-8 and Table 5-3). A further contrast was carried out to identify an effect of entropy on retrieval and comparison, but not encoding. No effect of entropy on retrieval and comparison was demonstrated.

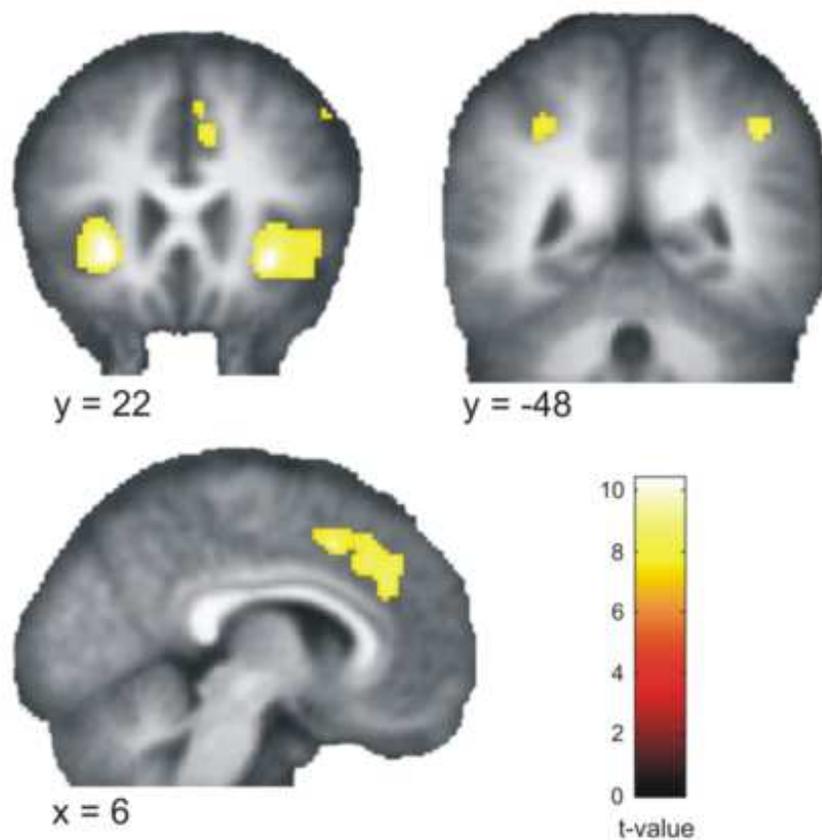


Figure 5-8 Retrieval and comparison results. Areas that show stronger activation ( $p < 0.05$ , FWE corrected) for retrieval and comparison than encoding, rendered on coronal ( $y = 22$  and  $y = -48$ , top left and right, respectively), and sagittal ( $x = 6$ , bottom left) sections of participants' normalised average structural scan. See also Table 5-3 for exact MNI coordinates.

Table 5-3 MNI coordinates of local maxima. Local maxima coordinates for the main effect of Retrieval and comparison (contrast: [second scan – first scan]) at  $p < 0.05$  (FWE corrected for multiple comparisons across the brain). VLPFC, ventrolateral prefrontal cortex; IPS, intraparietal sulcus; mSFG, medial superior frontal gyrus; DLPFC, dorsolateral prefrontal cortex; IFG, inferior frontal gyrus.

Region	Hemisphere	Retrieval - Encoding			
		x	y	z	t-value
mid-VLPFC	left	-28	22	-6	10.24
mid-VLPFC	right	30	22	-8	10.20
		32	26	2	8.70
		44	22	-4	7.64
IPS	right	46	-48	40	8.46
mSFG	medial	6	14	46	8.81
		-8	8	42	8.66
		-6	2	52	8.58
motor cortex	left	-36	-6	62	8.19
		-44	-28	48	7.82
		-48	-20	60	7.48
IPS	left	-30	-44	-40	7.89
		-36	-52	44	6.64
		-30	-56	52	6.50
mid-DLPFC	right	50	20	46	6.99
IFG	right	48	30	24	6.79
thalamus	left	-10	6	2	6.65

## 5.6 Further analysis considerations

For pitch sequences in Study 3 and Study 4 there was a significant effect of exponent  $n$  levels on the sample entropy estimates (ANOVA,  $F_{(5,204)} = 197.814$ , and  $F_{(4,145)} = 28.03$ , both  $p < 0.001$ , respectively). Post-hoc pairwise comparisons revealed that all levels except levels 1 to 3 were statistically different from each other (all  $p < 0.05$ ). The data were therefore analysed further by collapsing across those levels whose sample entropy estimates are not statistically different from each other (levels 1 to 3 in Study 3 and Study 4, resulting in four and three levels, respectively). The results are very similar to the results for both studies as reported in the main text, providing strong support for the original results (Figure 5-9).

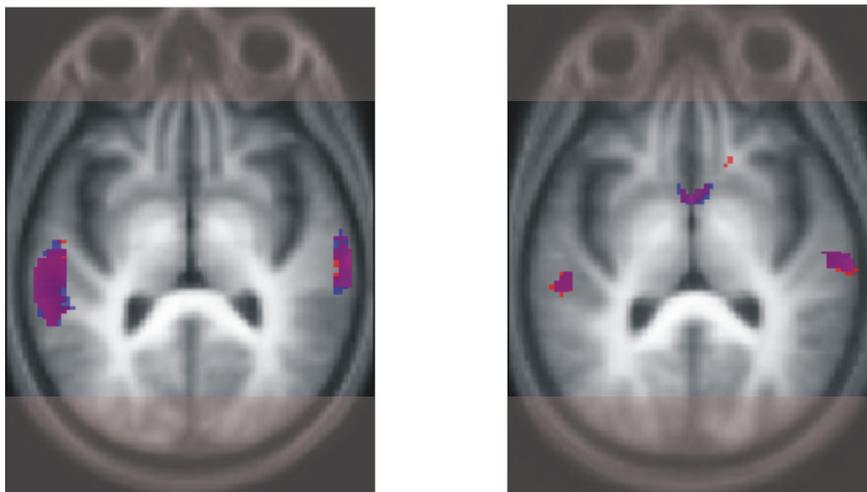


Figure 5-9 Analysis considerations I. (Left) Comparison of results for Study 3 when analysing the data with respect to the original 6 exponent  $n$  levels (red) or collapsing across levels 1 to 3, resulting in a total of 4 levels (blue). (Right) Comparison of results for Study 4 when analysing the data with respect to the original 5 exponent  $n$  levels (red) or collapsing across levels 1 to 3, resulting in a total of 3 levels (blue). Results are thresholded at  $p < 0.001$  (uncorrected) and rendered on the tilted (pitch =  $-0.5$ ) normalised mean structural of the 22 vs. 24 participants.

The data were further analysed by (a) parametrically modulating each individual pitch sequence with its specific sample entropy value and (b) classifying pitch sequences according to their sample entropy values (ignoring the exponent  $n$  value from which they were derived). The second classification method resulted in the following descriptive data for the two studies (Table 5-4).

Table 5-4 Descriptive data for the pitch sequences in the two studies with a classification scheme based on the sample entropy estimates.

level	Study 3			Study 4		
	mean (SEM)	range	# of exempl.	mean (SEM)	range	# of exempl.
1	1.61 (0.03)	$H > 1.5$	23	1.75 (0.04)	$H > 1.5$	40
2	1.35 (0.01)	$1.2 > H \leq 1.5$	87	1.35 (0.01)	$1.2 > H \leq 1.5$	49
3	1.08 (0.01)	$0.9 > H \leq 1.2$	37	1.09 (0.01)	$0.9 > H \leq 1.2$	38
4	0.74 (0.01)	$0.6 > H \leq 0.9$	35	0.73 (0.03)	$H \leq 0.9$	23
5	0.46 (0.01)	$H \leq 0.6$	28			

There was a significant effect of entropy level for Study 3 and Study 4 (ANOVA,  $F_{(4,205)} = 825.51$  and  $F_{(3,146)} = 241.71$ , both  $p < 0.001$ , respectively), and significant pairwise comparisons between all levels (all  $p < 0.05$ ). Importantly, the two methods for grouping the data into levels (i.e. with respect to exponent  $n$  or sample entropy) yielded very similar classifications, as indicated by highly significant Spearman rank correlations for Study 3 and Study 4 ( $\rho = 0.81$ , and  $\rho = 0.64$ , both  $p < 0.001$ , respectively).

In the case of Study 3, the results are almost identical to those reported in the main text, both when parametrically modulating each sequence with respect to its specific entropy value (analysis (a) above), as well as when parameterising using the mean sample entropy value for each of the five levels (analysis (b) above, Figure 5-10, left). In Study 4, the results of both re-analysis techniques are visible as a trend at a reduced significance threshold (Figure 5-10, right). The greater divergence between the original analysis and reanalysis of Study 4 is likely due to the variance of sample entropy estimates for short time series.

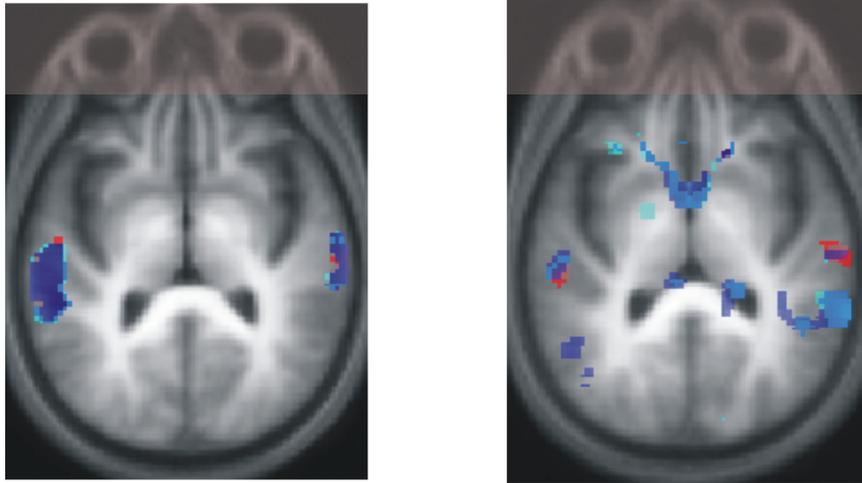


Figure 5-10 Analysis considerations II. Comparison of results for the three types of analyses in the two studies (Study 3, left; Study 4, right). Original analysis based on mean entropy value of the six levels derived from exponent  $n$  (red); analysis based on individual sample entropy value of each pitch sequence (analysis (a), blue); analysis based on categorisation derived from entropy values irrespective of the exponent  $n$  value from which the stimuli were derived (analysis (b), cyan). Results for Study 3 are thresholded at  $p < 0.001$  (uncorrected); results for Study 4 are thresholded at  $p < 0.005$  (red) and  $p < 0.05$  (blue and cyan) and are rendered on the tilted (pitch =  $-0.5$ ) normalised mean structural of the 22 vs. 24 participants.

## 5.7 Discussion

Both studies demonstrated an increase in the local neural activity as a function of the entropy of encoded pitch sequences in PT but not in HG. The results are consistent with a computational process in PT that requires increasing resource and energetic demands during encoding as the entropy of the sound stimulus is increased.

In the third study, the use of pure tones could not exclude a possible alternative explanation of the data in terms of sensory adaptation within cortical frequency representations. The existence of the relationship in PT, but not in HG, was indirect evidence against such sensory adaptation. However, the fourth study used broadband stimuli that continually activate a broad range of cortical frequency representations irrespective of pitch, rendering explanations based on sensory adaptation untenable.

Another interpretation of these results could be based on perceptual adaptation within cortical correlates of pitch (as opposed to sensory adaptation of the stimulus representation). Previous studies have demonstrated mapping of activity within secondary auditory cortex in lateral HG as a correlate of the perceived pitch salience, whether the stimulus mapping was in the temporal domain (Patterson *et al.*, 2002) or frequency domain (Penagos *et al.*, 2004). An explanation of the results of either study might therefore be based on adaptation within the ‘pitch centre’ in lateral HG for pitch sequences with higher fractal exponent  $n$ . The fourth study enabled the identification of separate activations in medial, central and lateral HG. Contrary to an interpretation based on adaptation in pitch sensitive channels, there was no relationship between the entropy and local activity in any of the subfields of HG that would have supported such an explanation. Furthermore, the interaction between HG and PT provides additional evidence for an effect of entropy that is specific to PT.

The most compelling explanation of these results is in terms of greater computational activity (and therefore local synaptic activity and BOLD signal: Goense & Logothetis, 2008; Logothetis *et al.*, 2001) as a function of the information content or entropy of the encoded sound. This is the first explicit demonstration of such a relationship. The results suggest an efficient form of encoding within PT whereby sequences are encoded by a mechanism that demands fewer computational resource for sequences carrying low information content and high redundancy (due to the predictability of the sequence) than that required to encode sequences with little or no redundancy. ‘Sparse’ (DeWeese & Zador, 2006; Friston, 2003c; Olshausen & Field, 2004) and ‘predictive’ (Baldeweg, 2006; Friston, 2005; von Kriegstein & Giraud, 2006) coding both constitute such mechanisms and bases for PT acting as a ‘computational hub’ (Griffiths & Warren, 2002).

In contrast, retrieval and comparison do not depend on entropy in the same way, which is proposed to reflect the decreased computational and energetic demands of retrieving and comparing stimuli at symbolic levels beyond stimulus encoding. The initial encoding process depends on a computationally expensive process that must abstract features from a complex spectrotemporal structure. Beyond this stage, the subsequent categorical retrieval and comparison mechanism does not depend on the detailed spectrotemporal structure. Indeed, the ‘computational hub’ model (Griffiths

& Warren, 2002) states that PT gates its output towards higher-order cortical areas that perform analyses at a symbolic and semantic level. It is suggested that at least part of the function of PT is to compress the neural code corresponding to the initial acoustic signal (e.g. via sparse or predictive coding), and that subsequent processing is not dependent on stimulus entropy.

That PT might even perform this type of analysis in more general or supramodal terms is suggested by work in the visual domain (Bischoff-Grethe *et al.*, 2000) demonstrating activation in Wernicke's area and its right-hemisphere homologue as a function of the entropy within a sequence of visually presented squares, irrespective of whether or not participants were aware of an underlying sequence. However, later studies using similar visual stimuli did not replicate this finding (Harrison *et al.*, 2006; Strange *et al.*, 2005).

The retrieval and comparison phase highlighted a fronto-parietal network consisting of the anterior insulae and frontal opercula, inferior parietal sulci, medial superior frontal cortex and DLPFC. This activation pattern is common in the retrieval and comparison phase of (auditory) delayed match-to-sample tasks (e.g. Arnott *et al.*, 2005; Zatorre *et al.*, 1998). In particular, the caudal part of mid-VLPFC is engaged in the top-down control of active retrieval processes (Kostopoulos & Petrides, 2008). In functional imaging studies, the mid-VLPFC is difficult to disambiguate from the anterior insula, which has been proposed as an additional auditory processing centre for allocating auditory attention, specifically with respect to sound sequences (for a review see Bamiou *et al.*, 2003). However, cytoarchitectonically, the granular prefrontal cortex can be readily disambiguated from the more dysgranular insula as part of the limbic cortex (Petrides & Pandya, 1994, 2002). Parietal cortex is generally regarded as being important for attention to and binding of sensory information (Cusack, 2005), while activity in prefrontal cortex is often associated with response preparation and selection (Passingham & Sakai, 2004).

While the main aim was to study generic neural mechanisms of sound encoding as a function of entropy, the range of pitch sequences used included those approximating  $f^{-1}$  ('one-over-f') power spectra, which resemble many naturally occurring acoustic phenomena (de Coensel *et al.*, 2003). Notably, music and speech display  $f^{-1}$  power spectra characteristics, reflecting the relative balance of 'surprises'

(e.g. musical transitions) and predictability in such signals (Voss & Clarke, 1975, 1978). Pertaining specifically to the signals used here falling in the range of  $f^{-1}$ , two recent electrophysiological studies demonstrated preference within primary sensory cortices for  $f^{-1}$  signals (Garcia-Lazaro *et al.*, 2006; Yu *et al.*, 2005). The data did not demonstrate any ‘tuning’ to particular values of exponent in HG (no main effect of Entropy level; Figure 5-2, Figure 5-5, and Figure 5-6). While a neuronal preference for particular natural sequence categories at the level of HG in humans is conceivable, the current studies addressed the computational and energetic demands of the perceptual encoding of sounds, rather than their sensory representation.

These studies used entropy to characterise pitch sequences here, but the information theoretic approach could be used to characterise sequences containing rhythm, or more complex natural sound sequences. The hypothesised mechanism in PT is not a specific pitch mechanism and also predicts a similar relationship between information content and the encoding of more natural stimuli. In summary, the present data implicate PT as a neural engine within which the computational and energetic demands of encoding are determined by the entropy of the acoustic signal.

## **Chapter 6. HIERARCHICAL ENCODING OF GLOBAL AND LOCAL INFORMATION IN PITCH PATTERNS**

### **Summary**

Pitch patterns, such as melodies, consist of two structural levels: a global level, comprising the pattern of ups and downs, or contour; and a local level, comprising the precise intervals that make up this contour. An influential neuropsychological model suggests that these two levels of processing are hierarchically linked, with processing of the global structure occurring within the right hemisphere in advance of local processing within the left. However, the predictions of this model and its anatomical basis have not been tested in neurologically normal individuals. Study 5 used fMRI and required participants to listen to consecutive pitch sequences while performing a same/different one-back task. Sequences, when different, either preserved (local) or violated (global) the contour of the sequence preceding them. When the activations for the local and global conditions were contrasted directly, additional activation was seen for local processing in right planum temporale and posterior superior temporal sulcus (pSTS). The presence of additional activation for local over global processing supports the hierarchical view that the global structure of a pitch sequence acts as a "framework" on which the local detail is subsequently hung. However, the lateralisation of activation seen in the present study, with global processing occurring in left pSTS and local processing occurring bilaterally, differed from that predicted by the neuropsychological model. A re-examination of the individual lesion data on which the neuropsychological model is based revealed that the lesion data equally well support the laterality scheme suggested by the current data. While Study 5 supports the hierarchical view of local and global processing, there is an evident need for further research, both in patients and neurologically normal individuals, before an understanding of the functional lateralisation of local and global processing can be considered established.

## 6.1 Introduction

Cognitive neuropsychological studies have demonstrated that pitch patterns, such as melodies, consist of two structural levels: the contour or pattern of ups and downs – synonymous with the ‘global’ level; and the precise intervals that make up this contour – synonymous with the ‘local’ level. Early behavioural support for this hierarchical model came from same/different tasks in which pairs of novel pitch sequences could differ at a local level, where contour is preserved, or at a global level, where the overall contour is violated (Dowling, 1978; Dowling & Fujitani, 1971; Dowling *et al.*, 1987). Individuals can reach high levels of accuracy in the detection of both types of change. However, if the second sequence is shifted in overall pitch, individuals are unable to detect differences where the contour is preserved. The dependence of participants’ accuracy on the presence or absence of a change in contour suggests that processing of contour provides a ‘scaffold’ on which the detail of the precise intervals are subsequently ‘hung’ (for further behavioural evidence of this model, see Bartlett & Dowling, 1980; Cuddy & Cohen, 1976; Trehub *et al.*, 1993).

Evidence for the neuroanatomical basis of this model has come from patient studies. Peretz (1990) tested patients with heterogeneous left or right hemispheric damage (LHD or RHD, respectively) on tasks similar to those described above. Deficits in the detection of differences involving a contour violation always co-existed with deficits in the detection of differences where the contour was preserved. In contrast, selective deficits in discriminating melodies that shared the same contour were seen without accompanying deficits in discriminating melodies that differed in contour. Moreover, this pattern was associated with damage to different hemispheres: RHD patients were worse than normal control (NC) participants for the detection of both types of differences, while LHD patients performed significantly better for contour-violated than contour-preserved differences.

A similar pattern of results was found by Liégeois-Chauvel and colleagues (1998) in patients with lesions confined to the temporal lobes. Lesions to right posterior temporal cortex were associated with deficits in the detection of contour-preserved and contour-violated differences, while lesions to left posterior temporal cortex were associated with selective impairments for the detection of differences

where the contour was preserved. Taken together, this pattern of results suggests a model of hierarchical co-operativity whereby contour processing precedes interval processing and these two stages of the hierarchy are right and left lateralised in posterior superior temporal cortex. However, in a study similar to Peretz (1990), Schuppert and colleagues (2000) confirmed the notion of a processing hierarchy in patients with heterogeneous cortical lesions, but the pattern of deficits did not support the proposed lateralisation dichotomy of global–right and local–left. Similarly, Zatorre (1985) found no statistically significant differences for detecting contour-preserved or contour-violated changes in simple melodies between patients with left or right temporal lobe lesions.

The neuropsychological approach in patients with brain lesions is of clear value in establishing the necessity of brain areas for given functions. However, several aspects of the approach caution against a sole reliance on lesion data to derive neuroanatomical models of cognitive processing. Brain lesions are rarely circumscribed, are heterogeneous across different patients, and may be functionally compensated for by other brain areas with a time-course that differs across patients. All these factors make assessment and interpretation of deficits challenging. Further, brain lesions occur within functional networks and particular damaged regions may not be sufficient in and of themselves to support the function, which may depend equally on other regions within a broader network. Functional imaging offers a valuable complement to the neuropsychological approach, providing a way to highlight the network of areas associated with the normal performance of a given function. The two approaches, when used in combination, provide a useful constraint on the interpretation of results and the formulation of new theories.

The present study used fMRI to test the model of Peretz and colleagues (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990) in neurologically normal individuals. The paradigm was modelled on the same/different tasks used in behavioural (Bartlett & Dowling, 1980; Cuddy & Cohen, 1976; Dowling, 1978; Trehub *et al.*, 1993) and patient studies (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990), but adapted to a one-back format which was more suited to fMRI. In addition, pitches were used that were drawn from a non-musical scale, so that findings could be generalised outside the purely musical domain (Foxton *et al.*, 2003; Patel & Iversen, 2008). The two

predictions arising from the model were tested: that the processing of pitch sequences involves a hierarchy (from a global to a local level) and that a different degree of lateralisation is seen for each of these stages (global – right; local – left).

## **6.2 Materials and Methods**

### **6.2.1 Participants**

Twenty four participants were recruited for the study. All participants (10 male, 14 female) reported an absence of any hearing or neurological disorder and gave their informed consent. The study was approved by the National NHS Research Ethics Committee.

### **6.2.2 Stimuli and Experimental Procedure**

Since it was the intention to investigate local and global levels of auditory processing at a generic level, and not only in music, the pitches were drawn from a set of frequencies that does not typically appear in combination in the Western musical tradition. Ten pitches, equally spaced in logarithmic steps, were taken from a two-octave range (120-480 Hz) (see also Chapter 5). Each pitch corresponded to a series of 30 harmonics with a trapezoidal spectral envelope, and a rise and decay time of 20 ms and 30 ms, respectively. Sounds were created digitally at a 44.1 kHz sampling rate and 16 bit resolution using Matlab ([www.mathworks.com](http://www.mathworks.com)). A pitch sequence consisted of four 300 ms long pitches, amounting to a duration of 1.2 seconds per sequence. Each trial was made up of four pitch sequences separated by an inter-sequence interval of 800 ms. There were two experimental trial types: Local and Global (Figure 6-1). For both local and global trials, consecutive pitch sequences were the same (Lsame or Gsame) or different (Ldiff or Gdiff), with equal probability. In the Local trials, consecutive sequences, when different, had a pitch change at either the second or third element of the sequence with the constraint that this change did not alter the contour. Correct performance depended on perceiving a difference in the

exact pitches or intervals in the two sequences. In the Global trials, consecutive sequences, when different, contained a pitch change brought about by reversing the order of the second and third elements, which always resulted in a difference in contour. Correct performance depended upon the perception of a difference in contour, in addition to any difference in the exact pitches or intervals in the two sequences. Participants performed a one-back task by pressing a key beneath their index or middle finger to indicate that the current sequence was either the same or different to the previous. Participants were trained on each trial type outside the scanner, to a criterion level of 80%. During scanning, their performance was recorded and analyzed off-line for accuracy. In addition to Local and Global trials, there were also Silent trials comprising a period of silence matched to the duration of the other trial types. Participants performed two experimental sessions in which the three trial types: Local, Global and Silence were pseudo-randomly intermixed, with 64 instances for each of the two sessions.

### 6.2.3 fMRI protocol and analysis

Gradient weighted echo planar images (EPI) (see Section 2.3.1) were acquired on a 1.5 Tesla Siemens Sonata system (Erlangen, Germany) using a sparse imaging protocol (repetition time 12.5 seconds), in order to temporally separate the scanner noise and the experimental sounds (Edmister *et al.*, 1999; Hall *et al.*, 1999) (see Section 2.4). A total of 48, 4mm axial slices were acquired, with an in-plane resolution of 3 x 3 mm. 192 brain volumes were acquired for each participant across the two sessions (64 for each condition). A high resolution T1 weighted structural image (1 x 1 x 1.5mm) was also obtained. During scanning, stimuli were presented using Cogent ([www.vislab.ucl.ac.uk/Cogent](http://www.vislab.ucl.ac.uk/Cogent)) and delivered via an external sound card ([www.edirol.com](http://www.edirol.com)) at a sound pressure level of 70 dB over a custom built electrostatic system based on Koss<sup>TM</sup> headphones.

Imaging data were processed and analysed using Statistical Parametric Mapping software (SPM5, [www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) (see also Sections 2.5-2.6). Scans from each participant were realigned to the first image of the time series and unwarped, spatially normalised to standard stereotactic space (Friston *et al.*, 1995a) and smoothed with an isotropic Gaussian kernel of 10 mm full-width-at-half-maximum.

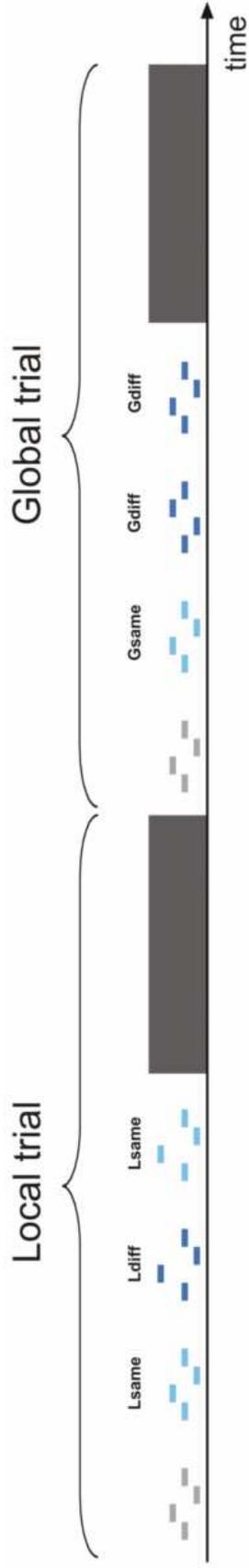


Figure 6-1 Experimental design. Schematic of two consecutive trials. Light blue indicates a pitch sequence that is identical to the previous sequence, dark blue indicates a pitch sequence that is different from the previous sequence; the first pitch sequence (grey) is neither the same nor different since there is no preceding pitch sequence. The scan period at the end of the trial is depicted in dark grey. Pitch sequences were 1200 ms long and separated by 800 ms gaps. Participants performed a one-back task, indicating whether a pitch sequence was same/different from the previous pitch sequence.

Population-level inferences were made through a two-stage procedure. First, the data from each participant were analysed within the context of the general linear model (Friston *et al.*, 1995b). Pitch sequences were categorised according to condition: Local Same (Lsame), Local Different (Ldiff), Global Same (Gsame) and Global Different (Gdiff) (Figure 6-1). Note that Lsame and Gsame sequences were identical and that the only difference was the context in which they were presented, either in a Local or a Global trial. Hence they were modelled separately to take account of potential ‘cognitive set’ effects. Each sequence was modelled as a short event of 1.2 seconds duration and was convolved with a haemodynamic response function (Figure 6-2). The first sequence of each trial was not modelled explicitly, since it was neither the same nor different. This approach explicitly models variance due to whether a given pitch sequence was same or different. From this model, parameter estimates for each condition were derived. Planned contrasts were applied to assess differences in activation between the conditions, resulting in a contrast image. These contrast images were used in a second level random effects analysis. For each contrast of interest, a one-sample t-test was performed to derive statistical parametric maps (SPMs). Since the focus was on areas with an *a priori* prediction, i.e. in auditory cortex, SPMs were thresholded at  $p < 0.001$  (uncorrected for multiple comparisons across the brain).

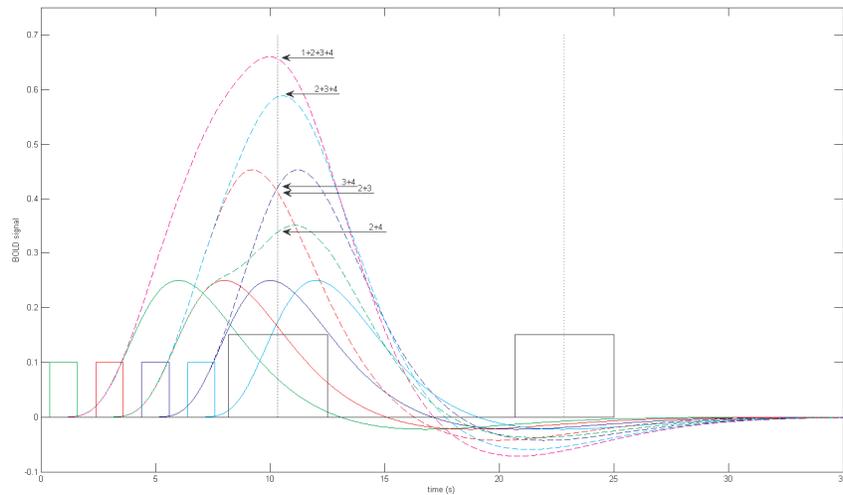


Figure 6-2 Schematic of the experimental design and modelled haemodynamic response functions (hrf's). The figure is similar to Figure 6-1; the four pitch sequences are depicted as four stimulus blocks (green, red, blue, cyan; respectively), while the two black boxes represent the time of the volume acquisition. Various haemodynamic response functions for the different permutations are plotted. For example, a trial in which the second and fourth sequence were *Ldiff* sequences, would lead to the green dashed hrf (labelled '2+4'). The dashed vertical line represents the onset of the reference time window at which the BOLD signal is sampled.

## 6.3 Results

### 6.3.1 Behavioural results

Twenty-four neurologically normal participants underwent behavioural testing. Four participants who showed a difference in accuracy of more than 10% between the Local and Global conditions were excluded, to avoid confounding the interpretation of the imaging findings by differential performance between the Local and Global conditions. Mean correct performance in the scanner for these 20 participants did not differ between Local (91.98%) and Global (93.15%) conditions (paired samples t-test,  $t_{19} = -1.17$ ,  $p > 0.1$ ) and was significantly above chance (50%) (one-sample t-test,  $t_{19} = 28.07$  and  $t_{19} = 39.12$ , both  $p < 0.001$ , for Local and Global conditions, respectively).

### 6.3.2 Effects of processing contour-preserved and contour-violated differences

Activation for *Lsame* and *Gsame* sequences did not differ from another, ruling out a potential ‘cognitive set’ effect, and these two conditions were therefore pooled as *Same*. In order to assess separately areas that are involved in the processing of local differences and global differences, the following contrasts were performed; *Local*: ( $[Ldiff - Same]$ ) and *Global*: ( $[Gdiff - Same]$ ). *Local* revealed bilateral activation in pSTS, while *Global* was lateralised to the left pSTS, even at a reduced statistical threshold of  $p < 0.05$  (Figure 6-3; see also Figure 6-4 and Table 1).

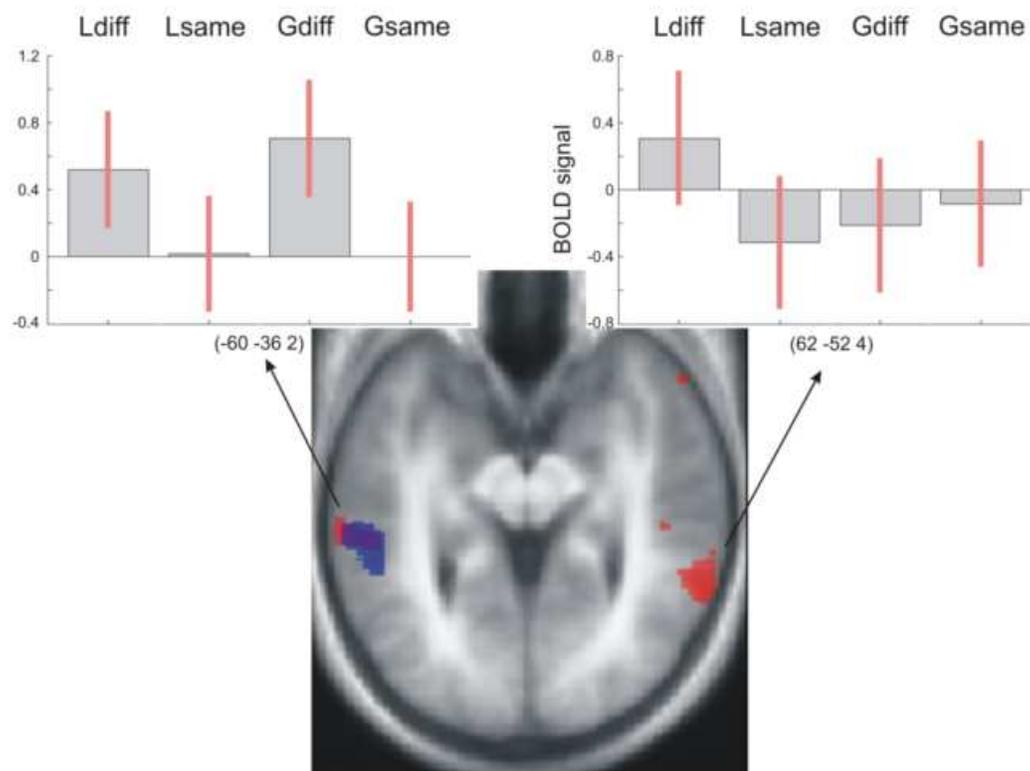


Figure 6-3 Main effects for *Local* and *Global*. Activations for the *Local* ( $[Ldiff - Same]$ ) (red) and *Global* ( $[Gdiff - Same]$ ) (blue) contrasts superimposed on a tilted (pitch:  $-0.5$  radians) normalised average structural scan covering STS. Activations are thresholded at  $p < 0.005$  (uncorrected), for display purposes. Plots show the BOLD signal at local maxima in left and right pSTS. See also Figure 6-4.

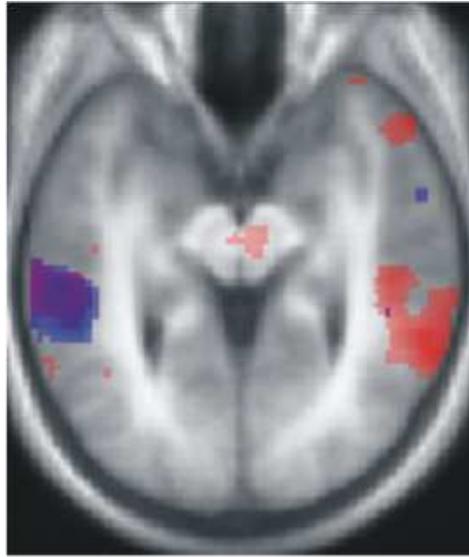


Figure 6-4 Main effects for *Local* and *Global* ( $p < 0.05$ ). Same as Figure 6-3, thresholded at  $p < 0.05$ , uncorrected for multiple comparisons. Note that *Global* does not reveal any activation in right pSTS.

Table 6-1 MNI coordinates of local maxima. Stereotactic coordinates for the three contrasts *Local*, *Global*, and *Local - Global*.

Contrast	x	y	z	t-value
<i>Local</i> [Ldiff - Same]	62	-52	4	4.31
	-68	-36	-8	4.00
	-68	-40	-8	3.92
	-58	-36	0	3.65
<i>Global</i> [Gdiff - Same]	-60	-36	-2	3.94
	-54	-44	0	3.48
<i>Local - Global</i> [Ldiff - Gdiff]	68	-46	4	5.03
	60	-30	-2	4.98
	62	-22	8	4.75

In order to test whether the activation patterns for these contrasts (*Local* and *Global*) were significantly lateralised, formal tests of lateralisation were performed (see also Section 3.2.4). A set of the original realigned and unwarped images and a set of ‘flipped’ left-right realigned and unwarped images were normalised to a symmetrical

template and subsequently smoothed before the statistical analysis. To test for statistical differences between the left and right hemispheres for each contrast (*Local* and *Global*), a voxel-by-voxel pairwise t-test between the original and the flipped images was then performed.

These tests of lateralisation confirmed that no areas showed any lateralisation for *Local*, while *Global* was significantly lateralised to the left pSTS (Figure 6-5).

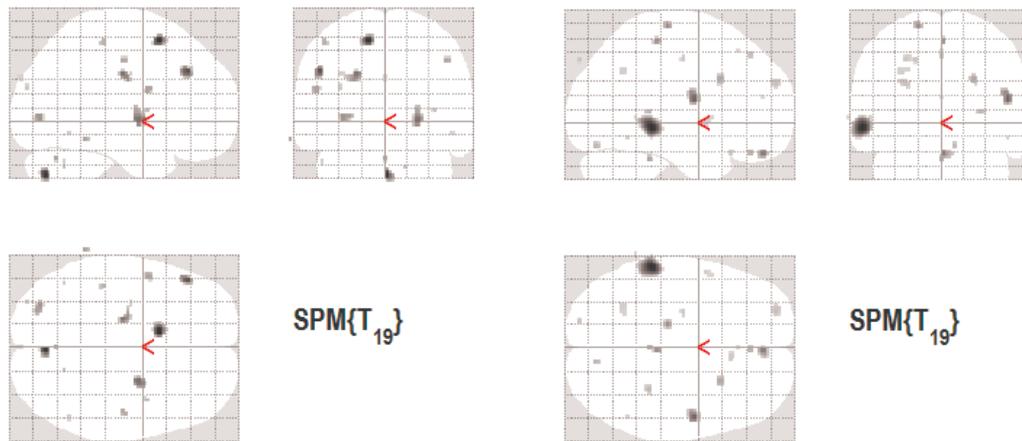


Figure 6-5 Results for the lateralisation test of (left) *Local* ([Ldiff – Same]) for original – flipped scans, and (right) *Global* ([Gdiff – Same]) for original – flipped scans.

### 6.3.3 Comparison of Local and Global Processing

It was examined whether the processing of local differences (contour-preserved) versus global differences (contour-violated) resulted in a distinct activation pattern via the contrasts *Local – Global* ([Ldiff – Gdiff]) and *Global – Local* ([Gdiff – Ldiff]). These contrasts directly compared activations corresponding to the detection of a contour-preserved difference versus a contour-violated difference and allowed to test for a hierarchical relation between these two processes. *Local – Global* revealed activations in the pSTS and planum temporale (PT) on the right, while there were no significant differences for the *Global – Local* contrast (Figure 6-6, see also Table 1).

A formal test of lateralisation confirmed these findings, showing right lateralised activations in pSTS and PT for *Local – Global* (Figure 6-7, see also Figure 6-4).

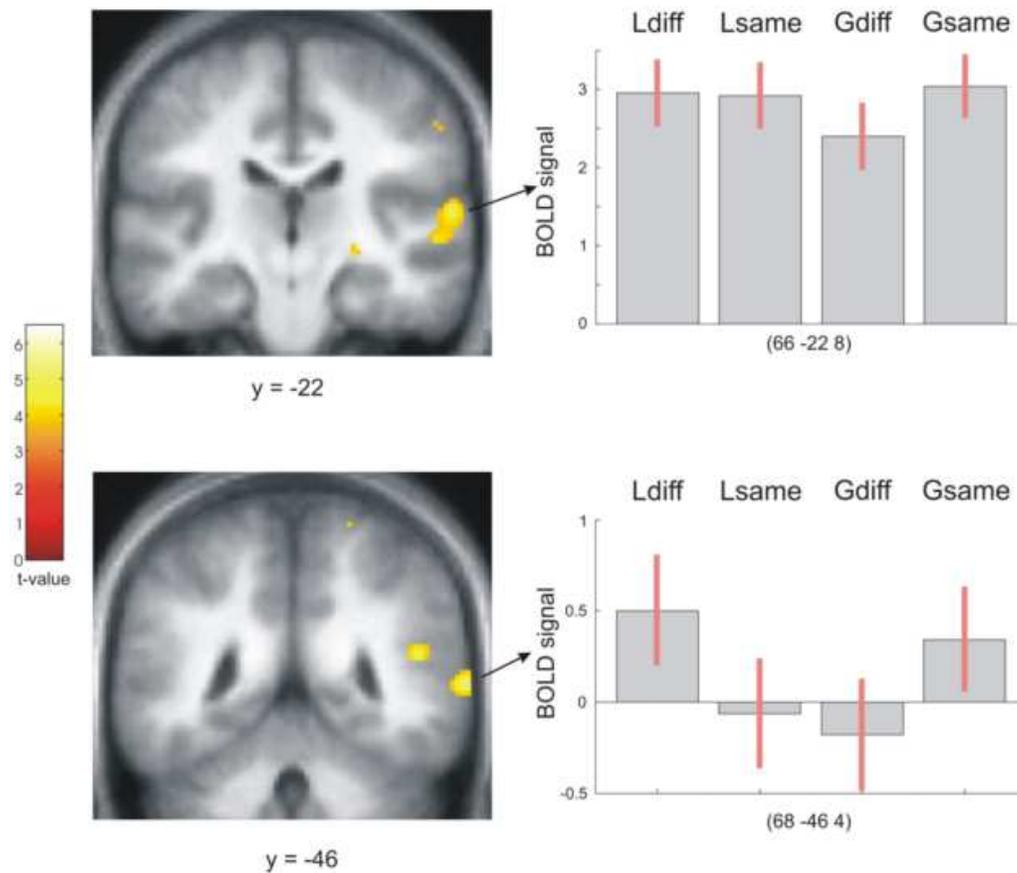


Figure 6-6 Activations for the *Local – Global* ([Ldiff – Gdiff]) contrast. The results are superimposed on coronal sections of participants' normalised average structural scan. Plots show the BOLD signal at local maxima in right PT (top right) and pSTS (bottom right).

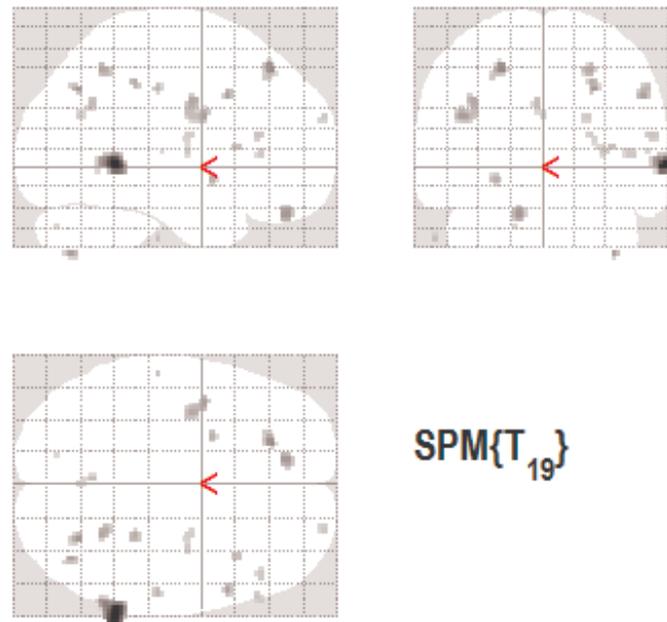


Figure 6-7 Results for the lateralisation test of *Local - Global* ([Ldiff - Gdiff]) for original - flipped scans.

## 6.4 Discussion

The aim of this study was to test both aspects of the model put forward by Peretz and colleagues (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990) which holds that the processing of pitch sequences involves a hierarchy (from global processing to local processing) and differential hemispheric lateralisation of these stages (global - right; local - left). The results of the present study confirm the hierarchy predicted by the model: a direct comparison of activation for the detection of a contour-preserved versus a contour-violated difference revealed greater activation for processing contour-preserved differences. No areas were more activated for processing of a contour-violated difference compared with a contour-preserved difference. The presence of additional activation for contour-preserved differences over and above those for contour-violated differences is consistent with a processing hierarchy in which local processing requires additional neural resources compared with global processing. However, these results contrast with the lateralisation account of the

model proposed by Peretz and colleagues (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990): rather than demonstrating an association of global and local processing with the right and left hemispheres respectively, processing of change at the global level was lateralised to the left posterior STS, while processing of change at the local level engaged bilateral posterior STS. The location of these activations is congruent with results in Liégeois-Chauvel and colleagues (1998), where damage to the posterior part of the superior temporal lobe (STL) was more detrimental for performance than anterior STL damage.

The processing hierarchy demonstrated here accords with cognitive neuropsychological and lesion-based evidence, and can be conceptualised as a fast serial search strategy whereby the first pitch sequence is encoded and provides a reference for the comparison of each of the constituent events of the second sequence. In such a scheme, incoming events are compared with the corresponding event in the first sequence, initially for contour direction (global) and then for the precise interval information (local). If a difference is detected in contour, the search is terminated, otherwise the search process continues at the interval level. While the temporal resolution of fMRI is insufficient to provide direct support for this serial model, data including faster reaction times as well as earlier and greater event-related potentials to contour violations compared with contour-preserved differences provide strong evidence for such a serial search strategy (Schiavetto *et al.*, 1999; Trainor *et al.*, 1999; Trainor *et al.*, 2002).

The results concerning hemispheric lateralisation of local and global processing are at first more difficult to reconcile with lesion data and the predictions of the model by Peretz and colleagues which suggest a pattern of laterality such that local processing occurs within the left hemisphere and global processing within the right (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990). However, as mentioned in the Introduction to Study 5 (Section 6.1), other neuropsychological studies (Schuppert *et al.*, 2000; Zatorre, 1985) did not support this hemispheric lateralisation account. Further, a close examination of the neuropsychological studies to date urges a more circumspect interpretation. Two of these studies (Peretz, 1990; Schuppert *et al.*, 2000) used unconventional cut-offs for defining impaired performance (the worst score and the mean score of the normal control (NC) groups, respectively), increasing the

likelihood of false positive results. In the study of Peretz (1990), although at a group level there was a pattern of deficits suggestive of a right –global; left – local dissociation, only five out of ten of the RHD patients had genuine global deficits (performance below cut-off), and only three out of ten of the LHD patients had genuine selective local deficits. Equally, in Liégeois-Chauvel and colleagues (1998), where lesion locations were confined to temporal cortex, three out of five patients with damage to right posterior temporal cortex had global deficits and one out of three patients with damage to left posterior temporal cortex had selective local deficits. Taken together, this more detailed picture suggest that the lateralisation scheme proposed by Peretz and colleagues (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990) can only partially account for the pattern of deficits observed in these patients.

It is suggest that the processing scheme suggested by the current data (global – left; local – bilateral) can account equally well for the pattern of results reported in previously published neuropsychological cases (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990; Schuppert *et al.*, 2000; Zatorre, 1985). For example, in Liégeois-Chauvel and colleagues (1998), two out of five cases with right posterior temporal lesions showed either no deficit for local and global tasks or selective deficits in the local task alone, while two out of three patients with left posterior temporal cortex lesions were below cut-off for both local and global tasks. Furthermore, while LHD patients in Peretz (1990) were better at global than local tasks, they nevertheless performed significantly worse than NC on both tasks.

The concept of local and global signal properties and the notion of their respective hemispheric lateralisation is not unique to pitch sequences or melodies, but also has a long tradition in vision research (for reviews, see Hellige, 1996; Hübner & Volberg, 2005). The defining characteristics of local and global stimulus properties as used in vision (e.g. Martin, 1979; Navon, 1977) are that (i) one is embedded within the other (local within global) and that (ii) both are orthogonal. In this way, local stimulus properties can be arranged to yield global stimulus properties, while both can be manipulated independently. In vision, the distinct processing of local and global stimulus properties likely reflects different spatial frequency filters (Robertson, 1996; Shulman *et al.*, 1986).

Recently, the representation of local and global stimulus properties as in Study 5 has been questioned for not reflecting the characteristic of orthogonality between local and global stimulus properties (Justus & List, 2005; List & Justus, 2007; List *et al.*, 2007; Sanders & Poeppel, 2007). These authors argue that a global violation in a pitch pattern also entails a local violation; this argument assumes that a pitch interval is not represented in terms of its absolute size, but rather together with its direction (e.g. a major third would be represented as an ascending major third or descending major third and not as a major third irrespective of its direction).

The original studies (Dowling, 1978; Dowling & Fujitani, 1971; Dowling *et al.*, 1987) conceived of the ‘ups’ and ‘downs’ of the global contour of melodies in terms of ‘+’ and ‘-’, while the intervals were denoted in semitones with absolute integers. For example, a simple melody such as *Frère Jacques* could then be described at the global level as ‘+ + - 0 + + -’, and at the local level as ‘2 2 4 0 2 2 4’; in this conceptualisation, the global structure can indeed be modified independently from the local level simply by substituting ‘+’ and ‘-’. However, previous studies (including Study 5) did not control the absolute interval size (i.e. presumably maintaining the same local level) when introducing a global violation. Furthermore, a ‘+2’ (ascending major second) pitch interval and a ‘-2’ (descending major second) interval are unlikely to be perceived as similar (Russo & Thompson, 2005). In contrast, with respect to music theory, it is the relative inversion of intervals rather than the absolute inversion that relates musical intervals: for example, an ascending major third is more closely related to its inverse, a descending minor sixth, than to a descending major third (Hindemith, 1940; Schönberg, 1911).

The current definition of local and global stimulus properties was originally motivated to capture two aspects of musical, or melody, information processing (Dowling, 1978; Dowling & Fujitani, 1971; Dowling *et al.*, 1987). In fact, the early studies generally referred to the two levels as ‘contour’ and ‘pitch’ (or ‘pitch interval’), instead of local and global levels. It is conceivable that the conceptualisation of local and global stimulus properties as applied in these studies is specific to structural information in the musical domain instead of in the auditory domain in general (however, see Foxton *et al.*, 2003; Patel & Iversen, 2008). In this case, a terminology using ‘contour’ and ‘pitch interval’ might be more appropriate, so

as to avoid possibly misleading comparisons with local and global stimulus properties used in vision research. While Study 5 based its stimulus manipulations on previous approaches derived from musical melodies (Dowling, 1978; Dowling & Fujitani, 1971; Dowling *et al.*, 1987), the 10-split octave and relatively short and generic pitch patterns (four pitches per pitch sequence) used in the study attempted to afford a more general interpretation of the results. For example, global and local levels in prosodic patterns have also been shown to be important aspects of speech perception (Foxton *et al.*, 2003; Patel & Iversen, 2008). It is possible, however, that the use of a 10-split octave and the one-back task design could account for the differences from the model put forward by Peretz and colleagues (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990), which is based on longer melodies and individual pairwise comparisons. Further studies are required to determine to what extent these experimental design manipulations might account for the different lateralisation results.

On a methodological note, the experimental design of this study (one-back task) allowed the presentation of two different stimulus types (*Lsame* and *Ldiff*, or *Gsame* and *Gdiff*) within one TR of a sparse imaging acquisition protocol (Figure 6-1 and Figure 6-2). This is the first time this approach has been implemented and it should lead to important advances and improved flexibility for sparse imaging protocols. Critically, the experimental design made it possible to investigate *Local* and *Global* processing as such (i.e. compared to a neutral baseline, *Same*), as opposed to with respect to each other (*Ldiff* versus *Gdiff*). This is an important advancement, since previous investigations have only been able to directly compare local versus global violations, but not their individual contributions.

Thus, while studies investigating local and global levels of auditory processing have generally confirmed the hierarchical account, evidence for hemispheric lateralisation of these levels has been more diverse and elusive (Fujioka *et al.*, 2004; Liégeois-Chauvel *et al.*, 1998; Peretz, 1990; Schiavetto *et al.*, 1999; Schuppert *et al.*, 2000; Trainor *et al.*, 1999; Trainor *et al.*, 2002). Clearly, further research using complementary experimental approaches and techniques is needed to refine the question of a lateralised hierarchy and to determine which parameters are relevant in driving the effect. In particular, there is a need for functional imaging studies of

patients with focal brain lesions to examine directly the distribution of processing following brain damage.

In conclusion, the present study is the first to demonstrate the neural bases of local and global levels of processing in pitch patterns in neurologically normal participants. The results show that local and global processing within pitch sequences differentially engage substrates in the posterior STS and that additional neural resources are required in the right posterior STS and PT for local pitch change processing. The findings support the notion of a pitch pattern processing hierarchy that is likely to be generic rather than specific to music. Furthermore, the data suggest an alternative lateralisation scheme at these two levels of analysis which, while different to the traditionally held view, is equally consistent with the neuropsychological data from which this previous model is derived. The present study urges caution in accepting the traditional view of lateralisation, based on neuropsychological studies of local and global pitch sequence processing, and emphasizes the need for further research, both with patients and neurologically normal individuals, before an understanding of the functional lateralisation of local and global pitch sequence processing can be considered established.

## Chapter 7. GENERAL DISCUSSION

This thesis investigated cortical correlates of processing higher-order statistical properties in complex acoustic signals. Specifically, it addressed fundamental cortical mechanisms for representing and segregating individual auditory objects, as well as grouped objects or object streams, by parametrically controlling higher-order statistical properties of acoustic signals in order to characterise auditory objects. The parametric nature of the synthesised stimuli, together with an information theoretic approach, enabled a detailed investigation of generic processes underlying auditory perception in complex and dynamically changing acoustic environments.

In particular, Study 1 (Chapter 3) assessed different temporal integration windows in auditory cortex by parametrically varying the spectrotemporal correlation in complex acoustic spectra. The results demonstrated increasing activity in auditory association cortex and a right-hemispheric lateralisation in STS as a function of spectrotemporal correlation, or, equivalently, increasing time windows.

Study 2 (Chapter 4) investigated distinct mechanisms for segregating and representing auditory objects by parametrically varying the spectrotemporal coherence of complex ‘auditory textures’. The results revealed a cortical processing hierarchy, in which primary and association areas detect statistical transitions at object boundaries, while the subsequent precise representation of the object properties occurs only later in auditory association cortex.

Studies 3 and 4 (Chapter 5) took an information theoretic approach to auditory encoding. By parametrically varying the entropy in complex pitch sequences, it was shown that the planum temporale of human auditory association cortex acts as a ‘computational hub’ (Griffiths & Warren, 2002), in which the encoding of stimuli with redundant information requires fewer computational resources than the encoding of those with high information content. Further, the results suggest a ‘sparse’ or ‘predictive’ coding scheme in PT that compresses the neural code such that subsequent stages of processing (e.g. the retrieval of pitch sequences) are independent of information entropy.

Finally, the fifth study (Chapter 6) investigated the cortical representation of two fundamental computational mechanisms for processing pitch transitions: encoding the direction (up/down) and precise size of intervals. The respective global and local levels (Dowling, 1978; Dowling & Fujitani, 1971; Dowling *et al.*, 1987) of pitch pattern processing were shown to be hierarchically linked, such that processing of local information requires additional resources over and above those required for processing global information.

The information theoretic approach employed in this thesis to control systematically statistical properties of acoustic signals conceptualises the brain as a dynamic system, which is constantly producing and testing hypotheses so as to optimise its coding (Friston, 2003a, 2005). Within this framework, cortical mechanisms compare the sensory input with pre-existing templates or priors and optimise or update the priors if they do not align with the sensory input (reducing the ‘prediction error’ between sensory information and priors). The hypothesis put forward here is that the statistical properties of generic acoustic signals provide a critical means by which the auditory system encodes the signal and optimises its coding based on such Bayesian principles. While the stimuli employed in this thesis used abstract sounds to characterise auditory objects or object streams, it is hypothesised that the underlying mechanisms are generic, and that they are applied to a variety of sound types, including those with semantic associations (e.g. voices or phonemes, musical instruments, environmental sounds, to name just a few). Indeed, it is argued that semantic associations are the result of repeated associations of certain spectrotemporal characteristics or statistical signal properties and thus represent special auditory objects that arise from these principles.

The following Sections 7.1–7.4 discuss the main implications and directions for future research raised by each of the five studies. Finally, a concluding section (Section 7.5) outlines a possible framework for processing statistical properties in a complex environment, informed by the results of this thesis.

## **7.1 The length of analysis time windows in auditory cortex increases from primary to association cortex**

Acoustic signals evolve over time and thus the auditory system has adapted to process information over a range of temporal windows simultaneously. The first study of this thesis (Chapter 3) investigated anatomically distinct representations for different time windows by parametrically controlling the spectrotemporal correlation in complex acoustic spectra. Such a stimulus manipulation requires the auditory system to apply different temporal windows so as to integrate the higher-order statistical properties across multiple frequency bands.

An important advantage of the approach taken in Study 1 is the explicit control of spectrotemporal correlation in a stimulus whose complexity approaches that of ethological sounds. Previous investigations (Boemio *et al.*, 2005; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001) manipulated the segment length in multi-segment sounds and thus introduced an arguably more artificial and more predictable stimulus manipulation. In contrast, the stimulus in Study 1 controlled the spectrotemporal correlation of the spectrum as a whole, mirroring continuous processes in naturally occurring sounds (instead of introducing abrupt segment boundaries). The different time windows within the stimulus emerged as a function of spectrotemporal correlation across multiple frequency bands; this is a fundamental property of natural sounds, in which the physical attributes of the sound-producing ‘instrument’ (e.g. the vocal folds or a violin) determine that adjacent time frames within a sound generally show a certain degree of correlation (instead of sudden changes).

However, an important exception to this rule are segment boundaries in speech sounds, such as those introduced by plosives, or stop consonants (Rogers, 2000). It could be argued that the previous investigations addressed processes more relevant to these speech attributes, while the stimulus in Study 1 was more suited to addressing longer time windows. Nevertheless, the probabilistic nature introduced by spectrotemporal correlation is a significant improvement compared to previous approaches (which used a more deterministic manipulation, with a fixed segment length for each parametric level), since it required the tracking of higher-level statistical properties so as to integrate complex sounds across multiple time windows.

Further studies will need to go beyond the description of different analysis time windows to elucidating the inner dynamics of the underlying network. The main reason for choosing a continuous acquisition paradigm in Study 1, despite its obvious caveats (see Section 2.4), was to achieve a better temporal resolution that would enable or facilitate a causal investigation of haemodynamic responses using DCM (Friston *et al.*, 2003; Penny *et al.*, 2004). DCM enables a more detailed investigation of hierarchical contributions between primary, secondary, and association cortex. For example, the model suggested by Boemio and colleagues (2005) posits that hemispheric lateralisations only emerge later in the auditory hierarchy (at the level of the STS); this is supported by the results in Study 1. Preliminary investigations of a dynamic network underlying the results on Study 1 using DCM have so far not yielded convincing results.

Future studies will also need to investigate the degree to which the differential representation of analysis time windows is pre-attentive (or intrinsic, Giraud *et al.*, 2007), or whether and to what degree it is influenced by attention. Studies to date have either had no experimental task (Boemio *et al.*, 2005; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001) or a stimulus irrelevant task (Study 1). It is conceivable that hemispheric preferences for different time windows emerge to a greater extent when they are task-relevant. For example, participants could be asked to rate the overall correlation within sounds such as those used in Study 1, while being scanned. DCM could then also address effects of attentional modulation within this network.

A necessary further step towards elucidating different temporal analysis windows in auditory cortex is the use of methodologies that have a far superior temporal resolution than fMRI (e.g. EEG or MEG). Luo & Poeppel (2007) demonstrated a sliding analysis window of roughly 200 ms (likely corresponding to theta band activity of 4-8 Hz) which was right-lateralised and that emerged when noise-vocoded speech was intelligible (see also Ahissar *et al.*, 2001; Elhilali *et al.*, 2003). Similarly, future investigations will need to study in detail such differentiations with respect to separate analysis time windows: one hypothesis would be that the power spectrum reveals entrainment effects to the different levels of spectrotemporal correlation.

Whether or not temporal analysis windows do indeed show a relative lateralisation preference (if not a selective preference) is currently still a matter of debate, given the often contradictory results between studies (Boemio *et al.*, 2005; Hickok & Poeppel, 2007; Schönwiesner *et al.*, 2005; Zatorre & Belin, 2001; Zatorre & Gandour, 2008). Furthermore, it is important to note that such lateralisation differences would reflect relative sensitivities instead of a categorical difference between left and right auditory cortex. For example, in the study by Obleser and colleagues (2008), the results show relatively small but nevertheless generally consistent lateralisation preferences for temporal and spectral resolution (especially in the case of a rightwards lateralisation of spectral resolution), in accordance with the spectrotemporal trade-off theory (Zatorre *et al.*, 2002a). However, in Obleser and colleagues (2008), the right hemisphere seemed more sensitive to the spectral and temporal variations in the stimuli, showing a weaker activation for high temporal resolution in particular, while the left hemisphere across participants did not show a clear preference for temporal resolution.

Here it is hypothesised that relative lateralisation preferences are an emergent phenomenon of a hierarchical processing scheme in which the length of analysis time windows increases as one progresses along the hierarchy in auditory cortex. Studies using simple sinusoidal amplitude modulations with different modulation rates (and thus different temporal windows) support this notion, demonstrating increasing sensitivity for decreasing modulation rates in the auditory system as one progresses from subcortical to higher order auditory cortical structures, but no lateralisation preferences (Giraud *et al.*, 2000; Harms & Melcher, 2002). This view of the representation of increasing temporal windows as one progresses in the cortical hierarchy would also be in agreement with recent conceptualisations in visual and theoretical neuroscience (Hasson *et al.*, 2008; Kiebel *et al.*, 2008), and would thus provide a general, modality-independent framework for processing temporal information in cortex.

## **7.2 The segregation and representation of auditory objects are hierarchically linked in auditory cortex**

Study 2 (Chapter 4) of this thesis investigated the segregation and integration of complex auditory objects and demonstrated a hierarchically organised network: the detection of auditory object boundaries based on transitions in higher-order statistical properties occurred earlier in the auditory hierarchy (primary and association auditory cortex) than the precise representation of auditory object properties (auditory association cortex). Importantly, these results revealed generic mechanisms underlying auditory object analysis, since the use of synthetic stimuli avoided semantic associations, suggesting that the auditory system abstracts statistical rules governing areas of frequency-time space; this provides a generic principle for auditory object analysis. At the same time, the experimental manipulation does not claim to address perceptual mechanisms for all auditory object classes; rather, it highlights one general acoustic dimension (spectrotemporal coherence) along which auditory objects can be differentiated and identified.

Study 2 is the first to address such a differentiation between auditory object segregation and representation. In fact, it can be argued that previous studies that employed sparse imaging designs and contrasted a condition in which the parameter of interest changed between different instantiations within a trial, with a condition in which that parameter remained fixed, are confounded by this change in the stimulus (e.g. von Kriegstein *et al.*, 2006; Warren *et al.*, 2005a; Zatorre *et al.*, 2004; pilot study to Study 2). From first principles, these designs reveal cortical substrates for the detection of changes in the parameter of interest, but not a cortical substrate for the parameter of interest as such. However, this does not exclude the possibility that these processes partially overlap, as was the case in Study 2 (in PT and TPJ).

The hierarchical organisation scheme proposed as a result of the data is based on anatomical evidence, given that processing in primary auditory cortex precedes that in auditory association cortex (the underlying temporal processes are likely below the resolution of fMRI). The anatomical scheme is in agreement with Schönwiesner and colleagues (2007), who found a similar and convergent pattern of results to parametrically increasing sound duration deviants in a mismatch paradigm (using both EEG and fMRI): the initial detection of deviants was reflected in primary auditory

cortex, while later stages of processing, e.g. a more precise spectrotemporal analysis and attentional resource allocation, engaged auditory association cortex and prefrontal cortex, respectively. Furthermore, there is evidence that such a processing stream holds across modalities, as texture boundary detection and scene segmentation in the visual system reveal a similar hierarchical organisation (Scholte *et al.*, 2008). Both studies (Scholte *et al.*, 2008; Schönwiesner *et al.*, 2007) report convergent findings from EEG and fMRI across modalities; similarly, a more precise understanding of the temporal hierarchy within the network demonstrated in study presented here will need to be addressed using methodologies with considerably higher temporal resolution, such as EEG or MEG.

A future complementary approach should investigate in more detail the functional organisation of this hierarchical network, for example using DCM (Friston *et al.*, 2003; Penny *et al.*, 2004). Such work would investigate the degree to which change detection in primary auditory cortex modulates the subsequent representation of auditory objects in auditory association cortex (e.g. Schönwiesner *et al.*, 2007). This bottom-up conceptualisation is based on the shorter response latencies in primary cortices reported in studies that had sufficient temporal resolution (Scholte *et al.*, 2008; Schönwiesner *et al.*, 2007); however, an alternative conceptualisation would predict top-down modulations of primary cortex, in which higher order areas provide rapid feedback. Yet another possibility is that recurrent feedback loops within primary cortex exist, similar to lateral or balanced inhibition for frequency selectivity (Wehr & Zador, 2003; Wu *et al.*, 2008), which in turn lead to differentiated feedforward propagations. This latter network architecture is perhaps least likely, given the complexity of the stimulus at hand and the need to integrate the spectrotemporal features across a broad frequency range and considerable time scales. Nevertheless, in DCM the Bayesian approach allows an explicit decision as to which of these models best fits the data and is thus a logical next step in delineating the network underlying auditory object segregation and integration.

A number of studies using electrophysiological recordings have demonstrated neurons in the lateral belt area of rhesus monkeys (Tian & Rauschecker, 2004) and posterior auditory field in cats (Tian & Rauschecker, 1998) that are sensitive to the direction of FM ramps, with a slight preference for ramps increasing in frequency.

Such a preference was also visible in lateral HG in Study 2. The ‘auditory textures’ stimulus developed in this thesis should contribute to this line of research in non-human primates, adding a critical new stimulus that moves beyond single FM ramps to the representation of statistical signal properties in a complex integrated whole.

While Study 2 controlled the spectrotemporal statistics of FM ramps to create auditory objects, alternative approaches are conceivable that focus on AM or combinations of AM and FM (Luo *et al.*, 2006) to generalise the present findings to other object classes. For example, it would be interesting to investigate whether a similar hierarchical network would be involved in detecting changes in spectrotemporal correlation as in Study 1, and to what degree this would be influenced by different analysis time windows in anatomically distinct regions of auditory cortex. It is important to note that defining auditory objects by their spectrotemporal statistics is a generic approach that lends itself to a whole range of other objects and object classes, thereby paving the way for further research investigating principles of auditory object analysis.

Finally, the experimental design in Study 2 was critical to disambiguating two fundamental perceptual processes in auditory scene analysis: the representation and segregation of auditory objects. By combining a parametric (different levels) and factorial (change vs. absolute coherence) approach in one design, it was possible to dissociate these two processes. This is an important innovation, since comparing changing object sequences with fixed object sequences cannot tease apart processes that signal a change in auditory objects from those that represent the statistical object properties as such (see for example the Pilot study to Study 2).

### **7.3 The planum temporale (PT) acts as a ‘computational hub’**

In their review of the function of the planum temporale (PT) of human auditory association cortex, Griffiths and Warren (2002) proposed that the PT acts as a ‘computational hub’, which compares incoming acoustic information to pre-existing templates and subsequently gates its output along the auditory hierarchy for further

processing. This model predicts an increased computational demand within PT as the information carried in the acoustic signal increases.

Studies 3 and 4 explicitly tested and demonstrated the dependency of PT activation on signal entropy. The results showed that the function of PT is different from that of preceding (e.g. HG and subcortical structures IC and MGB) and subsequent (e.g. IFG, DLPFC) processing stages; while the neural code of earlier structures is both more faithful and redundant (Chechik *et al.*, 2006), that of subsequent structures reflects a neural code that is independent of the original signal entropy. This is an intriguing result, as it highlights the computational power achieved in PT and suggests a division of labour that allows subsequent structures to represent different aspects associated with the acoustic signal (e.g. retrieving the signal).

It could be argued that the results reflect a perceptual (in contrast to sensory) representation of pitch interval size instead of signal entropy. The use of IRN pitch in Study 4 explicitly addressed this issue and made a sensory explanation of the results unlikely. Nevertheless, it is possible that the increased activation in PT reflects the processing of interval size (and not entropy as such), since interval size increased with entropy. Future investigations will need to address this potential confound. For example, one could use a more limited number of pitches and base pitch intervals on specified conditional probability matrices such that large intervals are just as likely as small intervals and would therefore convey an equal amount of information. The information theoretic properties would be based on specific pitch transition probabilities, irrespective of their size. However, one disadvantage of this approach would be a stimulus that is more deterministic than a stimulus derived from pitch sequences of specified random-phase power spectra. Furthermore, such pitch transition dependencies would first need to be established or learned, introducing additional and possibly confounding factors such as inter-subject learning rate variability.

Studies 3 and 4 controlled the signal entropy by means of pitch; future investigations should address whether the encoding mechanism in PT generalises to other sound features, such as rhythm or timbre. For example, just as the fractal time series derived from the IFFT's of specific power spectra was used to define pitch height, it could also be used to control the time interval between pitches (or any type

of sound), the number of harmonics of a complex sound (i.e. timbre variation) or spatial position. A further step would then be to test whether and in what form these different manipulations interact (e.g. pitch entropy with rhythm entropy).

One aspect that requires additional investigation is to which degree the effect in PT is specific to encoding and to what extent subsequent processes (such as retrieval) are indeed independent of the entropy in the original signal. Behavioural results in Study 4 did not reveal any effect of signal entropy on performance, and this was reflected in the functional imaging data, since no area showed an increase in activation as a function of signal entropy during the retrieval and comparison stage. It was proposed that PT acts as a computational hub that uses neural mechanisms such as ‘sparse’ (DeWeese & Zador, 2006; Friston, 2003a; Olshausen & Field, 2004) or ‘predictive’ (Baldeweg, 2006; Friston, 2005; von Kriegstein & Giraud, 2006) coding to compress the neural input and filter out redundancies (Chechik *et al.*, 2006); the output for subsequent processing stages (e.g. in presumptive ‘what’ and ‘where’ pathways) would then be less dependent on signal entropy. This hypothesis will need to be tested further using electrophysiological recordings and computational models.

#### **7.4 Local and global information in pitch patterns is hierarchically organised**

Study 5 (Chapter 6) tested both the hierarchical and lateralisation accounts for processing local and global information in pitch patterns in neurologically normal participants using fMRI. The results supported a hierarchy between local and global processing stages such that the processing of local violations recruited additional resources (in right PT and right posterior STS) over and above those needed for processing global violations. However, the lateralisation revealed by the results did not confirm a strict local–left and global–right dichotomy; rather, while the processing of local violations recruited both left and right STS, global violations were lateralised towards left STS. As outlined in the Discussion of Study 5, this lateralisation account is in fact equally well supported by a close examination of the pattern of brain lesions

and perceptual deficits reported in previous patient studies upon which the local–left and global–right dichotomy was based (Liégeois-Chauvel *et al.*, 1998; Peretz, 1990).

Importantly, previous designs were only able to make direct comparisons between violations of local and global levels. In other words, these designs only allowed a description of local processing with respect to global processing, and, similarly, a description of global processing with respect to local processing. In contrast, the one-back design in Study 5 enabled a selective investigation of the processing of local violations as such and global violations as such, without a direct comparison between the two. These were revealed in STS (bilaterally for processing local violations, left-lateralised for processing global violations). The results suggest a processing hierarchy, in which individual pitches are encoded in secondary cortex (Bendor & Wang, 2005; Griffiths, 2005; Griffiths *et al.*, 2001; Patterson *et al.*, 2002; Penagos *et al.*, 2004), before the representation of higher-order properties such as pitch interval contour and pitch interval size is achieved in higher order auditory association cortex. At the same time, the different experimental design (one-back design versus individual pair-wise comparisons) might account for the differences seen between previous studies and Study 5. One-back designs require a greater working memory load than individual pairwise comparisons, and it is possible that this contributed at least in part to the divergence in results. However, it is important to note that the pattern of results in Study 5 is equally well supported by the neuropsychological data from which the classical local-left; global-right dissociation was derived.

Study 5 did not properly control for potentially confounding effects of absolute pitch, since the stimuli did not transpose consecutive pitch patterns (this was also the case in Liégeois-Chauvel *et al.*, 1998; Peretz, 1990, who used melody pairs). The use of transposed pitch patterns is considered to be a less confounded test of the representation of global and local structural levels. However, the representation of the local level in particular is significantly diminished without absolute pitch information (Dowling *et al.*, 1987), introducing a potential task difficulty confound.

Study 5 was the first to probe the precise anatomical network participating in local and global pitch pattern analyses in the normal functioning brain (as opposed to patients with focal brain lesions), but there is an evident need for further studies.

These will need to further investigate the presumed hemispheric lateralisation scheme for processing local and global structural levels in pitch patterns and its conceptual usefulness. Similar to Study 1, it is conceivable that lateralisation preferences are at best relative or secondary effects of another hierarchical processing scheme. Similarly, it seems unlikely that the brain would have developed an absolute division of labour with respect to left and right auditory cortices. In this respect, the current experimental approach, which allowed a separate investigation of the cortical substrates processing local or global violations, gains importance, as direct comparisons between local and global processing emphasise their (potentially small) differences over their (potentially substantial) commonalities.

### **7.5 Key future problems and implications for auditory neuroscience**

The experimental work of this thesis investigated how higher-level statistical signal properties are represented in the auditory cortex. While there is considerable knowledge of relatively simple, deterministic sound features, this thesis focussed on complex, non-deterministic sound features that can nevertheless be controlled within probabilistic constraints. The generic nature of the synthetic stimuli, combined with a parametric approach to control various instantiations of statistical signal properties, ensured that the mechanisms addressed allow inferences about a variety of acoustic signals and are not dependent on semantic associations.

Critically, the acoustic stimuli employed in this thesis required mechanisms that evaluate the probabilistic properties of complex spectrotemporal signals. Within a Bayesian framework of probabilistic brain function (Friston, 2003a, 2005), sensory cortex evaluates the statistical properties of the signal so as to optimise its coding and make inferences based on previous experience or priors. The studies of this thesis support this view in the auditory domain, demonstrating neural mechanisms that encode, and in the case of PT compress, the information content of a variety of acoustic signals so as to facilitate subsequent processing stages. Furthermore, the results support the view of a hierarchical organisation of auditory perception that

allows dynamic processing including top-down modulations or the application of priors as in empirical Bayes (Friston, 2003a, 2005).

Despite the various types of acoustic stimuli used in each study, all studies revealed involvement of PT as part of a network comprising distinct parts of auditory cortex. This is in accord with the hypothesis that PT forms a generic ‘computational hub’ that is not selective to a specific acoustic feature, but instead encodes and evaluates complex statistical properties based on experience-dependent templates (Griffiths & Warren, 2002). Such a process then facilitates further computations in subsequent structures along the auditory hierarchy, such as STS and anterior STG, whose increased temporal analysis windows lend themselves to complex integrative processes. Finally, task and attention related processes in frontal cortex likely provide top-down modulatory input.

Constructing a more precise description of how PT achieves these computations remains a key challenge for research in auditory neuroscience. Up to now, the ‘computational hub’ model (Griffiths & Warren, 2002) has been primarily a descriptive hypothesis without explicit quantitative predictions amenable to computational models that consider neuronal dynamics and connectivity. Future collaborations between auditory and computational neuroscience may build a computational model of how PT achieves the compression of the neural code to facilitate subsequent processing. Such a model would be very useful to guide future research in auditory neuroscience at the level of stimulus complexity employed in this thesis.

One challenge to a computational model of PT function is its precise cytoarchitectonic, and consequently functional, parcellation. The PT occupies a large area posterior to the first transverse gyrus of Heschl (HG), and there are indications that it might divide into at least two subareas, one more rostral and one more caudal (or even additional gyri, Sweet *et al.*, 2005), of which the caudal part extends into the TPJ and is especially expanded compared to non-human primates (Galaburda & Sanides, 1980). Furthermore, the inter-individual variability is considerable (for example, in Westbury *et al.*, 1999, no individual voxel in PT was labelled with a probability of more than 65%), complicating any generalisations.

In functional imaging studies, this inter-individual variability is sometimes addressed by using fixed-effects statistics (see Section 2.6.3); indeed, random-effects statistics are relatively conservative and thereby increase the chance of falsely accepting the null hypothesis in a randomly selected sample of participants. Some studies (e.g. Patterson *et al.*, 2002) have made the case that fixed-effects statistics are more appropriate for auditory association cortex in particular, which shows a more pronounced inter-individual variability than primary and secondary auditory cortices. However, this is controversial since it increases the likelihood of false positives and one must show that the fixed-effects group results are not driven by a few outliers in the group. The novelty of the complex probabilistic stimuli employed in this thesis called for the more conservative approach (i.e. random-effects statistics).

In order to achieve a more explicit model of PT function, a necessary complement to neuroimaging techniques (such as fMRI, EEG, MEG) are invasive electrophysiological recordings, which can probe directly and with high temporal and spatial precision the degree to which PT compresses the incoming information, for example via ‘sparse’ (DeWeese & Zador, 2006; Friston, 2003a; Olshausen & Field, 2004) or ‘predictive’ (Baldeweg, 2006; Friston, 2005; von Kriegstein & Giraud, 2006) coding. However, in humans these are restricted mainly to pre-operative patients (Brugge *et al.*, 2003; 2008; Howard *et al.*, 1996; Liégeois-Chauvel *et al.*, 2001; 2004; 1994; 1991). Unfortunately, the significance of electrophysiological recordings from non-human primates, while undoubted with respect to primary and secondary auditory cortices, is limited with respect to auditory association cortex, as it is not clear whether PT has a homologue in non-human primates (Hackett, 2007). At the same time, the complex acoustic ecology of non-human primates is comparable to the sounds employed in this thesis, suggesting that auditory cortex of non-human primates might achieve similar computations, albeit possibly at a different stage in the auditory hierarchy (indeed, the generic approach of this thesis is readily applicable to non-human primates, enabling important direct comparisons between species with identical stimuli). Since the precise intra-cortical connectivity (e.g. between HG and PT through the injection of immunohistochemical tracers) cannot be determined experimentally in humans (with the possible exception of novel non-invasive techniques such as DTI, Upadhyay *et al.*, 2007, 2008), and because of the uncertainty whether PT has a homologue in non-human primates, a computationally sophisticated

model of PT function that takes into account its functional architecture with other cortical regions remains challenging.

The studies in this thesis aimed to further our understanding of how statistical signal properties are represented in the auditory cortex of humans. The results demonstrated distinct hierarchical mechanisms for auditory object analysis and segregation. Nevertheless, many questions remain, and just how these mechanisms are realised at the neuronal level remains a key challenge for future research using a variety of methodologies, models and species (Griffiths *et al.*, 2004). Thus, regardless of whether one's philosophy inclines one to view the multitude of outstanding questions in auditory neuroscience as a glass half filled or half empty, this thesis hopes to have contributed a few drops towards filling the glass (without simultaneously extending the volume of the glass itself or instead merely causing waves).

---

**BIBLIOGRAPHY**

1. 't Hart J, Collier R & Cohen AJ (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge, UK: Cambridge University Press.
2. Aertsen AM & Johannesma PI (1981a). A comparison of the spectro-temporal sensitivity of auditory neurons to tonal and natural stimuli. *Biol Cybern* **42**, 145-156.
3. Aertsen AM & Johannesma PI (1981b). The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol Cybern* **42**, 133-143.
4. Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahnkce H & Merzenich MM (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc Natl Acad Sci U S A* **98**, 13367-13372.
5. Alho K, Connolly JF, Cheour M, Lehtokoski A, Huottilainen M, Virtanen J, Aulanko R & Ilmoniemi RJ (1998a). Hemispheric lateralization in pre-attentive processing of speech sounds. *Neurosci Lett* **258**, 9-12.
6. Alho K, Winkler I, Escera C, Huottilainen M, Virtanen J, Jääskeläinen IP, Pekkonen E & Ilmoniemi RJ (1998b). Processing of novel sounds and frequency changes in the human auditory cortex: magnetoencephalographic recordings. *Psychophysiology* **35**, 211-224.
7. Amaro E, Williams CR, Shergill SS, Fu CHY, MacSweeney M, Picchioni MM, Brammer MJ & McGuire PK (2002). Acoustic noise and functional magnetic resonance imaging: current strategies and future prospects. *J Magn Reson Imaging* **16**, 497-510.
8. Andersson J, Hutton C, Ashburner J, Turner R & Friston KJ (2001). Modelling geometric deformations in EPI time series. *Neuroimage* **13**, 903-919.
9. ANSI (1994). *American National Standard - Psychoacoustical Terminology*. New York: American National Standards Institute.

10. Arnott SR, Grady CL, Hevenor SJ, Graham S & Alain C (2005). Functional organization of auditory working memory as revealed by fMRI. *J Cogn Neurosci* **17**, 819-831.
11. Attneave F (1959). *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results*. New York: Holt, Rinehart, and Winston.
12. Bachem A (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychol (Amst)* **7**, 80-88.
13. Bajo V, Nodal F, Overath T & King AJ (2007). Effect of auditory task on the expression of immediate-early gene c-Fos in the ferret auditory cortex. Poster presented at the *30th Midwinter Research Meeting of the Association for Research in Otolaryngology*, February 10-15, Denver, USA.
14. Baldeweg T (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends Cogn Sci* **10**, 93-94.
15. Bamiou D-E, Musiek FE & Luxon LM (2003). The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain Res Brain Res Rev* **42**, 143-154.
16. Bartlett JC & Dowling WJ (1980). Recognition of transposed melodies: a key-distance effect in developmental perspective. *J Exp Psychol Hum Percept Perform* **6**, 501-515.
17. Beauvois MW & Meddis R (1996). Computer simulation of auditory stream segregation in alternating-tone sequences. *J Acoust Soc Am* **99**, 2270-2280.
18. Beck E (1928). Die myeloarchitektonische Felderung des in der Sylvischen Furche gelegenen Teiles des menschlichen Schläfenlappens. *Z Neurol Psychiatr* **36**, 1-21.
19. Behrens TE & Johansen-Berg H (2005). Relating connectional architecture to grey matter function using diffusion imaging. *Philos Trans R Soc Lond B Biol Sci* **360**, 903-911.

20. Behrens TE, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CA, Boulby PA, Barker GJ, Sillery EL, Sheehan K, Ciccarelli O, *et al.* (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat Neurosci* **6**, 750-757.
21. Belin P (2006). Voice processing in human and non-human primates. *Philos Trans R Soc Lond B Biol Sci* **361**, 2091-2107.
22. Belin P & Zatorre RJ (2000). 'What', 'where' and 'how' in auditory cortex. *Nat Neurosci* **3**, 965-966.
23. Belin P, Zatorre RJ, Hoge R, Evans AC & Pike B (1999). Event-related fMRI of the auditory cortex. *Neuroimage* **10**, 417-429.
24. Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC & Samson Y (1998). Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* **10**, 536.
25. Bendor D & Wang C (2008). Neural response properties of primary, rostral, and rostrotemporal core fields in the auditory cortex of marmoset monkeys. *J Neurophysiol* **100**, 888-906.
26. Bendor D & Wang X (2006). Cortical representations of pitch in monkeys and humans. *Curr Opin Neurobiol* **16**, 391-399.
27. Bendor D & Wang XQ (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* **436**, 1161-1165.
28. Bischoff-Grethe A, Proper SM, Mao H, Daniels KA & Berns GS (2000). Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *J Neurosci* **20**, 1975-1981.
29. Bloch F, Hansen WW & Packard M (1946). The nuclear induction experiment. *Phys Rev* **70**, 474-485.
30. Boemio A, Fromm S, Braun A & Poeppel D (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* **8**, 389-395.

31. Bregman AS (1990). *Auditory Scene Analysis: The Perceptual Organisation of Sound*. Cambridge, MA: MIT Press.
32. Bregman AS, Ahad PA, Crum PAC & O'Reilly J (2000). Effects of time intervals and tone durations on auditory stream segregation. *Percept Psychophys* **62**, 626-636.
33. Bregman AS & Campbell J (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *J Exp Psychol Gen* **89**, 244-249.
34. Brodmann K (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Johann Ambrosius Barth Verlag.
35. Brosch M & Schreiner CE (1997). Time course of forward masking tuning curves in cat primary auditory cortex. *J Neurophysiol* **77**, 923-943.
36. Brugge JF, Volkov IO, Garell PC, Reale RA & Howard AM (2003). Functional connections between auditory cortex on Heschl's gyrus and the lateral superior temporal gyrus in humans. *J Neurophysiol* **90**, 3750-3763.
37. Brugge JF, Volkov IO, Oya H, Kawasaki H, Reale RA, Fenoy A, Steinschneider M & Howard AM (2008). Functional localization of auditory cortical fields of human: click-train stimulation. *Hear Res* **238**, 12-24.
38. Buxton RB, Wong EC & Frank LR (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magn Reson Med* **39**, 855-864.
39. Caclin A, Brattico E, Tervaniemi M, Näätänen R, Morlet D, Giard MH & McAdams S (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *J Cogn Neurosci* **18**, 1959-1972.
40. Caclin A, Giard MH, Smith BK & McAdams S (2007). Interactive processing of timbre dimensions: a Garner interference study. *Brain Res* **1138**, 159-170.

41. Caclin A, McAdams S, Smith BK & Giard MH (2008). Interactive processing of timbre dimensions: an exploration with event-related potentials. *J Cogn Neurosci* **20**, 49-64.
42. Caclin A, McAdams S, Smith BK & Winsberg S (2005). Acoustic correlates of timbre space dimensions: a confirmatory study using synthetic tones. *J Acoust Soc Am* **118**, 471-482.
43. Calford MB & Semple MN (1995). Monaural inhibition in cat auditory cortex. *J Neurophysiol* **73**, 1876-1891.
44. Carlyon RP (1991). Discriminating between coherent and incoherent frequency modulation of complex tones. *J Acoust Soc Am* **89**, 329-340.
45. Carlyon RP (2004). How the brain separates sounds. *Trends Cogn Sci* **8**, 465-471.
46. Carlyon RP, Cusack R, Foxton JM & Robertson IH (2001). Effects of attention and unilateral neglect on auditory stream segregation. *J Exp Psychol Hum Percept Perform* **27**, 115-127.
47. Carlyon RP, Plack CJ, Fantini DA & Cusack R (2003). Cross-modal and non-sensory influences on auditory streaming. *Perception* **32**, 1393-1402.
48. Celesia GG (1976). Organization of auditory cortical areas in man. *Brain* **99**, 403-414.
49. Chait M, Poeppel D, de Cheveigné A & Simon JZ (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *J Neurosci* **27**, 5207-5214.
50. Chait M, Poeppel D & Simon JZ (2008). Auditory temporal edge detection in human auditory cortex. *Brain Res* **1213**, 78-90.
51. Chambers S, Akeroyd MA, Summerfield Q & Palmer AR (2001). Active control of the volume acquisition noise in functional magnetic resonance imaging: method and psychoacoustical evaluation. *J Acoust Soc Am* **110**, 3041-3054.

52. Chechik G, Anderson MJ, Bar-Yosef O, Young ED, Tyshby N & Nelken I (2006). Reduction of information redundancy in the ascending auditory pathway. *Neuron* **51**, 359-368.
53. Cherry EC (1953). Some experiments on the recognition of speech, with one and with two ears. *J Acoust Soc Am* **25**, 975-979.
54. Clarke S & Rivier F (1998). Compartments within human primary auditory cortex: Evidence from cytochrome oxidase and acetylcholinesterase staining. *Eur J Neurosci* **10**, 741-745.
55. Coath M, Balaguer-Ballester E, Denham SL & Denham M (2008). The linearity of emergent spectro-temporal receptive fields in a model of auditory cortex. *Biosystems* **94**, 60-67.
56. Cowey A & Weiskrantz L (1976). Auditory sequence discrimination in macaca mulatta: the role of superior temporal cortex. *Neuropsychologia* **14**, 1-10.
57. Cuddy LL & Cohen AJ (1976). Recognition of transposed melodic sequences. *Q J Exp Psychol* **28**, 255-270.
58. Cusack R (2005). The intraparietal sulcus and perceptual organization. *J Cogn Neurosci* **17**, 641-651.
59. Cusack R, Brett M & Osswald K (2003). An evaluation of the use of magnetic field maps to undistort echo-planar images. *Neuroimage* **18**, 127-142.
60. Cusack R & Carlyon RP (2003). Perceptual asymmetries in audition. *J Exp Psychol Hum Percept Perform* **29**, 713-725.
61. Cusack R, Deeks J, Aikman G & Carlyon RP (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *J Exp Psychol Hum Percept Perform* **30**, 643-656.
62. Damadian R, Goldsmith M & Minkoff L (1977). NMR in cancer: XVI. FONAR image of the live human body. *Physiol Chem Phys Med NMR* **9**, 97-100.
63. Darwin CJ (1997). Auditory grouping. *Trends Cogn Sci* **1**, 327-333.

64. Darwin CJ & Carlyon RP (1995). Auditory Grouping. In, *Handbook of Perception and Cognition, Volume 6, Hearing*, BCJ Moore (ed.), London: Academic Press, pp. 387-424.
65. Davis MH & Johnsruide I (2003). Hierarchical processing in spoken language comprehension. *J Neurosci* **23**, 3423-3431.
66. de Boer E (1976). On the 'residue' and auditory pitch perception. In, *Handbook of Sensory Physiology (Vol. 3)*, WD Keidel & WD Neff (eds.), New York: Springer, pp. 479-583.
67. de Cheveigné A (2005). Pitch perception models. In, *Pitch. Neural Coding and Perception*, CJ Plack, AJ Oxenham, RR Fay & AN Popper (eds.), New York: Springer, pp. 169-233.
68. de Coensel B, Botteldooren D & de Muer T (2003). 1/f noise in rural and urban soundscapes. *Acta Acoustica* **89**, 287-295.
69. deCharms RC, Blake DT & Merzenich MM (1998). Optimizing sound features for cortical neurons. *Science* **381**, 1439-1443.
70. Deichmann R, Schwarzbauer C & Turner R (2004). Optimization of the 3D MDEFT sequence for anatomical brain imaging: technical implications at 1.5 T and 3 T. *Neuroimage* **21**, 757-767.
71. Deike S, Gaschler-Markefski B, Brechmann A & Scheich H (2004). Auditory stream segregation relying on timbre involves left auditory cortex. *Neuroreport* **15**, 1511-1514.
72. DeLaPaz R (1994). Echo-planar imaging. *Radiographics* **14**, 1045-1058.
73. Denham SL (2001). Cortical synaptic depression and auditory perception. In, *Computational Models of Auditory Function*, S Greenberg & M Slaney (eds.), Amsterdam: IOS Press, pp. 281-296.
74. Deouell L & Bentin S (1998). Variable cerebral responses to equally distinct deviance in four auditory dimensions: a mismatch negativity study. *Psychophysiology* **35**, 745-754.

75. Deouell L, Bentin S & Giard MH (1998). Mismatch negativity in dichotic listening: evidence for interhemispheric differences and multiple generators. *Psychophysiology* **35**, 355-365.
76. DeWeese MR & Zador AM (2006). Non-gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *J Neurosci* **26**, 12206-12218.
77. Dorsaint-Pierre R, Penhune VB, Watkins KE, Neelin P, Lerch JP, Bouffard M & Zatorre RJ (2006). Asymmetries of the planum temporale and Heschl's gyrus: relationship to language lateralization. *Brain* **129**, 1164-1176.
78. Dowling WJ (1978). Scale and contour: two components of a theory of memory for music. *Psychol Rev* **85**, 341-354.
79. Dowling WJ & Fujitani DS (1971). Contour, interval, and pitch recognition in memory for melodies. *J Acoust Soc Am* **49**, 524-531.
80. Dowling WJ, Lung KM & Herrbold S (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Percept Psychophys* **41**, 642-656.
81. Eckert MA, Leonard CM, Possing ET & Binder JR (2006). Uncoupled leftward asymmetries for planum morphology and functional language processing. *Brain Lang* **98**, 102-111.
82. Eden GF, Joseph JE, Brown HE, Brown CP & Cephire TA (1999). Using hemodynamic signal delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. *Magn Reson Med* **41**, 13-20.
83. Edmister WB, Talavage TM, Ledden PJ & Weisskoff RM (1999). Improved auditory cortex imaging using clustered volume acquisitions. *Hum Brain Mapp* **7**, 89-97.
84. Eggermont JJ (1999). The magnitude and phase of temporal modulation transfer functions in cat auditory cortex. *J Neurosci* **19**, 2780-2788.

85. Eggermont JJ, Aertsen AM, Hermes DJ & Johannesma PI (1981). Spectro-temporal characterization of auditory neurons: redundant or necessary. *Hear Res* **5**, 109-121.
86. Eke A, Herman P, Kocsis L & Kozak LR (2002). Fractal characterization of complexity in temporal physiological signals. *Physiol Meas* **23**, R1-R32.
87. Elhilali M, Chi T & Shamma SA (2003). A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. *Speech Commun* **41**, 331-348.
88. Elhilali M, Fritz JB, Tai-Shih C & Shamma SA (2007). Auditory cortical receptive fields: stable entities with plastic abilities. *J Neurosci* **27**, 10372-10382.
89. Erickson ML, Perry S & Handel S (2001). Discrimination functions: can they be used to classify singing voices? *J Voice* **15**, 492-502.
90. Erickson ML & Perry SR (2003). Can listeners hear who is singing? A comparison of three-note and six-note discrimination tasks. *J Voice* **17**, 352-368.
91. Evans AC, Collins DL, Mills SR, Brown RD, Kelly RL & Peters TM (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *Proc IEEE Nucl Sci Symp Med Imag Conf* **1-3**, 1813-1817.
92. Fastl H & Zwicker E (2007). *Psychoacoustics. Facts and Models (Third Edition)*. (TS Huang, T Kohonen & MR Schroeder, eds.) Berlin: Springer.
93. Feldman J (2003). What is a visual object? *Trends Cogn Sci* **7**, 252-256.
94. Fishman YI, Arezzo JC & Steinschneider M (2004). Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration. *J Acoust Soc Am* **116**, 1656-1670.
95. Fishman YI, Reser DH, Arezzo JC & Steinschneider M (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hear Res* **151**, 167-187.

96. Foundas AL, Leonard CM, Gilmore R, Fennell E & Heilman K (1994). Planum temporale asymmetry and language function. *Neuropsychologia* **32**, 1225-1231.
97. Fox PT & Raichle ME (1986). Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc Natl Acad Sci U S A* **83**, 1140-1144.
98. Foxton JM, Talcott JB, Witton C, Brace H, McIntyre F & Griffiths TD (2003). Reading skills are related to global, but not local, acoustic pattern perception. *Nat Neurosci* **6**, 343-344.
99. Friston K (2003a). Learning and inference in the brain. *Neural Networks* **16**, 1325-1352.
100. Friston K (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* **360**, 815-836.
101. Friston KJ (2003b). Experimental design and statistical parametric mapping. In, *Human Brain Function (Second Edition)*, RSJ Frackowiak, KJ Friston, CD Frith, RJ Dolan, CJ Price, J Ashburner, W Penny & S Zeki (eds.), London, UK: Academic Press, pp. 599-634.
102. Friston KJ (2003c). Learning and inference in the brain. *Neural Networks* **16**, 1325-1352.
103. Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD & Frackowiak RS (1995a). Spatial registration and normalisation of images. *Hum Brain Mapp* **2**, 165-189.
104. Friston KJ, Harrison LM & Penny W (2003). Dynamic causal modelling. *Neuroimage* **19**, 1273-1302.
105. Friston KJ, Holmes AP & Worsley KJ (1999). How many subjects constitute a study? *Neuroimage* **10**, 1-5.
106. Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD & Frackowiak RS (1995b). Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* **2**, 189-210.

107. Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS & Dolan RJ (1996a). The trouble with cognitive subtraction. *Neuroimage* **4**, 97-104.
108. Friston KJ, Williams S, Howard R, Frackowiak RSJ & Turner R (1996b). Movement related effects in fMRI time series. *Magn Reson Med* **35**, 346-355.
109. Fritz J, Shamma SA, Elhilali M & Klein D (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* **6**, 1216-1223.
110. Fritz JB, Elhilali M, David SV & Shamma SA (2007). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear Res* **229**, 186-203.
111. Fritz JB, Elhilali M & Shamma SA (2005). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci* **25**, 7623-7635.
112. Frodl-Bauch T, Kathmann N, Möller H-J & Hegerl U (1997). Dipole localization and test-retest reliability of frequency and duration mismatch negativity generators. *Brain Topogr* **10**, 3-8.
113. Fujioka T, Trainor LJ, Ross B, Kakigi R & Pantev C (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *J Cogn Neurosci* **16**, 1010-1021.
114. Gaab N, Gabrieli JD & Glover GH (2007). Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Hum Brain Mapp* **28**, 721-732.
115. Galaburda AM & Pandya DN (1983). The intrinsic architectonic and connective organization of the superior temporal region of the rhesus monkey. *J Comp Neurol* **221**, 169-184.
116. Galaburda AM & Sanides F (1980). Cytoarchitectonic organization of the human auditory cortex. *J Comp Neurol* **190**, 597-610.

117. Garcia-Lazaro JA, Ahmed B & Schnupp JWH (2006). Tuning to natural stimulus dynamics in primary auditory cortex. *Curr Biol* **16**, 264-271.
118. Garrido MI, Friston KJ, Kiebel SJ, Stephan KE, Baldeweg T & Kilner JM (2008). A functional anatomy of the MMN: a DCM study of the roving paradigm. *Neuroimage* **42**, 936-944.
119. Giard MH, Lavikahen J, Reinikainen K, Perrin F, Bertrand O, Pernier F & Näätänen R (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J Cogn Neurosci* **7**, 133-143.
120. Giraud AL, Kell C, Thierfelder C, Sterzer P, Russ MO, Preibisch C & Kleinschmidt A (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cereb Cortex* **14**, 247-255.
121. Giraud AL, Kleinschmidt A, Poeppel D, Lund TE, Frackowiak RSJ & Laufs H (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron* **56**, 1127-1134.
122. Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R & Kleinschmidt A (2000). Representation of the Temporal Envelope of Sounds in the Human Brain. *J Neurophysiol* **84**, 1588.
123. Goense JD & Logothetis NK (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr Biol* **18**, 631-640.
124. Greenberg S, Carvey H, Hitchcock L & Chang S (2003). Temporal properties of spontaneous speech - a syllable-centric perspective. *J Phonetics* **31**, 465-485.
125. Griffiths TD (2001). The neural processing of complex sounds. *Ann N Y Acad Sci* **930**, 133.
126. Griffiths TD (2005). Functional imaging of pitch processing. In, *Pitch: Neural Coding and Perception*, C. J. Plack, A. J. Oxenham, R. R. Fay & AN Popper (eds.), Berlin: Springer, pp. 147-168.

127. Griffiths TD, Büchel C, Frackowiak RS & Patterson RD (1998). Analysis of temporal structure in sound by the human brain. *Nat Neurosci* **1**, 422-427.
128. Griffiths TD, Johnsrude I, Dean JL & Green GG (1999). A common neural substrate for the analysis of pitch and duration pattern in segmented sound? *Neuroreport* **10**, 3825-3830.
129. Griffiths TD, Kumar S, Warren JD, Stewart L, Stephan KE & Friston KJ (2007). Approaches to the cortical analysis of auditory objects. *Hear Res* **229**, 46-53.
130. Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O & Patterson RD (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci* **4**, 633-637.
131. Griffiths TD & Warren JD (2002). The planum temporale as a computational hub. *Trends Neurosci* **25**, 348-253.
132. Griffiths TD & Warren JD (2004). What is an auditory object? *Nat Rev Neurosci* **5**, 887-892.
133. Griffiths TD, Warren JD, Scott SK, Nelken I & King AJ (2004). Cortical processing of complex sound: a way forward? *Trends Neurosci* **27**, 181-185.
134. Grimault N, Bacon SP & Micheyl C (2002). Auditory stream segregation on the basis of amplitude-modulation rate. *J Acoust Soc Am* **111**, 1340-1348.
135. Grimault N, Micheyl C, Carlyon RP, Arthaud P & Collet L (2000). Influence of peripheral resolvability on the perceptual segregation of complex tones differing in fundamental frequency. *J Acoust Soc Am* **108**, 263-271.
136. Grooten S, Hutton C, Howseman AM, Ashburner J, Josephs O, Rees G, Friston KJ & Turner R (2000). Characterisation and correction of interpolation effects in the realignment of fMRI time series. *Neuroimage* **11**, 49-57.
137. Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M & Oxenham AJ (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J Neurosci* **25**, 5382-5388.

138. Gutschalk A, Micheyl C & Oxenham AJ (2008). Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol* **6**, e138.
139. Gutschalk A, Oxenham AJ, Micheyl C, Wilson EC & Melcher J (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *J Neurosci* **27**, 13074-13981.
140. Hackett TA (2007). Organization and correspondence of the auditory cortex of humans and nonhuman primates. In, *Evolution of Nervous Systems*, JH Kaas (ed.), Oxford: Elsevier, pp. 109-119.
141. Hackett TA, Preuss TM & Kaas JH (2001). Architectonic identification of the core region in auditory cortex of macaques, chimpanzees, and humans. *J Comp Neurol* **441**, 197-222.
142. Hackett TA, Stepniewska I & Kaas JH (1998a). Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* **394**, 475-495.
143. Hackett TA, Stepniewska I & Kaas JH (1998b). Thalamocortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* **400**, 271-286.
144. Hackett TA, Stepniewska I & Kaas JH (1999). Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* **817**, 45-58.
145. Hajda JM (1999). *The effect of time-variant acoustical properties on orchestral instrument timbres*. PhD Thesis (University of California, Los Angeles).
146. Hajda JM, Kendall RA, Carterette EC & Harschberger ML (1997). Methodological issues in timbre research. In, *Perception and Cognition of Music*, I Deliège & J Sloboda (eds.), Hove, UK: Psychology Press, pp. 253-307.
147. Halgren E, Baudena P, Clarke JM, Heit G, Liégeois C, Chauvel P & Musolino A (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. I. Superior temporal plane and parietal lobe. *Electroencephalogr Clin Neurophysiol* **94**, 191-220.

148. Halgren E, Marinkovic K & Chauvel P (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalogr Clin Neurophysiol* **106**, 156-164.
149. Hall DA, Barrett DJK, Akeroyd MA & Summerfield AQ (2005). Cortical representations of temporal structure in sound. *J Neurophysiol* **94**, 3181-3191.
150. Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM & Bowtell RW (1999). "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp* **7**, 213-223.
151. Handel S (1995). Timbre perception and auditory object identification. In, *Hearing*, BCJ Moore (ed.), San Diego: Academic Press, pp. 425-461.
152. Handel S & Erickson ML (2001). A rule of thumb: the bandwidth for timbre invariance is one octave. *Music Percept* **19**, 121-126.
153. Handel S & Erickson ML (2004). Sound source identification: the possible role of timbre transformations. *Music Percept* **21**, 587-610.
154. Hari R, Hämäläinen M, Ilmoniemi RJ, Kaukoranta E, Reinikainen K, Salminen J, Alho K, Näätänen R & Sams M (1984). Response of the primary auditory cortex to pitch changes in a sequence of tone pips: neuromagnetic recordings in man. *Neurosci Lett* **50**, 127-132.
155. Harms MP & Melcher JR (2002). Sound repetition rate in the human auditory pathway: representations in the waveshape and amplitude of fMRI activation. *Journal of Neurophysiology* **88**, 1433-1450.
156. Harrison LM, Duggins A & Friston KJ (2006). Encoding uncertainty in the hippocampus. *Neural Networks* **19**, 535-546.
157. Hart HC, Hall DA & Palmer AR (2003). The sound-level-dependent growth in the extent of fMRI activation in Heschl's gyrus is different for low- and high-frequency tones. *Hear Res* **179**, 104.
158. Hartmann WM (2000). *Signals, Sound, and Sensation*. New York: Springer.

159. Hartmann WM & Johnson D (1991). Stream segregation and peripheral channelling. *Music Percept* **9**, 155-184.
160. Hasson U, Yang E, Vallines I, Heeger DJ & Rubin N (2008). A hierarchy of temporal receptive windows in human cortex. *J Neurosci* **28**, 2539-2550.
161. Haykin S & Chen Z (2005). The cocktail party problem. *Neural Comput* **17**, 1875-1902.
162. Heffner HE & Heffner RS (1986). Hearing loss in Japanese macaques following bilateral auditory cortex lesions. *J Neurophysiol* **55**, 256-271.
163. Heil P, Rajan R & Irvine DR (1992). Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. I: Effect of variation of stimulus parameters. *Hear Res* **63**, 108-134.
164. Hellige JB (1996). Hemispheric asymmetry for visual information processing. *Acta Neurobiol Exp (Warsz)* **56**, 485-497.
165. Helmholtz HLF (1875). *The sensations of tone as a physiological basis for the theory of music*. (AJ Ellis, ed.) London: Longman's.
166. Henson RNA (2006). Efficient experimental design for fMRI. In, *Statistical Parametric Mapping: The analysis of functional brain images.*, KJ Friston, J Ashburner, S Kiebel, T Nichols & W Penny (eds.), London: Elsevier, pp. 193-210.
167. Hickok G & Poeppel D (2007). The cortical organization of speech processing. *Nat Rev Neurosci* **8**, 393-402.
168. Hindemith P (1940). *Unterweisung im Tonsatz, 1. Theoretischer Teil*. Mainz: Schott.
169. Hochberg I & Tamhane Y (1987). *Multiple Comparison Procedures*. New York: John Wiley and Sons Inc.

170. Howard MA, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC & Granner MA (1996). A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* **724**, 260-264.
171. Howard MA, Volkov IO & Mirsky R (2000). Auditory cortex on the human posterior superior temporal gyrus. *J Comp Neurol* **416**, 79-92.
172. Hübner R & Volberg G (2005). The integration of object levels and their content: a theory of global/local processing and related hemispheric differences. *J Exp Psychol Hum Percept Perform* **31**, 520-541.
173. Hudspeth AJ (1989). How the ear's works work. *Nature* **341**, 397-404.
174. Hulse HS, MacDougall-Shackleton SA & Wisniewski AB (1997). Auditory scene analysis by songbirds: stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *J Comp Psychol* **111**, 3-13.
175. Huron D (2006). *Sweet Anticipation. Music and the Psychology of Expectation*. Cambridge, MA: MIT Press.
176. Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J & Turner R (2002). Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage* **16**, 217-240.
177. Izumi A (2002). Auditory stream segregation in Japanese monkeys. *Cognition* **82**, 113-122.
178. Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, *et al.* (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A* **101**, 6809-6814.
179. Jamison H, Watkins KE, Bishop DVM & Matthews PM (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb Cortex* **16**, 1266-1275.

180. Jemel B, Achenbach C, Müller BW, Röpke B & Oades RD (2002). Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. *Brain Topogr* **15**, 13-27.
181. Johnsrude IS, Penhune VB & Zatorre RJ (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* **123**, 155-163.
182. Johnsrude IS, Zatorre RJ, Milner BA & Evans AC (1997). Left-hemisphere specialization for the processing of acoustic transients. *Neuroreport* **8**, 1761.
183. Jones EG (2003). Chemically defined parallel pathways in the monkey auditory system. *Ann N Y Acad Sci* **999**, 218-233.
184. Jones EG, Dell'Anna ME, Molinari M, Rausell E & Hashikawa T (1995). Subdivisions of macaque auditory cortex revealed by calcium-binding proteins immunoreactivity. *J Comp Neurol* **362**, 153-170.
185. Joris PX, Schreiner CE & Rees A (2004). Neural processing of amplitude-modulated sounds. *Physiol Rev* **84**, 541-577.
186. Justus T & List A (2005). Auditory attention to frequency and time: An analogy to visual local-global stimuli. *Cognition*, 31-51.
187. Kaas JH & Hackett TA (1999). Editorial. 'What' and 'where' processing in auditory cortex. *Nat Neurosci* **2**, 1045-1047.
188. Kaas JH & Hackett TA (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A* **97**, 11793-11799.
189. Kaczmarek L & Robertson HA (eds.) (2002). *Immediate Early Genes and Inducible Transcription Factors in Mapping the Central Nervous System Function and Dysfunction*. Amsterdam: Elsevier.
190. Kendall RA, Carterette EC & Hajda JM (1999). Perceptual and acoustical features of natural and synthetic orchestral instrument tones. *Music Percept* **16**, 327-363.

191. Kiebel SJ, Daunizeau J, Frith CD & Friston KJ (2008). A hierarchy of time-scales and the brain. *PLoS Comp Biol* **4**, e1000209.
192. Kikuchi Y, Horwitz B & Mishkin M (2007). Auditory response properties in the rostral and caudal stations of the auditory stimulus processing stream of the macaque superior temporal cortex. Poster presented at the *Society for Neuroscience Conference, Nov 3-7, San Diego, USA*.
193. Kim DO & Molnar CE (1979). A population study of cochlear nerve fibres: Comparisons of spatial distributions of average-rate and phase-locking measures of responses to single tones. *J Neurophysiol* **42**, 16-30.
194. Kosaki H, Hashikawa T, He J & Jones EG (1997). Tonotopic organization of auditory cortical fields delineated by parvalbumin immunoreactivity in Macaque monkeys. *J Comp Neurol* **386**, 304-416.
195. Kostopoulos P & Petrides M (2008). Left mid-ventrolateral prefrontal cortex: underlying principles of function. *Eur J Neurosci* **27**, 1037-1049.
196. Krimphoff J (1993). *Analyse acoustique et perception du timbre. Unpublished DEA thesis*. Thesis (Le Mans, Université du Maine).
197. Krimphoff J, McAdams S & Winsberg S (1994). Caractérisation du timbre des sons complex. II Analyses acoustiques et quantification psychoacoustique [Characterization of the timbre of complex sounds. II Acoustic analyses and psychophysical quantification]. *J Phys (Paris)* **4**, 625-628.
198. Kropotov JD, Alho K, Näätänen R, Ponomarev VA, Kropotova OV, Anichkov AD & Nechaev VB (2000). Human auditory-cortex mechanisms of preattentive sound discrimination. *Neurosci Lett* **280**, 87-90.
199. Kropotov JD, Näätänen R, Sevostianov AV, Alho K, Reinikainen K & Kropotova OV (1995). Mismatch negativity to auditory stimulus change recorded directly from the human temporal cortex. *Psychophysiology* **32**, 418-422.

200. Krumhansl CL (1989). Why is musical timbre so hard to understand? In, *Structure and Perception of Electroacoustic Sound and Music*, S Nielzen & O Olsen (eds.), Amsterdam: Excerpta Medica, pp. 43-53.
201. Krumhansl CL (1990). *Cognitive Foundations of Musical Pitch*. New York: Oxford University Press.
202. Kubovy M & Van Valkenburg D (2001). Auditory and visual objects. *Cognition* **80**, 97-126.
203. Kujala A, Tervaniemi M & Schröger E (2007). The mismatch in cognitive and clinical neuroscience: theoretical and methodological considerations. *Biol Psychol* **74**, 1-19.
204. Kumar S, Stephan KE, Warren JD, Friston KJ & Griffiths TD (2007). Hierarchical processing of auditory objects in humans. *PLoS Comp Biol* **3**, e100.
205. Lake DE, Richman JS, Griffin MP & Moorman JR (2002). Sample entropy analysis of neonatal heart rate variability. *Amer J Physiol Regul Integr Comp Physiol* **283**, R789-R797.
206. Laureys S, Salmon E, Goldman S & Majerus S (2003). Functional neuroimaging of auditory processing. *Acta Otorhinolaryngol Belg* **57**, 267-273.
207. Lauterbur PC (1973). Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature* **242**, 190-191.
208. Leonard CM, Puranik C, Kuldau JM & Lombardino LJ (1998). Normal variation in the frequency and location of human auditory cortex landmarks. Heschl's gyrus: where is it? *Cereb Cortex* **8**, 397-406.
209. Levänen S, Ahonen A, Hari R, McEvoy L & Sams M (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cereb Cortex* **6**, 288-296.

210. Levänen S, Hari R, McEvoy L & Sams M (1993). Responses of the human auditory cortex to changes in one versus two stimulus features. *Exp Brain Res* **97**, 177-183.
211. Lewicki MS (2002). Efficient coding of natural sounds. *Nat Neurosci* **5**, 356-363.
212. Liégeois-Chauvel C, Giraud AL, Badier JM, Marquis P & Chauvel P (2001). Intracerebral evoked potentials in pitch perception reveal a functional asymmetry of the human auditory cortex. *Ann N Y Acad Sci* **930**.
213. Liégeois-Chauvel C, Lorenzi C, Trébuchon A, Régis J & Chauvel P (2004). Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* **14**, 731-740.
214. Liégeois-Chauvel C, Musolino A, Badier JM, Marquis P & Chauvel P (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalogr Clin Neurophysiol* **92**, 204-214.
215. Liégeois-Chauvel C, Musolino A & Chauvel P (1991). Localization of the primary auditory area in man. *Brain* **114**, 139-153.
216. Liégeois-Chauvel C, Peretz I, Babai M, Laguitton V & Chauvel P (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain* **121**, 1853-1867.
217. List A & Justus T (2007). Auditory priming of frequency and temporal information: effects of lateralised presentation. *Laterality* **12**, 1-30.
218. List A, Justus T, Robertson LC & Bentin S (2007). A mismatch negativity study of local-global auditory processing. *Brain Res* **1153**, 122-133.
219. Logan BR & Rowe D (2004). An evaluation of thresholding techniques in fMRI analysis. *Neuroimage* **22**, 95-108.

220. Logothetis NK (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos Trans R Soc Lond B Biol Sci* **357**, 1003-1037.
221. Logothetis NK (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci* **23**, 3936-3971.
222. Logothetis NK (2004). Interpreting the BOLD signal. *Annu Rev Physiol* **66**, 735-769.
223. Logothetis NK (2008). What we can do and what we cannot do with fMRI. *Nature* **453**, 869-878.
224. Logothetis NK, Pauls J, Augath M, Trinath T & Oeltermann A (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* **412**, 150-157.
225. Lu T, Liang L & Wang X (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci* **4**, 1131-1138.
226. Luo H & Poeppel D (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* **54**, 1001-1010.
227. Luo H, Wang Y, Poeppel D & Simon JZ (2006). The concurrent encoding of frequency and amplitude modulation in human auditory cortex: MEG evidence. *J Neurophysiol* **98**, 3473-3485.
228. MacDougall-Shackleton SA, Hulse HS, Gentner TQ & White W (1998). Auditory scene analysis by European starlings (*Sturnus Vulgaris*): perceptual segregation of tone sequences. *J Acoust Soc Am* **103**, 3581-3587.
229. Maess B, Jacobsen T, Schröger E & Friederici AD (2007). Localizing pre-attentive auditory memory-based comparison: magnetic mismatch negativity to pitch change. *Neuroimage* **37**, 561-571.
230. Malone BJ, Scott BH & Semple MN (2007). Dynamic amplitude coding in the auditory cortex of awake rhesus macaques. *J Neurophysiol* **98**, 1451-1474.

231. Malonek D & Grinvald A (1996). Interactions between electrical activity and cortical microcirculation revealed by imaging spectroscopy: implications for functional brain mapping. *Science* **272**, 551-554.
232. Mansfield P (1977). Multi-planar image formation using NMR spin echoes. *J Phys [C]* **10**, L55-L58.
233. Mansfield P & Grannel PK (1973). NMR 'diffraction' in solids? *J Phys [C]* **6**, L422-L426.
234. Martin M (1979). Hemispheric specialization for local and global processing. *Neuropsychologia* **17**, 33-40.
235. Mazziotta JC, Toga A, Evans A, Fox P & Lancaster J (1995). A probabilistic atlas of the human brain: theory and rationale for its development. *Neuroimage* **2**, 89-101.
236. McAdams S (1989). Segregation of concurrent sounds. I. Effects of frequency-modulation coherence. *J Acoust Soc Am* **86**, 2148-2159.
237. McAdams S (1993). Recognition of sound sources and events. In, *Thinking in Sound: The Cognitive Psychology of Human Audition*, S McAdams & E Bigand (eds.), Oxford: Oxford University Press, pp. 146-198.
238. McAdams S (1996). Audition: Cognitive psychology of music. In, *The Mind-Brain Continuum: Sensory Processes*, R Llinàs & PS Churchland (eds.), Cambridge, MA: MIT Press, pp. 251-279.
239. McAdams S & Cunibile JC (1992). Perception of timbral analogies. *Philos Trans R Soc Lond B Biol Sci* **336**, 383-389.
240. McAdams S, Winsberg S, Donnadieu S, De Soete G & Krimphoff J (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res* **58**, 177-192.
241. McAlpine D (2005). Creating a sense of auditory space. *J Physiol (Lond)* **566**, 21-28.

242. McDermott J & Hauser M (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Percept* **23**, 29-59.
243. Mendelson JR & Cynader MS (1985). Sensitivity of cat primary auditory-cortex (A1) neurons to the direction and rate of frequency modulation. *Brain Res* **327**, 331-335.
244. Menon V, Levitin DJ, Smith BK, Lembke A, Krasnow BD, Glazer D, Glover GH & McAdams S (2002). Neural correlates of timbre change in harmonic sounds. *Neuroimage* **17**, 1742-1754.
245. Micheyl C, Carlyon RP, Gutschalk A, Melcher JR, Oxenham AJ, Rauschecker JP, Tian B & Courtenay Wilson E (2007). The role of auditory cortex in the formation of auditory streams. *Hear Res* **229**, 116-131.
246. Micheyl C, Tian B, Carlyon RP & Rauschecker JP (2005). Perceptual organization of tone sequences in the auditory cortex of awake macaques. *Neuron* **48**, 139-148.
247. Middlebrooks JC (2002). Auditory space processing: here, there or everywhere? *Nat Neurosci* **5**, 824-826.
248. Middlebrooks JC (2008). Auditory cortex phase locking to amplitude-modulated cochlear implant pulse trains. *J Neurophysiol* **100**, 76-91.
249. Miller LM, Escabi MA, Read HL & Schreiner CE (2001). Functional convergence of response properties in the auditory thalamocortical system. *Neuron* **32**, 151-160.
250. Milner AD & Goodale MA (1995). *The Visual Brain in Action*. Oxford: Oxford University Press.
251. Moelker A & Pattynama PMT (2003). Acoustic noise concerns in functional magnetic resonance imaging. *Hum Brain Mapp* **20**, 123-141.
252. Moore BCJ (1999). Modulation minimizes masking. *Nature* **397**, 108-109.

253. Moore BCJ (2003). *An Introduction to the Psychology of Hearing (Fifth Edition)*. London: Academic Press.
254. Moore BCJ & Gockel H (2002). Factors influencing sequential stream segregation. *Acta Acoustica* **88**, 320-333.
255. Morel A, Garraghty PE & Kaas JH (1993). Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol* **335**, 437-459.
256. Morosan P, Rademacher J, Schleicher A, Amunts K, Schormann T & Zilles K (2001). Human primary auditory cortex: cytoarchitectonic subdivisions and mapping into a spatial reference system. *Neuroimage* **13**, 684-701.
257. Morosan P, Schleicher A, Amunts K & Zilles K (2005). Multimodal architectonic mapping of human superior temporal gyrus. *Anat Embryol (Berl)* **210**, 401-406.
258. Näätänen R (1995). The mismatch negativity: a powerful tool for cognitive neuroscience. *Ear Hear* **16**, 6-18.
259. Näätänen R (2000). Mismatch negativity (MMN): perspectives for application. *Int J Psychophysiol* **37**, 3-10.
260. Näätänen R (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* **38**, 1-21.
261. Näätänen R (2003). Mismatch negativity: clinical research and possible applications. *Int J Psychophysiol* **48**, 179-188.
262. Näätänen R, Gaillard AWK & Mäntysalo S (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol (Amst)* **42**, 313-329.
263. Näätänen R, Jacobsen T & Winkler I (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* **42**, 25-32.

264. Näätänen R, Lehtokoski A, Lennes M, Cheour M, Huotilainen M, Iivonen A, Vainio M, Alku P, Ilmoniemi RJ, Luuk A, *et al.* (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* **385**, 432-434.
265. Näätänen R, Paavilainen P, Rinne T & Alho K (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol* **118**, 2544-2590.
266. Näätänen R & Picton TW (1987). The N1 wave of the human electric and magnetic response to sound: a review and analysis of component structure. *Psychophysiology* **24**, 375-425.
267. Näätänen R & Winkler I (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol Bull* **125**, 826-859.
268. Nahum M, Nelken I & Ahissar M (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS Biol* **6**, e126.
269. Nandy R & Cordes D (2007). A semi-parametric approach to estimate the family-wise error rate in fMRI using resting-state data. *Neuroimage* **34**, 1562-1576.
270. Narain C, Scott SK, Wise RJ, Rosen S, Leff A, Iversen SD & Matthews PM (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb Cortex* **13**, 1362-1368.
271. Navon D (1977). Forest before trees: the precedence of global features in visual perception. *Cognit Psychol* **9**, 353-383.
272. Nelken I (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol* **14**, 474-480.
273. Nelken I & Chechik G (2007). Information theory in auditory research. *Hear Res* **229**, 94-105.
274. Nelken I, Rotman Y & Bar Yosef O (1999). Responses of auditory-cortex neurons to structural features of natural sounds. *Nature* **397**, 154-157.

275. Niessing J, Ebisch B, Schmidt KE, Niessing M, Singer W & Galuske RA (2005). Hemodynamic signals correlate tightly with synchronized gamma oscillations. *Science* **309**, 948-951.
276. Obleser J, Eisner F & Kotz SA (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J Neurosci* **28**, 8116-8123.
277. Ogawa S & Lee TM (1990). Magnetic resonance imaging of blood vessels at high fields: in vivo and in vitro measurements and image stimulation. *Magn Reson Med* **16**, 9-18.
278. Ogawa S, Lee TM, Kay AR & Tank DW (1990a). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc Natl Acad Sci U S A* **87**, 9768-9872.
279. Ogawa S, Lee TM, Nayak AS & Glynn P (1990b). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Magn Reson Med* **14**, 68-78.
280. Ohl FW, Scheich H & Freeman WJ (2001). Change in pattern of ongoing neural activity with auditory category learning. *Nature* **412**, 733-736.
281. Olshausen BA & Field DJ (2004). Sparse coding of sensory inputs. *Curr Opin Neurobiol* **14**, 481-487.
282. Opitz B, Mecklinger A, von Cramon DY & Kruggel F (1999). Combining electrophysiological and hemodynamic measures of the auditory oddball. *Psychophysiology* **36**, 142-147.
283. Opitz B, Rinne T, Mecklinger A, Von Cramon DY & Schröger E (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage* **15**, 167-174.
284. Overath T (2004a). *Auditory chimaeras. Music-speech laterality revisited*. MSc Thesis (Oxford, University of Oxford).

285. Overath T (2004b). *Expression of the immediate-early gene proteins Fos and Egr-1 in ferret auditory cortex*. MSc Thesis (Oxford, University of Oxford).
286. Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP & Griffiths TD (2007). An information theoretic characterisation of auditory encoding. *PLoS Biol* **5**, e288.
287. Oxenham AJ, Bernstein JG & Penagos H (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc Natl Acad Sci U S A* **101**, 1421-1425.
288. Pandya DN (1995). Anatomy of the auditory cortex. *Rev Neurol (Paris)* **151**, 486-494.
289. Passingham D & Sakai K (2004). The prefrontal cortex and working memory: physiology and brain imaging. *Curr Opin Neurobiol* **14**, 163-168.
290. Patel AD (2008). *Music, Language, and the Brain*. New York: Oxford University Press.
291. Patel AD & Balaban E (2000). Temporal patterns of human cortical activity reflect tone sequence structure. *Nature* **404**, 80-84.
292. Patel AD & Iversen JR (2008). The linguistic benefits of musical abilities. *Trends Cogn Sci* **11**, 369-372.
293. Patel AD, Peretz I, Tramo M & Labreque R (1998). Processing prosodic and musical patterns: a neuropsychological investigation. *Brain Lang* **61**, 123-144.
294. Patterson RD, Handel S, Yost WA & Datta AJ (1996). The relative strength of the tone and the noise components in iterated rippled noise. *J Acoust Soc Am* **100**, 3286-3294.
295. Patterson RD, Uppenkamp S, Johnsrude IS & Griffiths TD (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* **36**, 767-776.

296. Pauling L & Coryell C (1936). The magnetic properties and structure of hemoglobin. *Proc Natl Acad Sci U S A* **22**, 210-216.
297. Pearce MT & Wiggins GA (2004). Improved methods for statistical modelling of monophonic music. *J New Music Res* **33**, 367-385.
298. Pearce MT & Wiggins GA (2006). Expectation in melody: The influence of context and learning. *Music Percept* **23**, 377-405.
299. Pelleg-Toiba A & Wollberg Z (1989). Tuning properties of auditory cortex cells in awake squirrel monkeys. *Exp Brain Res* **74**, 353-364.
300. Penagos H, Melcher JR & Oxenham AJ (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci* **24**, 6810-6815.
301. Penny WD, Stephan KE, Mechelli A & Friston KJ (2004). Comparing dynamic causal models. *Neuroimage* **22**, 1157-1172.
302. Peretz I (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain* **113**, 1185-1205.
303. Peretz I, Kolinsky R, Tramo M, Labreque R, Hublet C, Demeurisse G & Belleville S (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain* **117**, 1283-1302.
304. Petkov CI, Kayser C, Augath M & Logothetis NK (2006). Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* **4**, e215.
305. Petrides M & Pandya D (1994). Comparative architectonic analysis of the human and macaque frontal cortex. In, *Handbook of Neuropsychology*, F Boller & J Grafman (eds.), Amsterdam: Elsevier, pp. 17-58.
306. Petrides M & Pandya D (2002). Comparative cytoarchitectonic analysis in the human and in the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur J Neurosci* **16**, 291-310.

307. Plack CJ & Oxenham AJ (2005). The psychophysics of pitch. In, *Pitch. Neural Coding and Perception*, CJ Plack, AJ Oxenham, RR Fay & AN Popper (eds.), New York: Springer, pp. 7-55.
308. Plack CJ, Oxenham AJ, Fay RR & Popper AN (eds.) (2005). *Pitch. Neural Coding and Perception*. New York: Springer.
309. Poeppel D (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Commun* **41**, 245-255.
310. Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P & Mishkin M (2004). Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* **427**, 448-451.
311. Pressnitzer D, Sayles M, Micheyl C & Winter IM (2008). Perceptual organization of sound begins in the auditory periphery. *Curr Biol* **18**, 1124-1128.
312. Price CJ (2000). The anatomy of language: contributions from functional imaging. *J Anat* **197**, 335-359.
313. Price DL, De Wilde JP, Papadaki AM, Curran JS & Kitney RI (2001). Investigation of acoustic noise on 15 MRI scanners from 0.2 to 3 T. *J Magn Reson Imaging* **13**, 288-293.
314. Purcell EM, Torrey HC & Pound RV (1946). Resonance absorption by nuclear magnetic resonance. *Magn Reson Q* **5**, 263-281.
315. Rademacher J, Caviness VSJ, Steinmetz H & Galaburda AM (1993). Topographical variation of the human primary cortices: implications for neuroimaging, brain mapping, and neurobiology. *Cereb Cortex* **3**, 313-329.
316. Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ & Zilles K (2001). Probabilistic mapping and volume measurement of human primary auditory cortex. *Neuroimage* **13**, 669-683.

- 
317. Rauschecker JP (1998). Cortical processing of complex sounds. *Curr Opin Neurobiol* **8**, 516-521.
318. Rauschecker JP & Tian B (2000). Mechanisms and streams for processing "what" and "where" in the auditory cortex. *Proc Natl Acad Sci U S A* **97**, 11800-11806.
319. Rauschecker JP, Tian B & Hauser M (1995). Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* **268**, 111-114.
320. Rauschecker JP, Tian B, Pons T & Mishkin M (1997). Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol* **382**, 89-103.
321. Ravicz ME & Melcher JR (2001). Isolating the auditory system from acoustic noise during functional magnetic resonance imaging: examination of noise conduction through the ear canal, head, and body. *J Acoust Soc Am* **109**, 216-231.
322. Ravicz ME, Melcher JR & Kiang NY (2000). Acoustic noise during functional magnetic resonance imaging. *J Acoust Soc Am* **108**, 1683-1696.
323. Recanzone GH, Guard DC & Phan ML (2000). Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *J Neurophysiol* **83**, 2315-2331.
324. Rees A & Malmierca MS (2005). Processing of dynamic spectral properties of sounds. *Int Rev Neurobiol* **70**, 299-330.
325. Richman JS & Moorman JR (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am J Physiol Heart Circ Physiol* **278**, H2039-H2049.
326. Ricketts C, Mendelson JR, Anand B & English R (1998). Responses to time-varying stimuli in rat auditory cortex. *Hear Res* **123**, 27-30.
327. Rilling JK & Seligman RA (2002). A quantitative morphometric comparative analysis of the primate temporal lobe. *J Hum Evol* **42**, 505-533.

328. Rinne T, Degerman A & Alho K (2005). Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study. *Neuroimage* **26**, 66-72.
329. Rivier F & Clarke S (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: Evidence for multiple auditory areas. *Neuroimage* **6**, 288-304.
330. Rizzolatti G, Fogassi L & Gallese V (1997). Parietal cortex: from sight to action. *Curr Opin Neurobiol* **7**, 562-567.
331. Roberts B, Glasberg BR & Moore BCJ (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J Acoust Soc Am* **112**, 2074-2085.
332. Robertson LC (1996). Attentional persistence for features of hierarchical patterns. *J Exp Psychol Gen* **125**, 227-249.
333. Rogers H (2000). *The Sounds of Language. An Introduction to Phonetics*. Harlow, Essex: Pearson.
334. Roland PE & Zilles K (1994). Brain atlases - a new research tool. *Trends Neurosci* **17**, 458-467.
335. Romanski LM, Bates JF & Goldman-Rakic PS (1999). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *J Comp Neurol* **403**, 141-157.
336. Rosburg T (2003). Left hemispheric dipole locations of the neuromagnetic mismatch negativity to frequency, intensity and duration deviants. *Cognitive Brain Research* **16**, 83-90.
337. Rosburg T, Trautner P, Dietl T, Korzyukov OA, Boutros NN, Schaller C, Elger CE & Kurthen M (2005). Subdural recordings of the mismatch negativity (MMN) in patients with focal epilepsy. *Brain* **128**, 819-828.
338. Rosen S (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci* **336**, 367-373.

339. Russo FA & Thompson WF (2005). The subjective size of melodic intervals over a two-octave range. *Psychon Bull Rev* **12**, 1068-1075.
340. Sams M, Hämäläinen M, Antervo A, Kaukoranta E, Reinikainen K & Hari R (1985a). Cerebral neuromagnetic responses evoked by short auditory stimuli. *Electroencephalogr Clin Neurophysiol* **61**, 254-266.
341. Sams M, Kaukoranta E, Hämäläinen M & Näätänen R (1991). Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. *Psychophysiology* **28**, 21-29.
342. Sams M, Paavilainen P, Alho K & Näätänen R (1985b). Auditory frequency discrimination and event-related potentials. *Electroencephalogr Clin Neurophysiol* **62**, 437-448.
343. Sanders LD & Poeppel D (2007). Local and global auditory processing: behavioral and ERP evidence. *Neuropsychologia* **45**, 1172-1186.
344. Scheich H, Baumgart F, Gaschler-Markefski B, Tegeler C, Tempelmann C, Heinze HJ, Schindler F & Stiller D (1998). Functional magnetic resonance imaging of a human auditory cortex area involved in foreground-background composition. *Eur J Neurosci* **10**, 803-809.
345. Scherg M, Vajsar J & Picton TW (1989). A source analysis of the late human evoked potentials. *J Cogn Neurosci* **1**, 336-355.
346. Schiavetto A, Cortese F & Alain C (1999). Global and local processing of musical sequences: an event-related brain potential study. *Neuroreport* **10**, 2467-2472.
347. Schmuckler MA & Gildea DL (1993). Auditory perception of fractal contours. *J Exp Psychol Hum Percept Perform* **19**, 641-660.
348. Schnupp JW (2001). Linear processing of spatial cues in primary auditory cortex. *Nature* **414**, 200-204.

349. Schnupp JW, Hall TM, Kokelaar RF & Ahmed B (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* **26**, 4785-4795.
350. Scholl BJ (2001). Objects and attention: state of the art. *Cognition* **80**, 1-46.
351. Scholte HS, Jolij J, Fahrenfort JJ & Lamme VA (2008). Feedforward and recurrent processing in scene segmentation: electroencephalography and functional magnetic resonance imaging. *J Cogn Neurosci* **20**, 1-13.
352. Schönberg A (1911). *Harmonielehre*. Leipzig: Universal Edition.
353. Schönwiesner M, Novitski N, Pakarinen S, Carlson S, Tervaniemi M & Näätänen R (2007). Heschl's gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of changes. *J Neurophysiol* **97**, 2075-2083.
354. Schönwiesner M, Rübsem R & von Cramon DY (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci* **22**, 1521-1528.
355. Schönwiesner M, von Cramon DY & Rübsem R (2002). Is it tonotopy after all? *Neuroimage* **17**, 1144-1161.
356. Schreiner CE & Langner G (1988). Periodicity coding in the inferior colliculus of the cat. II. Topographical organization. *J Neurophysiol* **60**, 1823-1840.
357. Schreiner CE & Langner G (1997). Laminar fine structure of frequency organization in auditory midbrain. *Nature* **388**, 383-386.
358. Schuppert M, Münte TM, Wieringa BM & Altenmüller E (2000). Receptive amusia: evidence for cross-hemispheric neural networks underlying music processing strategies. *Brain* **123**, 546-559.
359. Scott SK (2005). Auditory processing - speech, space and auditory objects. *Curr Opin Neurobiol* **15**, 197-201.

360. Scott SK, Blank CC, Rosen S & Wise RJ (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* **123**, 2400-2406.
361. Scott SK, Rosen S, Lang H & Wise RJ (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech - a positron emission tomography study. *J Acoust Soc Am* **120**, 1075-1083.
362. Shamma SA (2008). On the emergence and awareness of auditory objects. *PLoS Biol* **6**, e155.
363. Shamma SA & Symmes D (1985). Patterns of inhibition in auditory cortical cells in awake squirrel monkeys. *Hear Res* **19**, 1-13.
364. Shannon CE (1948). A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379-423 and 623-656.
365. Shannon CE (1949). Communication in the presence of noise. *Proc IRE* **37**, 10-21.
366. Shannon RV (2005). Speech and music have different requirements for spectral resolution. *Int Rev Neurobiol* **70**, 121-134.
367. Shannon RV, Zeng FG, Kamath V, Wygonski J & Ekelid M (1995). Speech recognition with primarily temporal cues. *Science* **270**, 303-304.
368. Shulman GL, Sullivan MA, Gish K & Sakoda WJ (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception* **15**, 259-273.
369. Smith ZM, Delgutte B & Oxenham AJ (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature* **416**, 87-90.
370. Snyder JS & Alain C (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychol Bull* **133**, 780-799.
371. Snyder JS, Alain C & Picton TW (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *J Cogn Neurosci* **18**, 1-13.

372. Steinmetz H, Rademacher J, Huang YX, Hefter H, Zilles K, Thron A & Freund HJ (1989). Cerebral asymmetry: MR planimetry of the human planum temporale. *J Comput Assist Tomogr* **13**, 996-1005.
373. Stewart L, von Kriegstein K, Warren JD & Griffiths TD (2006). Music and the brain: disorders of musical listening. *Brain* **129**, 2533-2553.
374. Strange BA, Duggins A, Penny W, Dolan RJ & Friston KJ (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks* **18**, 225-230.
375. Summerfield AQ & Culling J (1992). Auditory segregation of competing voices: absence of effects of FM or AM coherence. *Philos Trans R Soc Lond B Biol Sci* **336**, 357-366.
376. Sweet RA, Dorph-Petersen KA & Lewis DA (2005). Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *J Comp Neurol* **491**, 270-289.
377. Talairach P & Tournoux J (1988). *A Stereotactic Coplanar Atlas of the Human Brain*. Stuttgart: Thieme.
378. Talavage TM & Edmister WB (2004). Nonlinearity of fMRI responses in human auditory cortex. *Hum Brain Mapp* **22**, 216-228.
379. Talavage TM, Edmister WB, Ledden PJ & Weisskoff RM (1999). Quantitative assessment of auditory cortex responses induced by imager acoustic noise. *Hum Brain Mapp* **7**, 79-88.
380. Talavage TM, Ledden PJ, Benson RR, Rosen BR & Melcher JR (2000). Frequency-dependent responses exhibited by multiple regions in human auditory cortex. *Hear Res* **150**, 225-244.
381. Temple E, Poldrack RA, Protopapas A, Nagarajan S, Salz T, Tallal P, Merzenich MM & Gabrieli JD (2000). Disruption of the neural response to rapid acoustic stimuli in dyslexia: evidence from functional MRI. *Proc Natl Acad Sci U S A* **97**, 13907-13912.

- 
382. Terhardt E (1974). Pitch, consonance, and harmony. *J Acoust Soc Am* **55**, 1061-1069.
383. Tian B & Rauschecker JP (1998). Processing of frequency-modulated sounds in the cat's posterior auditory field. *J Neurophysiol* **79**, 2629-2642.
384. Tian B & Rauschecker JP (2004). Processing of frequency-modulated sounds in the lateral auditory belt cortex of the Rhesus monkey. *J Neurophysiol* **92**, 2993-3013.
385. Tian B, Reser D, Durham A, Kustov A & Rauschecker JP (2001). Functional specialization in rhesus monkey auditory cortex. *Science* **292**, 290-293.
386. Tiitinen H, May P, Reinikainen K & Näätänen R (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* **370**, 90-92.
387. Toga A, Ambach K, Quinn B, Hutchin M & Burton J (1994). Postmortem anatomy from cryosectioned whole human brain. *J Neurosci Methods* **54**, 239-252.
388. Trainor LJ, Desjardins RN & Rockel C (1999). A comparison of contour and interval processing in musicians and nonmusicians using event-related potentials. *Australian J Psychol* **51**, 147-153.
389. Trainor LJ, McDonald KL & Alain C (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *J Cogn Neurosci* **14**, 430-442.
390. Trehub SE, Unyk AM & Trainor LJ (1993). Adults identify infant-directed music across cultures. *Inf Behav Dev* **16**, 193-211.
391. Treisman AM & Gelade G (1980). A feature-integration theory of attention. *Cognit Psychol* **12**, 97-136.
392. Ungerleider LG & Haxby JV (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol* **4**, 157-165.

393. Upadhyay J, Ducros M, Knaus TA, Lindgren KA, Silver A, Tager-Flusberg H & Kim DS (2007). Function and connectivity in human auditory cortex: a combined fMRI and DTI study at 3 Tesla. *Cereb Cortex* **17**, 2420-2432.
394. Upadhyay J, Silver A, Knaus TA, Lindgren KA, Ducros M, Kim DS & Tager-Flusberg H (2008). Effective and structural connectivity in the human auditory cortex. *J Neurosci* **28**, 3341-3349.
395. van Noorden LPAS (1975). *Temporal Coherence in the Perception of Tone Sequences*. PhD Thesis (Eindhoven, University of Technology).
396. van Zuijen TL, Sussman E, Winkler I, Näätänen R & Tervaniemi M (2004). Grouping of sequential sounds--an event-related potential study comparing musicians and nonmusicians. *J Cogn Neurosci* **16**, 331-338.
397. Vliegen J, Moore BCJ & Oxenham AJ (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *J Acoust Soc Am* **106**, 938-945.
398. Vliegen J & Oxenham AJ (1999). Sequential stream segregation in the absence of spectral cues. *J Acoust Soc Am* **105**, 339-346.
399. von Békésy G (1960). *Experiments in Hearing*. (EG Wever, ed.) New York: McGraw-Hill.
400. von Economo C & Horn L (1930). Über Windungsrelief, Masse und Rindenarchitektur der Supratemporalfläche, ihre individuellen und ihre Seitenunterschiede. *Z Neurol Psychiatr* **130**, 678-757.
401. von Economo C & Koskinas G (1925). *Die Cytoarchitektur der Hirnrinde des erwachsenen Menschen*. Berlin: Julius Springer Verlag.
402. von Kriegstein K, Eger E, Kleinschmidt A & Giraud AL (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res Cogn Brain Res* **17**, 48-55.

403. von Kriegstein K & Giraud AL (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* **22**, 948-955.
404. von Kriegstein K & Giraud AL (2006). Implicit multisensory associations influence voice recognition. *PLoS Biol* **4**, 1809-1820.
405. von Kriegstein K, Warren JD, Ives DT, Patterson RD & Griffiths TD (2006). Processing the acoustic effect of size in speech sounds. *Neuroimage* **32**, 368-375.
406. Voss RF & Clarke J (1975). 1/f noise in music and speech. *Nature* **258**, 317-318.
407. Voss RF & Clarke J (1978). 1/f noise in music: music from 1/f noise. *J Acoust Soc Am* **63** 258-263.
408. Vuust P, Pallesen KJ, Bailey C, van Zuijen TL, Gjedde A, Roepstorff A & Ostergaard L (2005). To musicians, the message is in the meter. Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* **24**, 560-564.
409. Wallace MN, Johnston PW & Palmer AR (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Exp Brain Res* **143**, 499-508.
410. Wang X (2000). On cortical coding of vocal communications sounds in primates. *Proc Natl Acad Sci U S A* **97**, 11843-11849.
411. Wang X & Kadia SC (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J Neurophysiol* **86**, 2616-2620.
412. Wang X, Liu T & Liang L (2003). Cortical processing of temporal modulations. *Speech Commun* **41**, 107-121.
413. Warren JD, Jennings AR & Griffiths TD (2005a). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage* **24**, 1052-1057.

414. Warren JD, Uppenkamp S, Patterson RD & Griffiths TD (2003). Separating pitch chroma and pitch height in the human brain. *Proc. Natl. Acad. Sci. USA* **100**, 10038-10042.
415. Warren JE, Wise KJ & Warren JD (2005b). Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci* **28**, 636-643.
416. Warren RM (2008). *Auditory Perception. An Analysis and Synthesis (Third Edition)*. Cambridge: Cambridge University Press.
417. Wehr M & Zador AM (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442-446.
418. Wertheimer M (1922). Untersuchungen zur Lehre von der Gestalt. I. Prinzipielle Bemerkungen. *Psychol Forsch* **1**, 47-58.
419. Wertheimer M (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychol Forsch* **4**, 301-350.
420. Westbury CF, Zatorre RJ & Evans AC (1999). Quantifying variability in the planum temporale: a probability map. *Cereb Cortex* **9**, 392-405.
421. Wichmann FA & Hill NJ (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys* **63**, 1314-1329.
422. Wilson EC, Melcher J, Micheyl C, Gutschalk A & Oxenham AJ (2007). Cortical fMRI activation to sequences of tones alternating in frequency: relationship to perceived rate and streaming. *J Neurophysiol* **97**, 2230-2238.
423. Winer JA & Lee CC (2007). The distributed auditory cortex. *Hear Res* **229**, 3-13.
424. Winkler I, Lehtokoski A, Alku P, Vainio M, Czigler I, Csépe V, Aaltonen O, Raimo I, Alho K, Lang AH, *et al.* (1999). Pre-attentive change detection utilizes both phonetic and auditory memory representations in across- and within-category vowel contrasts. *Cognitive Brain Research* **7**, 357-369.

425. Winkler I, Tervaniemi M & Näätänen R (1997). Two separate codes for missing-fundamental pitch in the human auditory cortex. *J Acoust Soc Am* **102**, 1072-1082.
426. Winkler I, van Zuijen TL, Sussman E, Horvath J & Näätänen R (2006). Object representation in the human auditory system. *Eur J Neurosci* **24**, 625-634.
427. Wu GK, Arbuckle R, Liu BH & Zhang LI (2008). Lateral sharpening of cortical frequency tuning by approximately balanced inhibition. *Neuron* **58**, 132-143.
428. Yost WA, Patterson R & Sheft S (1996). A time domain description for the pitch strength of iterated rippled noise. *J Acoust Soc Am* **99**, 1066-1978.
429. Yu Y, Romero R & Lee TS (2005). Preference of sensory neural coding for 1/f signals. *Phys Rev Lett* **94**, 108103.
430. Zatorre RJ (1985). Discrimination and recognition of tonal melodies after unilateral cerebral excisions. *Neuropsychologia* **23**, 31-41.
431. Zatorre RJ (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am* **84**, 566-572.
432. Zatorre RJ & Belin P (2001). Spectral and temporal processing in human auditory cortex. *Cereb Cortex* **11**, 946-953.
433. Zatorre RJ, Belin P & Penhune VB (2002a). Structure and function of auditory cortex: music and speech. *Trends Cogn Sci* **6**, 37-46.
434. Zatorre RJ, Bermudez P, Warrier CM & Evans AC (1998). Cerebral mechanisms associated with encoding and recognition of melodies. *Neuroimage* **7**, S849.
435. Zatorre RJ, Bouffard M, Ahad P & Belin P (2002b). Where is 'where' in the human auditory cortex? *Nat Neurosci* **5**, 905-909.
436. Zatorre RJ, Bouffard M & Belin P (2004). Sensitivity to auditory object features in human temporal neocortex. *J Neurosci* **24**, 3637-3642.

437. Zatorre RJ, Evans AC & Meyer E (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J Neurosci* **14**, 1908-1919.
438. Zatorre RJ & Gandour J (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philos Trans R Soc Lond B Biol Sci* **363**, 1087-1104.
439. Zhang LI, Tan AY, Schreiner CE & Merzenich MM (2003). Topography and synaptic shaping of direction selectivity in primary auditory cortex. *Nature* **424**, 201-205.

## APPENDIX I: AUTHOR CONTRIBUTIONS

**Study 1:** The author was involved in the design of the experimental stimulus, designed the study, acquired and analysed the data, and was involved in writing the published manuscript (see Appendix II). Sukhbinder Kumar was involved in the design of the experimental stimulus and data analysis, and Katharina von Kriegstein helped with data acquisition and data analysis. Tim Griffiths was involved in the design of the experimental stimulus, analysis of the data, and writing the manuscript.

**Study 2:** The author was involved in adapting the experimental stimulus and implementing the appropriate experimental design, acquired and analysed the data, and was involved in writing the submitted manuscript (see Appendix II). Sukhbinder Kumar created the experimental stimulus, helped with data analysis and wrote the submitted manuscript. Katharina von Kriegstein was involved with data acquisition and data analysis. Lauren Stewart performed pilot studies with the stimulus. Tim Griffiths created the stimulus, was involved in the design study, analysis of the data, and writing the submitted manuscript. Rhodri Cusack gave valuable advice on data analysis. Adrian Rees was involved in writing the submitted manuscript.

**Studies 3 & 4:** The author was involved in adapting the experimental stimulus and the design of the study, acquired and analysed the data and was involved in writing the published manuscript (see Appendix II). Rhodri Cusack was involved in the design of the study, analysis of the data, and writing of the manuscript. Sukhbinder Kumar was involved in the design of Study 3 and information theoretic analysis of the data. Katharina von Kriegstein helped acquiring and analysing the data. Jason Warren was involved in the conceptualisation of the stimulus. Manon Grube contributed to the design of Study 3. Bob Carlyon contributed to the design of Study 3 and the interpretation of the results. Tim Griffiths created the experimental stimulus, designed the studies, and was involved in writing the manuscript.

**Study 5:** The author analysed the data and was involved in writing the published manuscript (see Appendix II). Lauren Stewart designed the study, acquired and analysed the data, and was involved in writing the manuscript. Jason Warren and Jessica Foxton were involved in the design of the study. Tim Griffiths created the stimulus, designed the study, and was involved in writing the manuscript.

## APPENDIX II: PUBLICATIONS ARISING FROM THIS THESIS

Overath T, Cusack R, Kumar S, von Kriegstein K, Warren JD, Grube M, Carlyon RP, Griffiths TD (2007) An information theoretic characterisation of auditory encoding. *PLoS Biology* **5**: e288.

Stewart L, Overath T (joint first author), Warren JD, Foxton JM, Griffiths TD (2008) fMRI evidence for a cortical hierarchy of pitch pattern processing. *PLoS ONE* **3**: e1470.

Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. *J Neurosci* **28**: 13268-13273.

Overath T, Kumar S (joint first author), Stewart L, von Kriegstein K, Cusack R, Rees A, Griffiths TD (under review) Cortical mechanisms for the segregation and representation of auditory objects.