The Perception and Cortical Processing of Communication Sounds



Kerry M. M. Walker, M.Sc.

Brasenose College, University of Oxford

A thesis submitted for the degree of

Doctorate of Philosophy

Trinity 2008

To Nan, Pop and Dad.

Acknowledgements

If a doctorate student is very lucky, they will happen to pick an excellent supervisor. I was fortunate enough to choose two - Prof. Andrew King and Dr. Jan Schnupp.

Jan, your most off-the-wall ideas have so often turned out to be shear brilliance. I hope I've managed to inherit some of your creativity, your now famous ability to inspire an audience, and, of course, your uncanny ability to make TDT kit run. Thanks for your constant encouragement and enthusiasm.

Andy, I'll never understand how you manage to hold an impossibly thorough grasp on the current literature, remain dedicated to running ferret behaviour, and always find time to "talk science" with us graduate students in face of a million other deadlines that keep the lab running. We appreciate everything you do to make our science possible, and I continue to learn from you with every conversation we share.

Huge thanks are due to my mentor and dear friend, Jenny Bizley. I'm sure it's no secret what an inspiration you are to me. Working alongside you has truly made me a better scientist, and the giggles have made our work a joy. This work would not have been possible without you, to say the least.

I am highly grateful to Brasenose College, for giving my graduate career that unique and wonderful Oxford experience. This was made possible due to the generous support of the Rothermere Fellowship, the Hector Pilling Scholarship of Brasenose College, and research grants from the Biotechnology and Biological Sciences Research Council and the Wellcome Trust.

Finally, thank you Simon, for always believing in me and for your crazy, physics-based point-of-view.

The Perception and Cortical Processing of Communication Sounds

Kerry M. M. Walker, M.Sc.

Brasenose College, University of Oxford

A thesis submitted for the degree of Doctorate of Philosophy

Trinity 2008

The neural processes used to extract perceptual features of vocal calls, and subsequently to re-integrate those features to form a coherent auditory object, are poorly understood. In this thesis, extracellular recordings were carried out in order to investigate how the temporal envelope, pitch, timbre and spatial location of communication sounds are represented by neurons in two core and three belt areas of ferret (*Mustela putorius furo*) auditory cortex. Potential neural underpinnings of auditory perception were tested using neurometric analysis to relate the reliability of neural responses to the performance of ferret and human listeners on psychophysical tasks.

I found that human listeners' discrimination of the temporal envelopes of vocalization sounds matched the best neurometrics calculated from the temporal spiking patterns of ferret cortical neurons. Neurometric scores based on the spike rates of cortical neurons accounted for ferrets' discrimination of the pitch of artificial vowels. I show that most auditory cortical neurons are modulated by a number of stimulus features, rather than being tuned to only one feature. Neurons in the core auditory cortical fields often respond uniquely to particular combinations of pitch and timbre features, while those in belt regions respond more linearly to feature combinations. Subtle differences in the sensitivity of neurons to pitch, timbre and azimuthal cues were found across cortical areas and depths. These results suggest that auditory cortical neurons provide widely distributed representations of vocalizations, and a single neuron can often use combinations of spike rate and temporal spiking responses to encode multiple sound features.

List of Abreviations

2AFC	Two-Alternative Forced Choice
A1 or AI	Primary Auditory Cortex
A2 or AII	Secondary Auditory Cortex
AAF	Anterior Auditory Field
ADF	Anterior Dorsal Field
AEG	Anterior Ectosylvian Gyrus
aes	anterior actosylvian sulcus
AES	Auditory field of the anterior Ectosylvian Sulcus
AVF	Anterior Ventral Field
ANOVA	Analysis Of Variance
CF	Characteristic Frequency
CN	Cochlear Nucleus
dPE	dorsal Posterior Ectosylvian gyrus
DZ	Dorsal Zone
fAES	Anterior Ectosylvian Sulcal Field
FDL	Frequency Difference Limen
fMRI	functional Magnetic Resonance Imaging
FRA	Frequency Response Area
HRTF	Head-Related Transfer Function
IC	Inferior Colliculus
ILD	Interaural Level Difference
IN	Insular region
iPE	intermediate Posterior Ectosylvian gyrus
ISI	Inter-Stimulus Interval
ITD	Interaural Timing Difference
MEG	Middle Ectosylvian Gyrus
MGB	Medial Geniculate Body

PAF	Posterior Auditory Field
PEG	Posterior Ectosylvian Gyrus
pes	posterior ectosylvian sulcus
PPF	Posterior Pseudosylvian Field
PSF	Posterior Suprasylvian Field
pss	pseudosylvian sulcus
PC	Principal Component
PCA	Principal Components Analysis
PSTH	Post-Stimulus Time Histogram
ss	suprasylvian sulcus (cat)
sss	suprasylvian sulcus (ferret)
T	Temporal region
VAF	Ventral Auditory Field
VAS	Virtual Acoustic Space
VP	Ventroposterior area
VPAF	Ventral Posterior Auditory Field
vPE	ventral Posterior Ectosylvian gyrus

Contents

1 General Introduction **2**

1.1 The big questions 2

1.2 What characterizes a vocal call? 4

1.3 Subcortical processing of complex sounds 8

1.4 Auditory cortical processing of complex sounds 12

1.5 Thesis overview 17

2 Neurophysiological Methods **21**

2.1 Introduction 21

2.2 Anaesthetised preparation 21

2.3 Unanaesthetised preparation 23

2.4 Stimulus presentation 24

2.5 Stimuli used to identify and characterize neural units 24

2.6 Spike sorting 26

3 Discrimination of the Temporal Envelopes of Vocal Calls **28**

3.1 Introduction 28

3.2 Methods 30

3.3 Results 35

3.4 Discussion 53

3.5 Acknowledgements 60

4 Interdependent Encoding of Stimulus Features in Auditory Cortex **62**

4.1 Introduction 62

4.2 Methods 64
4.3 Results 68
4.4 Discussion 93
4.5 Acknowledgements 105

5 Discrimination by Ferrets of the Direction of Pitch Changes 107
5.1 Introduction 107
5.2 Methods 110
5.3 Results 118
5.4 Discussion 127
5.5 Acknowledgements 134

6 Behavioural and Physiological Measures of Ferrets' Pitch Judgments 137
6.1 Introduction 137
6.2 Methods 140
6.3 Results 151
6.4 Discussion 182
6.5 Acknowledgements 189

7 General Discussion 191
7.1 Representations of multiple vocalization features in cortical neurons . . . 191
7.2 Identifying the neural basis of what we hear 193
7.3 Functional organization in auditory cortex 194
7.4 Using more than one neuron to solve a task 196
7.5 Conclusions 197

Appendices 199

A Code for "Flipping" Sounds 199

B Further Examples of Cortical Responses to Artificial Vowels 202

References 215

List of Figures

1.1 Three vocalization sounds 6
1.2 Ferret vocalizations 7
1.3 Ferret auditory cortex 15

2.1 Frequency response areas of 2 neurons 25

3.1 Natural and flipped twitter stimuli 31
3.2 Human discrimination performance 36
3.3 Raster plots for 2 units 38
3.4 Eigenvalues of principal components 39
3.5 Calculating neurometrics for 2 units 40
3.6 All neurometrics (anaesthetised) 42
3.7 Count of significant neurometric scores 43
3.8 Neurometrics as a function of the number of PCs 44
3.9 Best envelope neurometric 45
3.10 Pooled neurometrics for each penetration 47
3.11 Neurometrics pooled across penetrations 48
3.12 Temporal resolution of neurometrics 49
3.13 Neurometric integration windows 51
3.14 All neurometrics (unanaesthetised) 52
3.15 Stimulus-response relation 53

4.1 Sixteen artificial vowel stimuli 66
4.2 Raster plot for 64 vowel sounds: unit A 73
4.3 Raster plot for 64 vowel sounds: unit B 74

4.4 PSTHs for 16 vowel combinations 75
4.5 CF maps of individual animals 78
4.6 Composite CF mapping across auditory cortex 79
4.7 Cortical sensitivity to 3 stimulus parameters - composite maps 80
4.8 Significance of cortical sensitivity to stimulus parameters 81
4.9 Stimulus-related variance in responses across cortical areas. 82
4.10 Cortical sensitivity to 3 stimulus parameters - individual maps 83
4.11 Cortical sensitivity to feature combinations - composite maps 85
4.12 Cortical sensitivity to feature combinations - histograms 86
4.13 Cortical sensitivity to combinations of spatial and non-spatial features . . 86
4.14 Stimulus feature preferences across units. 87
4.15 Azimuth, pitch and timbre sensitivity across cortical depth. 89
4.16 Azimuth, pitch and timbre tuning across CF. 90
4.17 CFs of neural units excluded from analyses. 91
4.18 Stimulus feature sensitivity across the response duration. 92
4.19 Auditory cortex of the cat and ferret 99

5.1 Training apparatus 112
5.2 Discrimination trial schematics. 114
5.3 Effect of pitch on /i/ spectrum. 116
5.4 Ferret pitch psychometric and choice probabilities 119
5.5 Complex pitch psychometrics 120
5.6 Ferrets' pitch difference limens and Weber fractions (vowels) 121
5.7 Bias in pitch psychometrics of ferrets (vowels) 122
5.8 Ferrets' pitch psychometrics (tones) 124
5.9 Ferrets' pitch difference limens and bias (tones) 125
5.10 Ferret and human pitch difference limens (vowels and tones) 127

6.1 Constructing neurometrics using ROC analysis 144
6.2 Neurometrics using average response to reference 147
6.3 Pooling neural units 149
6.4 Raster plots of responses to vowel pitch 152

6.5 Three classes of pitch-sensitive units . . . 154
6.6 Composite CF maps of auditory cortex . . . 155
6.7 Pitch modulation across cortical fields . . . 156
6.8 Neurometrics across reference pitches . . . 158
6.9 Median reference response neurometrics of one unit . . . 159
6.10 Pitch neurometrics for all units . . . 160
6.11 ROC and median reference response neurometrics . . . 161
6.12 Best reference pitches and response durations . . . 162
6.13 Best single unit neurometrics . . . 163
6.14 Neurometric slope by CF . . . 164
6.15 Neurometric slopes across 3 types of pitch-sensitive units . . . 164
6.16 Ferret pitch discrimination performance . . . 165
6.17 Pitch discrimination by ferrets and cortical units . . . 166
6.18 All population neurometrics . . . 168
6.19 Individual unit vs. population neurometrics . . . 169
6.20 Population neurometrics across reference pitches . . . 170
6.21 Best population neurometrics . . . 170
6.22 Neurometrics across cortical fields . . . 171
6.23 Neurometrics of simultaneous and shuffled populations . . . 173
6.24 Pitch discrimination by ferrets and cortical populations . . . 175
6.25 Neurometrics of neural subpopulations . . . 176
6.26 Pitch discrimination by ferrets and cortical subpopulations . . . 177
6.27 Population neurometrics using 4 spiking codes . . . 180
6.28 Pitch discrimination of ferrets and 4 neural codes . . . 181

B.1 Rasterplot for an A1 neuron . . . 204
B.2 PSTH plot for an A1 neuron . . . 205
B.3 Rasterplot for an AAF neuron . . . 206
B.4 PSTH plot for an AAF neuron . . . 207
B.5 Rasterplot for an PPF neuron . . . 208
B.6 PSTH plot for an PPF neuron . . . 209
B.7 Rasterplot for an PSF neuron . . . 210

B.8 PSTH plot for an PSF neuron 211
B.9 Rasterplot for an ADF neuron 212
B.10 PSTH plot for an ADF neuron 213

Chapter 1

General Introduction

1.1 The big questions

A major goal of auditory neuroscience is to provide a biological explanation for our conscience perception of sound. Hearing is the primary sense used for communication in humans and many other animals, and so the study of vocalization perception is of particular interest. How is it that a feeble, linear pressure wave in the air surrounding us can be parsed, processed and re-integrated by our brains to unlock the beauty of a piece of music or the tell-tale growl of a predator hidden in the distance? An understanding of how the brain gives rise to our ability to identify and discriminate vocal calls not only offers insight into the nature of our mental world, but can also ultimately lead to new and refined communication technologies.

While past research has led to a more thorough understanding of how structures early in the auditory nervous system encode acoustically simple stimuli, such as pure-tone frequencies, relatively little is known about how auditory cortical neurons encode more spectrally and temporally complex sounds, like those encountered in natural environments. For example, there are a number of encoding strategies that auditory cortical neurons might employ to represent complex sounds - are rate, place and temporal codes all utilized? Along the auditory pathway, from early subcortical to higher cortical structures, there is a gradual decrease in the ability of neurons to fire rapid trains of phase-locked spikes in response to periodic stimuli (Joris *et al.*, 2004; Liu *et al.*, 2006; Miller *et al.*, 2002; Wallace *et al.*, 2002). This calls into question the role of temporal

spiking patterns as a key encoding strategy at the level of cortex. However, the latency of the first spike in response to the onset of a sound can be as temporally reliable in cortex as in the auditory nerve (Phillips and Hall, 1990), and cortical neurons can lock to the fine structure of spectrotemporally complex sounds at much faster rates than those previously observed with temporally regular stimuli (Elhilali *et al.*, 2004). Therefore, fine temporal spiking patterns in cortical cells might be used to encode information about spectrotemporally complex vocal calls.

There are also unresolved issues concerning the distribution of feature sensitivity across the anatomically and physiologically distinct areas of mammalian auditory cortex. Human imaging studies (Alain *et al.*, 2001; Barrett and Hall, 2006; Maeder *et al.*, 2001; Warren and Griffiths, 2003) and studies of selective cortical deactivation in mammals (Kudoh *et al.*, 2006; Lomber and Malhotra, 2008; Malhotra *et al.*, 2004; Malhotra and Lomber, 2007; Malhotra *et al.*, 2008) provide compelling evidence that some auditory cortical fields are specialized to process particular features of vocalization sounds, such as pitch and spatial location. However, the degree of regional specialization found in these types of studies is in contrast to the more widely distributed cortical sensitivity observed in investigations of the firing properties of individual neurons (Furukawa and Middlebrooks, 2002; Stecker *et al.*, 2005b; Stecker and Middlebrooks, 2003). The functional and physiological organizations of auditory cortex must be reconciled in a coherent model of complex auditory processing.

This thesis addresses the above issues in its investigation of the representation of communication sounds in neurons throughout the auditory cortex of the ferret (*Mustela putorius furo*). Findings from physiological and behavioural studies of hearing are integrated in order to explore how neural representations might support animals' discrimination of vocal calls. The goal is to further our understanding of how complex sounds are encoded in auditory cortex, how these neural representations are distributed across a number of auditory cortical areas, and how this neural activity relates to behavioural decisions made by listeners.

1.2 What characterizes a vocal call?

Six general classes of perceptual attributes of sounds have been described: pitch, loudness, timbre, spatial location, duration and reverberant environment (Levitin, 1999). Of these, pitch and timbre are likely to be most revealing about the identity of a sound source. Timbre has been broadly defined as the quality that allows us to distinguish between two sounds when the pitch and loudness are constant. For instance, it is the clearly audible difference between a violin and a piano playing the same musical note. Timbre is determined by both the spectral (Helmholtz, 1954) and temporal (Grey, 1977) aspects of a sound, but its primary acoustical correlate is the spectral envelope of the sound, which describes how the sounds' acoustical energy is distributed across the frequency domain. The American National Standards Institute has defined pitch as "that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high" (ANSI, 1994).Pitch is determined by a pure tone's frequency, or, in the case of more complex sounds, the periodicity of the sound. In other words, it is determined by the shortest period over which there is temporal repetition of the acoustic waveform. Two adjacent keys on a piano produce sounds of similar timbre but that differ in pitch, and we can order these pitches along a musical scale, from low to high.

To appreciate the importance of pitch and timbre to vocalization perception, consider the example of human speech. Speakers generate a rapid train of glottal pulses with their vocal cords, and the rate of the pulse train determines the pitch of the vocalization. The pitch of speech conveys information about speaker identity, since men typically have lower pitched voices than women or children (Gelfer and Mikos, 2005). Pitch also conveys the emotional state of the speaker - excited speakers tense their vocal folds, thereby raising the fundamental frequency of their vocalizations (Fuller *et al.*, 1992; Reissland *et al.*, 2003). The glottal pulse train travels through, and is filtered by, a number of resonant cavities in the nasopharynx, which impart specific resonant peaks, known as "formants", onto the speech sound. The formants depend on the size and shape of the vocal tract, which are specific to the individual speaker. Furthermore, by moving articulators, like the lips, jaws and tongue, the speaker can alter formant frequencies and generate different speech sounds that are identifiable by their characteristic spectral profile. Distinguishing the vowel /ɑ/ from the vowel /i/ is therefore a form of timbre discrimination (Peterson

and Barney, 1952).

Other mammalian species often generate vocalizations in an entirely analogous pulse-train-resonance fashion to humans (Figure 1.1), and these animals similarly use periodicity and spectral cues to interpret vocalizations (Capranica, 1966; Koda and Masataka, 2002; Kojima *et al.*, 2003; Nelson, 1989). This makes processing the pitch and timbre of vocal calls an important task for the mammalian auditory system, and the ferret is no exception. Although there has not yet been a scientific study of the full set of vocalizations used by the domestic ferret, a variety of ferret vocalizations have been informally observed by pet owners, veterinarians and scientists. A sample of these calls is presented in Figure 1.2. Ferret kits produce a distinctive, high-pitched call for their mothers, and the pitch of this call drops systematically as the kit matures toward adulthood (unpublished observations from the laboratory of Didier Depireux, University of Maryland). The "bark" is a loud chirp-like call, produced when the ferret is excited. The "hiss", which sounds similar to the hiss of a domestic cat, is produced when a ferret is irritated or desires to assert its dominance over another ferret. The "dook", also known as "clucking" or "chuckling", is produced when a ferret is happy, commonly during play or petting. The quality of the sound has been compared to that of a chicken clucking. The dook evokes a clear pitch percept, and this pitch tends to vary from syllable to syllable. Finally, the "screech" is strongly periodic like the dook, but is higher in pitch and is a much louder call. This is a sign of fear, pain or irritation produced during rough play, fighting or during a test of dominance between two ferrets. Zoological studies have described a similar set of vocal calls for the Siberian ferret (*Mustela eversmanni*) and other mustelid species (Farley *et al.*, 1987). Some of the vocalizations of these species, particularly the chuckle- and screech-like calls, are produced at a characteristic pitch that differs across species, so determining the pitch of the call may be particularly important in identifying conspecific callers.

Another feature of vocal calls that is often crucial for intelligibility is the temporal unfolding of acoustic events over short time scales. That is, the shape of the temporal envelope of the sound. In human speech, for example, the temporal resolution of sequential acoustic events at fine scales (*ca.* 20 - 50 ms) is necessary for speech intelligibility (Liberman *et al.*, 1967; Rosen, 1992; Saberi and Perrott, 1999). Temporal fidelity also

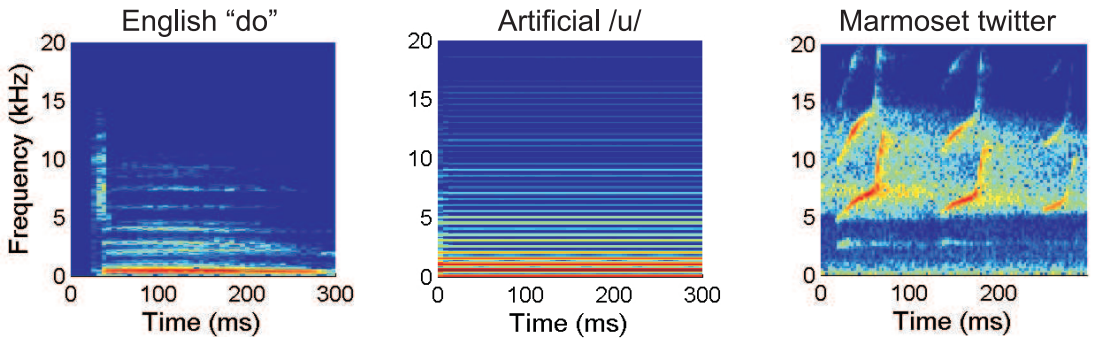


Figure 1.1: Spectrograms of three vocalization sounds. Spectrograms show the spectral profile of sounds across time. The color scale runs from blue to red, indicating low to high energy. From left to right, the spectrograms describe: the word “do” spoken by an adult British female; an artificial version of the vowel sound /u/, generated with a periodicity of 500 Hz (as used in Chapters 4,5, and 6); and a marmoset “twitter” call (as used in Chapter 3).

appears to be necessary for the interpretation of non-human vocal calls. “Call detector” neurons have been identified in the auditory cortex of marmosets. These cells fire at higher rates in response to the presentation of natural conspecific vocalizations than to the same vocalizations presented in reverse (Wang and Kadia, 2001). This type of temporal envelope selectivity for conspecific calls has also been demonstrated behaviourally in monkeys (Ghazanfar *et al.*, 2001).

Determining the spatial location of a speaker is not critical for interpreting the semantics of the call, but it nevertheless can contribute to intelligibility. Spatial separation of multiple sound sources has been shown to aid stream segregation (Boehnke and Phillips, 2005) and helps listeners distinguish auditory figures from the background noise (Eramudugolla *et al.*, 2008; O’Connor and Sutter, 2000). Thus, spatial localization may help to process calls in an acoustically cluttered environment (i.e. solving the “cocktail party” problem).

The acoustical cues for localizing sounds in the vertical plane (“elevation”) are primarily spectral. These are provided by the external ear, which applies a direction-dependent filter to sounds waves before they enter the ear (Musicant *et al.*, 1990). In the horizontal plane (“azimuth”), three cues are used to determine the location of sound sources: interaural level differences (ILDs); interaural timing differences (ITDs); and, to avoid

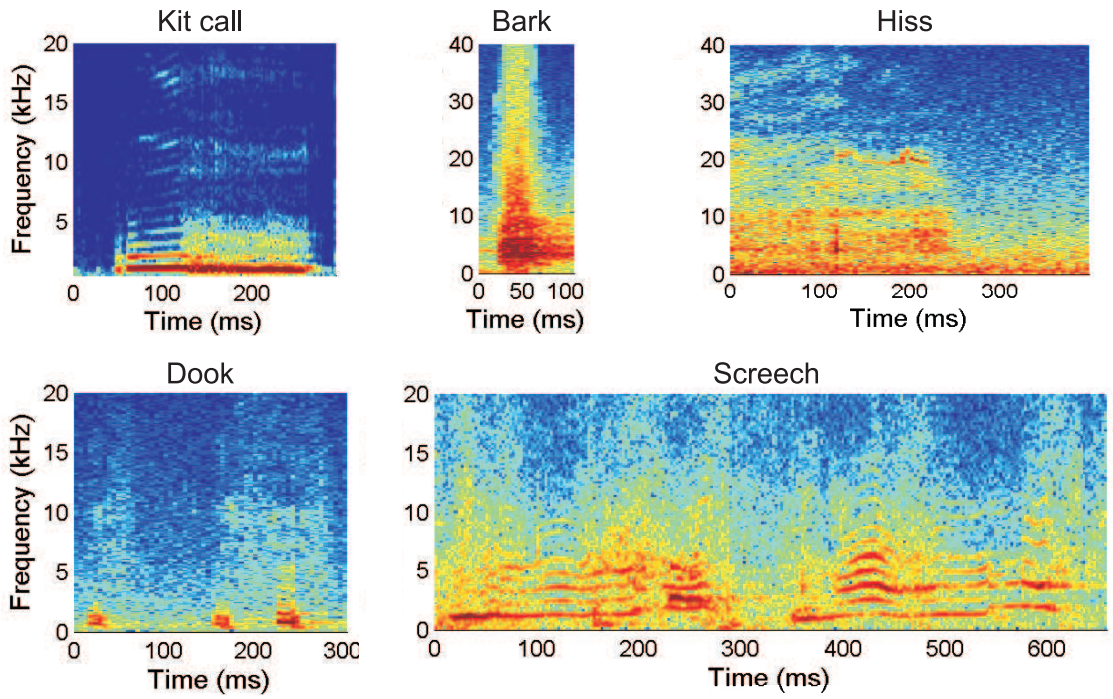


Figure 1.2: Spectrograms of ferret vocalizations. Top, left panel shows a ferret kit call, recorded by Didier Depireux (University of Maryland). I recorded the remaining samples from adult male and female ferrets in our lab. The color scales run from blue to red, indicating low to high energy.

front-back confusions, spectral cues similar to those used for elevation perception (King *et al.*, 2001). Unless a sound source is positioned directly in front of or behind a listener (i.e. at 0°or 180° azimuth), a sound wave emanating from the source will reach one ear of the listener earlier than it reaches the other ear. In addition to this interaural timing difference in the onset of the sounds, the phase of the sound wave will also vary across ears, depending on the position of the source. The interaural phase difference provides a cue for azimuthal localization throughout the duration of the sound, provided that the frequency of the sound is low enough for the auditory nerve to phase-lock to it. In mammals, the upper limit of phase locking is *ca.* 3.5 - 5 kHz, but phase-locking is best below about 1.5 kHz (Johnson, 1980; Palmer and Russell, 1986). Sounds may also be more intense at the ipsilateral ear, since the head will "shadow" the wave at the contralateral ear. The ILD is most effective if the wavelength is shorter than the listeners' head diameter, to avoid diffraction around the head. Thus, there are multiple cues to

the position of sounds around the azimuth, and these complement each other at different ends of the frequency range.

1.3 Subcortical processing of complex sounds

Sound signals are processed at a number of stations throughout the peripheral and subcortical auditory system before they arrive at the auditory cortex. This textbook account of a linear procession of information from the ear to auditory cortex is actually an oversimplification of the true auditory system, since there exist multiple pathways of afferent and efferent projections among subcortical and cortical areas (reviewed by Winer and Lee, 2007). The precise form and function of these multiple pathways are poorly understood and are outside the scope of the present thesis. To orient the reader, this section provides a brief overview of sound processing in the mammalian auditory system from ear to cortex.

1.3.1 Sound transduction in the ear

Vibrating objects generate sounds in the form of pressure waves. The properties of these waves depend on the physical properties of the objects, and so the auditory system can make use of the information in these waves to recognize objects. When the pressure waves reach the tympanic membrane (also known as the ear drum), they are transformed into mechanical displacements in the membrane. This motion is transferred and amplified by three small bones ("ossicles") in the middle ear. This energy is then applied to the oval window of the fluid-filled cochlea, causing waves in the fluid. These, in turn, lead to travelling waves on the basilar membrane within this chamber. The resonant frequency of the basilar membrane changes logarithmically along its length. Consequentially, the position of displacement along the length of the basilar membrane is proportional to the frequency of acoustical energy in the signal applied to it. In this manner, the basilar membrane essentially performs a Fourier transform on the incoming sound, thereby setting up a topographic map of sound frequency (i.e. a "tonotopic" map). The frequency acuity of this map is refined by active mechanisms in the outer hair cells of the cochlea. The array of hair cells that transduce the local mechanical perturbations of the basilar

membrane into graded membrane potentials inherit this tonotopic map, and the array of auditory nerve fibers that are excited by these hair cell potentials are also tonotopically organized.

1.3.2 The auditory nerve

Each auditory nerve fibre is most sensitive (i.e. has its threshold) at a specific frequency - the characteristic frequency (CF) of the cell. The tonotopic arrangement of CFs is preserved along the auditory pathway, up to the level of auditory cortex. In addition to holding a "place code" for frequency, auditory nerve fibers use a "rate code" to represent the intensity of sounds. That is, they fire more action potentials per unit of time in response to higher sound levels, up to a certain sound level where the firing rate will saturate. These neurons also exhibit phase-locked firing in response to sound frequencies up to about 1.5 kHz. For these frequencies, each neuron has a higher probability of firing at a particular phase of the sine wave. This provides a "temporal code" of the phase of sounds. Due to the combined efforts of these three encoding strategies, the distribution of activity across the population of auditory nerve fibers provides a sort of "neurogram" of the sound. By this we mean that the firing activity across the bank of auditory nerve fibers resembles the spectrogram of the sound. The spectral profile of vocalizations is preserved in the relative firing rates across auditory nerve fibers with low spontaneous firing rates (i.e. high firing thresholds), and pitch can be calculated from the peak of the all-order interspike interval histogram (i.e. a histogram of spike autocorrelations) of these neurons (Cariani, 1999).

1.3.3 The cochlear nucleus

Auditory nerve fibers synapse onto cells in the cochlear nucleus (CN) of the brainstem. There are three distinct regions of the CN, each of which is tonotopically arranged. The CN contains a large number of cell types (up to 22) which are distinguishable based on their morphologies and firing properties (Brawer *et al.*, 1974; Pfeiffer, 1966; Young and Brownell, 1976). The functional significance of all of the CN cell types is not yet known in full, but some of these exhibit firing properties that seem ideally suited to process particular features of communication sounds. Primary-like cells in the CN have

response properties that closely resemble auditory nerve fibers, and so they preserve information about the fine temporal structure of sounds that might be necessary for pitch extraction by higher regions (Blackburn and Sacks, 1990; Winter and Palmer, 1990). Primary-like-with-notch neurons appear to be sensitive to more complex spectral profiles than auditory nerve fibers, with inhibitory "notches" that might be used to encode spectrally-based location and timbre cues (Blackburn and Sacks, 1990; Reiss and Young, 2005). Finally, onset units in the CN respond to periodic inputs with temporally precise, periodic spikes, and the probability of firing in these units is proportional to the number of synchronous inputs. These cells represent the fundamental frequency of a complex sound as the reciprocal of their interspike intervals, thereby converting the all-order interspike interval code of stimulus periodicity in auditory nerve fibers into a first-order interspike interval pitch code (Winter *et al.*, 2001).

1.3.4 The superior olive

Auditory signals from the two ears are first combined at the level of the superior olivary nucleus of the brainstem. ITD and ILD cues are processed separately in the medial and lateral nuclei of the superior olive, respectively. ITD cues are processed in the superior olive using the relative timing of inhibitory inputs from cochlear nuclei on both sides of the brainstem (Brand *et al.*, 2002). This requires a faithful representation of the timing of peripheral events on a microsecond time scale, which is made possible by low-threshold voltage-gated potassium channels, unusually rapid-acting transmitter-gated channels, and exceptionally large nerve terminals (Trussell, 1999). This computation results in a periodic change in spike counts across the range of ITD values (Brugge and Merzenich, 1973; McAlpine *et al.*, 2001). ILDs are computed by integrating excitatory projections from ipsilateral CN neurons and inhibitory inputs from the contralateral CN (Sanes, 1990). Neurons in the lateral superior olive respond preferentially to interaural disparities corresponding to sounds in contralateral space (Phillips and Irvine, 1981). The response properties of cells in both the medial and lateral superior olive preserve spatial information best for sounds in the frontal hemifield of space, particularly near the midline (McAlpine *et al.*, 2001; Phillips and Irvine, 1981).

In addition to processing binaural cues, neurons here might also help to preserve the

dynamic range of CN neurons via descending inhibitory projections (May and Sachs, 1992; Winslow and Sacks, 1988).

1.3.5 The inferior colliculus

There are as many synaptic connections between the primary fibers and inferior colliculus (IC) of the auditory system as there are between the primary fibers and primary visual cortex of the visual system. Accordingly, signals arriving to the IC are already considerably preprocessed, as we have seen. One consequence of the spectral processing in the IC and earlier nuclei might be to sharpen the representation of spectral peaks in complex sounds (Shamma, 1999). This is accomplished via lateral inhibition across neurons, according to their frequency selectivity. In the case of periodic vocalizations, this would correspond to a sharpening of formants, which could aid timbre discrimination.

The first evidence of non-monotonic rate-level functions in the auditory system are observed in IC (Irvine and Gago, 1990). Many neurons here show bandpass sensitivity to the sinusoidal amplitude modulation of sounds. This sensitivity is evident in phase-locked responses to stimuli with modulations of up to 500 hertz, but is encoded with unsynchronized spike rates for faster modulations (up to 1 kHz) (Langner and Schreiner, 1988). ILD and ITD representations established in the superior olive also undergo further processing in the IC. These binaural cues remain to be encoded in largely segregated neural populations (Park, 1998), although recent studies have shown that some IC neurons are sensitive to both ILD and ITD cues (Chase and Young, 2005). Langner *et al.* (2002; 1988) have suggested that a "periodotopic" map of best modulation frequency exists in IC, running orthogonal to the direction of the tonotopic map.

1.3.6 The thalamus

From the IC, neurons project to the medial geniculate body (MGB) of the thalamus, and from there, onto the auditory cortex. The MGB is often viewed as a simple "relay station" for signals ascending to cortex, but growing evidence suggests that specialized thalamic circuits can play an important role in determining the behavioural significance of auditory stimuli. For instance, neurons in the MGB directly project onto the amygdala, modulating amydala activity. This projection is thought to underlie the conditioning of

emotional reactions to acoustic stimuli (Clugnet *et al.*, 1990; Farb and LeDoux, 1997). Additionally, while most ascending thalamic inputs come from IC, there also exists a direct, non-tonotopic, short-latency projection from the dorsal cochlear nucleus to the medial devision of the MGB (Anderson *et al.*, 2006; Malmierca *et al.*, 2002). The function of this neural network remains speculative, but it may modulate auditory information processing based on the integration of multisensory cues.

1.4 Auditory cortical processing of complex sounds

Data about the function of the primary auditory cortex (A1) are inconclusive. Some investigators argue that A1 functions as an auditory object aggregator or novelty detector (Nelken and Bar Yosef, 2008; Nelken *et al.*, 2003). Others suggest that A1 holds multiple maps of complex stimulus attributes, such as pitch (Schulze *et al.*, 2002; Schulze and Langner, 1997a). The plasticity of A1 suggests that it is likely to play a key role in directing attention (Fritz *et al.*, 2007a) and/or perceptual learning (Dahmen and King, 2007). The response properties of A1 neurons suggest that it is tuned specifically to encode sound dynamics that occur in natural scenes and vocal calls (Garcia-Lazaro *et al.*, 2006). Clearly, there is still much to learn about the functional roles of auditory cortex, even at the level of the primary field. Such research questions are met with great enthusiasm by the naive graduate student.

1.4.1 Receptive fields and organization of A1

The tonotopic map established at the level of the cochlea is preserved in A1 (Kelly *et al.*, 1986a; Merzenich *et al.*, 1975). The upper limit of phase-locking is decreased at the cortical level (Joris *et al.*, 2004; Liu *et al.*, 2006; Miller *et al.*, 2002; Wallace *et al.*, 2002), suggesting that the A1 integrates acoustical information across relatively wide temporal windows. An imaging study has suggest that in the human primary auditory cortex, the temporal integration window may be on the order of 25-50 ms, while integration windows widen to up to 200-300 ms in higher auditory cortical regions (Boemio *et al.*, 2005). This time-scale is consistent with studies of single neurons in marmoset cortex, which have shown that some neurons in A1 can produce spikes that are synchronized with

click trains for rates up to about 50 Hz, while these "fast" synchronizing neurons are less common in the higher rostral fields of auditory cortex (Bendor and Wang, 2007; Lu *et al.*, 2001b). Many A1 neurons have been shown to apply a complex spectrotemporal filter to its input, which results in a preference for particular spectral and temporal features of acoustical signals (Depireux *et al.*, 2001; Linden *et al.*, 2003; Theunissen *et al.*, 2000). The receptive fields of cortical neurons can be described along at least five dimensions: the temporal modulations of the sound envelope, the characteristic frequency (CF) along the tonotopic axis, the spectral bandwidth around the CF, the symmetry of the spectral bandpass filter around the CF, and the spatial location of the sound source. The degree of spectral specialization in A1 should make it particularly sensitive to the timbre of vocalization sounds.

A number of studies have found evidence of non-random distributions of receptive field properties across the surface of A1. The CF-normalized spectral bandwidth has been shown to change gradually and reliably along the axis that runs orthogonal to the CF gradient (Heil *et al.*, 1992; Recanzone *et al.*, 1999; Schreiner and Mendelson, 1990; Shamma *et al.*, 1993). The bandwidth appears to be arranged such that the widest excitatory bandwidths are found near the middle of the isofrequency band, and the bandwidth narrows at both ends of the isofrequency band. Changes in spectral bandwidth along isofrequency contours are smoother in some species than in others (Cheung *et al.*, 2001). The symmetry of spectral response filters around the CF for each neuron have also been observed to be ordered along isofrequency bands (Recanzone *et al.*, 1999; Shamma *et al.*, 1993). Within this organization, neurons at one end of the isofrequency band are more inhibited by frequencies below their CF, while those at the other end are more inhibited by frequencies above their CF. Non-random organizations of stimulus intensity and first-spike latency have also been demonstrated within isofrequency bands of A1 (Cheung *et al.*, 2001; Heil *et al.*, 1992, 1994; Recanzone *et al.*, 1999).

Schulze and colleagues (Schulze *et al.*, 2002; Schulze and Langner, 1997a,b) have suggested that periodotopic maps of preferred modulation rates exist in the gerbil A1, similar to those found by the same group in IC. Two studies have also found evidence of such maps in human A1 using magnetoencephalography (Langner *et al.*, 1997; Pantev *et al.*, 1989). However, the existence of periodotopic maps remain controversial, since the choice

of stimuli (discussed in Winter, 2005) and methodology (Lütkenhöner, 2003; discussed in Griffiths, 2005) in these studies may have been susceptible to artifactual findings, and other groups have failed to replicate these results (Fishman *et al.*, 1998; Lütkenhöner, 2003; Nelken *et al.*, 2008; Steinschneider *et al.*, 1998; Tomlinson and Schwarz, 1988).

While no topographic map of space exists in auditory cortex, the distribution of binaural sensitivities across A1 is nevertheless non-random. Bands of neurons with shared binaural response properties can be found, lying in parallel to the tonotopic axis and traversing the cortical depth (Imig and Adrian, 1977; Kelly and Judge, 1994; Middlebrooks *et al.*, 1980). This being said, Campbell *et al.* (2006) have shown that the traditional binaural response categories used to classify A1 neurons fail to reflect the more continuous distribution of spatial preferences of these neurons. Cells in A1 tend to be tuned to particular directions along the azimuthal plane derived from a number of integrated binaural and monaural directional cues, rather being selectively tuned to any one particular spatial cue alone (Hall and Goldstein, 1968; Middlebrooks and Pettigrew, 1981; Mrsic-Flogel *et al.*, 2001; Schnupp *et al.*, 2001). The spatial tuning of auditory cortical cells tends to be quite broad and to prefer contralateral directions. Spatial information is encoded in both the spike rate and first-spike latencies of neural responses (Mickey and Mddlebrooks, 2003; Mrsic-Flogel *et al.*, 2005; Nelken *et al.*, 2005).

1.4.2 Response properties of higher auditory cortical areas

A number of higher auditory cortical areas have been identified in several mammalian species, but the present review will focus on findings regarding my experimental animal of choice - the ferret. Ferrets are a popular animal model in auditory science, as they are well suited to both behavioural and physiological studies of hearing. They can be trained to perform a variety of sound localization (Bizley *et al.*, 2007b; Kacelnik *et al.*, 2006; Kelly and Kavanagh, 1994; Nodal *et al.*, 2008) and detection tasks (Fritz *et al.*, 2005b, 2007b; Hine *et al.*, 1994; Kelly *et al.*, 1986b, 1996). Their low-frequency tone detection thresholds are much better than those of rodents (Kelly and Kavanagh, 1994), and their vocalizations contain periodicities similar to those of human speech sounds, suggesting that this is an appropriate species in which to study the perception of periodic vocal calls.

The auditory cortex of the ferret lies on the lateral surface of the ectosylvian gyrus (Figure 1.3). The physiological properties of ferret A1 have been extensively characterized (Depireux *et al.*, 2001; Garcia-Lazaro *et al.*, 2006; Kelly and Judge, 1994; Phillips *et al.*, 1988; Schnupp *et al.*, 2001; Shamma *et al.*, 1993; Versnel *et al.*, 2002) and the higher auditory cortical fields of this species has been more thoroughly studied in recent years (Bizley and King, 2008; Bizley *et al.*, 2007a, 2005; Kowalski *et al.*, 1995; Nelken *et al.*, 2008, 2004; Wallace *et al.*, 1997). The anatomical projections among ferret auditory cortical fields and subcortical structures are also being elucidated (Bajo *et al.*, 2007; Bizley *et al.*, 2007a), making across-species comparisons more feasible.

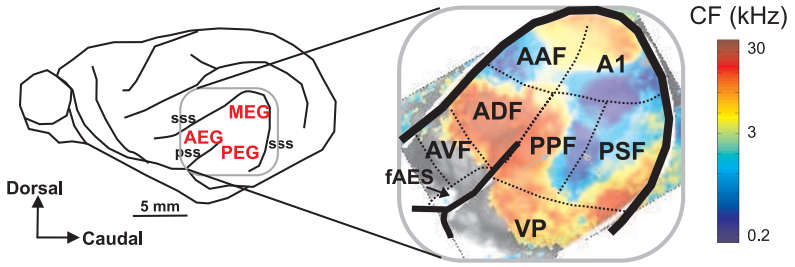


Figure 1.3: Auditory cortex of the ferret. Location of ferret auditory cortex on the middle, anterior and posterior ectosylvian gyri (MEG, AEG and PEG, respectively). The suprasylvian (sss) and pseudosylvian (pss) sulci are also labelled. The inset shows the location of eight auditory cortical fields: the primary and anterior auditory fields (A1 and AAF), the tonotopically organized posterior pseudosylvian and suprasylvian fields (PPF and PSF), and the non-tonotopic anterior dorsal and ventral fields (ADF and AVF), ventroposterior area (VP) and anterior ectosylvian sulcal field (fAES). The color scale shows the tonotopic organization of these auditory cortices, as visualized using optical imaging of intrinsic signals (modification of a figure presented by Nelken *et al.*, 2004).

Previous anatomical and physiological investigations in our laboratory have identified six distinct areas of auditory cortex in the ferret: the primary and anterior auditory field (A1 and AAF), the posterior pseudosylvian and suprasylvian field (PPF and PSF), and the anterior dorsal and ventral field (ADF and AVF). Nelken *et al.* (2004) classified four of these regions based on their tonotopicity, and Bizley and colleagues (Bizley and King, 2008; Bizley *et al.*, 2007a, 2005) have described the responses of neurons in these and two further fields to tones, noise bursts, and points of light. The results of these investigations suggest that areas A1 and AAF, found on the middle ectosylvian gyrus, make up the

core of auditory cortex. This classification is based on physiological and anatomical properties of A1 and AAF. For example, neurons in "core" areas typically respond to pure tones with short latencies, are tuned to a restricted range of frequencies, and are tonotopically organized (Bizley *et al.*, 2005). Anatomical studies of the cytoarchitecture (Bajo *et al.*, 2007), cortico-collicular connections (Bajo *et al.*, 2007), and cortico-cortical connections (Bizley *et al.*, 2007a) also support the classifaction of A1 and AAF as the auditory core. These areas contain parallel tonotopic maps with similar frequency acuity. In these maps, higher frequencies are represented more dorsally on the gyrus than lower ones. Perhaps unintuitively, the response latencies of neurons in AAF and ADF are shorter than the latencies of A1 neurons (Bizley *et al.*, 2005). These physiological results are in general agreement with a previous comparison of ferret A1 and AAF, although the frequency tuning of AAF neurons was previously reported to be twice as wide as those in A1 (Kowalski *et al.*, 1995). Kowalski *et al.* (1995) additionally demonstrated that the spectral asymmetries (i.e. preferences for sounds that are upward- or downward-modulated in frequency) of A1 and AAF neurons were similar.

PPF and PSF are tonotopically organized belt areas found on the posterior ectosylvian gyrus, and ADF is a non-tonotopic belt region of the anterior ectosylvian gyrus. The response latencies of PPF and PSF are longer than those of the more anterior and dorsal fields, and these responses are also more sustained (that is, they last longer throughout the duration of a steady-state sound). Many neurons in these posterior fields have non-monotonic rate-level functions, resulting in diamond-shaped frequency response areas. ADF may play a special role in sound localization, as neurons here encode more information about the azimuthal location of noise bursts than those in more posterior fields. ADF and AVF are highly modulated by visual stimulation as well as sounds, and AVF is considered to be a multisensory area (Bizley and King, 2008; Bizley *et al.*, 2007a). In fact, it is often difficult to drive neural responses in AVF with pure tones, or even noise bursts, in the absence of visual stimulation (Bizley *et al.*, 2007a, 2005). Most ADF neurons are poorly driven by tones that are below 8 kHz, and the frequency tuning of these cells is often very broad.

Two additional cortical regions have been identified in anatomical studies of the ferret ectosylvian gyrus: the anterior ectosylvian sulcal field (fAES) and the ventroposterior

area (VP) (Bajo *et al.*, 2007; Bizley *et al.*, 2007a; Manger *et al.*, 2005; Pallas and Sur, 1993; Ramsay and Meredith, 2004). However, the response properties of neurons in these fields for auditory stimulation have not yet been systematically studied.

Nelken *et al.* (2008) used optical imaging of intrinsic signals to investigate the topography of responses to complex sounds across the two core and three belt areas of ferret auditory cortex. They found distributed sensitivity to the periodicity and binaural properties of sounds across all five fields, but there was no evidence of topographic maps of pitch or space.

1.5 Thesis overview

This thesis builds on our understanding of cortical physiology, with an emphasis on the processing of communication sounds like those encountered in natural environments. In particular, it aims to identify the types of neural codes that might be used to encode different features of vocal calls across a number of cortical fields. Neural responses to sounds are measured in anaesthetised and unanaesthetised animals, and are compared to psychophysical results using the same stimuli in order to assess the potential behavioural significance of neural representations of acoustical properties. Experimental Chapters 3 to 6 progress chronologically through investigations of these research questions that I have undertaken throughout my D.Phil. studies.

- **Chapter 3: Discrimination of the temporal envelopes of vocal calls.** The first experimental chapter investigates how the temporal envelopes of spectrotemporally rich vocal calls (marmoset vocalizations) are represented by neurons in primary auditory cortex. Neural responses are measured in both anaesthetised and unanaesthetised ferrets. "Neurometric analysis", a statistical approach previously used only to examine spike rates and phase-locked responses, is modified here to allow me to examine more complex spiking patterns. Using this analysis, I ask if the cortical representations of the temporal unfolding of vocalizations can support listeners behavioural discrimination of the same sounds.

- **Chapter 4: Encoding pitch, timbre and azimuth in auditory cortex.** While the above chapter investigates how mammals process temporal variations in

vocalization sounds, this chapter and those that follow investigate how the mammalian brain processes perceptual features of steady-state vocalizations (i.e. those with envelopes that do not vary in time). "Artificial vowel" sounds are used as stimuli, since they model the spectral properties of vocal calls but can be parametrically varied in timbre and pitch. Previous functional imaging and cortical deactivation studies suggest that separate cortical processing streams are specialized for processing spatial and non-spatial features of sounds. The pitch, timbre and spatial location of a voice are in some sense seperable concepts. A speaker's timbre can remain constant while they moves around the room, or, similarly, they could sing the same vowel at different musical notes while standing still. It therefore seems natural to suppose that the cortical processing of vocal calls might anatomically segregate these attributes. In this chapter, neural activity in the two core and three belt areas of ferret auditory cortex is measured in response to vowels that simultaneously vary in pitch, timbre and azimuthal location. This stimulus set allows me to ask unique questions concerning the degree to which these stimulus features are represented independently in neural spiking patterns. I also examine the distribution of neural sensitivity to these stimulus attributes across the cortical surface and through the cortical depth.

- **Chapter 5: Ferrets' discrimination of the direction of pitch changes.** The previous chapter reveals the degree to which ferret cortical neurons are sensitive to the pitch of artificial vowels, but this alone does not demonstrate that ferrets can perceive these changes. In fact, while many past studies have examined animals' ability to detect a frequency change in a sequence of tones, few have measured animals' discrimination of the periodicity of complex, naturalistic stimuli. Furthermore, the ability of animals to report the direction of periodicity changes in complex sounds has not previously been investigated. Thus, it has not been demonstrated empirically that animals perceive the pitch of periodic sounds along an ordered scale, as humans do. Here, four ferrets were trained on a two-alternative forced choice task to discriminate sounds that were higher or lower in pitch than a reference, using both simple tones and artificial vowel sounds. Their performance is compared to that of naive human listeners on a similar task.

- **Chapter 6: Behavioural and physiological measures of ferrets' pitch judgments.** The final experimental chapter relates the behavioural measures of ferrets' pitch judgments collected in Chapter 5 to the pitch acuity of neurons in ferret auditory cortex. I am particularly interested in whether the activity of neurons that are sensitive to sound periodicity, but not necessarily selective for this feature, is sufficient to account for ferrets' psychometric performance curves. As in Chapter 3, neurometric analyses are used to compare behavioural and physiological data. A number of neural encoding strategies are examined.

Chapter 2

Neurophysiological Methods

2.1 Introduction

Chapters 3, 4 and 6 describe electrophysiological data that were collected using a common set of surgical and extracellular recording procedures. These methods are detailed here to avoid repetition. Psychophysical methods are described in detail in the experimental chapters. All animal procedures were approved by the local ethical review committee and performed under licence from the UK Home Office in accordance with the Animal (Scientific Procedures) Act 1986. Experiments were carried out on adult, pigmented ferrets. The animals in these experiments received regular otoscopic examinations prior to experimentation, to ensure that both their ears were clean and disease-free.

2.2 Anaesthetised preparation

2.2.1 Surgical procedures

Anaesthesia was induced by a single, intramuscular dose of medetomidine (Domitor; 0.022 mg/kg/h; Pfizer, Sandwich, UK) and ketamine (Ketaset; 5 mg/kg/h; Fort Dodge Animal Health, Southampton, UK). Once anaethesia was confirmed in the animal, the radial vein was cannulated, and a continuous infusion (5 ml/h) of a mixture of medetomidine and ketamine in physiological saline containing 5% glucose was provided throughout the experiment. The ferret also received a single, subcutaneous dose of 0.06 mg/kg/h atropine sulphate (C-Vet Veterinary Products, Leyland, UK) at the beginning of the

surgery, and again as neccessary (either subcutaneously or intraveinously) throughout the experiment to treat bradycardia and excess bronchial secretions, as these symptoms presented. Finally, subcutaneous doses of 0.5 mg/kg dexamethasone (Dexadreson; Intervet UK Ltd., Milton Keynes, UK) were given every 12 hours to reduce cerebral oedema.

The ferret was intubated, placed on a ventilator (7025 respirator, Ugo Basile, Milano, Italy) and supplemented with oxygen-enriched air. Body temperature, end-tidal CO_2, and the electrocardiogram (AS/3 monitor, Datex, Ohmeda, Finland) were monitored throughout the experiment to ensure stable anaesthesia.

Next, the animal was placed in a stereotaxic frame and the temporal muscles on both sides were retracted to expose the dorsal and lateral parts of the skull. A metal bar was fixed to the right side of the skull using dental cement (Simplex Rapid, Austenal Dental, Harrow, UK) and stainless steel screws, holding the head without further need of a stereotaxic frame. On the left side, the temporal muscle was largely removed, and the suprasylvian and pseudosylvian sulci were exposed through a craniotomy (Figure 1.3). The dura was removed and the cortex covered with silicon oil or 1% agar. The animal was then transferred to a small table in an anechoic chamber (Industrial Acoustics Company Ltd., Winchester, UK).

2.2.2 Extracellular recordings

Spike activity was recorded extracellularly using silicon array Michigan probes with impedences near 1-2 MΩ (Centre for Neural Communication Technology, Univ. of Mich., USA). The voltage signals recorded were bandpass filtered (500 Hz - 5 kHz in Chapters 4 and 6; 300 Hz - 3 kHz in Chapter 3), amplified up to 20,000 times, and digitized at 25 kHz. Data acquisition was performed using TDT System 3 multichannel recording systems (RX5 and/or RX8 processors, with high impedance headstages and Medusa preamplifiers; Tucker-Davis Technologies, Alachua, FL) running customized RPvds programs (Tucker-Davis Technologies), together with personal computers running BrainWare (Tucker-Davis Technologies) and customized MATLAB scripts (MathWorks Inc., Natick, MA).

2.3 Unanaesthetised preparation

2.3.1 Surgical procedures

In one normal adult ferret, extracellular recordings were carried out while the animal was unanaesthetised and passively listening to stimuli. A cranial mount was surgically implanted a month prior to the first recording session. During this procedure, surgical anaesthesia was induced with an intramuscular injection of medetomidine (0.022 mg/kg) and ketamine (5 mg/kg). The animal was intubated and anaesthesia was maintained with 1-2% isoflourine (Isoflo; Abbott Laboratories, Maidenhead, UK) in oxygen-enriched air. The animal was placed in a stereotaxic frame and the temporal muscles were retracted and partially removed bilaterally to expose the dorsal and lateral skull. The auditory cortex was exposed by craniotomy and a cranial mount of dental cement was fixed to the skull. This mount contained a metal well above the craniotomy that allowed access to the dura, and could be sealed with a rubber stopper to protect the underlying tissue between recording sessions. It also contained a metal fitting which allowed the head to be restrained by fixing the head mount to a solid recording frame. During the month following implant surgery, the animal was allowed to recover, and was trained with positive food reinforcement to accept head restraint.

2.3.2 Extracellular recordings

Recordings were carried out on 14 days over the period of one month. The animal's head was fixed during recording sessions, and its body was held in place loosely in a plexiglass tube. Each recording session lasted approximately three to four hours. The animal was habituated to sit in the plexiglass tube during the month prior to implant surgery, and was rewarded for sitting still with food treats. Therefore, the animal was generally relaxed during recording sessions, but the session was ended early if the animal started to exhibit signs of discomfort, such as fidgeting.

To record neural responses, quartz/platinum-tungsten electrodes (Thomas Recording, Giessen, Germany) were lowered through the dura using a Mini Matrix System microdrive (Thomas Recording). Up to five electrodes (linearly configured, with 305 μm spacing) could be lowered into the brain at once, and each contained a single 1-2 MΩ recording

site. The microdrive contained a five-channel preamplifier. All other signal acquisition details were as described for the anaesthetized preparation above (Section 2.2.2).

2.4 Stimulus presentation

All stimuli were generated and presented using TDT System 3 hardware (RP2 and RP2.1 real-time processor, along with HB7 headphone drivers; Tucker-Davis Technologies) running RPvds programs, and MATLAB scripts running on personal computers. Sounds were presented to anaesthetised animals in an anechoic room through pairs of customized Panasonic earphone drivers (RPHV297, Panasonic, Bracknell, UK), which were mounted on plastic otoscope speculae inserted into each ear canal. The earphones were closed-field calibrated using a 1/8 inch condenser microphone (Type 4138; Brüel and Kjær, Naerum, Denmark) placed at the end of a model ferret ear canal. In the unanaesthetised, head-fixed preparation, stimuli were presented in the free-field in the same anechoic room via an Audax TWO26M0 speaker (Audax Industries, Chateau du Loir, France), located approximately 80 cm from the animal's head at approximately 30° contralateral from the midline and 0° elevation. The free-field speaker was calibrated using a 1/4 inch condenser microphone (Type 4143; Brüel and Kjær) placed near the position of the ferret's head within the recording set-up in the anechoic room. All speaker and headphone driver calibrations measured for these experiments were used to create inverse filters to ensure that the devices produced flat ($< \pm 5$ dB SPL) outputs across 100 - 24000 Hz.

2.5 Stimuli used to identify and characterize neural units

Noise bursts (100 ms in duration; 70-80 dB SPL) were used as a search stimulus during recording sessions, to identify acoustically-responsive neural activity. When neural activity was identified that appeared to be aligned to the onset of noise bursts, these bursts were presented more systematically across a wider range of pseudorandomized levels (10-80 dB, in 10 dB steps). Post-stimulus time histograms (PSTHs) and raster plots of neural responses to 10-30 repetitions of these noise bursts were then inspected to confirm the presence of driven responses.

Next, pure tone stimuli were presented so that the Frequency Response Areas (FRAs)

of neural units could be constructed. The tones ranged, in 1/3 octave steps, from 200 Hz to 24 kHz, and they were 100 ms in duration (5 ms cosine ramped). The intensities of tones ranged from 10-80 dB SPL, in 10 dB increments. Each frequency-level combination was presented in pseudorandom order for at least 3 repetitions, at a rate of 1 Hz. The total number of spikes that a neural unit fired in response to each frequency-level combination (within 200-ms response window) was averaged across all repetitions. These average spike rates were arranged in a matrix, across tone level (rows) and frequency conditions (columns). This matrix was then smoothed with a nine-point running hamming window, and visualized as an FRA plot (Figure 2.1).

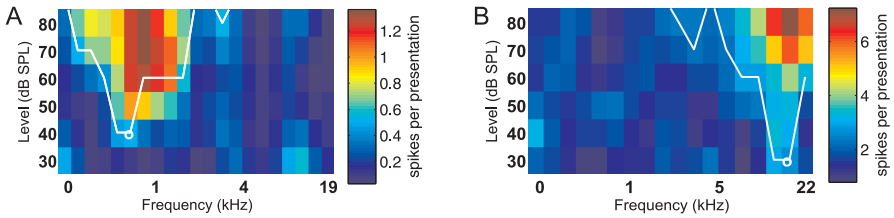


Figure 2.1: Frequency Response Areas (FRAs) of two neurons. (A) Average spike rate (color bar) of a ferret auditory cortical neuron in response to pure tones, as a function of the tone level (y-axis) and frequency (x-axis). The tuning curve (white line) and characteristic frequency (white circle) of the neuron are shown. (B) FRA of a second neuron, which has with a higher CF than the neuron in A.

The spontaneous firing rate of a neuron was calculated as its' average spike count during the 200-ms-long period of silence beginning 800 ms after the onset of each tone. FRA spike rates were compared to a response criterion that was defined as being higher than the neurons' spontaneous firing rate and at least 20% of the maximum FRA firing rate. The frequency tuning curve was constructed from the lowest sound levels neccessary to reach this criterion at each frequency (Figure 2.1, white lines). The Characteristic Frequency (CF) was calculated as the minimum of this tuning curve (Figure 2.1, white circle). If more than one frequency was effective at the minimum level, CF was calculated as the logarithmically-weighted mean of these frequencies. All units with a bandwidth of less than two octaves at 10 dB above threshold were assigned a CF, and those with wider bandwidths were labelled untuned.

2.6 Spike sorting

Spike sorting was performed offline. Single units were isolated from the digitized signal either by manually "clustering" spike shapes according to spike features such as amplitude, width and area, or by using an automated k-means clustering algorithm, in which the voltage potential at seven points across the duration of the spike window served as variables. I also inspected auto-correlation spike histograms, and only cases in which the inter-spike-interval histograms revealed a clear refractory period and the spike shape was clearly defined were classed as single neurons. All other neural units were assumed to be multi-unit clusters. PSTHs and rasterplots of each neural unit were visually inspected in BrainWare. Spike timing data from all single and multi-unit neural clusters that looked like they might be driven by auditory stimuli were exported to MATLAB for further analysis.

Chapter 3

Discrimination of the Temporal Envelopes of Vocal Calls

3.1 Introduction

Neurometric analysis has been a powerful tool for linking electrophysiological activity with perception, and has made important contributions to our understanding of vision and somatosensation (Britten *et al.*, 1992b; Hernandez *et al.*, 2000; Krug *et al.*, 2004; Liu and Newsome, 2005; Luna *et al.*, 2005; Merchant *et al.*, 1997; Parker *et al.*, 2002; Purushothaman and Bradley, 2005), but has been used far less in auditory research (Fay and Coombs, 1992; Javel and Viemeister, 2000; Narayan *et al.*, 2006; Relkin and Pelli, 1987). Classically, neurometric studies investigate whether stimulus-related changes in firing rate within a population of sensory neurons could provide the "psychophysical signal" on which sensory discriminations are based. In practice, firing rates are measured over relatively long time periods, and are subjected to receiver operating characteristic (ROC) analysis to construct neurometric discrimination functions which can be directly compared to psychometric functions measured in two-alternative forced choice (2AFC) behavioral experiments.

This approach has proved useful for the study of stimuli which vary along a single perceptual dimension. However, many natural stimuli are characterized by their complex and individual temporal structure, which cannot easily be represented by a single spike count variable. For example, the characteristic temporal features of many animal

vocalizations are encoded by time-locked responses in the auditory nerve. Early stages of the auditory pathway faithfully preserve this temporal pattern code, using special adaptations like unusually large and secure synapses (Trussell, 2002). As information about a stimulus ascends the auditory pathway towards the primary auditory cortex, individual neurons are increasingly less able to synchronize their discharges to very rapid temporal modulations in their inputs (Cariani, 1999; Joris *et al.*, 2004; Liu *et al.*, 2006; Miller *et al.*, 2002; Wallace *et al.*, 2002; Wang *et al.*, 2008) and the relationship between discharge patterns and temporal features of the stimulus becomes more complicated.

Nevertheless, temporal spiking patterns appear to continue to play a role in representing features of acoustic stimuli in A1 (DeWeese *et al.*, 2003; Lu *et al.*, 2001b; Narayan *et al.*, 2006; Wang, 2000; Winter and Funkenstein, 1973), giving rise to neural sensitivities to stimuli that a spike-rate-based analysis might miss. For example, Wang *et al.* (1995) compared responses of neurons in marmoset A1 to natural and time-reversed marmoset "twitter" vocalizations. They found that marmoset A1 neurons often responded much more vigorously (i.e. with a higher spike rate) to natural than to time-reversed vocalizations. Such a spike rate preference for natural calls was not observed in A1 when the same stimuli were presented to cats (Wang and Kadia, 2001). It is tempting to interpret the selectivity for the natural marmoset twitter calls in marmoset A1 as a neural specialization used to recognize the species-specific calls of the marmoset. However, if such a spike rate preference was *required* to recognize or discriminate natural marmoset calls, this would lead to the uncomfortable conclusion that the A1 neurons of carnivores, such as the cat, do not support discrimination of these unfamiliar vocalizations. However, humans can easily distinguish forward from reversed marmoset calls, as can trained ferrets (Schnupp *et al.*, 2006). How can this discrimination arise in the absence of a spike rate code for maroset call directionality? Schnupp *et al.* (2006) have shown that the temporal discharge patterns of ferret A1 neurons encode information about the identity of natural and time reversed marmoset twitter calls even though spike rate codes read out over time scales coarser than 40 ms are entirely uninformative.

If temporal coding at a timescale of tens of milliseconds in A1 indeed plays an important role in the perception of complex sounds, then one would predict that the temporal patterns should become increasingly harder to discriminate as the stimuli become per-

ceptually more difficult to distinguish. Here, I have tested this prediction by developing a spike-pattern-based neurometric analysis that enabled me to compare the reliability of ferret A1 spike patterns directly with the psychophysical performance of human listeners asked to detect small changes in the temporal structure of a unfamiliar, complex sound - the marmoset twitter call.

3.2 Methods

3.2.1 Stimuli

For this study I required a complex, natural sound with a rich temporal structure. My aim was to alter the temporal structure of the stimulus in a manner that would be progressively more difficult to detect perceptually. I used a digitized recording of a marmoset "twitter" call kindly provided by Dr X. Wang (Wang and Kadia, 2001). I manipulated the temporal structure of this sound using a procedure similar to that previously used in a study of speech recognition (Saberi and Perrott, 1999). Saberi and Perrott (1999) locally time-reversed consecutive segments of speech recordings and found that speech intelligibility decreased when the width of the time-reversed segments was 50 ms or longer. Their method involved simply cutting the sound sample into segments, time reversing each, and then rejoining the segments. Unfortunately, this leads to amplitude transient artifacts at the segment boundaries since amplitudes transitions at the newly apposed segment boundaries are unlikely to be smooth. These transient artifacts are audible as a "buzzing" that is superimposed on the manipulated speech. These artifacts do not affect speech intelligibility, but in the present study I set out to measure whether the temporally manipulated sounds could be distinguished from the original signal, and transient artifacts would have given a strong and rather trivial cue as to which signals had been manipulated by local time reversals. To avoid these artifacts, I cross-faded between consecutive time reversed segments, thereby smoothing out transient artifacts and discontinuities at the boundaries of the segments. I generated the sound segments to be reversed by first modulating the amplitude of the original signal within each segment with a cosine window function. The sequential cosine windows had 50% temporal overlap so that by simply adding each time-reversed segment across the duration

of the stimulus, there was no net amplitude modulation of the original signal. The time-reversed windows were either 10, 20, 40, or 80 ms wide. For decreasing window durations, the locally time-reversed sound more and more closely approximates the original stimulus waveform, converging to the original sound in the limit as the window length approaches zero. For brevity, I refer to the temporally manipulated stimuli as "flipped twitters". Figure 3.1 illustrates these stimuli, and the MATLAB script used to implement these temporal manipulations is provided in Appendix A.

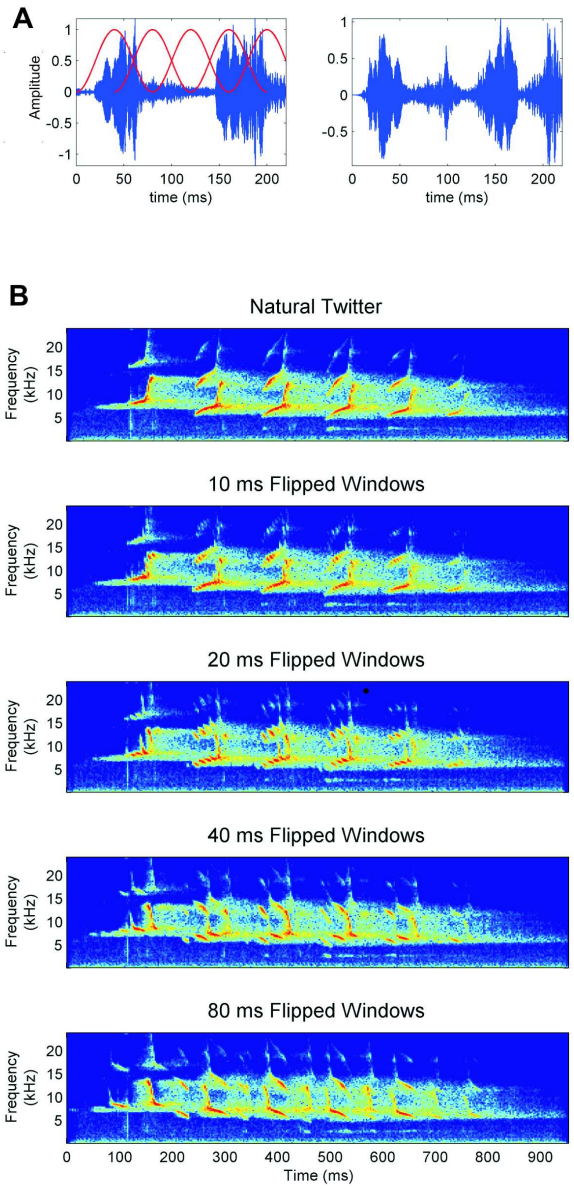


Figure 3.1: Natural and flipped twitter stimuli. (A) Schematic of local time reversals of the acoustic stimulus. In the above panel, the natural stimulus waveform is shown in blue with overlapping cosine functions superimposed in red. The "window" of the stimulus falling within each period of the waveforms was amplified by the cosine function and locally reversed in time. (B) Spectrograms of the natural stimulus (top panel), and the "flipped" stimuli with 10, 20, 40 and 80 ms flipped window widths (below).

3.2.2 Psychophysics

Human listeners ($n = 6$) performed a 2AFC oddity task in a soundproof booth. The psychophysical task was run using custom MATLAB code running on a personal computer, and stimuli were presented at a comfortable listening level over headphones (Sony MDR-CD250, Surrey, UK) using a TDT RM1 mobile processor. On each trial, three twitter stimuli were presented in sequence. The sequences, chosen at random, were of the form "ABB", "BBA", "AAB" or "BAA", where "A" stands for the natural twitter and "B" stands for one of the locally time reversed stimuli. Thus, either the first or the last stimulus in the sequence (but not both) differed from the stimulus in the middle, and the middle stimulus served as the standard for comparison. Listeners were instructed to indicate by key-press whether the first or the last stimulus differed from the middle standard. At least one of the stimuli in the sequence was always a natural twitter stimulus so that the experiment measured the perceptual distance of each of the temporally manipulated stimuli from the original sound. Mutual perceptual distances between pairs of different flipped twitter stimuli were not assessed. Catch trials with "0 ms time reversal" windows (i.e. an "AAA" sequence), were also included in this paradigm to determine if first/third guesses were balanced. Subjects completed ten demonstration trials (two trials per stimulus condition) prior to testing in order to familiarize themselves with the experimental procedure. Training prior to data collection was deliberately kept at a minimum so that the human volunteers, like the animals in the electrophysiology experiments, had very limited exposure to this particular type of stimulus prior to data collection. The subjects then completed four blocks of ten randomly interleaved test trials per stimulus condition, and were permitted to rest between blocks.

Informed consent was received from each subject before they began the experiment and the psychophysical procedures were carried out under the guidelines of the Ethical Review Committee of the University of Oxford.

3.2.3 Electrophysiological recordings

In three adult ferrets, recordings were carried out in primary auditory cortex under general anaesthesia, as per Section 2.2. Spike activity was recorded using 2 MΩ, 4 x 4 silicon array Michigan probes. To characterize the responses of A1 units, the natural and

flipped twitter stimuli were presented at 1.5 sec intervals. Each stimulus was presented between 10-40 times at a comfortable listening level (70-85 dB SPL) in a randomly interleaved order. Noise bursts were used as a search stimulus to find responsive units, and sets of noise bursts and pure tones were used to characterize neurons in area A1 (as per Section 2.4). In one adult, female ferret, extracellular recordings were carried out while the animal was unanaesthetised and passively listening to the stimuli. The awake, head-fixed recording methodology was fundamentally similar to that used in the laboratory of Fritz and Shamma (Fritz *et al.*, 2003), and methodological details are given in Section 2.3. The same stimulus sets were presented to the anaesthetised animal over earphones and to the unanaesthetised animal via a free-field speaker in an anechoic room. The responses of single neurons and small groups of neurons were isolated (Section 2.6), and these are referred to collectively here as "neural units".

3.2.4 Data analysis

3.2.4.1 Preliminary exploration of neural activity

Spike-timing data from acoustically responsive units were exported to MATLAB for analysis. Each neural unit's responses to twitter stimuli were visualized as raster plots, which displayed spike events as a function of time for each stimulus presentation (a response "sweep"). Based on a visual inspection of the raster plots, only units which exhibited stimulus-driven responses were included in further analyses. The first 1280 ms of response data following onset of each twitter stimulus was used in all further analyses because the vast majority of spikes occurred within the first 800 - 1200 ms of the response, and 1280 was a convenient logarithmic step value for the sampling rate analyses.

3.2.4.2 Principal components analysis

To prepare the neural response data for statistical analysis, post-stimulus time histograms (PSTHs) of the responses of each unit were generated so that each 1280-ms-long response sweep was represented as a time-ordered array of spike counts. Each array formed a row in a unit's PSTH matrix. The length of this array was dependent upon the bin widths (or temporal resolution) at which spike counts were sampled over the duration of the stimulus presentation. A bin width of 5 ms, which divided the 1280 ms response into

256 time bins, was used for all of the following analyses except those which specifically examine the effects of bin sizes.

The array of spike counts across time bins can be conceptualized as a point in a 256-dimensional space. Principal Components Analysis (PCA) was carried out to reduce the dimensionality of this response space. PCA transforms data in a high-dimensional vector space onto a new orthogonal coordinate system such that the smallest number of these new principal component dimensions can explain the largest proportion of the variance in the data (Hernandez *et al.*, 2000). Often most of the structure of the data can be captured by a relatively small number of high-order principal components, while many of the low-order principal component dimensions are largely attributable to noise and can be discarded.

Here, I kept only the first five principal components to represent each response pattern. Thus, each response pattern was first quantified as a PSTH of 256 bins, each 5 ms wide, and then represented as a 5-dimensional point in principal component space (pc-space). Finally, each principal component value was multiplied by the corresponding eigenvalue for that principal component, in order to weight the principal components by the amount of response variance they explained.

3.2.4.3 Receiver operating characteristics

The Euclidean distance between two 5-element response vectors in pc-space provided a measure of similarity between these responses, because neural responses with more similar temporal discharge patterns would be closer in pc-space. The Euclidean distance between two responses to the same stimulus was referred to as a "within-category" distance, and "across-category" distances were measured between responses to natural and flipped twitters. For each unit, the distributions of all within-category and across-category Euclidean distances were collected for each flipped twitter condition.

In the psychometric oddball task described above, human subjects had to compare one natural and one flipped target stimulus to a standard, and we can assume that these stimuli evoked responses in their A1 which are essentially similar to those I recorded in the auditory cortex of the ferrets. The subjects thus had to decide which of the target responses was more similar to the standard, and here I model this decision process

by comparing one within-category and one across-category response distance drawn at random from the observed distributions. Consequently, a unit should support good psychometric discrimination performance if the across-category distances were reliably greater than the within-category distances (i.e. if there is little overlap in the within- and across-category distributions).

Similarly to previous authors (Britten *et al.*, 1992b), I used ROC analysis to construct neurometric curves for each unit. I constructed ROC curves for each flipped twitter condition by comparing the differences in within-category and across-category distances against a wide range of possible criterion values. The area under the resulting ROC curves provides a measure of discrimination performance (expected proportion of correct responses) that is independent of potential criterion bias and can be directly compared to behavioural performance on the present psychophysical task (Green and Swets, 1974).

The standard error of a receiver operator characteristic curve (Hanley and McNeil, 1982) was calculated as:

$$SE = \sqrt{\frac{A(1-A) + (n_a - 1)(Q_1 - A^2) + (n_w - 1)(Q_2 - A^2)}{n_a n_w}}, \tag{3.1}$$

where A is the area under the curve, n_a and n_w are the number of across- and within-category distances respectively, and $Q1$ and $Q2$ were estimated by:

$$Q_1 = \frac{A}{2 - A}, \tag{3.2}$$

and

$$Q_2 = \frac{2A^2}{1 + A}. \tag{3.3}$$

3.3 Results

3.3.1 Psychophysics

Six human subjects completed a 2AFC oddity task in which they were asked to discriminate a natural marmoset twitter call from locally time reversed ("flipped") copies of the stimulus (see Figure 3.1). Flipped twitters with 0 ms reversed window widths are natural, unmanipulated twitters. Discrimination scores for each subject across flipped

twitter conditions are shown in Figure 3.2. The performance of human listeners in discriminating natural and flipped twitter stimuli declined as the width of the flipped time bins decreased. Performance on the discrimination task was near perfect for all subjects when the reversed windows were 80 ms wide, and was impaired for most subjects when the windows were 40 ms wide. When the windows were 20 ms wide, the mean discrimination performance was near chance, and subjects could not discriminate 10 ms flipped twitters from natural ones ($p > 0.05$, based on a binomial distribution with 20 trials). Performance did not significantly differ between trials on which the non-standard stimulus came first in the sequence verses those in which it was the last in the sequence (Wilson score interval, $p > 0.05$).

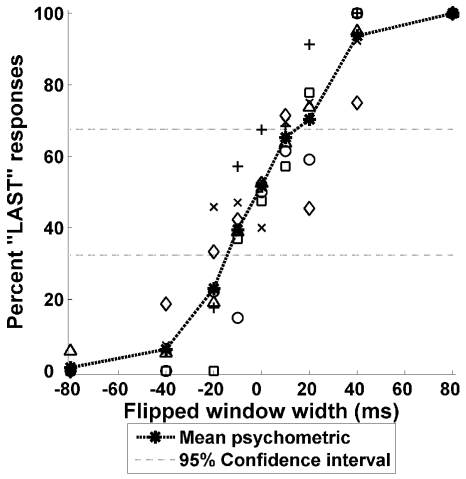


Figure 3.2: Human discrimination performance on the twitter discrimination task. Flipped window widths of +20 ms and -20 ms both refer to oddball stimuli which were locally time reversed with 20 ms windows, but a negative value indicates that the oddball stimulus was the first of the three stimuli presented in the trial, while positive values indicate that the oddball stimulus was presented last. The six different types of symbols show the mean percentage of trials on which each subject reported that the last sound in the sequence was different, as a function of stimulus condition. The continuous line shows the mean of all six subjects.

3.3.2 Calculating neurometric performance

Multi-electrode extracellular recordings of responses to natural and flipped twitters were collected in the left primary auditory cortex of three adult, anaesthetised ferrets. In total, 142 stimulus-driven units were identified: 77 in the first ferret, 28 in the second, and 37 in the third. These included 120 single units and 22 multi-unit clusters.

Visual inspection of standard raster plots indicated that a subset of the units in A1 responded to twitter stimuli with reliable and stimulus-specific temporal discharge patterns (Figure 3.3), as observed previously (Schnupp *et al.*, 2006; Wang *et al.*, 1995). While the spike timings and spike counts of these units varied somewhat as the same stimulus was repeatedly presented, and while their temporal discharge patterns appeared to contain noise from spontaneous activity across trials, this subset of units nevertheless exhibited reproducible temporal discharge patterns which, if appropriately decoded, ought to support the discrimination of the natural from some of the flipped twitter calls. PSTHs were generated for each unit, in which the response to each twitter presentation was represented as an array of spike counts in 5-ms time intervals following stimulus onset. For each unit, PCA was then carried out to reduce the dimensionality of these response vectors (see Section 3.2.4.2). The first five principal components together were sufficient to explain between 13 - 60% ($mean = 27.3\%$) of each units' response variance (Figure 3.4).

A discrimination algorithm was designed to attempt to distinguish natural from flipped twitters based on the information available in the temporal discharge patterns of A1 units (see Section 3.2.4.3). If the temporal discharge pattern of a unit was reliably similar for repeated presentations of the same stimulus and sufficiently different when another stimulus was presented, then two responses of that unit to the same twitter stimulus (a "within" twitter comparison) should be closer in pc-space than two responses to different twitter stimuli (an "across" twitter comparison). But if the responses lacked reproducible and stimulus-selective spiking patterns, then their distances in pc-space would be uninformative about the categorization of the twitter stimuli to which they were responding. Histograms of within- and between-category Euclidean distances for each flipped twitter condition for the two units illustrated in Figure 3.3A and B are shown in Figure 3.5A and B, respectively. Figure 3.5A clearly illustrates that the distribution

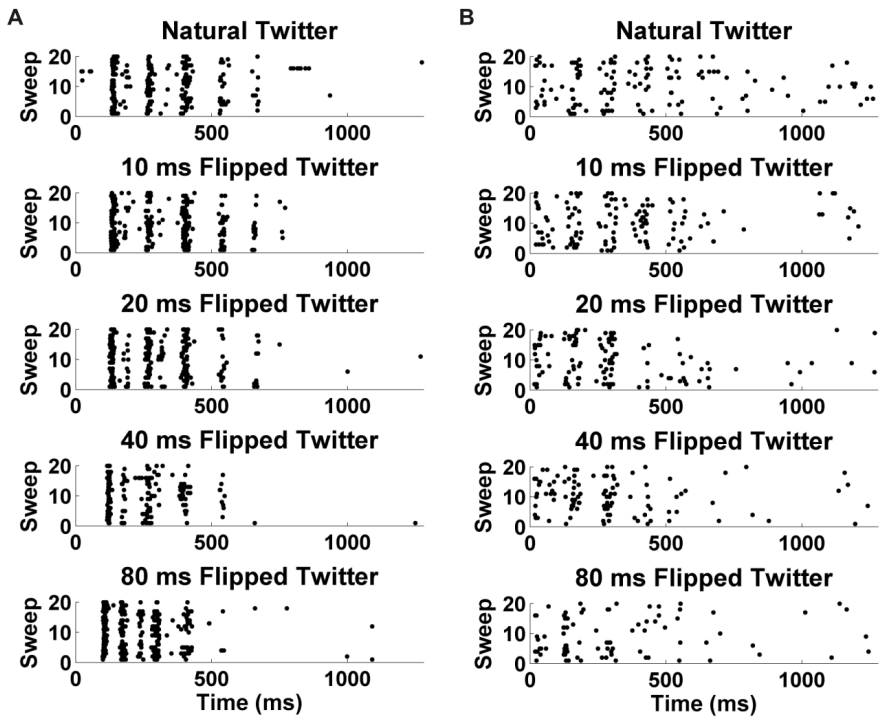


Figure 3.3: Responses of two ferret A1 units to the presentation of twitter stimuli. (A) Responses of one neural unit to the natural twitter call (top panel) and the four locally time-reversed calls (below). Each row of dots shows the response to a single presentation of the stimulus, with each dot indicating the time of occurrence of an action potential following stimulus onset (time 0 ms). (B) Raster plots showing the responses of a different unit to the same stimuli.

of this neurons' responses to flipped and natural twitters becomes highly separable when the flipped window is at least 40 ms wide. Using analyses derived from Signal Detection Theory, ROC curves were calculated for each flipped twitter condition (Figure 3.5C-D; Green and Swets, 1974), and a neurometric discrimination curve for each unit was derived from these ROC curves (Figure 3.5E-F).

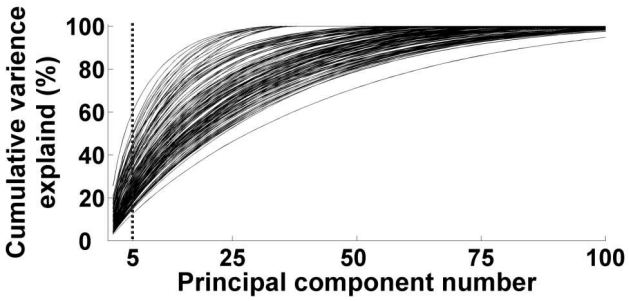


Figure 3.4: Amount of response variance explained by principal components across units. Percentage of response variance captured by principal components analysis as a function of the number of principal components, for each of the 142 units in my sample (black lines). The first five components (indicated by the dashed, vertical line) explain between about 15 - 60% of the variance, depending on the unit.

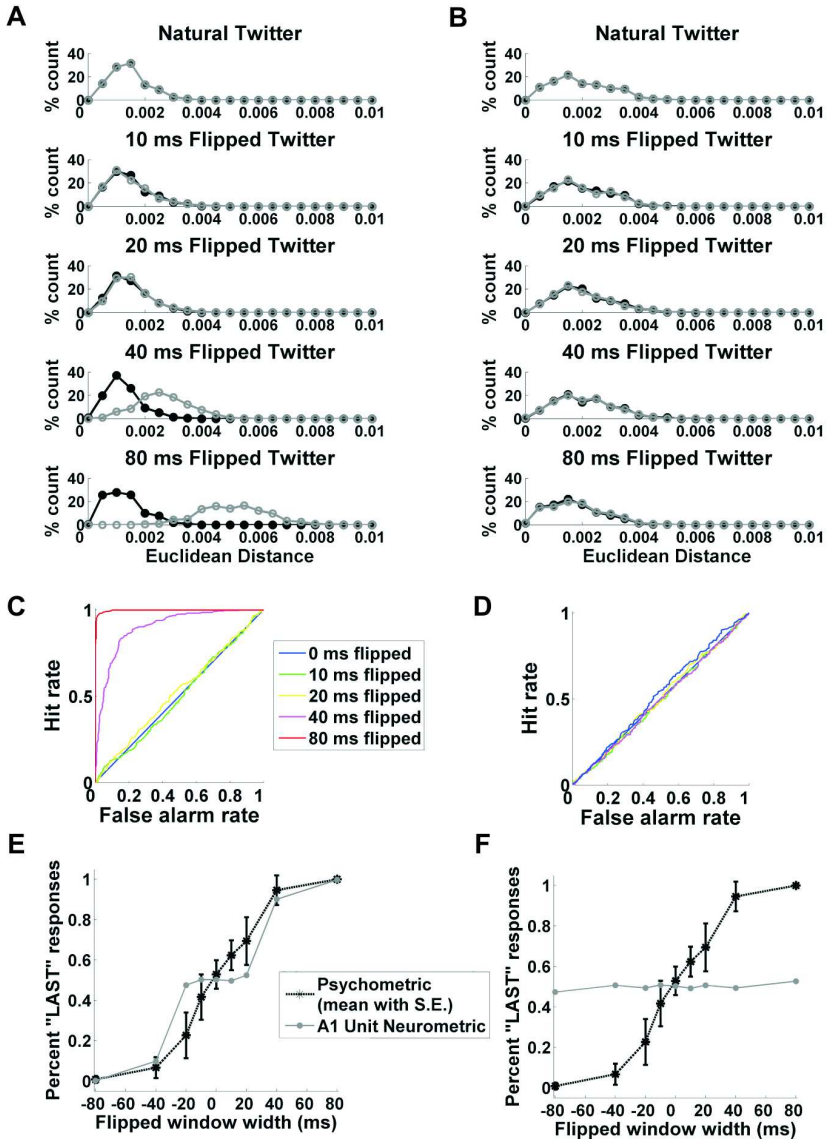


Figure 3.5: Calculating neurometrics for two ferret A1 units. (A & B) Distributions of within-category response distances (filled black circles) and across-category distances (open grey circles) for the units shown in Figures 3.3A and B, respectively. (C & D) ROC curves constructed from the distributions shown in A and B, for each stimulus condition (see legend). (E & F) Neurometric scores (grey circles) for the two units, calculated by plotting the area under the ROC curves in C and D. The average psychometric performance of the human volunteers with standard error bars is also shown for comparison (black asterisks).

3.3.3 Discrimination performance of individual A1 units

The neurometric functions of all 142 units are plotted along with the psychometric function in Figure 3.6A. From this figure, it is clear that neurometric discrimination curves varied widely from unit to unit. Neurometric performance at each flipped twitter condition ranged from chance performance up to psychometric averages. As is illustrated in Figure 3.7, the neurometric scores for many units were significantly above chance (i.e. at least two standard errors of the ROC curve above chance, $p < 0.05$), and the discrimination performance of these units improved as flipped window widths increased.

Figure 3.6B shows the results of the same neurometric algorithm when a single spike rate variable, calculated over the whole response period, was used instead of temporal pattern information. These neurometrics did not reach psychometric performance values, and they did not show systematic improvement in discrimination as the flipped windows widened. Of course one might ask whether the much poorer performance of the single spike rate neurometric compared with the spike pattern neurometric might not simply be attributable to the fact that the bin width of 5 ms used for the temporal pattern analysis was a more fortuitous choice then the 1280 ms wide window used in the spike rate analysis shown in Figure 3.6B. In other words, if one were to use a single spike rate measure obtained from a single, suitably chosen 5-ms interval, would one be able to match the neurometric performance reached by the pattern analysis? To test this, I calculated single spike rate neurometrics for each 5 ms wide time bin, from 0 to 1280 ms in 5 ms steps, and for each unit chose the 5 ms wide time bin that gave the best neurometric performance (i.e. the greatest area under the neurometric curve). I refer to this as the "best bin" neurometric analysis.

The best bin neurometrics obtained for each of the 142 units are shown in Figure 3.6C. While these neurometrics performed considerably better overall than the average spike rate neurometrics shown in Figure 3.6B, they nevertheless failed to reach the same level of performance as the full pattern neurometrics shown in Figure 3.6A, nor do they approximate human psychometric performance as well as the pattern neurometrics. Figure 3.6D plots the difference between the spike pattern neurometric and the best bin neurometric for all units that reached at least 70% performance in either neurometric analysis, for any of the stimulus conditions. Such a comparison is of little interest for units which

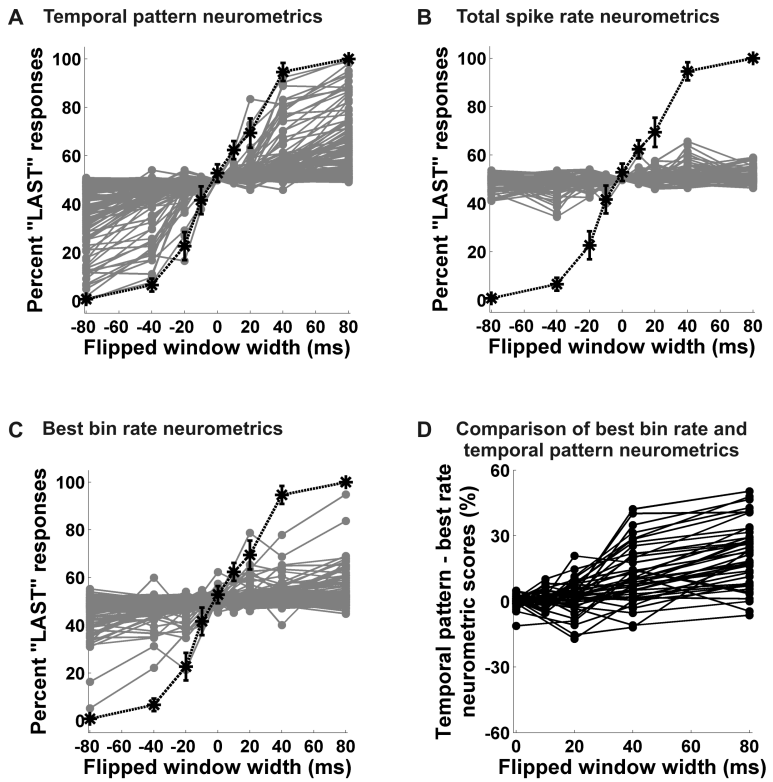


Figure 3.6: Neurometric functions for a population of A1 units in anaesthetised ferrets. (A) Temporal pattern neurometrics for all 142 units are plotted superimposed (grey circles). The average psychometric performance (with standard error bars) of the human listeners is also shown for comparison (black asterisks). (B) Spike rate neurometrics for the same 142 units (grey circles) are plotted, along with the average psychometric curve (black asterisks with standard error bars). (C) Spike rate neurometrics calculated for each of the 142 units using the spike rates observed in a single, optimally chosen, 5-ms-wide time bin. (D) Difference between temporal pattern neurometric curves shown in A and "best bin" neurometric curves shown in C, for all units whose neurometric performance reached at least 70% in either analysis, for any stimulus condition.

performed poorly with all neurometric algorithms, and they were therefore not plotted in Figure 3.6D in order not to clutter the graph. These differences were positive (i.e. pattern neurometric performance exceeded the best bin neurometric performance) in the large majority of cases. I also repeated this analysis using 10, 20 and 40 ms wide optimally chosen time bins and obtained very similar results (not shown). Full pattern neurometrics consistently outperformed single best bin neurometrics, and single best bin neurometrics never closely matched the human psychometric curve. These results suggest

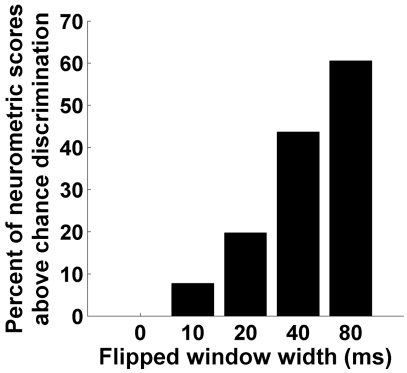


Figure 3.7: More A1 units show significant neurometric discrimination for wider flipped windows. Histogram shows the number of units for which neurometric discrimination performance was significantly above chance (at $p < 0.05$, standard error of the ROC curve), for each stimulus condition.

that in A1, the discharge pattern, and not simply the number of spikes in any one time window, carries important information about the identity of temporally varying sounds.

In the neurometric algorithm above, the temporal discharge patterns across 256 time bins was quantified using only the first five principal components. It is possible that incorporating more information about response variation into the decision algorithm by including more principal components would improve neurometric performance. An investigation of this question demonstrated the importance of weighting the principal components by their eigenvalues in the algorithm. When neurometrics were calculated using the raw principal component values, performance varied considerably depending upon the number of principal components used to represent the response (Figure 3.8A). Initially, using more than one principal component in the algorithm improved neurometric performance for some units, but increasing the number of principal components beyond the first three caused a decline in performance for most units. This interpretation of neural responses would require an observer to consider an optimum number of principal components for each cell in order to maximize discrimination performance. In contrast, when the principal components were weighted by their eigenvalues to account for their relative importance in explaining the response variance, using only the first five principal components yielded near-maximum discrimination performance for all neural units (Figure 3.8B). Statistically, the optimal number of principal components varied more across units in the unweighted condition than in the weighted condition (paired t-test, $p < 0.05$).

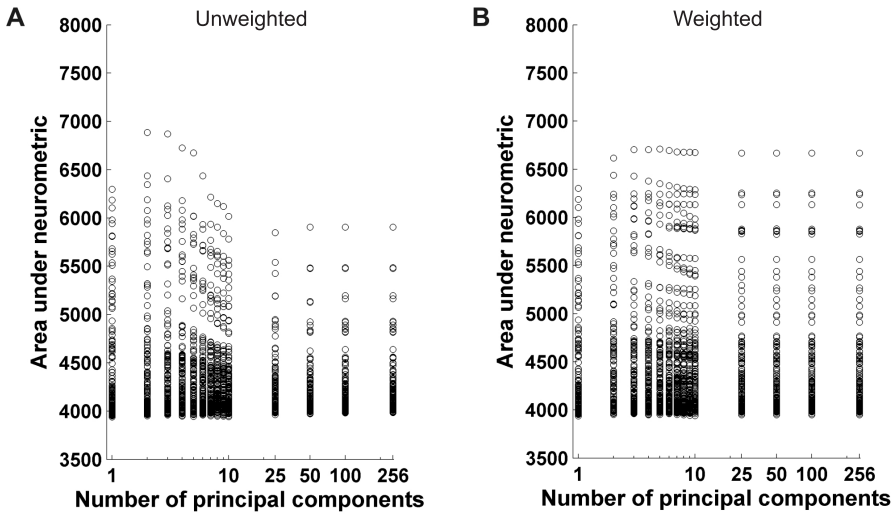


Figure 3.8: The number of principal components used to represent neural response patterns affects neurometric performance. (A) The area under a neurometric curve is used here as an estimate of neurometric performance for each unit across all stimulus conditions. The area was calculated across the positive end of the x-axis (0 - 80ms), such that perfect discrimination on all conditions would produce a rectangle with an area of 8000 ms% (i.e. 80 ms times 100%), and chance performance would produce a rectangle of 4000 ms% (80 ms times 50%). For each neuron (open circles), the area under the neurometric is indicated on the y-axis, when calculated using a variable number of principal components to quantify the temporal spiking pattern (x-axis). Here, the neurometric calculations were based on the distances between raw principal component values (i.e. PCA eigenvectors). (B) Results of the neurometric algorithm when each principal component was weighted by its corresponding eigenvalue.

3.3.4 Twitter discrimination based on the pooled activity of A1 populations

From Figure 3.6A, it is clear that most units' neurometric curves were considerably flatter than the psychometric curve, and only a small proportion of neurometric curves approximated the psychometric curve closely. One might expect that the perceptual decision making process operating in an observer engaged in a sensory discrimination task might pool or combine information across a number of neurons rather than relying on the output of just one single neuron in auditory cortex. It is therefore of interest to ask what sort of neural pooling mechanisms might be able to account for psychometric performance. In past studies, there have been at least two commonly used methods of combining in-

formation across units: "enveloping" neurometric curves, and pooling responses prior to neurometric analysis (Parker and Newsome, 1998). Here, both techniques were used to examine the neurometric performance of groups of units.

If a distinction between several stimuli is required, behavioral decisions might be based on the outputs of only those neurons that respond most selectively to these stimuli (Barlow, 1972). The consequences of this process can be represented graphically by drawing an optimum or "best" envelope along the contour of the best neurometric curves (Parker and Newsome, 1998). By applying this best envelope approach to my data, I found that the three best units in the measured sample were sufficient to obtain a neurometric curve that closely matched psychometric discrimination performance (Figure 3.9).

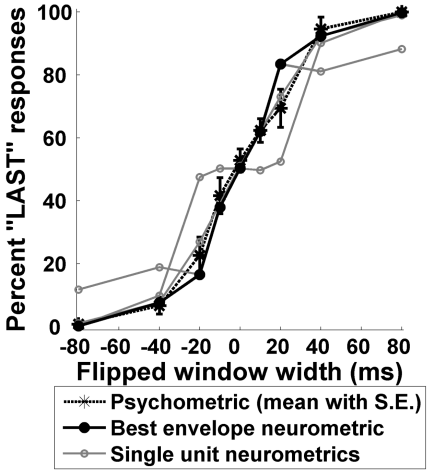


Figure 3.9: Upper limit of neurometric performance curves across individual A1 units. The "best envelope" neurometric curve (black circles) was constructed from the three most informative units. The individual neurometrics of these three best units (grey circles) and the average psychometric performance (black asterisks, with standard error bars) of the human listeners are also shown.

In the past, enveloping has most often been used to investigate more peripheral areas of the nervous system (Parker and Newsome, 1998). Generally speaking, how the brain might implement an algorithm that is equivalent to the enveloping of neurometric functions is as yet unclear. Thus, perhaps a more appropriate alternative to enveloping best neurometrics is to pool the responses of several simultaneously active neurons to determine how the algorithm would perform using the temporal patterns of summed spike inputs. This is especially relevant in A1, where complex stimuli are thought to be represented in the activity of populations of neural units distributed throughout the tonotopic map (Wang *et al.*, 1995).

The 4 x 4 Michigan electrodes used in the anaesthetised ferret experiment can simultaneously collect responses from several neural units at each of 16 electrode recording sites during each penetration into the brain. This allowed me to examine the temporal spiking patterns of neural populations by pooling the responses of neural units in each penetration. To preserve spike timing information, the 256-element response vectors for each trial were pooled by summing the spike rates in each time bin across all units. Once the responses had been pooled for the complete set of trials, they were transposed using PCA and temporal discharge neurometrics were calculated as for individual units. Neurometrics of pooled units across 14 multi-electrode penetrations are shown in Figure 3.10. The neurometric curves of each individual unit are also shown for comparison, as are the human psychophysics data.

In all 14 penetrations, the pooled neurometric discriminated twitters better than most individual units alone, demonstrating that the contribution of simultaneous activity from multiple neurons tends to improve neurometric performance even if some of these units themselves have noisy, poorly informative responses to the stimuli. The performance of the pooled neurometrics varied considerably from site to site. In penetration two, neurometrics based on each individual unit performed near chance, but the neurometric performance of the pooled activity of all these units closely matched the human psychometric curve. In four other penetrations in which the individual neurometrics performed poorly, the activity of pooled units also performed near chance. The collection of pooled neurometrics varied in performance from chance up to human psychometric values (Figure 3.11A), demonstrating again a correspondence between the discrimination capacity of A1 responses and the behavioural performance of human listeners.

In some previous neurometric investigations of spike rate codes, (e.g. Britten *et al.*, 1992b), the responses of single units recorded on separate occasions have been pooled by summing their spike counts. Correlations in firing across neurons may limit the amount of independent information that a further neuron can add to the pool, so these studies incorporated an inter-neuronal correlation term into their pooling models to account for this effect. When neural responses are pooled for the purposes of examining temporal spiking patterns, the effects of inter-neuronal correlations are less clear. Each spike in a response may be a part of the neural "signal" that represents the sound if its timing is

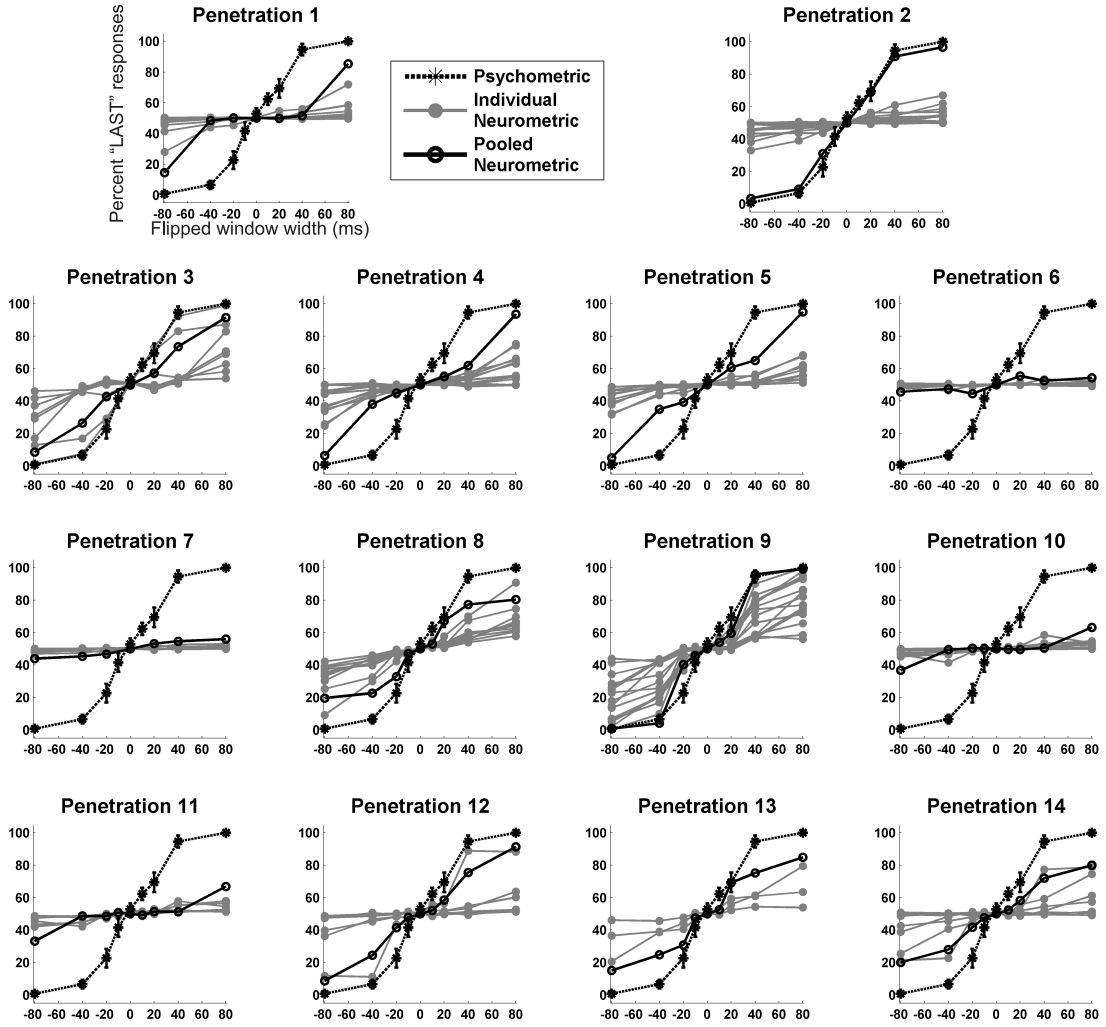


Figure 3.10: Neurometric curves of pooled activity within each electrode penetration. Each plot shows the neurometric curve obtained from the pooled simultaneous responses of all the units recorded in a given electrode penetration (black circles), along with the individual neurometrics of each of these units (grey circles). The average psychometric performance (with standard error bars) of the human listeners is also shown (black asterisks).

determined by the stimulus. Or it may contribute to the neural "noise" if its timing is unrelated to the sound being presented. Pooling information across a neural population can enhance the signal-to-noise ratio and lead to better neurometric performance only if noise correlations in the pooled population are small, so that each neuron in the pool gives a more-or-less independent look at the stimulus.

To directly investigate whether noise correlations affected the pooled neurometric

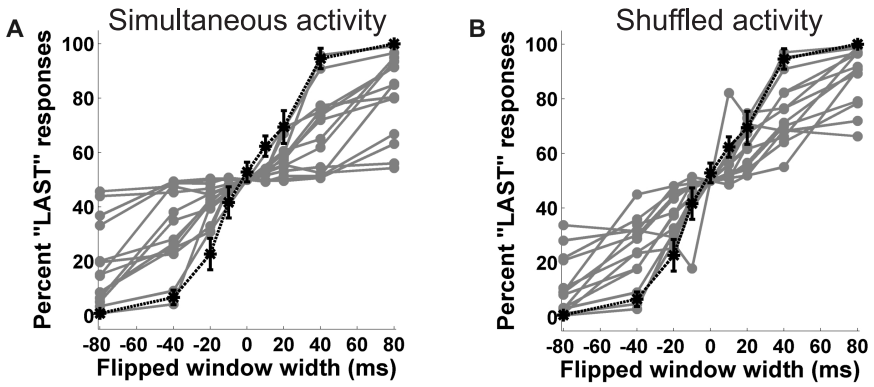


Figure 3.11: Summary of pooled neurometric curves across all penetrations. (A) Neurometric curves derived by pooling simultaneously recorded activity in each multi-electrode penetration are shown (grey circles) along with the human psychometric curve (black asterisks with standard error bars). (B) Neurometric curves derived from the same data, but activity was pooled after randomly shuffling the order of responses for each unit to remove trial-to-trial correlations in the activity of neurons in each penetration (grey circles). The human psychometric curve is again shown for comparison.

performance obtained from my data, I removed correlations between neural responses in the pool by shuffling responses across repetitions of the same stimulus. These "shuffled", pooled neurometrics were compared to those derived from pooling only simultaneously recorded activity. The rationale behind this approach is similar to that motivating the use of shift-predictors in cross-correlation histogram analyses (Moore *et al.*, 1966). For each unit in a pool, the order of responses was randomized within each stimulus condition prior to summing the responses across units. Therefore, the response sweeps of each neuron were added to responses of other neurons in the pool that were evoked by the same stimulus, but at different points in time.

The neurometric performance of the shuffled pools for each of the 14 penetrations are shown in Figure 3.11B. The pooled neurometric performance for most penetrations improved with shuffling, indicating that noise correlations across units do reduce the benefit obtained from pooling neural responses. The upper envelope of these pooled neurometrics was largely similar to the psychometric curve, with the exception of the 10 ms condition, where the neurometric of one penetration performed better than human listeners.

3.3.5 Effects of temporal resolution and integration windows

Given that the discrimination algorithm used to calculate neurometric curves makes use of temporal pattern information, one would expect its performance to decline if the temporal resolution at which these patterns are read out is inappropriate. Binning spike data can reduce noise that results from temporal jitter in the stimulus-locked responses of auditory neurons (Hill *et al.*, 1989) and provide a more efficient encoding of information about the stimulus, but if the binning is too coarse important temporal patterning can be lost. Therefore, there should be a range of resolutions at which A1 responses can be binned to provide the most efficient spike pattern code without disrupting discrimination performance.

To explore the possible effects of sampling rate, enveloped neurometrics were constructed as for Figure 3.9, but the spike patterns were calculated using a range of bin sizes in logarithmic steps from 2.5 to 1280 ms. The results of this analysis, shown in Figure 3.12, indicated that the discrimination algorithm performed best when the spike data were sampled at the finest temporal resolutions, as might be expected since these contain the most information about spike timing. However, neurometric performance remained high when the resolution was dropped to 10 ms. For resolutions of 20 - 40 ms, neurometric performance remained near perfect for the easiest discrimination task (i.e. 80 ms flipped twitters), but was impaired for more difficult twitter discriminations. At still coarser resolutions, performance was severely impaired across the entire stimulus set.

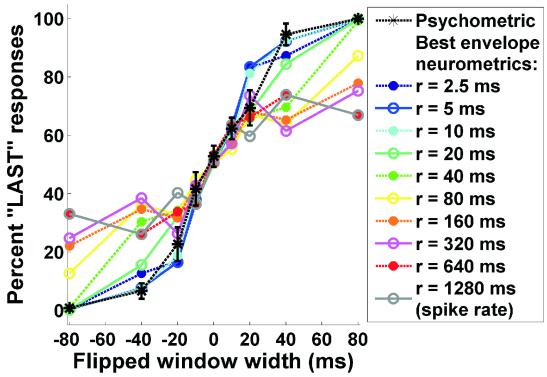


Figure 3.12: Sensitivity of neurometric curves to changes in temporal resolution. Best envelope neurometrics were constructed using a range of temporal bin widths (coloured lines, see legend). Note that some lines are partially occluded by others. The average psychometric performance (with standard error bars) of the human listeners is also shown (black asterisks).

Analyzing the responses at an appropriately fine temporal resolution clearly seems

important, but it may not be necessary to subject the whole neural response to neurometric analysis. Analyzing only the first few 100 ms of the response after stimulus onset might suffice to achieve good neurometric performance. I repeated the neurometric analysis described above at a 5 ms temporal resolution, but here I progressively shortened the length of the analysed response from the full 1280 ms to 800, 400, 200, 100 and 50 ms. The results of this analysis are shown in Figure 3.13.

Reducing the analyzed response duration to 400 ms can be seen to have little impact on the neurometrics, and even if only the first 200 ms of the response are analyzed a proportion of the units are able to maintain good neurometric performance. However, when the analyzed response periods were reduced further to 100 ms or less neurometric performance collapsed. The present results are therefore broadly similar to those observed by Narayan *et al.* (2006), who reported optimal temporal resolutions for the analysis of song bird auditory forebrain responses to complex vocalizations of about 10-20 ms, and best temporal integration windows in the order of several hundred milliseconds (597 ms median).

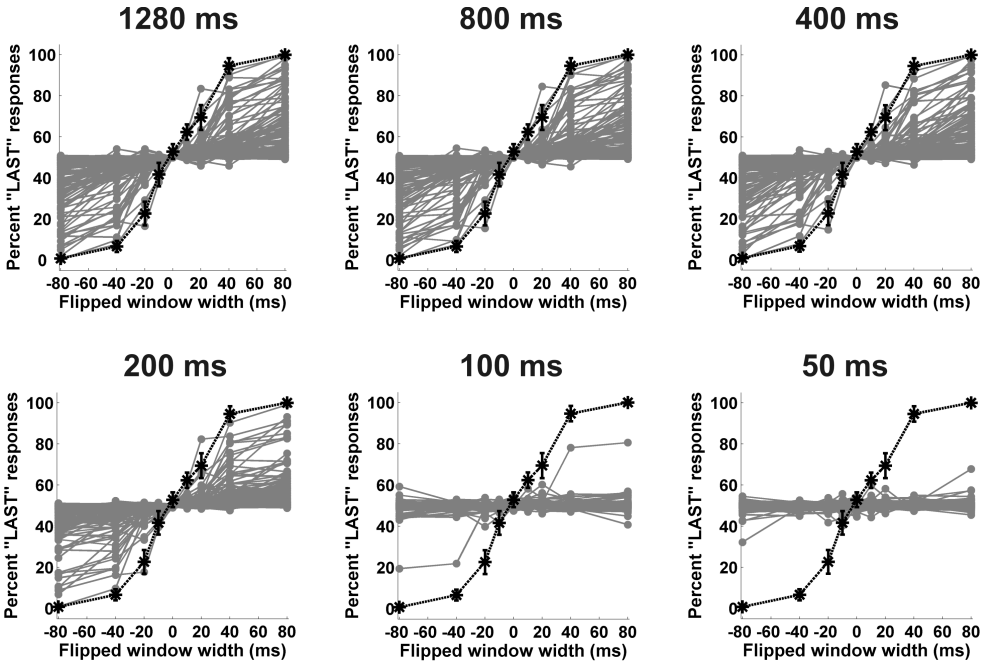


Figure 3.13: Effect of reducing the response integration window on neurometric performance. Neurometric curves were calculated at a 5 ms temporal resolution and plotted as in Figure 3.6A, but the duration of the analyzed response period was progressively reduced from the full 1280 ms duration to the integration window indicated above each panel.

3.3.6 Neurometrics in the unanaesthetised animal

If the above similarities between psychometric and A1 neurometric curves indicate that perceptual performance is limited by the resolution of temporal spiking patterns in A1, these similarities should hold when the same neurometric algorithm is used to interpret the auditory cortical responses of an unanaesthetised animal. An unanaesthetised animal more closely resembles the behavioural state of the human listeners.

Figure 3.14 shows the neurometric performance of 29 auditory cortical single units recorded in an unanaesthetised ferret. As in the anaesthetised data, neurometrics based on the temporal discharge patterns of auditory cortical neurons in the awake, passively listening ferret varied from chance performance up to human psychometric values. Again, the best envelope neurometric closely matched psychometric performance (Figure 3.14A). In contrast, neurometrics based on the spike rates of these responses did not reach behavioural performance levels (Figure 3.14B).

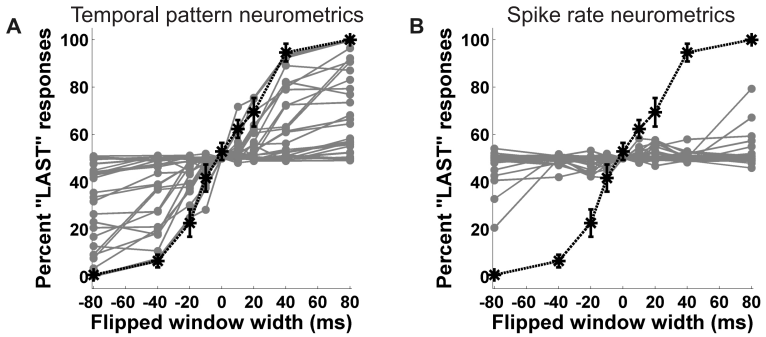


Figure 3.14: Neurometric performance of a population of A1 units in an unanaesthetised ferret. (A) Temporal pattern neurometrics for all 29 units are plotted superimposed (grey circles). The average psychometric performance (with standard error bars) of the human listeners is also shown for comparison (black asterisks). (B) Spike rate neurometrics for the same 29 units (grey circles) are plotted, along with the average human psychometric curve.

3.3.7 A note on the stimulus-response relation

Many previous studies have investigated the temporal and spectral characteristics of complex stimuli that influence neural firing responses (Narayan *et al.*, 2006; Theunissen *et al.*, 2000; Wang *et al.*, 1995). While significant progress has been made in this area, some aspects of the stimulus-response relation remain unclear. In the present study, rather than focus on the nature of the stimulus-response relation, I have focused on the relation between neural responses and psychophysical performance. However, a simple visual inspection of the pooled temporal discharge patterns of A1 units in anaesthetised ferrets suggests that the cortical population as a whole responded with correlated firing rates that were synchronized to discrete events in the temporal envelope of the stimulus (Figure 3.15A). The pooled PSTH was similar for auditory cortical neurons recorded in the unanaesthetised animal, but with stronger responses to the onset and offset of the sound (Figure 3.15B). This type of synchronization of auditory cortical firing to transients in complex vocalizations has been previously described in other species (Wang *et al.*, 1995; Winter and Funkenstein, 1973). At the single unit level, the stimulus-response relationship may be more complex, but a more complete description of this relation was not undertaken here.

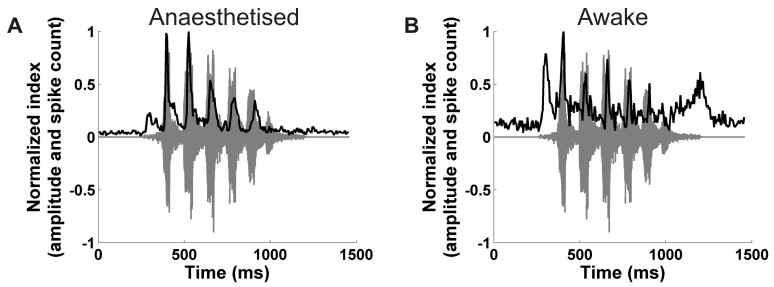


Figure 3.15: Pooled neural responses are locked to discrete stimulus events. (A) The normalized post-stimulus time histogram of the pooled responses of 142 units from the anaesthetised dataset is shown (black line) superimposed on the waveform of the natural twitter stimulus (grey line) to which it is responding. Stimulus onset occurs at 250 ms. (B) Normalized PSTH of the pooled responses of 29 units recorded in the unanaesthetised animal.

3.4 Discussion

3.4.1 Examining temporal spiking patterns in auditory cortex

Previous studies that have compared neurometric functions in sensory cortex to psychometric performance have focused almost exclusively on spike rates (Britten *et al.*, 1992b; Krug *et al.*, 2004; Liu and Newsome, 2005; Tolhurst *et al.*, 1983). However, Hernandez *et al.* (2000) explored the links between neurometrics based on phase-locked, periodic firing patterns and psychometric performance. Hernandez and colleagues investigated the respective roles that spike rate and periodic discharge patterns in the primary somatosensory cortex might play in encoding the frequency of a vibrating tactile stimulus. They found that spike rate based neurometrics were adequate to account for psychophysical thresholds, while neurometrics derived from the periodicity of neural discharge resulted in thresholds that were much lower than those obtained behaviorally. These results served to reinforce the notion that, in sensory cortex, mean spike rates correlate with perception while temporal discharge patterns do not. The present results suggest that this may not be the general case for all stimuli. I found that spike rate based neurometrics failed completely to account for behavioral performance, while the spike pattern based neurometric analysis devised here closely matched the psychometric curve. There are of course many important differences between the present study and that of Hernandez *et al.* (2000), the most important of which is likely to be the nature of the temporal change in the stimuli

that is to be detected. Hernandez *et al.* investigated a tactile frequency discrimination task, while in my case subjects were required to detect a change in the temporal structure of a complex natural sound, which might be best described perceptually as a change in the "rhythm" of the sound. In addition, the neurometric analysis used by Hernandez *et al.* was limited to the examination of temporal discharge patterns that are periodic, whereas the present approach allows one to investigate temporal discharge patterns of any form.

I envisage that the adapted neurometric method introduced here may prove to be a valuable tool in future studies of auditory cortex function. Despite dramatic progress in our understanding of the functional organization of auditory cortex in recent years, many aspects of the roles played by auditory cortex in perception remain uncertain. Early lesion studies on monkeys suggested that damage to auditory cortex impaired an animal's ability to perform cognitive tasks involving sound localization (Heffner and Masterton, 1975) or the discrimination of conspecific vocalizations (Heffner and Heffner, 1986), but these studies also noted that cortical lesions did not impair reflex-like orientation towards unexpected sounds. More recent observations on a human patient who had suffered extensive bilateral auditory cortex damage appear to confirm that such lesions lead to a total loss of auditory awareness, a "cortical deafness", even though accurate reflexive orienting behavior to sudden sounds remains intact (Garde and Cowey, 2000). These observations suggest that auditory cortex may play a critical role in every aspect of the conscious perception of sound, even if sophisticated sub-conscious auditory processing seems to be possible without it.

In recent years, a number of researchers have investigated whether particular aspects of auditory perception can be ascribed to particular auditory cortical areas. A commonly used approach is to investigate stimulus-induced changes in discharge patterns across various cortical areas (Middlebrooks, 1999; Narayan *et al.*, 2006; Rauschecker, 1998; Recanzone, 2000; Wallace *et al.*, 2005), but to link this cortical activity to perception one must also demonstrate that the neural responses correlate with perceptual decisions as assessed in psychophysical experiments. Ideally, this should be done on a trial-by-trial basis in an animal engaged in the perceptual task. Studies which have managed to achieve this in the visual (Britten *et al.*, 1992b) and somatosensory (Hernandez *et al.*,

2000) systems have been very influential, but similar experiments have not yet become common in the auditory system. This could perhaps be due in part to the fact that the simple rate coding assumption inherent in conventional neurometric analysis appears inappropriate for auditory cortex. Many recent studies have emphasized the importance of the temporal patterns of responses in auditory cortex in representing acoustic stimuli (Brugge *et al.*, 1996; Furukawa and Middlebrooks, 2002; Gehr *et al.*, 2000; Narayan *et al.*, 2006; Nelken *et al.*, 2005; Schnupp *et al.*, 2006). Consequently, the neurometric approach may have to be adapted to be sensitive to differences in spike patterns before it can usefully be applied to auditory cortex. An important contribution made by the present study is to show that this can indeed be done successfully.

This claim has recently been corroborated by experiments from another lab. Near the time of publication of the present study, Wang *et al.* (2007) published the results of a neurometric examination of vocalization representations in the homologue of primary auditory cortex in songbirds. Having independently devised a neurometric approach similar to ours, which could be used to examine temporal spiking patterns, they presented findings that are in agreement with my own. As in the present study, they showed that discriminations based on the temporal discharge patterns, but not the spike rates, of the "best" cortical neurons can account for listeners' discrimination of the temporal envelopes of vocalizations. They also found that pooling the correlated activity of populations of neurons led to better discrimination scores than those of single units alone.

3.4.2 Psychophysical acuity for the temporal order of acoustic events

The psychophysical task used here probes a form of temporal pattern judgment that is sometimes necessary to discriminate natural sounds. Listeners discriminated natural from "flipped" twitters with near perfect accuracy when local time reversals extended over 80 ms, but performed near chance with time reversals of 20 ms or below. There are several other lines of evidence to suggest that the ordering of auditory events is limited to time frames of 20 ms or above. In humans, accurate temporal order judgments of clicks with varying amplitudes require inter-click intervals greater than 20 ms (Hirsh, 1959), and those of two pure tones presented simultaneously but with asynchronous onsets require stimulus onset asynchronies of approximately 20 ms (Efron, 1963; Pastore, 1983;

Stevens and Weaver, 2005). Interestingly, it has been suggested that the range of rates of change of stimulus events in spoken language, which is in the order of 2 - 50 Hz (Rosen, 1992), contains temporal modulation frequencies that the auditory system can detect. The limited temporal resolution of these perceptual phenomena may be attributable to limitations in the accuracy of temporal pattern codes in the underlying neural circuitry.

My analysis of temporal discharge patterns of ferret A1 units is consistent with this hypothesis. As time-reversed windows become narrower, inaccuracies in stimulus-locked firing resulted in a failure to represent fine differences in the temporal structure of stimuli. Other previous studies of the responses of cat, monkey and human auditory cortex have observed similar limitations in the temporal resolution of neural response patterns to represent discrete auditory events. For example, responses to stop-consonant syllables exhibit characteristic temporal response patterns when voice-onset times are longer than about 20 - 40 ms, but not for shorter voice-onset times (Eggermont, 1995; Sharma *et al.*, 2000; Steinschneider *et al.*, 2003, 2005). Here, the temporal discharge patterns of a subset of A1 units were found to discriminate natural from flipped twitters if the latter contained temporal manipulations that were at least 20 ms long.

In addition, discrimination based on the stimulus-synchronized firing of A1 units approximated human psychometric performance only when the responses were read-out in time bins that were 20 ms or finer. Schnupp *et al.* (2006) found similar optimal temporal resolutions in sampling A1 discharge patterns for twitter classification, in their case between 10 - 50 ms. Taken together, these results suggest that psychophysical thresholds for ordering sequential acoustic events may often lie near 20 ms because the temporal discharge patterns that auditory cortical neurons use to encode these stimuli are unable to represent finer time scales accurately.

3.4.3 Different brands of temporal processing

The limited temporal resolution studied here may not apply in higher-order cortical areas or in the context of other types of auditory tasks. For example, at the level of auditory cortex, pitch judgments may be performed through a specialized set of rate coding pitch neurons (Bendor and Wang, 2005). Nor might this temporal resolution apply to the perception of transients within a frequency channel, such as ramped and damped tones

(Lu *et al.*, 2001a) or within-channel gap detection (Phillips, 1999). Some physiological results suggest that auditory processes operating at time scales either below or above the lower pitch limit (i.e. *ca.* 30 - 50 Hz) may be processed by separate subgroups of A1 neurons. More rapid (i.e. 50 Hz) transients within an auditory perceptual channel (i.e. within a pure-tone frequency) are encoded in the spike rates of some A1 neurons (Lu *et al.*, 2001a,b), while slower modulations elicit stimulus-synchronized discharges in other A1 neurons (Lu *et al.*, 2001b). This is consistent with the present findings, where among a sample of A1 units that were driven by complex stimuli, only a subset exhibited temporal discharge patterning that supported discrimination of the stimuli. Therefore, while temporal pattern based neurometrics have been shown to perform better than spike rate neurometrics on a task of discriminating relatively slow (< 50 Hz) changes in the envelopes of complex stimuli, spike count based neurometrics might be expected to perform better on tasks that require the discrimination of very fast (> 50 Hz) transients. This idea does not require that these two types of encoding be carried out in separate neural populations, although the work of Lu *et al.* (2001b) suggests that this might be the case. The pitch discrimination performance supported by spike rates and spike patterns are explored in Chapter 6.

It should also be noted that a 10 - 20 ms limit in the temporal resolution at which sequential auditory events can elicit spikes in A1 neurons does not imply that the temporal precision of spikes in representing the onset of each of these auditory events is limited to 50 - 100 Hz. On the contrary, the temporal jitter in evoked spikes across repeated presentations of a stimulus can be considerably less than 10 ms for some A1 neurons (Elhilali *et al.*, 2004; Wehr and Zador, 2003).

3.4.4 Responses of populations of A1 neurons

I have shown that the temporal discharge patterns of some A1 units were sufficiently distinguishable that the output of the three best discriminating units was sufficient to match human psychophysical performance. Therefore, the activity of a small number of cortical units is in principle sufficient to account for psychophysical data, provided these units are pre-selected to be the most informative. However, it is far from clear whether perceptual decision making processes do indeed rely on only a small number of units,

and how the most appropriate neurons for the particular task at hand would be chosen if this were the case. Results such as those of Britten and colleagues (1992b) question the relevance of this "best envelope" approach to cortical processing. They show that many MT neurons have spike rates that outperform the behaving animal on a visual discrimination task, and psychometric performance in this case is closely approximated by the median of neurometric performance across units. Therefore, approaches that attempt to interpret the activity of pooled populations of simultaneously active neurons may be a preferable way to interpret cortical data.

Using a simple, linear pooling of simultaneously active neurons, I have shown that the responses of even independently uninformative neurons can provide a temporal activation pattern that supports behavioural measures of the discrimination of complex sounds. Further, the temporal spiking patterns of A1 neurons appear to be quite robust in that the neurometric performance of informative neurons was largely unimpaired when their responses were pooled with noisy units. It would certainly be possible to conceive of alternative and potentially more powerful pooling algorithms than the one used here, however the algorithm chosen is a biologically plausible one. A convergence of connections with simple synaptic summation would achieve this type of pooling of raw spike responses. Past neurometric studies have often used optimized weighting rules to pool neural responses in order to match psychometric findings. For instance, Purushothaman and Bradley (2005) offered a detailed examination of the effects of pool size, weighting of neural inputs in the pool, and pooling non-linearity on the neurometric performance of populations of neurons in the middle temporal region of visual cortex. Furukawa *et al.* (2000) found that the responses of at least 128 cat A2 neurons were needed to reduce the median errors in a sound localization task to psychoacoustic thresholds.

Although these analyses are interesting in their own right, the interpretation of such models can be problematic. For any one task, optimal pooling will require only a few cells and optimization of performance on a slightly different task will likely require a different pool. This returns to a problem with enveloping the responses of a few best neurons. That is, it is unclear how the brain might isolate the activity of these pools from task to task. I have shown that neurometrics based on the pooled activity of A1 neurons using a simple linear summation can closely approximate behavioural performance, provided

that the neural activity is recorded simultaneously.

3.4.5 Studying cortical processing in anaesthetised animals

One might expect response properties of auditory cortical neurons to differ in anaesthetised and unanaesthetised animals, but there are as yet no published studies which have carried out carefully controlled comparisons of anaesthetised and unanaesthetised recordings to demonstrate clear and fundamental differences in response properties. In the present study, I recorded responses of auditory cortex neurons to the same stimulus set in both anaesthetised and unanaesthetised ferrets. While data from these two preparations were not fundamentally different, there nevertheless appeared to be subtle differences. For example, there seemed to be a somewhat higher proportion of neurons with good neurometric performance in the unanaesthetised dataset (compare Figures 3.6 and 3.14). I also observed that the PSTHs of pooled firing across units showed stronger onset and offset response in the awake cortex (compare Figure 3.15).

However, caution must be exercised in comparing the results of these two types of experiments. The unanaesthetised data were recorded using different types of recording electrodes from those used in the anaesthetised preparation, since the silicon probes used could not penetrate ferret dura. One can therefore not be certain to what extent any quantitative differences in the two datasets may reflect differences in anaesthetic state or artifacts of the sampling and recording methodologies. This potential confound makes quantitative differences in the awake and anaesthetised datasets difficult to interpret, which is why I did not examine these differences here in greater detail. I focus instead on the clear and unambiguous similarities in the two datasets that appear despite methodological variations. In both the anaesthetised and the unanaesthetised data, the best envelope and pooled neurometrics matched human psychometric performance. Stimulus-locked temporal discharge patterns are present in both states, and these firing patterns are able to support the discrimination of complex sounds where a simple spike rate based analysis fails. The fact that these features are observed regardless of anaesthetic state suggests that the temporal discharge patterns described here are a fundamental aspect of complex sound encoding in the auditory cortex.

3.5 Acknowledgements

Dr. Bashir Ahmed supervised the surgical procedures during the anaesthetised electrophysiological experiments, and assisted with the subsequent recordings. Dr. Fernando Nodal developed the technique for head-fixed, awake recordings in our lab, Dr. Victoria Bajo and Dr. Fernando Nodal supervised the surgery and recovery of the animal for unanaesthetised electrophysiology, and Johannes Dahmen assisted in carrying out electrophysiological recordings in this animal. Dr. Xiaoqin Wang of Johns Hopkins University School of Medicine, Baltimore, MA provided the recordings of marmoset vocalizations. Dr. Jan Schnupp supervised this work, and Prof. Andrew King provided essential consultation and experimental facilities. This research has been published in the Journal of Cognitive Neuroscience (Walker *et al.*, 2008).

Chapter 4

Interdependent Encoding of Stimulus Features in Auditory Cortex

4.1 Introduction

Building on earlier studies in the visual system (Mishkin and Ungerleider, 1982), it has been proposed that a separation of function exists within human auditory cortex, such that more posterior and/or dorsal cortical areas facilitate sound localization, whereas more anterior and/or ventral areas are responsible for object identification (Alain *et al.*, 2001; Barrett and Hall, 2006; Maeder *et al.*, 2001; Warren and Griffiths, 2003). Regional differences in neural sensitivities to the identity or location of sounds have also been demonstrated in the auditory cortex of non-human primates (Kaas and Hackett, 2000; Rauschecker *et al.*, 1997; Romanski *et al.*, 1999; Tian *et al.*, 2001), and this has been interpreted as evidence of parallel cortical processing streams for determining the "what" (i.e. identity) and "where" (i.e. spatial location) of sounds - a "dual stream" hypothesis of auditory cortical function (Kaas and Hackett, 1999; Rauschecker and Tian, 2000). A recent study has demonstrated a double-dissociation of pattern recognition and localization function in the anterior and posterior areas of cat auditory cortex, respectively, using reversible cortical deactivation (Lomber and Malhotra, 2008). This result is

consistent with the dual stream hypothesis. However, other models of auditory cortical organizations have been offered to explain these experimental results, including a dorsal/ventral segregation for processing "what" the sound is and "how" it is produced (Belin and Zatorre, 2000; Griffiths, 2008). On the whole, it remains unclear whether the functional organization of auditory cortex includes separate what/where processing streams, or if some other division of function is implemented along these pathways (discussed in Sumner *et al.*, 2008).

One of the most important functions of the auditory system is to identify and discriminate vocal calls, such as speech sounds. This task requires the listener to process several complex perceptual properties of a single auditory object, and so it is likely to engage a number of functionally distinct cortical areas in parallel. To identify a spoken vowel, for example, the auditory system must determine the positions of formant peaks in the spectral envelope of the vowel sound (Peterson and Barney, 1952). Vowel discrimination is therefore a timbre discrimination task. Meanwhile, the pitch of the vowel may convey information about the speaker's identity (Gelfer and Mikos, 2005) and emotional state (Fuller *et al.*, 1992; Reissland *et al.*, 2003), and so the periodicity of the vowel must also be analyzed. Finally, localization of the speaker requires processing interaural difference cues and monaural spectral cues that are independent of the timbre and pitch of the vowel.

If separate processing streams for auditory localization and identification exist, one might expect the pitch and timbre of vowel sounds to be represented together, in a separate region from its spatial location. On the other hand, it is not uncommon for listeners to find themselves in cluttered acoustic environments, where the pitch, timbre and spatial location of several sound sources may have to be tracked simultaneously. A thorough separation of the different perceptual attributes of each sound might make this task more difficult, creating a sort of "binding problem". Previous cellular recording studies have focused almost exclusively on identifying differences across cortical areas in the representation of just one stimulus parameter, such as either the location (Stecker *et al.*, 2005a,b, 2003; Stecker and Middlebrooks, 2003) or pitch (Bendor and Wang, 2005) of sounds. To what extent these perceptual attributes are encoded independently within neurons has not previously been investigated, and so I address this questions in the

present study.

I measured cortical sensitivity to features of "artificial vowel" sounds, which were constructed by band-pass filtering click trains. This choice of stimulus allowed me to systematically and independently manipulate three perceptual attributes of the sound in a fully parametric fashion. The band-pass filters imposed peaks ("formants") in the energy spectrum of the stimuli, which determined the stimulus timbre. The click rate determined the pitch of the sound, and virtual acoustic space was used to introduce sound-source location cues. I investigated how pitch, timbre and sound source location are encoded within five areas of the ferret auditory cortex. While the present stimulus set was based on sounds that model human speech, ferret vocalizations have similar periodicities and spectral profiles (see Section 1.2), and these animals have been demonstrated to perceive changes in each of the three perceptual dimensions explored here. My aims were to determine whether, or to what extent, these perceptual attributes are represented in a mutually independent fashion in the two primary and three secondary cortical fields that cover the ferret ectosylvian gyrus, and to look for evidence of feature specialization across these five fields.

4.2 Methods

4.2.1 Stimuli

Sounds were generated using TDT System 3 hardware and MATLAB software, and were presented through customized earphones, as described in Section 2.4. Noise bursts (100 ms in duration) were presented at 70-80 dB as search stimuli with which I could identify acoustically-responsive neural activity. Pure tone stimuli were used to obtain Frequency Response Areas (FRAs), both to characterize individual units and to classify their characteristic frequencies (CFs; Section 2.5).

Artificial vowel stimuli were created in MATLAB, using an algorithm adapted from Malcolm Slaney's Auditory Toolbox[1]. Click trains were first produced, with a repetition rate corresponding to the desired fundamental frequency and a duration of 150 ms. These were then passed through a cascade of four bandpass filters to impart spectral peaks at

[1] http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/

the desired formant frequencies, thereby modelling vowel sounds. These artificial vowel sounds were root-mean-square normalised to ensure that changes in pitch or timbre did not influence the overall sound pressure level. Virtual Acoustic Space (VAS) techniques were then used to add sound source direction cues to the artificial vowel sounds. Measurements including head size, sex, body weight and pinna size, were taken from each ferret in order to select the best match from an extensive library of ferret head-related transfer function recordings. Members of our laboratory have shown previously that localization cues scale with the size of the head and external ears (Schnupp *et al.*, 2003). Sound source direction cues were generated by convolving the artificial vowel sounds with minimum phase filters which imparted the appropriate interaural level differences and spectral cues, and which equalized out any differences in the headphone transfer functions that had been revealed during headphone calibration. Small delays were then introduced in the sound waveforms to generate appropriate interaural time differences.

The sounds were presented in VAS at -45°, -15°, 15° and 45° azimuth, at 0° elevation. Negative azimuths denote locations to the animal's right, contralateral to the recording sites. Fundamental frequencies of 200, 336, 565 and 951 Hz were used. The four timbres presented corresponded to the vowels: /ɑ/ with formant frequencies F1-F4 at 936, 1551, 2815 and 4290 Hz; /ɛ/ with formant frequencies at 730, 2058, 2979 and 4294 Hz; /u/ with formant frequencies at 460, 1105, 2735 and 4115 Hz; and /i/ with formant frequencies at 437, 2761, 3372 and 4352 Hz. The combinations of these four pitches, four timbres and four sound source directions resulted in a stimulus set of 64 artificial vowels.

These parameter ranges were chosen to make the stimuli easily discriminable along each perceptual dimension, both for human listeners (based on informal observations), and, as far as I can tell from the available psychoacoustical data, also for my animal model species, the ferret. The separation of my stimuli in both azimuth (Parsons *et al.*, 1999) and pitch (Chapter 5) is large and of similar magnitude compared to behaviourally measured psychoacoustic difference limens. Furthermore, pilot experiments in our lab suggest that ferrets can rapidly learn to discriminate between /ɑ/ and /i/ versions of these artificial vowels (unpublished observations).

Figure 4.1 shows the frequency spectra of artificial vowel sounds with each of the 16 combinations of pitch and timbre used in this experiment. This figure illustrates that

both timbre and pitch changes affect the spectral envelope of the sounds, and presenting these sounds in VAS can introduce further changes in the spectral envelope. Yet while changes in location, pitch and timbre all affect the sound spectrum, the perceptual consequences of these changes are quite distinct. For example, the pitch percept evoked, in human listeners, by the stimuli within any one row of Figure 4.1 is remarkably constant, despite clear differences in the spectral envelopes of these sounds. If the perceptual distinction between pitch, timbre and location is reflected at the level of neuronal discharges in auditory cortex, then this stimulus set ought to reveal this separation according to perceptual categories.

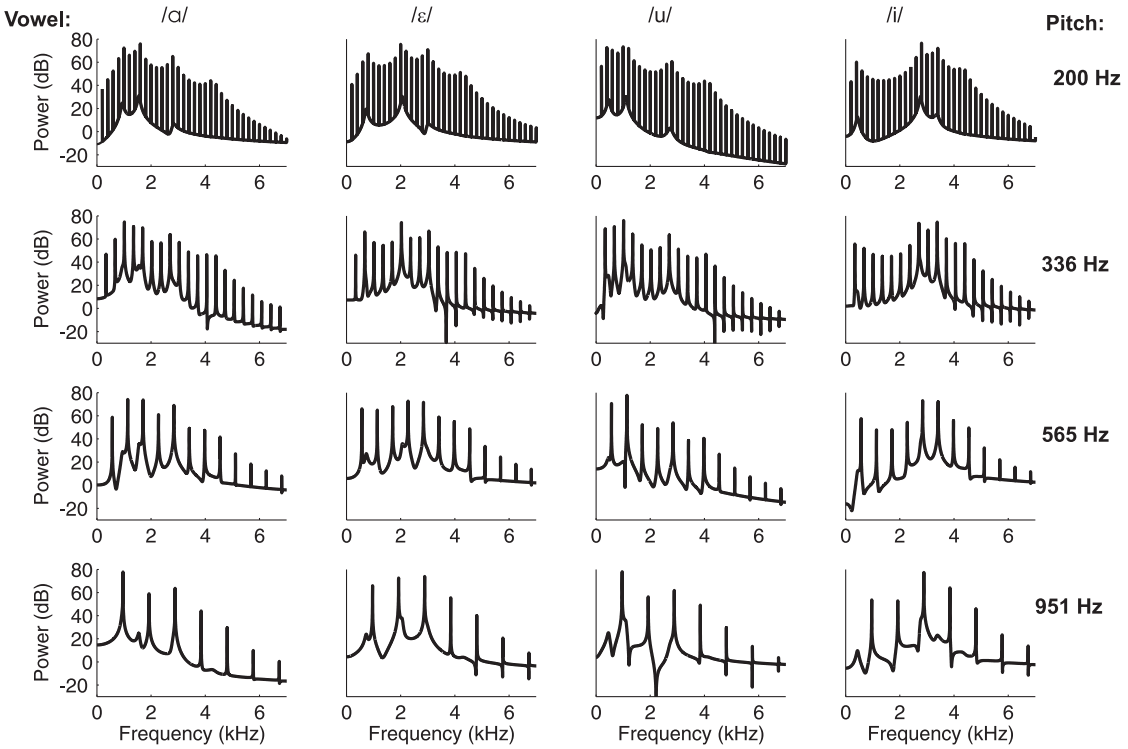


Figure 4.1: Frequency spectra of 16 artificial vowels used in this experiment. The four timbres (corresponding to the vowels /ɑ/, /ɛ/, /u/ and /i/) are shown in different columns, with the four pitches (F0 of 200, 336, 565 and 951 Hz) shown in different rows. Each of these stimuli was presented at four different locations in VAS (-45°, -15°, 15° and 45° azimuth).

4.2.2 Electrophysiological recordings

Extracellular recordings of the responses of neural units in the auditory cortices of five adult, female, pigmented ferrets were carried out under general anaesthesia, according to the procedures outlined in Section 2.2. Silicon electrodes were inserted into five cortical areas to record extracellular potentials. These areas included the primary auditory cortex (A1), the anterior auditory field (AAF), the posterior pseudosylvian and suprasylvian fields (PPF and PSF), and the anterior dorsal field (ADF) (recall Figure 1.3). In two animals, electrodes with an 8 x 4 configuration (eight active sites on four parallel probes, with a vertical spacing of 150 μm) were used. In a small number of recordings in one of these animals, and in another experimental animal, electrodes with a 16 x 2 configuration (16 active sites, spaced at 100 μm intervals on each of two probes) were used. In the final two animals, electrodes with 4 x 4 and 16 x 1 configurations were used (100-150 μm spacing of active sites on each probe). The electrodes were positioned so that they entered the cortex approximately orthogonal to the surface of the ectosylvian gyrus. A photographic record was made of each electrode penetration to allow later reconstruction of the location of each recording site relative to anatomical landmarks (e.g. surface blood vessels and sulcal patterns).

The responses of single neurons and multi-unit clusters of neurons were isolated (as per Section 2.6), and all further data analyses were carried out in MATLAB. Ketamine anaesthesia can often result in spontaneous bursts of spikes, and I did not want these artifacts to affect the results. Therefore, any presentation sweeps on which the spike count (calculated over the 300 ms post-stimulus-onset) of a neural unit was more than six times the units' average spike count were excluded from the analyses.

4.2.3 Constructing a composite map of auditory cortex

Before examining how neural units' responses to vowel stimuli varied as a function of their cortical location, each penetration was first assigned to a cortical field on the basis of the responses of neurons within that penetration to simple stimuli (Section 2.5). Penetrations were assigned to a cortical field using a combination of the CF and tuning characteristics obtained from the FRAs, as well as neural response latencies and response durations (Bizley *et al.*, 2005), photographs of the site of each penetration on the cortical surface,

and the overall frequency map obtained for each animal.

Having established the recording locations within each animal, I noted that there were consistent trends in the responses to vowel sounds within cortical fields across animals. I then created a "composite" map of auditory cortex, on which I could import data from all animals. Field boundaries were determined using physiological criteria, in particular on the basis of responses to pure tone and noise burst stimuli, but blind to the responses to vowel sounds. This approach has previously been used to investigate the representation of multisensory responses in ferret auditory cortex (Bizley and King, 2008). The penetration locations and cortical field boundaries for each individual animal were projected onto a single animal's frequency map derived using optical imaging of intrinsic signals (Figure 1.3). This frequency map was a representative example taken from Nelken *et al.* (2004). The projections were performed separately for each animal, so that the data across animals were integrated in a bias-free fashion.

4.3 Results

4.3.1 Preliminary inspections of responses to stimulus pitch, timbre and location

Extracellular recordings of neural responses to artificial vowel sounds were made in the auditory cortex of ferrets. The vowel sounds varied across four azimuthal locations, four pitches and four timbres (Section 4.2.1). Over 900 acoustically-responsive neural units were sampled, 615 of which were "driven" by the vowel stimuli and were sufficiently stable to record throughout 30-40 repetitions of each of the 64 pseudorandomly presented vowels. Driven neural units were those in which the average spike count during stimulus presentation (7-200 ms period following vowel onset) differed from the average spike count during silent intervals (657-850 ms period following vowel offset; t-test, $alpha = 0.05$). All further analyses were carried out on these 615 neural units, which included 324 single neurons and 292 small clusters of neurons. Because I was unable to find any systematic difference in the response properties of single neurons and small neural clusters, the term "neural unit" is used to refer to both types of recordings.

I observed a rich variety of response types across units, many of which were clearly

modulated by more than one, and often all three, dimensions of the vowel stimuli. Figures 4.2 and 4.3 show the responses of two different neural units. In each case, the three panels show the same data plotted three times, with the 64 stimuli ordered according to azimuth, pitch or timbre (first, second and third panel, respectively). These examples illustrate two common findings. First, tuning to stimulus pitch, timbre or azimuth alone did not adequately describe the responses of these units, and second, their responses could not be captured satisfactorily with a single spike count value. Rather, neurons typically showed a degree of sensitivity to each parameter (Figure 4.2), and/or to particular combinations of parameters (Figure 4.3), in a time-dependent manner. Further examples are illustrated in Appendix B.

In order to examine these effects, I constructed arrays of Post-Stimulus Time Histogram (PSTH), which were sorted according to two of the three stimulus parameters and data were pooled across the third. Figures 4.4A and C show such PSTH arrays for the neural unit illustrated in Figure 4.2. The 16 panels to the top left of Figure 4.4A show the data arranged according to all 16 timbre and pitch combinations. The first four columns show the PSTHs of responses to timbres corresponding to /i/, /u/, /ɛ/ and /ɑ/, while the top four rows show PSTHs for pitches at 200, 336, 565 and 951 Hz, respectively. The rightmost column in Figure 4.4A shows the mean response to each pitch, averaged across all timbres and azimuths, while the bottom row shows the mean response to each timbre, and the bottom right PSTH illustrates the grand average of responses to all 64 vowel types.

Displaying the data in this manner makes it easier to appreciate the effect of varying either stimulus pitch or timbre on a neural response. To describe these effects, I adopt the terminology used in ANOVA-type linear statistical models, where the spike rate during some small interval in time is the observed variable, while pitch, timbre, location and post-stimulus time serve as explanatory variables. I treat these as categorical variables as one cannot expect the relationship between stimulus parameter value and spike rate to be linear or even monotonic (especially in the case of timbre). Within that conceptual framework, comparing the top four panels of the rightmost column with the bottom right panel therefore reveals the "main effect" of varying pitch on the discharge pattern on this unit. To make it easier to visualize the main effects of each stimulus parameter,

I plotted the individual PSTHs on top of a color scale that shows the difference between that particular PSTH and the grand average at each time bin. Red colours indicate that the neural firing rate is, at that time point, larger than the grand average, blue indicates that it is below the average, and the saturation of the color scales with the magnitude of the difference.

The grand average PSTH in the bottom right panel in Figure 4.4A shows that this unit responded to artificial vowel sounds with an initial increase in firing rate, which peaked at a rate of 50 Hz at around 50 ms after stimulus onset, followed by a smaller second peak at *ca.* 180 ms. The main effect of presenting a relatively low pitch (200 Hz, top panel of the last column) was to decrease the size of the first peak in the PSTH and to increase that of the second. Conversely, the main effect of higher pitches (951 Hz, fourth cell in the last column) was to increase the size of the first response peak and to decrease the second. Similarly, the main effect of varying timbre can be appreciated by comparing the panels in the bottom row of the PSTH matrix in Figure 4.4A. A timbre corresponding to the vowel /u/ strongly enhanced the initial response peak, while the vowel /i/ suppressed it. This neural unit was therefore sensitive to both pitch and timbre, and the effects of changing pitch and timbre can be expressed at different latencies after stimulus onset.

In the conceptual framework of an ANOVA style analysis, the simplest prediction for responses to a particular pitch and timbre combination would be that the main effects of pitch and timbre are additive. To identify non-linear interactions between two stimulus parameters, the PSTHs in the top-left 16 cells of the array were compared against the hypothetical PSTHs that would be predicted from the linear sum of the main effects of pitch and timbre. The differences between these predictions and the observed responses (i.e. two-way interactions throughout time) are shown by the color scales in these 16 cells. For the unit in Figure 4.4A, these interaction effects were relatively small, not exceeding ±33 spikes/s compared to ±86 spikes/s for the largest main effect.

Figure 4.4C shows the azimuth-by-pitch main effects and interactions for the same unit, while Figures 4.4B and D show the pitch-by-timbre and pitch-by-azimuth main effects and interactions, respectively, for a second sample unit (the same as that illustrated in Figure 4.3). Although this second neural unit exhibited rather different temporal dis-

charge patterns, like the first unit it was clearly influenced by more than one stimulus dimension. As we shall see, the effects shown in Figure 4.4 were fairly typical of many of the units recorded throughout all cortical areas characterized. Thus, most units were sensitive to more than one stimulus dimension, their response patterns could change at various times post-stimulus onset, and interactions between stimulus features were not uncommon.

To quantify the magnitude and significance of these main effects and interactions, I performed 4-way ANOVAs on the average spike counts for each neural unit. The spike counts in this analysis were averaged across the 30-40 repeated presentations of each of the 64 stimuli, in 20-ms-long discretized time bins throughout the first 300 ms following stimulus onset. In this manner, each response was represented as an time-ordered array of 15 average spike counts. The choice of 20 ms bin widths was based on previous studies of ferret auditory cortex, which indicate that this is likely to be a suitable temporal resolution for decoding neural responses (Schnupp *et al.*, 2006, and Chapter 3). In this ANOVA, the three stimulus parameters (azimuth, pitch and timbre) and time post-stimulus onset served as factors (i.e. predictors of the variance in spike counts). To quantify the relative strength with which one of the three stimulus dimensions influenced the firing of a particular unit, I calculated the proportion of spiking variance explained by each of azimuth, pitch and timbre as:

$$V_p = \frac{SS_{p\cdot t} - SS_{Error} df_{p\cdot t}}{SS_{Total} - SS_t}, \tag{4.1}$$

where V_p is the proportion of variance explained by the stimulus parameter p of interest (pitch, timbre or azimuth), $SS_{p\cdot t}$ is the Sum of Squares for the interaction of the stimulus parameter and time bin t, SS_{Error} is the Sum of Squares for the error term, $df_{p\cdot t}$ is the degrees of freedom for the stimulus parameter and time bin interaction, SS_{Total} is the total Sum of Squares, and SS_t is the Sum of Squares for the time bin factor. The SS_t term reflects the average change in spike counts over the duration of the 300-ms response. For instance, a neuron that is driven by the vowel sounds might show higher firing rates, averaged across all stimulus conditions, in time bins that occur just following the onset of the stimulus (i.e. a classic onset response). A significant effect of time on spike rate, indicating that the PSTH was not flat, is an unsurprising finding in itself. But by examining the interactions of a given stimulus parameter with time, I was able

to determine the influence of the stimulus parameter not simply on the overall spike rate of the neuron, but on the temporal pattern of the neural response.

Significant interactions between time and stimulus parameters were common in these ANOVAs, and they were useful in revealing how a particular stimulus parameter influenced the shape of the PSTH. Subtracting $SS_{error}df_{p\cdot t}$ from the $SS_{p\cdot t}$ term allowed me to calculate the proportion of response variance attributable to each of the stimuli, while controlling for the additional variance explained simply due to adding that parameter to the regression model. To assess the significance of the effects of stimulus parameters, an alpha of 0.05 was Bonferroni-corrected for multiple comparisons across 15 time bins. This value ($0.00\overline{3}$) was conservatively approximated with an alpha of 0.001.

For the responses shown in Figure 4.4A and C, the percentage of variance in PSTH shape explained by the "main effects" of stimulus parameters was 5% for azimuth, 17% for timbre and 56% for pitch, while for the unit shown in Figure 4.4B and D, the main effects of azimuth, timbre and pitch accounted for 3%, 19% and 7%, respectively, of the variance in spiking patterns. Thus, while the responses of both units were significantly influenced by all three stimulus dimensions, one might describe the first as being predominantly sensitive to pitch, and the second to timbre. Only 23% of units were significantly ($p < 0.001$) modulated by either azimuth, pitch or timbre alone. In contrast, 36% of neural responses were dependent on two of the three stimulus dimensions, and 29% of units were influenced by all three.

I treated data from single units and from small multi-unit clusters similarly in my analyses. Therefore, in order to verify that the joint sensitivity observed was not merely a result of recording from more than one neuron, I calculated the proportion of single neurons sensitive to combinations of two parameters. Sensitivity to both pitch and timbre was observed in 56% of all recording sites and in 52% (169/324) of single units. Thirty per cent of all recordings and 33% of single units were sensitive to both pitch and azimuth, while 34% of all recordings and 36% of single units were sensitive to timbre and azimuth. Thus, combination sensitivity was observed with equal likelihood in both multi-unit clusters and well separated single neurons.

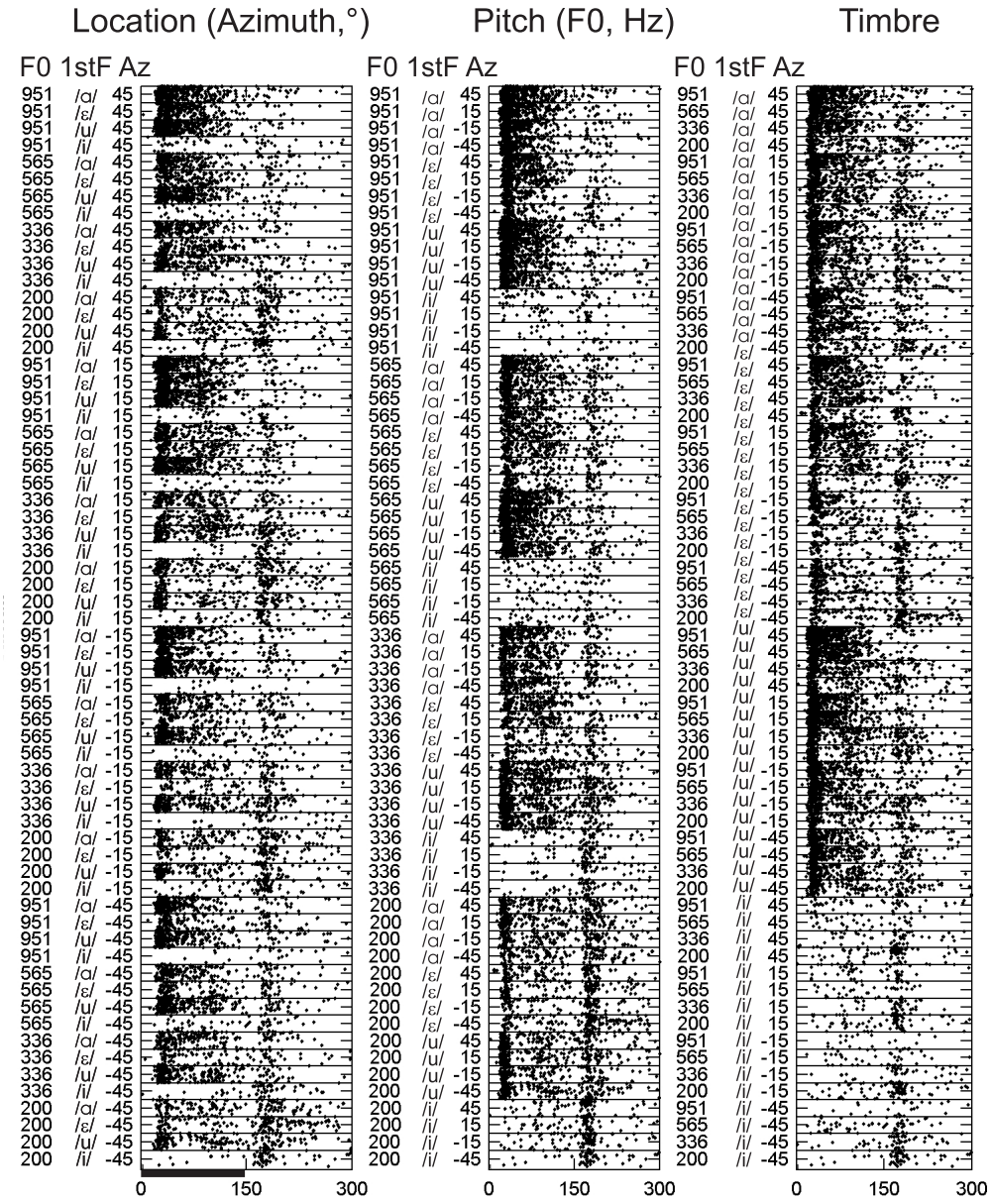


Figure 4.2: Raster plots of responses to artificial vowels: unit A. Action potentials (black dots) fired by a cortical neural unit in response to 64 artificial vowel stimuli. The complete dataset are re-plotted three times, organized either according to stimulus azimuth (Az, in degrees), pitch (F0, in Hz), or timbre (arranged by first formant frequency). Negative azimuth values indicate sounds designed to be lateralized to the ipsilateral (left) side of the animal. The neural unit produced responses that were clearly dependent on at least two of the three stimulus dimensions.

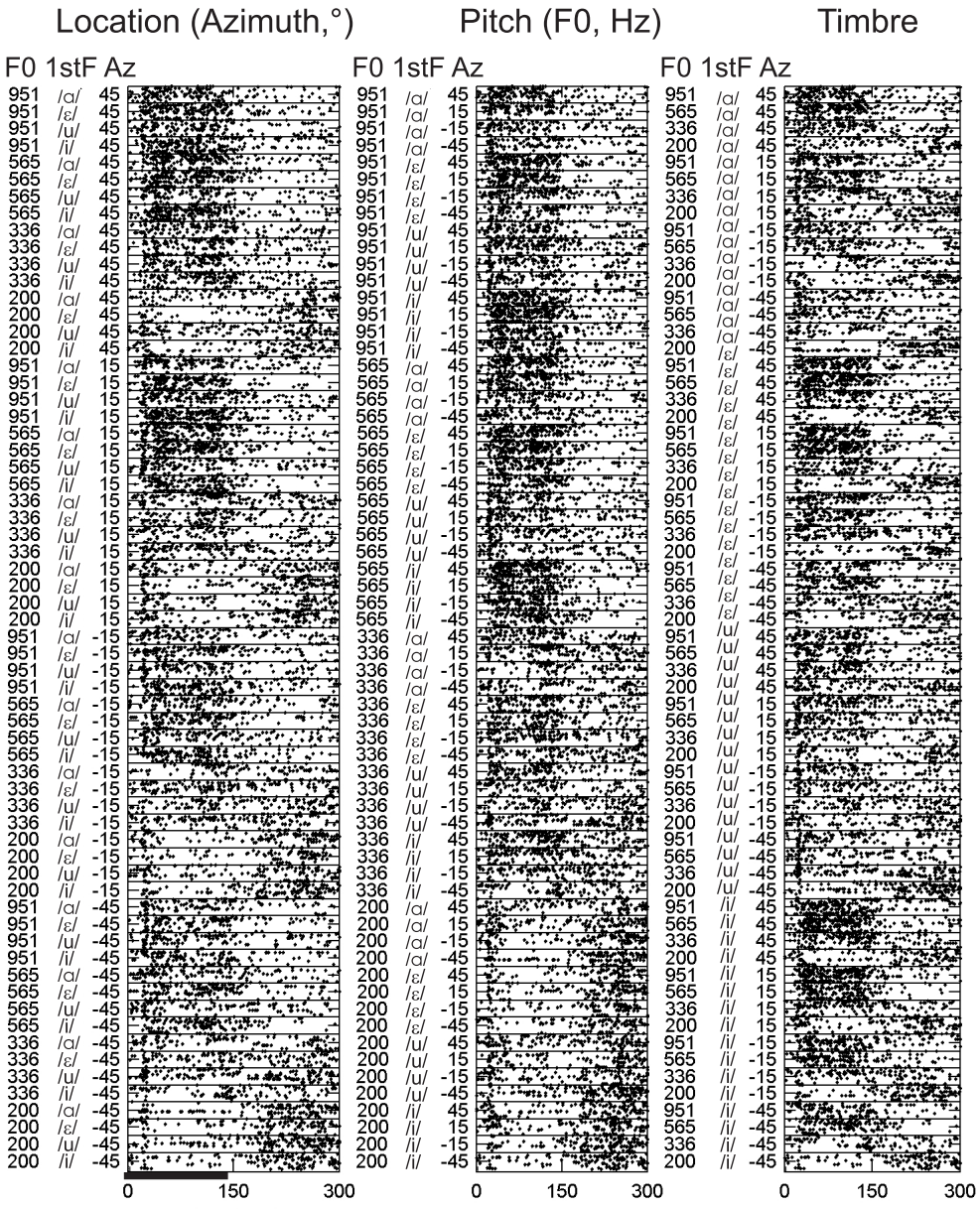


Figure 4.3: Raster plots of responses to artificial vowels: unit B. Responses of a second neural unit to the 64 artificial vowel stimuli, shown as in Figure 4.2. This neural unit also produced responses that were clearly dependent on at least two of the three stimulus dimensions.

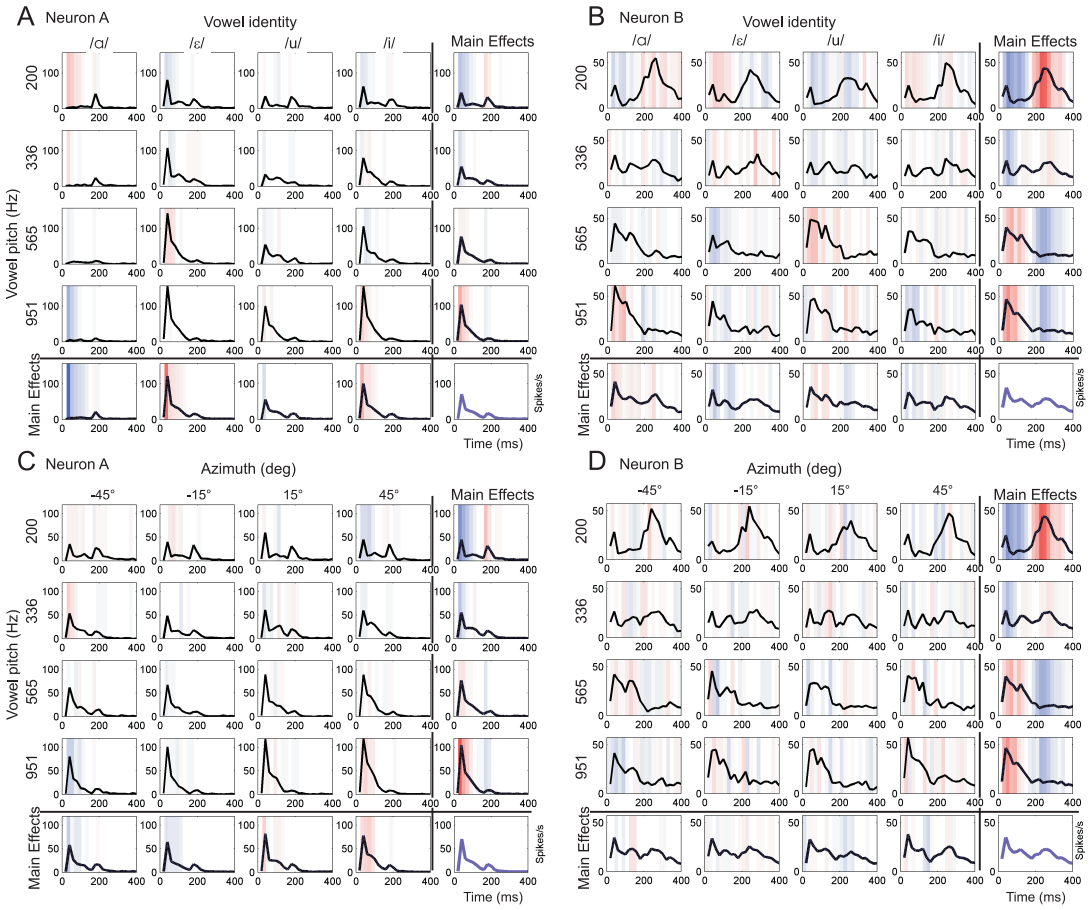


Figure 4.4: PSTHs illustrating the main effects and 2-way interactions of timbre, pitch and azimuth on the responses of two cortical neurons. (A) Black lines trace the mean firing rate (Hz) of the neural unit described in Figure 4.2. The first four rows and columns of the array of PSTHs show the responses to 16 combinations of four timbres and four pitches, averaged across all azimuth conditions. The cells to the right of each row and at the bottom of each column show "main effects" - the average response to each pitch (across rows) and each timbre (across columns). The bottom right PSTH shows the grand average response (purple line), across all 64 stimuli. The color scale underlay in each main effect PSTH represents the difference between that stimulus condition and the grand average PSTH, with blue indicating a relative decrease in firing rate, and red showing a relative increase. The color scale underlay in the 16 "interaction" PSTHs show non-linear combination sensitivity. (C) Responses of the neural unit shown in A, across azimuth and pitch conditions. (B & D) Similar PSTH arrays for the neural unit described in Figure 4.3.

4.3.2 Spike count measures do not fully capture response complexity

To examine the importance of the temporal pattern in the neural response, I also performed an ANOVA that was restricted to the overall spike counts calculated over the first 75 ms after stimulus onset. This ANOVA was performed on these spike counts using pitch, azimuth and timbre as predictor variables. This analysis resulted in far fewer units exhibiting significant response modulation as a function of the stimulus attributes. For example, using a single spike count measure, only 11% of all units exhibited significant ($p < 0.05$) sensitivity to sound azimuth, although 39% of units had shown a significant ($p < 0.001$) time bin and azimuth interaction. The same comparison for pitch and timbre when total spike count was used in the ANOVA, 18% of units were significantly sensitive to stimulus pitch and 29% to timbre ($p < 0.05$), while these values were 66% and 75% for pitch and timbre in the binned ANOVA ($p < 0.001$).

Over the single time window of 75 ms, the discharge rates of 60% of units no longer exhibited any significant stimulus main effects, and only 13% of units were modulated by more than one parameter. Analyzed in this manner, the unit shown in Figure 4.2, for example, was found to be sensitive only to timbre, while the unit depicted in Figure 4.3 was no longer sensitive to any of the three stimulus dimensions. When I performed the analysis using a longer response window (300 ms, data not shown) there were even fewer neurons whose responses were significantly modulated by these stimuli, and neither of the units shown in Figures 4.2 and 4.3 was found to be sensitive to pitch, azimuth or timbre.

These results demonstrate that highly significant but transient stimulus effects can be easily missed in an analysis that is temporally too coarse-grained. I therefore adopted the time-binned analysis described in Equation 4.1 to investigate neural sensitivity in the analyses that follow.

4.3.3 Cortical distribution of sensitivity to pitch, timbre and azimuth

To examine the distribution of stimulus sensitivity across auditory cortex, recordings were made in five previously identified auditory cortical areas on the ferret ectosylvian gyrus: A1 and AAF on the tip of the gyrus, the tonotopically organized PPF and PSF located on the posterior bank of the gyrus, and the non-tonotopic ADF on the anterior

bank (Bizley *et al.*, 2005; Kowalski *et al.*, 1995; Nelken *et al.*, 2004). These five areas, shown in Figure 1.3, make up the auditory core (A1 and AAF) and belt (PPF, PSF, ADF) regions in this species. Previous studies from our lab have reported that about 60% of neurons in the anterior ventral field (AVF) are visually sensitive (Bizley *et al.*, 2007a). This field and the ventroposterior (VP) area are likely to be para-belt areas and were not included in the present study. In four out of five animals studied, recordings were made in all five cortical fields, and in the remaining animal responses were recorded in three fields.

To investigate the anatomical distribution of the stimulus feature sensitivities described above, I generated a composite map of auditory cortex (described in Section 4.2.3). Because the electrodes were inserted into the brain approximately orthogonal to the auditory cortical surface, I assumed that all neurons recorded simultaneously on a single electrode probe were located at a single position on the rostrocaudal-dorsoventral plane. Pure-tone FRAs were recorded at all recording sites, to confirm the assignment of recording sites to cortical fields using documented physiological criteria (Bizley *et al.*, 2005). Locations of penetration sites recorded in each animal, colour-coded according to average CF, are mapped in Figure 4.5A-E. Figure 4.5F shows the projection of penetration sites from one animal (Figure 4.5E) onto the composite auditory cortical map. The template image for this map (shown alone in Figure 4.6A) was a representative example of the tonotopic arrangement of ferret auditory cortex derived in a previous optical imaging study (Nelken *et al.*, 2004).

Multi-site electrode arrays were used to measure neural responses, so I commonly recorded from several neural units at a single rostrocaudal-dorsoventral co-ordinate, at different depths below the cortical surface. To visualize the range of CFs observed in these cases, I made Voronoi tessellation maps of auditory cortex, with one Voronoi tile for each unit recorded (as used in Bizley *et al.*, 2005)[2]. Tiles from a common electrode penetration were arranged in a circular fashion around the penetration site, with the most shallow recorded unit shown rightmost and proceeding counter-clockwise from superficial

[2]Given a set of N penetration points on the cortical surface (which is assumed to be a plane), the Voronoi tessellation divides the surface into a set of polygonal regions, the boundaries of which are the perpendicular bisectors of the lines joining the points. The result being that each penetration is represented by one polygon, or "tile".

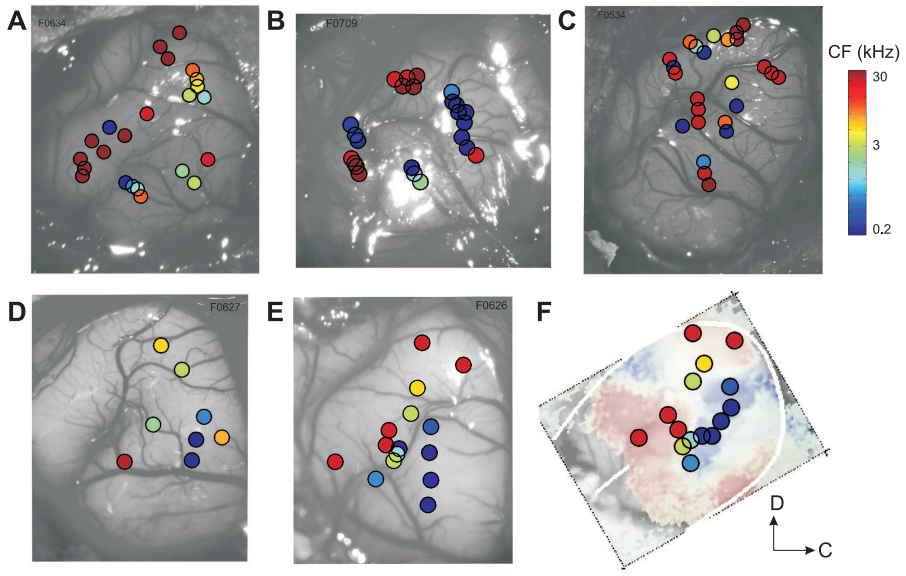


Figure 4.5: Maps of CFs measured across the auditory cortices of five animals. (A - E) Greyscale pictures of the auditory cortices of five animals. In each panel, the locations of recording sites within each cortex are shown as circles superimposed on the cortical surface, with the average CF of units recorded in each recording site (i.e. electrode penetration) indicated by the colour of the circles (colour bar, right). (F) Results of projecting the penetrations from the animal in panel E onto the frequency tuning map of ferret cortex from Nelken *et al.* (2004).

to deep. The composite frequency map obtained in this manner is shown in Figure 4.6B. The low frequency regions that mark the boundaries of fields AAF with ADF, of A1 with PPF and PSF, and of PPF with PSF, are readily apparent.

Figure 4.7A-C maps the proportion of spike pattern variance that is explained (Eqtn 4.1) by the "main effects" of azimuth, pitch and timbre, respectively, onto the surface of the auditory cortex. Each tile in these Voronoi tessellations shows the average V_p value obtained for all units recorded during a single electrode penetration at the indicated recording site. The colour scale shows the proportion of variance explained, with darker red colours indicating small values, and lighter yellow colours indicating larger values. The color scales saturate at values less than the maximum V_p in order to better display the entire distribution of results across cortical sites. These plots suggest that there are areas in which clusters of units have a greater sensitivity to particular stimulus parameters. Units that were highly modulated by the azimuth of sound sources were

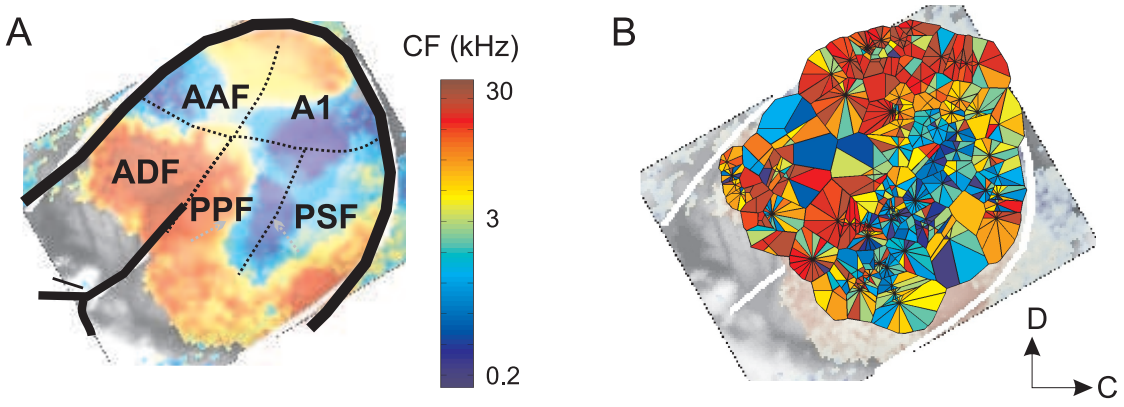


Figure 4.6: Map of the CFs of units across auditory cortex. (A) Frequency map of auditory cortex in the ferret, based on optical imaging of intrinsic signals. Frequency preferences are indicated by the colour scale (colour bar). As a reminder, the five auditory cortical areas from which recordings were made are labelled and field boundaries are drawn (dotted lines). (B) Voronoi tessellations show the CFs of all units from which I measured FRAs ($n = 811$). These data were collected from five animals and are projected onto the map of auditory cortex shown in A. Each tile of the tessellation represents one neural unit recorded in that area, with the colour indicating the CF of the unit (colour scale as in A). Tiles representing units from a single electrode penetration are arranged counter-clockwise according to depth, with the most shallow unit at the far right.

particularly common in A1, a dorsal rostral section of AAF, and the area encircling the tip of the pseudosylvian sulcus. Units that were highly sensitive to vowel pitch were most commonly found in the middle of auditory cortex, around the point at which the low frequency edges of the tonotopic core and belt areas converge. Finally, timbre sensitivity was highest in the core fields and along the low frequency ridge that separates the two posterior fields, PPF and PSF.

To illustrate the distribution of the percentage of response variance explained by these stimulus parameters across individual neural units, Figure 4.7D-F plots V_p values (again, as a tessellated colour map) for each unit. In these maps, each tile represents a recorded unit, and multiple units from a single penetration are arranged in a circular fashion, as in Figure 4.6, around the site of the penetration. Figure 4.8A-C shows neural units for which each of the stimulus parameters had a statistically significant main effect on temporal spiking patterns ($p < 0.001$; significantly modulated units are shaded black).

Overall, Figures 4.7 and 4.8 suggest that there are local clusters of neurons that are

relatively more sensitive to stimulus azimuth, pitch or timbre. However, these clusters of sensitivity are not restricted to particular cortical fields, and within each of the five fields I observed considerable unit-to-unit variability in the sensitivity to pitch, timbre and location.

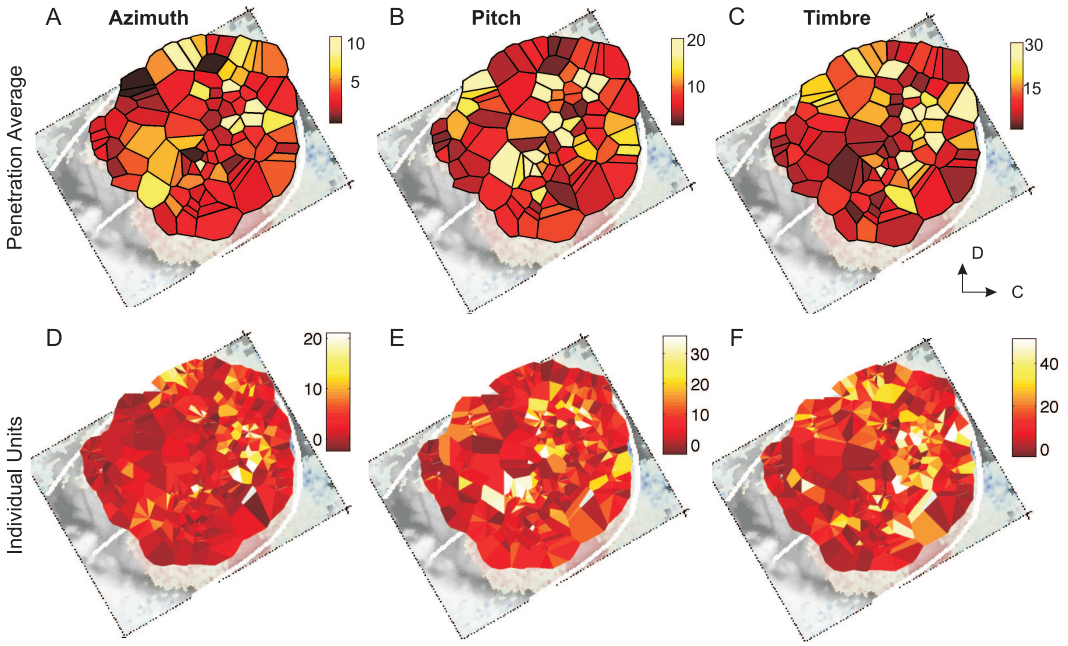


Figure 4.7: Maps of azimuth, pitch and timbre sensitivity across auditory cortex. Voronoi tessellation maps show the proportion of spike pattern variance that is explained by each of the three stimulus attributes: azimuth (A), pitch (B) and timbre (C). Each tile shows the average V_p value obtained across all neural units recorded in a penetration centred at that location (see colour bars). Panels D-F are as A-C, but here the sensitivities of individual units are plotted ($n = 615$), with units from a single penetration arranged counter-clockwise by increasing depth around the penetration site (with the most shallow unit at the far right).

The distribution of feature sensitivity for each cortical area was also visualized in box plot format (Figure 4.9A-C). Despite the wide variation in stimulus feature sensitivity across units, some degree of specialization within cortical areas was nevertheless apparent. All three stimulus attributes showed significant differences across cortical fields in the proportion of neural response variance they explained (Kruskal-Wallis test, $\chi^2 = 49$, 78, and 67 for azimuth, pitch and timbre, respectively; $p < 0.0001$). Tukey-Kramer post-hoc comparisons ($p < 0.01$) revealed that azimuth sensitivity was, on average, significantly

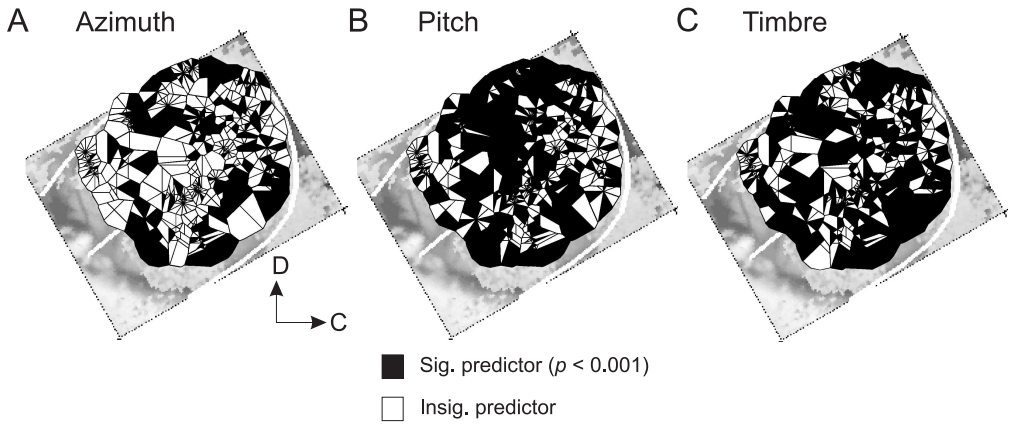


Figure 4.8: Maps of significant azimuth, pitch and timbre sensitivity across auditory cortex. Neural units with spiking responses that were significantly modulated by the azimuth (A), pitch (B), and timbre (C) of vowel sounds are shown in black, while all other neural units are shaded white ($n = 615$). Units with significantly modulated responses are those for which the interaction of time and the given stimulus parameter were statistically significant in the 4-way ANOVA ($p < 0.001$).

higher in A1 than in AAF, PSF and ADF (but not PPF), pitch sensitivity tended to be lower in ADF than in all other fields, and timbre sensitivity was higher in A1 and AAF than in the belt regions.

A jackknife test was used to determine if the above trends were seen consistently in each of the animals, and not just in the pooled data. The same statistical tests (i.e. Kruskal-Wallis and Tukey Kramer post-hoc tests) were implemented five times, excluding one of the five animals from the dataset on each occasion. In each of these analyses, the trends in the proportion of spiking variance explained by stimulus parameters across cortical areas were well preserved. Naturally, the decrease in the sample sizes of these subsets of data meant that some significant effects observed in the complete dataset were not observed in some of the jackknife analyses. However, in no case was there a significant difference observed in a data subset which was not present in the analysis of all animals, so there was no evidence of effects arising in some groups of animals that were masked by the inclusion of an additional animal. Furthermore, all of the significant differences observed with the complete dataset were present in at least three out of the five jackknife analyses, so the effects observed in the complete dataset were not attributable to results

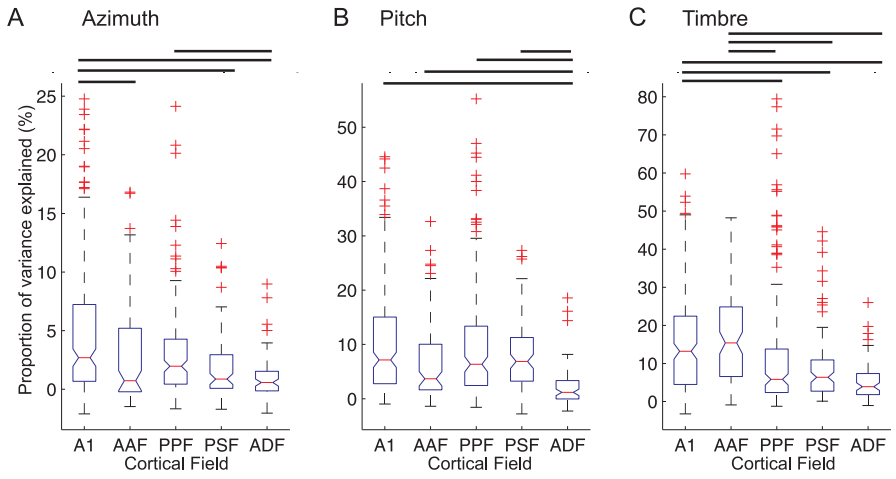


Figure 4.9: Stimulus-related variance in response patterns across cortical areas. (A-C) Box plots show the proportion of neural firing variance attributable to changes in the azimuth (A), pitch (B) and timbre (C) of vowels, across five cortical areas. The boxes show the upper and lower quartile values, and the horizontal line at their notch indicates the median. The whiskers extend to 1.5 times the interquartile range of the data, and the crosses indicate outlier values. The horizontal lines above the plots indicate pairwise significant differences between cortical fields ($p < 0.01$, Tukey-Kramer post-hoc comparisons).

from any one animal. As a final check, I visually inspected individual animal cortical maps of V_p values (Figure 4.10). From these maps, the same general trends as seen in composite auditory cortex (Figure 4.7) are apparent upon close inspection.

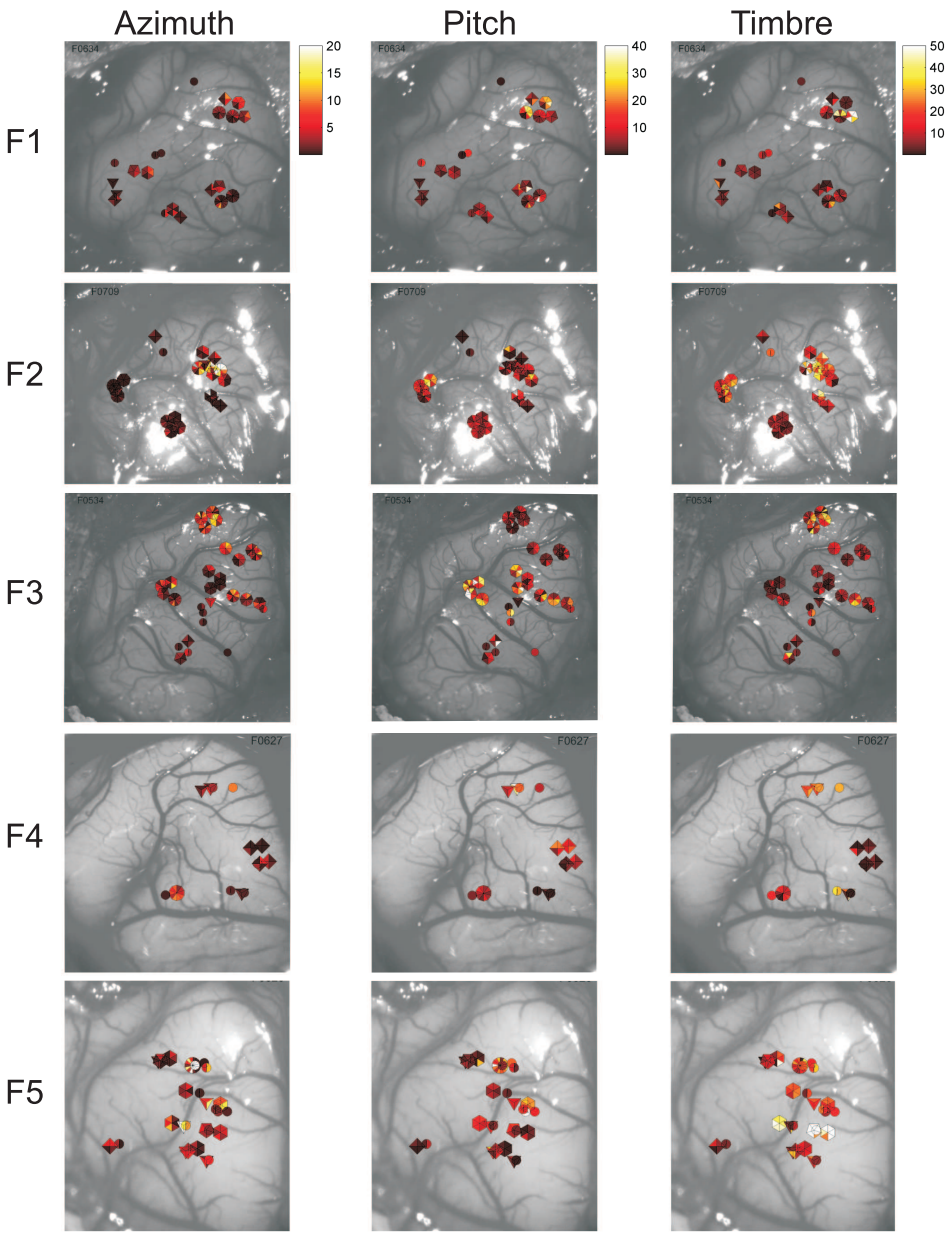


Figure 4.10: Maps of azimuth, pitch and timbre sensitivity across the auditory cortices of five animals. Similarly to Figure 4.7D-F, each map is superimposed with tiles, representing recorded neural units. The tile colour indicates the proportion of neural spiking variance explained by each stimulus attribute (azimuth, pitch and timbre, from left to right). Each row shows the data for one of the five ferrets examined. Tiles representing units from a single penetration are arranged counter-clockwise by increasing depth around the penetration site, with the most shallow unit at the far right.

4.3.4 Cortical sensitivity to combinations of stimulus features

My analysis quantified the linear contribution of each of the three stimulus dimensions to neurons' responses, as well as the non-linear effects of particular stimulus combinations. This non-linear effect is quantified by the interaction of the time bin factor with two stimulus parameters, and is indicative of a type of combination sensitivity. Such effects result from multiplicative interactions between categorical variates (such as the vowel /i/ at a pitch of 200 Hz), and are analogous to a logical AND operation. I calculated the proportion of a units' response variance explained by interactions between two stimulus parameters as:

$$V_{p_{ij}} = \frac{SS_{p_i \cdot p_j \cdot t} - SS_{Error} df_{p_i \cdot p_j \cdot t}}{SS_{Total} - SS_t}, \tag{4.2}$$

where $V_{p_{ij}}$ is the proportion of temporal spiking variance explained by interactions of stimulus parameters i and j (say, pitch and timbre), $SS_{p_i \cdot p_j \cdot t}$ is the Sum of Squares for the three-way interaction of the time bin variable with the two stimulus parameters i and j, and $df_{p_i \cdot p_j \cdot t}$ is the degrees of freedom for the interaction of time, stimulus parameter i, and stimulus parameter j.

Over two thirds of neural units were sensitive to more than one stimulus dimension, and 41% of all neurons showed significant ($p < 0.001$) non-linear interactions for at least one pair of stimulus parameters (i.e. pitch and timbre, timbre and azimuth, or azimuth and pitch). The distribution of sensitivity to stimulus parameter combinations (i.e. $V_{p_{ij}}$ values) across the cortical surface is shown in Figure 4.11, and the number of significant non-linear interactions is compared across cortical fields in Figure 4.12. Non-linear interactions were most commonly observed in fields A1 and AAF, where 54% and 56% of neural units, respectively, were significantly influenced by one or more of the two-stimulus combinations ($p < 0.001$). The most common non-linear interactions were attributable to pitch and timbre combinations, however azimuth-by-timbre and azimuth-by-pitch combination sensitivity were also observed. The proportion of neurons exhibiting significant non-linear interactions fell to 36%, 30% and 15% in fields PPF, PSF and ADF, respectively.

While the number of neurons showing significant non-linear interactions varied across cortical fields (Figure 4.12), the average magnitude of the interaction terms did not vary across fields for either the azimuth-by-pitch or azimuth-by-timbre combinations. The

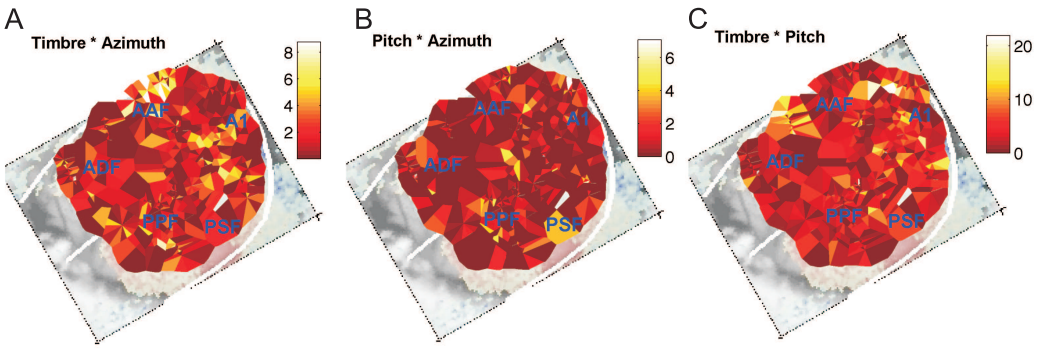


Figure 4.11: Maps of azimuth, pitch and timbre non-linear combination sensitivity across auditory cortex. Voronoi tessellations show the distributions of non-additive neural sensitivities to timbre-by-azimuth (A), pitch-by-azimuth (B), and timbre-by-pitch (C) combinations.

distribution of the sum of these spatial-by-non-spatial interaction terms is plotted in Figure 4.13A (Kruskal-Wallis test, $\chi^2 = 5.5$, $p = 0.24$). In contrast, the proportion of response variance explained by the interaction of the two non-spatial stimulus parameters (namely, pitch and timbre) did differ significantly across cortical fields (Figure 4.13B; Kruskal-Wallis test, $\chi^2 = 48.2$, $p < 0.001$). In the core areas A1 and AAF, a greater proportion of neural response variance was explained by the pitch-by-timbre interaction term than in the posterior fields PPF and PSF (Tukey-Kramer post-hoc test, $p < 0.05$). In turn, this interaction term was, on average, larger in PPF and PSF neurons than in ADF ($p < 0.05$).

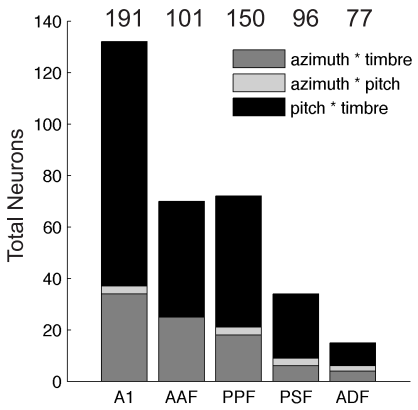


Figure 4.12: Histograms of non-linear combination sensitivity across auditory cortical areas. Stacked histograms of the number of neural units in each field in which each of the two-feature interactions were statistically significant ($p < 0.001$). Note that some units show significant interactions for more than two different combinations of features. The total number of units analysed for each cortical field are listed above.

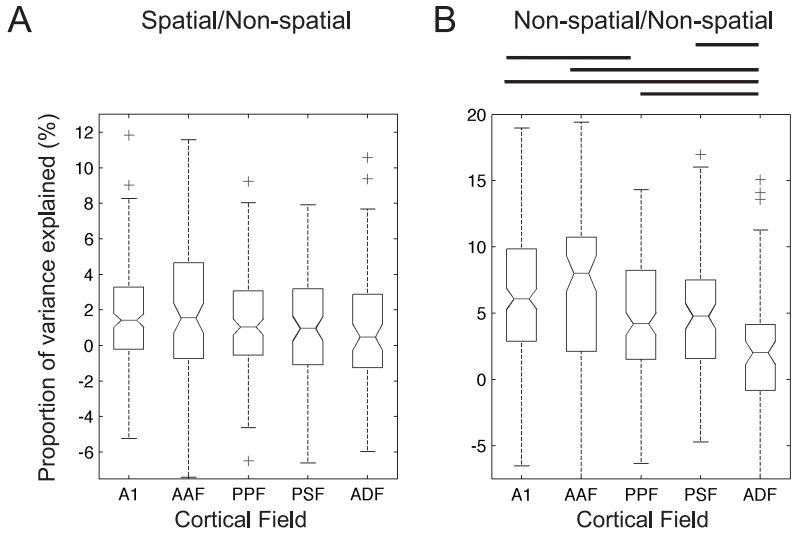


Figure 4.13: Neural sensitivity to combinations of spatial and non-spatial stimulus parameters across fields. (A) Distributions of the summed response variance explained by azimuth-by-timbre and azimuth-by-pitch interactions across cortical areas. (B) Distributions of pitch-by-timbre interaction effects across cortical areas. Horizontal lines above the plot highlight distributions that were significantly different in pairwise, post-hoc comparisons (Tukey-Kramer post-hoc test, $p < 0.05$).

4.3.5 Stimulus feature preferences of cortical neurons

To look for evidence of cortical "what" and "where" processing streams in these data, it would be helpful to categorize each unit as encoding predominantly one stimulus attribute, which might be spatial (azimuth) or non-spatial (pitch or timbre). A simple approach would be to label a unit a "timbre unit" if the proportion of its response variance that is explained by timbre is higher than that explained by pitch or azimuth. Figure 4.14A shows the distribution of neural units that were categorized as being mainly azimuth-, pitch- or timbre-sensitive using this approach. As might be predicted on the basis of the results in Figure 4.7, the emerging map has patches of "predominantly" timbre, pitch and azimuth units, which are not confined to specific cortical areas. There were a large number of units with a timbre classification in area AAF, but otherwise the distributions of these units across cortical fields were rather uniform. The y-axes in Figure 4.9 indicate that timbre and pitch accounted for, on average, a greater proportion of stimulus-related response variance than azimuth. Consistent with this, neurons that were predominantly sensitive to azimuth were less common and occupied a smaller area of the cortical map than those which were most strongly sensitive to pitch or timbre.

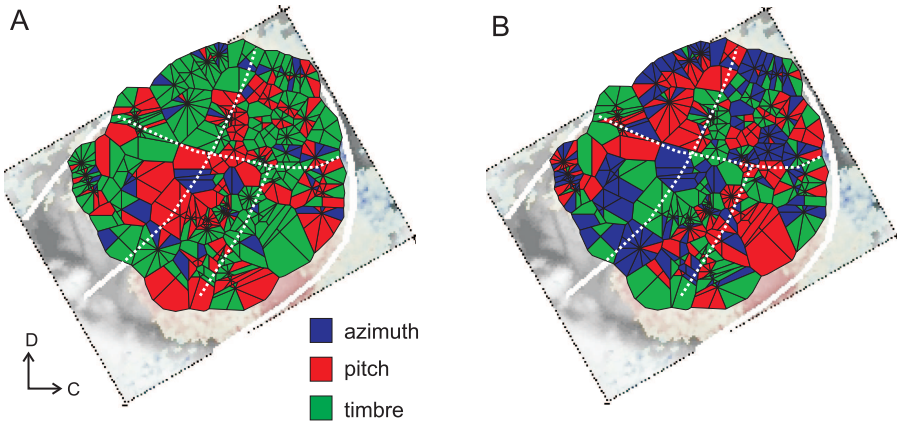


Figure 4.14: Stimulus feature preferences of units across the surface of auditory cortex. Veronoi tessellation maps, with tile colors indicating the categorized feature sensitivity of each neural unit. (A) Neural units were categorized as being mainly sensitive to the azimuth (blue), pitch (red) or timbre (green) of vowels, according to which of these stimulus parameters explained the largest proportion of the unit's spiking variance. (B) Categorizations based on the rank ordering of neural units according to the spiking variance explained by each of the three stimulus features.

This may reflect the underlying psychophysiology of the central auditory system. Processing perceptual properties, like timbre and pitch, which are related to object identity, may occupy more cortical circuitry than processing the spatial properties of a stimulus. However, given the very large proportion of units that were sensitive to multiple stimulus attributes, the results of a simple assignment based on the maximum proportion of variance explained could be biased to some extent by the choice of parameter ranges used in the experiments. If I were to choose, for example, a set of stimuli that differed greatly in pitch but hardly at all in azimuth, so that pitch was easily discriminable but azimuth was not, then it would not be surprising if pitch explained a larger proportion of the stimulus-related variance. This would reflect the properties of the stimulus set rather than the organization of the neural circuitry. While the four values of each perceptual dimension were chosen to be easily discriminable, sufficient ferret psychophysical data are not available to precisely match these stimulus features for their discriminability.

I therefore re-categorized neural units based on a ranking of their sensitivities to the azimuth, pitch and timbre of vowels. For each stimulus parameter, the entire set of neural units was ranked according to the proportion of their response variance explained by that parameter. Then, each unit was labelled as being relatively most sensitive to the stimulus parameter in which it ranked highest. This approach guaranteed a more equal number of units across the three feature categories. The results of this categorization are shown in Figure 4.14B. The organization of this map appears to have less structure than that in Figure 4.14A. The five cortical areas all included units with preferences for each of the three stimulus parameters, and no compelling evidence of area specialization was found.

4.3.6 Sensitivity to stimulus features across cortical depth

The recording electrodes used in this experiment made simultaneous recordings at 8-16 depths within the cortex. The electrode penetrations were made orthogonal to the surface of the cortex, and electrode insertion was observed under a microscope to ensure that the top site of the electrode array was always placed just beneath the cortical surface. I was therefore able to estimate the depths at which each neural unit was recorded, which enabled me to ask whether there were systematic differences in response properties across

different cortical layers.

Recordings were grouped into "superficial" or "deep" layers based simply on whether the recording site was less than 800 μm below the cortical surface. These divisions coarsely divide units into supra- and infra-granular cortical layers. The distribution of the proportion of response variance explained by each stimulus parameter is plotted across these two depth ranges in Figure 4.15. Azimuth sensitivity was found to be greatest in the deeper cortical layers (t-test, $p = 0.01$), while pitch and timbre sensitivity were found to be greater in the more superficial layers ($p = 0.004$ and $p < 0.001$, respectively). Pitch-by-timbre interactions were also more common in the superficial layers ($p = 0.002$), while pitch-by-azimuth and timbre-by-azimuth interactions were equally distributed across the two depths ($p = 0.33$, $p = 0.35$).

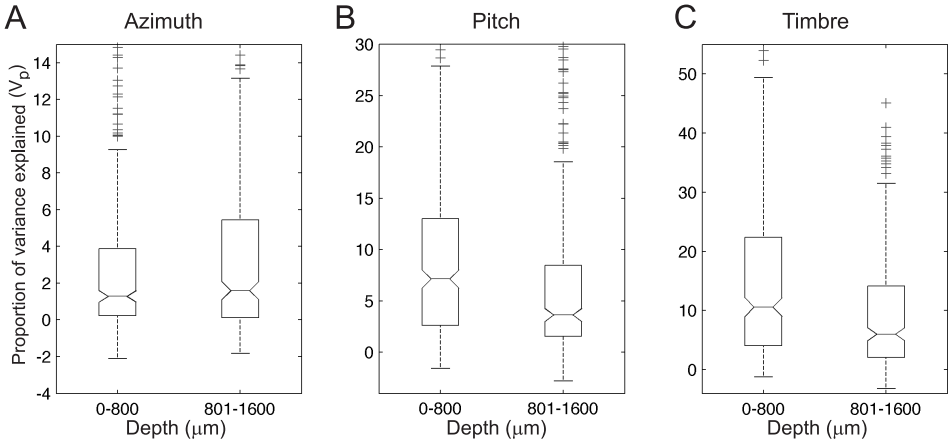


Figure 4.15: Azimuth, pitch and timbre sensitivity across cortical depth. Distributions of neural response variance attributable to the azimuth (A), pitch (B) and timbre (C) of vowel sounds in superficial ($< 800\mu$m) and deep ($> 800\mu$m) cortical layers. Higher sensitivities to pitch and timbre were more common in superficial layers, and azimuth showed the opposite trend.

Data were pooled across all cortical areas for these analyses, but similar trends were observed in each of the five cortical fields individually (data not shown). Unfortunately, a histological analysis of the precise depth of recording electrodes was not feasible in this study, mainly because it was not possible to make small, electrolytic lesions with the silicone probes. Therefore, the results of the present analysis of stimulus sensitivity across cortical depths may underestimate laminar differences due to the broad depth

categories used.

4.3.7 Azimuth, pitch and timbre sensitivity do not depend on unit CF

Figure 4.7F shows that clusters of units with high timbre sensitivity were commonly found in the low-frequency border region between fields A1, PPF and PSF. This raises the question of whether sensitivity to timbre, or indeed to other stimulus parameters, varies systematically with unit CF. Scatterplots of unit CF against the proportion of variance explained (Figure 4.16A-C, left panels) showed no systematic relationship between the frequency preferences and azimuth, pitch or timbre sensitivity of units (Pearson correlation, $p > 0.05$). Furthermore, pitch, timbre and azimuth sensitivities were just as common among units that were unresponsive or untuned to pure tones as in units with clearly defined CFs (Figure 4.16A-C, right panels). There were no CFs at which it was particularly common or uncommon to obtain driven vowel responses (Figure 4.17). In addition, I found no significant correlation between CF and best pitch (defined as the pitch that elicited the most spikes per presentation) for pitch-sensitive units with a CF below 1 kHz (Pearson correlation, $p > 0.05$).

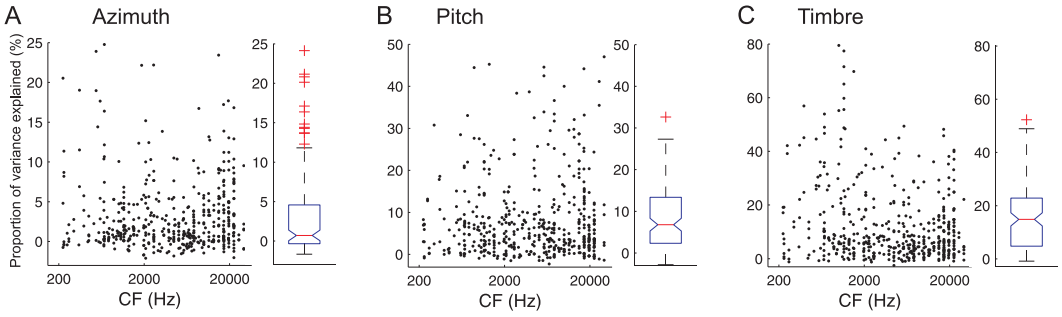


Figure 4.16: No dependence of azimuth, pitch or timbre sensitivity on pure tone frequency tuning. Left panel scatterplots show the proportion of each unit's response variance explained by azimuth (A), pitch (B) and timbre (C), as a function of the unit's characteristic frequency. In the right panels of A-C, box plots show the proportion of response variance explained by azimuth (A), pitch (B) and timbre (C) for units that could not be assigned a CF.

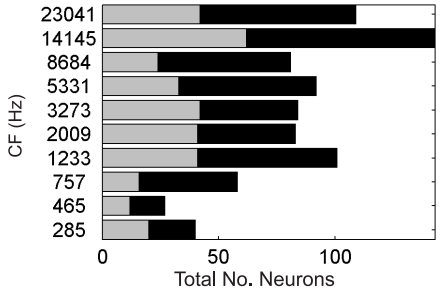


Figure 4.17: Histograms of unit CFs. Distribution of CFs for all recorded units (black), and for the subset of these units for which vowel responses were *not* analysed (grey). Responses to vowels in the later group of units were usually not analyzed because these neurons were not driven by vowels, but also occasionally because the recordings were not stable across at least 30 repetitions of the vowel sounds.

4.3.8 Time course of stimulus sensitivity

A visual inspection of the PSTHs of several neural units (as in Figure 4.4) suggested that different stimulus parameters were encoded at different times throughout a neuron's response. Specifically, changes in stimulus azimuth commonly altered the amplitude of the onset response, whereas stimulus pitch and timbre often influenced the amplitude or time-course of the onset response and also affected later parts of the more sustained responses. I carried out Tukey-Kramer post-hoc tests on the results of the 4-way ANOVAs to determine in which time bins the mean response rate was significantly affected by each stimulus parameter. As might be expected, there was considerable variability from one unit to another. To see whether there were nonetheless common trends within, (and systematic differences across), cortical areas, I plotted the percentage of units in each cortical field for which the spike rates in each 20-ms time bin showed a significant effect of azimuth (blue), pitch (red) or timbre (green) (Figure 4.18). An alpha of 0.05 was entered into the analysis to test for significance, but as this was a Tukey test, a "Tukey's Honestly Significant Different" alpha correction for multiple comparisons was utilized.

Generally, the results obtained for each stimulus dimension were rather similar. At *ca.* 20 ms after stimulus onset, the firing rates of around 15-40% of units were significantly modulated by each of the three stimulus dimensions in all five cortical areas. There were,

nevertheless, some differences across areas. In PPF and PSF, for example, some units were sensitive to pitch and timbre during the later stages of sustained onset responses (*ca.* 150-250 ms post stimulus onset), but fewer were significantly sensitive to azimuth during this period. Responses to the offset of the stimulus (*ca.* 180-200) were also sensitive to the pitch and timbre of vowels in the posterior fields (PSF and PPF) and, to a lesser extent, in A1 and AAF. Sensitivity of the offset response to azimuth was far less common in all of these fields.

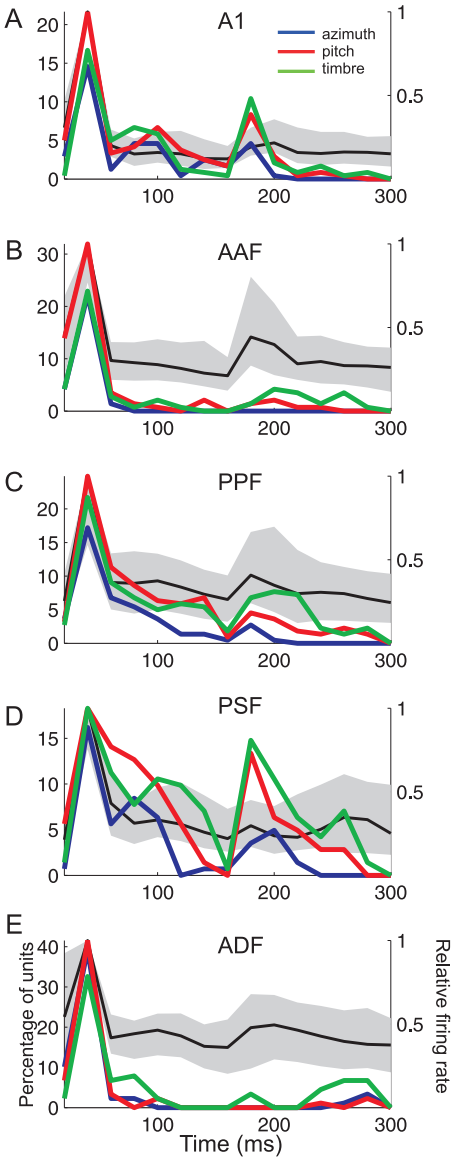


Figure 4.18: Azimuth, pitch and timbre sensitivity across the duration of neural responses in each cortical area. (A) The percentage of neural units in A1 with spike rates that were significantly sensitive (Tukey-Kramer post-hoc test, $p < 0.05$) to azimuth (blue line), pitch (red line) and timbre (green line) are shown for each of fifteen 20-ms-wide time bins across the duration of the neural response to vowels. For comparison, the black lines show the median PSTH, normalized relative to each unit's maximum firing rate in each cortical field. The grey bands extend from the 75th to the 25th percentile of the distribution of these normalized firing rates. (B-E) As in A, but showing data from the four other cortical fields, as indicated.

One might expect the firing rate to vary particularly strongly as a function of the stimulus parameters during periods of elevated average firing rate. To facilitate that

comparison, the black lines in Figure 4.18 show the median normalized PSTH across all stimulus parameters and all units in each cortical area, and the grey bands show the 75th to 25th percentile range. Although high proportions of significantly modulated responses usually coincided with periods of elevated firing rates, this was not universally the case. For example, the population-averaged PSTHs for AAF and ADF exhibited an offset response peak at *ca.* 180-200 ms, but this activity discriminated poorly between different stimulus conditions, since the firing of very few units varied significantly as a function of the stimulus parameters during this period. Conversely, in 12-15% of units in PSF, firing rates were significantly modulated by pitch and timbre during this offset response period, but this was not accompanied by a corresponding peak in the overall average firing rate.

4.4 Discussion

It has been suggested that a division of labor exists across the multiple areas that constitute the auditory cortex, with information parsed into anatomically separate processing streams for recognizing and localizing auditory objects (Rauschecker and Tian, 2000). This concept is inspired by earlier studies that postulated the existence of distinct hierarchies for processing different visual features, such as color or motion, in extrastriate visual cortex (Ungerleider and Haxby, 1994). Functional feature selectivity across auditory cortical areas has been strongly suggested by demonstrations of behavioral deficits following damage to, or inactivation of, specific cortical regions (Heffner, 1997; Kavanagh and Kelly, 1987; Lomber and Malhotra, 2008; Malhotra and Lomber, 2007), functional imaging studies (Griffiths *et al.*, 1998; Hall *et al.*, 2006), and by anatomical evidence of differences in the connectivity of these regions (Bizley *et al.*, 2007a; Romanski *et al.*, 1999). Physiological investigations of the response properties of cortical neurons, on the other hand, have demonstrated only modest, relative degrees of regional specialization in the spatial and non-spatial feature sensitivity that is found distributed throughout auditory cortex (Stecker *et al.*, 2005a, 2003; Tian *et al.*, 2001).

Most of these past physiological studies compared the sensitivity of neurons in different auditory cortical areas in response to only one stimulus parameter, such as its spatial location or pitch. Here I adopted a different approach, based around a stimulus

set that could be parametrically varied along several dimensions, in order to: (a) explore the relative sensitivity of neurons in different cortical fields to azimuth, pitch and timbre, and (b) ask, for the first time, whether neural information processing in auditory cortex serves to separate out these perceptual features of sounds. The responses of most neurons were significantly modulated by at least two of these stimulus parameters, arguing against the existence of cortical areas at this level that are specialized for processing a single stimulus feature. Although I observed clear differences in the extent to which individual neurons and auditory cortical areas encode these stimulus dimensions, the data show that sensitivity to azimuth, pitch and timbre is distributed throughout all five cortical areas studied. My findings are therefore not consistent with a simple division of cortical areas into those that contribute exclusively to either sound source localization or identification.

4.4.1 Temporal spike codes for stimulus features

The "proportion of response variance" calculation implemented in this study quantified the extent to which pitch, timbre, azimuth, and combinations of these features modulated neural responses. These methods were sensitive to changes in neural firing that occurred throughout the duration of a 300 ms response window, and were able to capture effects of stimulus parameters on spiking responses that a simple spike count measure failed to identify. Average spike count analyses tended to under-estimate neural modulations as a function of stimulus parameters, and so my findings emphasize the importance of the timing of spiking events in representing complex sounds in cortex. Previous studies have also shown that the timing of spikes in auditory cortex carries information about the pitch (Steinschneider *et al.*, 1998) and spatial location (Furukawa and Middlebrooks, 2002; Nelken *et al.*, 2005) of sounds.

The temporal-based analysis also showed that different stimulus attributes are often encoded at different times during a neural response. Information about sound-source location was almost exclusively encoded in the early phase of the response, whereas sensitivity to pitch and timbre continued to present later into the response, especially in area A1 and on the posterior bank (PPF and PSF). This is consistent with the findings of Ahveninen *et al.* (2006), who used a combination of functional Magnetic Resonance

Imaging (fMRI) and magnetoencephalography to investigate the processing of localization and phonetic cues in human auditory cortex. These two cues not only activated different higher order cortical areas, but did so over a subtly different time courses. Localization-sensitive cortical voxels were activated 30 ms earlier than those involved in processing phonetic information.

4.4.2 Feature specialization or integration?

Neurons sensitive to any one stimulus parameter were found in all five of the cortical fields examined, but I did observe small but significant differences in neurons' average sensitivity to each of the three parameters across cortical areas. The large variation in stimulus sensitivity seen across neurons within a given cortical field (Figure 4.7) and the small differences seen across neural populations in different areas (Figure 4.9) do not lend strong support to a dual stream hypothesis of cortical organization. However, the two core and three belt auditory cortical areas examined here represent still relatively early stages of the cortical auditory processing stream, and even more pronounced differences might emerge in downstream, higher-order areas.

While it is possible that cortical areas other than those examined here exhibit greater stimulus feature specialization, it is important to note that there is a potential trade-off between the advantages of anatomically localized, cortical specialization, and the necessity that spatial and non-spatial processing streams often have to interact intimately. The auditory system may frequently need to be able to consider both spatial and non-spatial aspects of sounds together, rather than separately. This need arises, for example, in acoustic environments which contain several sound sources. It is quite possible that, in order to operate effectively in such environments, the central auditory system must be able to track particular pitch, timbre and sound source direction combinations over time (e.g. *'was the speaker with the high pitched, nasal voice to my left, and the speaker with the low baritone voice on my right, or vice versa?'*). Interactions between spatial and non-spatial processing streams have been observed in the visual domain. For example, researchers have long been interested in so-called visual "form-from-motion" stimuli (Britten *et al.*, 1992a; Gulyas *et al.*, 1994), in which cues to the shape of an object (a "what" task) are provided exclusively through the concerted and coherent motion of iso-

lated, bright dots moving on a computer screen (a "where" task). One might expect spatial/non-spatial integration to be of even greater importance in auditory perception, since auditory objects can be attended to at any spatial location and we can hear multiple sounds in several locations simultaneously, whereas in the visual domain, we foveate on objects in only one location at a time.

In accordance with this argument, my analyses revealed that neurons' representations of pitch, azimuth and timbre cues in auditory cortex are highly interdependent. Neurons throughout all five cortical areas are often sensitive to more than one of these stimulus features, and their firing properties can often respond non-additively to feature combinations. Non-linear combination sensitivity was especially prevalent in the core fields A1 and AAF. It has been argued that the role of A1 is to form "auditory objects" by grouping physical stimulus attributes from a common source together, prior to later cortical areas extracting perceptual features such as pitch and location (Nelken and Bar Yosef, 2008). Grouping stimulus attributes is essential for tracking a sound source through a potentially cluttered acoustic environment, and the non-linear sensitivity observed in auditory cortex may be ideally suited to achieving this. Many cues are important for grouping physical stimulus attributes, such as common onset timing, pitch and location (Bregman, 1990). I found a decrease in non-linear feature interactions in cortical areas beyond A1, suggesting an increase in the independence of these perceptual dimensions in higher auditory cortex. Beyond A1, neurons are still highly sensitive to multiple features of complex sounds, but this sensitivity is well described by linear interactions of feature effects. This linearity is likely to preserve information and may be computationally more efficient than more selective "grandmother cells".

4.4.3 Can ferrets perceive these changes in stimulus features?

It could be argued that these low-frequency sounds were not optimally designed for probing auditory localization, since cortical neurons are most sensitive to spectral localization cues that are most clearly expressed at higher frequencies (Mrsic-Flogel *et al.*, 2005). However, ferrets have no trouble localizing low-frequency narrowband noise bursts (Kacelnik *et al.*, 2006) and they rapidly learn to accurately lateralize the same synthetic vowels that were used in these recording experiments (Chapter 5). While the range

of azimuthal positions used here covered only the frontal quadrant of auditory space, this range was deliberately chosen to span the region over which behavioural and electrophysiological studies have shown particularly high sensitivities to changes in sound source direction (Stecker *et al.*, 2005b). The majority of ferret cortical neurons respond broadly to sounds in contralateral space and exhibit the greatest alteration in spike rate in response to location changes across the midline (Campbell *et al.*, 2006; Mrsic-Flogel *et al.*, 2005), and ferrets' behavioural auditory localization is also most accurate in the anterior quadrant of space (Nodal *et al.*, 2008).

The existence of periodicities in ferret vocalizations suggests that these animals are likely be sensitive to pitch cues. At the time of these experiments, the pitch acuity of ferrets had not been investigated, but we now know that ferrets can be trained to discriminate the pitch of artificial vowel sounds (Chapter 5). The pitch steps used here correspond to approximately two ferret pitch "just noticable differences", and ferrets can similarly discriminate sounds in azimuthal space that are about twice as proximate as the azimuthal steps presented here.

There have been few psychophysical studies of timbre discrimination in animals, but Macaque monkeys have been shown to use spectral envelopes to discriminate vowel sounds in a manner similar to human listeners (Sommers *et al.*, 1992), and chinchillas can discriminate the vowels /i/ and /a/ across a range of speakers and pitches (Burdick and Miller, 1975). Ferrets' ability to discriminate the spectral envelopes associated with the four different vowels has not yet been formally investigated, but preliminary experiments in our lab have demonstrated that ferrets can rapidly learn to discriminate the identity of these vowel sounds. I am therefore confident that the chosen stimulus parameters tested the sensitivity of cortical neurons to each acoustic dimension over behaviourally relevant and broadly comparable ranges.

4.4.4 Azimuth representations in auditory cortex

Neurons that were sensitive to the azimuth of vowels were found in all five cortical fields. However, neural units in A1 were, on average, more sensitive to spatial location than those in other areas. Behavioral studies testing the effects of cortical manipulations have shown that A1 is necessary for normal sound localization in the ferret (Bizley *et al.*,

2007b; Kavanagh and Kelly, 1987; Smith *et al.*, 2004), just as it is in the cat (Malhotra *et al.*, 2004; Malhotra and Lomber, 2007; Malhotra *et al.*, 2008) and monkey (Heffner, 1997). But other cortical fields have also been shown to contribute to sound source localization. Reversible deactivation studies have suggested that the posterior auditory field (PAF), the anterior ectosylvian sulcal (AES) field, and, (to a lesser extent), the dorsal zone (DZ) of the auditory cortex of the cat all play an important role in localizing sounds (Lomber and Malhotra, 2008; Malhotra *et al.*, 2004; Malhotra and Lomber, 2007; Malhotra *et al.*, 2008). In addition, extracellular recordings of neural responses have demonstrated that neurons in cat PAF and DZ convey more information about the spatial location of sound sources than neurons in AI and AII (Stecker *et al.*, 2005a,b, 2003). Regional differences in spatial sensitivity have also been reported in the auditory cortex of ferrets (Bizley *et al.*, 2007a) and non-human primates (Recanzone, 2000; Recanzone *et al.*, 2000; Tian *et al.*, 2001). Human functional imaging data suggest a specialization for spatial processing in caudal auditory cortex, although these data have also been interpreted within a framework of distributed processing of spatial cues (Zatorre *et al.*, 2002).

The ferret homologues of cat auditory cortical areas remain unclear, making direct comparisons between these two species problematic. The auditory cortices of both species are displayed in Figure 4.19. Comparative anatomical layout, cytoarchitecture, and corticocollicular projections suggest that the ferret PPF is analogous to cat AII (Bajo *et al.*, 2007). On the other hand, the physiological properties of ferret PPF and PSF, including their tonotopic organization, more closely resemble neurons in cat PAF, while neurons in ferret ADF respond more similarly to cat AII neurons (Bizley *et al.*, 2005). The analogies of ferret A1 and AAF to cat AI and AAF, respectively, are less ambiguous. These areas share anatomical locations, cytoarchitectonic features (Bajo *et al.*, 2007; Mellot *et al.*, 2005; Winer, 1992), corticocollicular projections (Bajo *et al.*, 2007; Winer *et al.*, 1998), and physiological properties (Bizley *et al.*, 2005; Winer *et al.*, 2005) in both species, pointing to a clear homology. Finally, the cat AES has been likened to ferret fAES, which lies along the pseudosylvian sulcus in the latter species (Bizley *et al.*, 2007a; Manger *et al.*, 2005; Ramsay and Meredith, 2004). A ferret homologue of DZ has not yet, to my knowledge, been proposed.

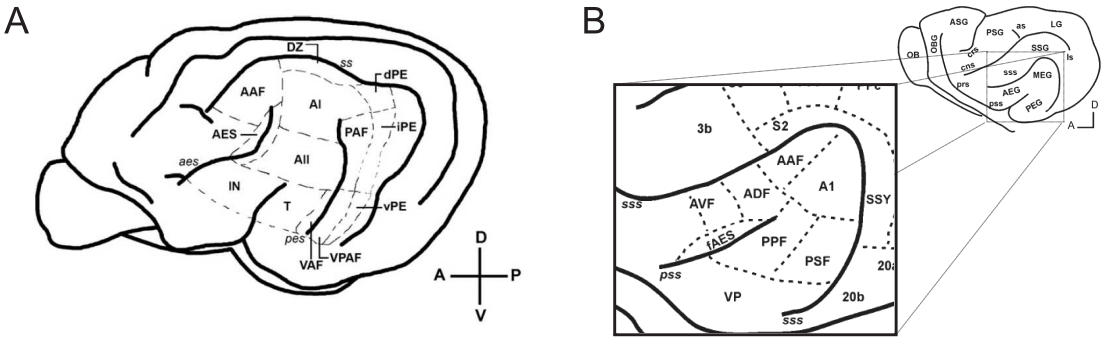


Figure 4.19: Auditory cortex of the cat and ferret. (A) Schematic of the lateral view of the left cat cortex, with dotted lines showing the boundaries of auditory cortical fields (reproduced from Lomber *et al.*, 2007). Three sulci are labelled in italics: the anterior ectosylvian sulcus (aes), posterior ectosylvian sulcus (pes), and suprasylvian sulcus (ss). Thirteen auditory cortical fields are also labelled: primary auditory cortex (AI, or A1), secondary auditory cortex (AII, or A2), anterior auditory field (AAF), auditory field of the anterior ectosylvian sulcus (AES), dorsal posterior ectosylvian gyrus (dPE), dorsal zone (DZ), insular region (IN), intermediate posterior ectosylvian gyrus (iPE), posterior auditory field (PAF), temporal region (T), ventral auditory field (VAF), ventral posterior auditory field (VPAF), and ventral posterior ectosylvian gyrus (vPE). (B) Schematic of the lateral view of the left side of ferret cortex (top right), with a magnified view of the ectosylvian gyrus inset (modified from Bizley *et al.*, 2007a). Again, field boundaries are drawn with dotted lines and sulci are labelled in italics (suprasylvian sulcus - sss; pseudosylvian sulcus - pss). Eight auditory cortical fields are indicated: A1, AAF, PPF, PSF, ADF, anterior ventral field (AVF), anterior ectosylvian sulcal field (fAES), and ventroposterior area (VP).

Based on the cat cortical deactivation studies described above, I might have expected to find an increased sensitivity to azimuth in ferret A1 and near the pseudosylvian sulcus (the ferret analog of cat AES). My results were consistent with this hypothesis (Figure 4.7D). However, if this homology was complete, I would have also expected spatial sensitivity in the posterior fields to be higher than that in the other four cortical fields studied, and this was not born out in the data. Neural units in PPF were significantly more sensitive to azimuthal location than those in ADF (Figure 4.9A), although this difference was a modest one. This finding is inconsistent with the results of Bizley *et al.* (2007a), who showed that neurons in ADF carried more information about sound source location than those in more posterior regions. However, the differences in stimuli used

in Bizley *et al.* (2007a) and my study might explain this apparent discrepancy. Bizley *et al.* (2007a) presented wide-band noise bursts, which will optimally stimulate area ADF (Bizley *et al.*, 2005), while my stimuli contained most energy in the low frequency range. ADF neurons are less likely to respond to artificial vowel sounds than noise bursts, although my stimuli are more representative of the spectra of ferret vocal calls.

The present results are best described as showing widespread sensitivity to azimuthal sound source location, in agreement with the distributed localization code proposed by previous electrophysiological studies of cat auditory cortex (Furukawa and Middlebrooks, 2002; Middlebrooks, 2002; Middlebrooks *et al.*, 1998; Stecker and Middlebrooks, 2003). Subtle differences in the spatial tuning of neurons in the posterior and more anterior-dorsal fields of the ferret might nevertheless exist, and if so, these might be demonstrated using more spatially dense stimuli and sensitive analytical approaches (as used by Stecker *et al.*, 2005a,b, 2003). However, the present results emphasize that in the ferret, as in the cat, the differences in sensitivity to sound source azimuth across cortical areas is merely a matter of degree, and there is no evidence of strict regional selectivities for location features.

Widespread sensitivity to azimuthal sound source location contrasts sharply with recent psychophysical data from Lomber and Malhorta (2008), who report that deactivation of cat PAF produces dramatic impairments in sound localization behaviour, while deactivating AAF has no measurable effect on this task. However, complex behaviours, like remembering to approach a sound source for a reward, are bound to require cognitive control from high-order cortical areas, most likely in frontal cortex. Thus, the profound differences observed by Lomber and Malhorta (2008) in response to cooling PAF and AAF may have little to do with physiological response properties in these areas, but instead reflect potentially very large differences in their respective anatomical projections patterns to higher brain regions (Romanski *et al.*, 1999).

4.4.5 Pitch representations in auditory cortex

The representation of pitch in auditory cortex is not uncontroversial, with some studies stressing the role of a localized area for pitch processing and others emphasizing distributed pitch sensitivity. My results are consistent with distributed pitch represen-

tations, but this finding does not contradict the existence of a specialized pitch centre, as we will see.

Bendor and Wang (2005) investigated pitch encoding in the auditory cortex of the marmoset. Their study reserved the term "pitch neuron" to describe only those which were tuned to the fundamental frequency of sounds irrespective of the spectral content (i.e. timbre) of sounds. They found only a small number of these pitch neurons, confined to the low-frequency, anterio-lateral border of A1 with field R. My "pitch sensitivity" measure is far less stringent than that used by Bendor and Wang. This study aimed to identify all neurons that were sensitive to pitch cues, rather than the subset of these that were selective for pitch. Nevertheless, in agreement with their findings, I found that pitch-sensitive neurons were commonly (although not exclusively) found near the low-frequency border of A1 with the belt areas (namely, PPF and PSF) in ferret cortex. It is therefore possible that a ferret homologue of the marmoset pitch area is included within the neurons that were identified here as pitch-selective. Unlike Bendor and Wang, I did not find a correlation between unit CF and "best" pitch, nor did I observe a tendency for low CF neurons to be more sensitive to pitch cues. Further investigation will be necessary to determine the relationship between pitch-sensitive and pitch-selective neurons in auditory cortex, and how these neural populations might contribute to pitch perception. My results suggest that if a ferret "pitch centre" does exist, it is perhaps unlikely to be found in anterior auditory cortex, especially ADV, where neurons were most poorly modulated by the pitch of vowels.

A recent optical imaging study of ferret auditory cortex showed that sensitivity to pitch cues is distributed across cortical fields (Nelken *et al.*, 2008), similarly to the present findings. Several papers have shown compelling fMRI evidence of a localized pitch-sensitive region in Heschl's gyrus of the human auditory cortex (Griffiths, 2003; Griffiths *et al.*, 1998; Hall *et al.*, 2006), and some have further specified that a human pitch centre exists in secondary auditory cortex on this gyrus (Patterson *et al.*, 2002; Penagos *et al.*, 2004). Electromagnetic studies of human cortex have produced findings consistent with this functional organization (Gutschalk *et al.*, 2002; Krumbholz *et al.*, 2003), but the interpretation of these results is not uncontroversial. A very recent fMRI investigation has cast doubt on whether human cortex features just a single, universal pitch centre

(Hall and Plack, 2007), and different aspects of pitch perception (namely, pitch chroma and pitch height) have been shown to be processed in distinct cortical regions beyond A1 (Warren *et al.*, 2003). While electrophysiological studies like the present one are necessary to demonstrate if and how single neurons encode pitch in their firing rates, human imaging studies offer the desirable option of studying pitch-dependent activity while subjects are actively engaging in listening (Jäncke *et al.*, 1999). Therefore, to ultimately answer the question of how auditory cortical neurons contribute to pitch perception, their spike outputs must be investigated in the context of listening behaviour. For instance, how acutely do ferrets perceive pitch changes, and how does this relate to neural representations of pitch across auditory cortical areas? Studies of complex, periodicity pitch perception in non-humans are rare, more so those that use naturalistic stimuli. In fact, it has not been previously demonstrated that animals perceive the periodicity of complex, harmonic stimuli along an ordered, low-to-high pitch scale, as humans do. An understanding of the behavioural relevance of pitch cues to animals will be essential in order to integrate the results of human imaging and animal electrophysiological studies into a general theory of mammalian pitch processing. Some of these experimental questions are addressed in Chapter 5, and more in Chapter 6.

4.4.6 Timbre representations in auditory cortex

The neural basis for timbre processing has been much less widely studied. This is perhaps in part because of the difficulty in defining this perceptual attribute and the parameters that correlate with it, but in most cases (and certainly in the case of vowel sounds) timbre perception is be dominated by the spectral envelope of sounds (discussed in Shamma, 1999). Previous studies have demonstrated that the response properties of neurons in the primary auditory cortex of cats (Schreiner and Calhoun, 1994) and ferrets (Shamma and Versnel, 1995) are well suited to select for these types of spectral envelope cues. Versnel and Shamma (1998) estimated the excitatory and inhibitory receptive fields of ferret A1 neurons based on their responses to sinusoidally-modulated broadband sounds ("ripple" stimuli), and showed that neurons' responses to vowels of different timbres were well predicted by these receptive field properties. Predictions based on the pure-tone frequency tuning of neurons could not match this predictive power. This suggests

that rather than simply responding to the energy of broadband sounds at a narrow characteristic frequency range (as takes place at the level of the auditory nerve), A1 neurons respond to the wider spectral profile of broadband sounds, making them highly sensitive to timbre cues. Consistent with this hypothesis, I found that sensitivity to vowel timbre was greatest in the core auditory cortical areas of the ferret - A1 and AAF. This is also in agreement with the proposed role of auditory cortex as an auditory object filter (Nelken and Bar Yosef, 2008). Keeling *et al.* (2008) have demonstrated task-relevant changes in the spectral response properties of A1 neurons in cats while they learned to discriminate vowel-like sounds, suggesting that these receptive properties play a key role in behavioural discrimination of timbre.

On the other hand, timbre processing is not likely to be confined to A1 in mammals. As in the cases of azimuth and pitch cues, I found neurons that were sensitive to the timbre of vowels in all five areas of auditory cortex. Kudoh *et al.* (2006) showed that lesions of the dorsal/rostral auditory association cortex, but not A1, in rats impaired their performance on a multi-formant vowel discrimination task. It is possible that this region is analogous to AAF in ferrets, which is also dorsal/rostral to A1 and which showed improved sensitivity to timbre features when compared to more posterior fields in my study. Regions beyond the primary auditory cortex of the human brain have also been implicated in timbre perception. Kumar *et al.* (2007) carried out probabilistic tests of dynamic causal models of fMRI data and demonstrated that the planum temporale may be particularly suited to abstract the spectrotemporal template of sounds. Patients with lesions of the temporal lobe, particularly in the right hemisphere, often have impaired timbre perception (Samson *et al.*, 2002).

4.4.7 Stimulus modulation across the cortical depth

My analysis also provided some indication that sensitivities to different stimulus features may be distributed differently throughout the cortical depth. Pitch and timbre predicted significantly more spiking variance in the responses of neurons recorded at more superficial recording sites, corresponding to the upper cortical layers. Bendor and Wang (2005) also reported that most of the pitch neurons they identified were located in the upper layers.

Here I observed, for the first time, a small but statistically significant tendency for neurons in the deeper layers to be more sensitive to stimulus location. There is some evidence that response properties such as bandwidth, inhibitory sideband structure, preferred modulation rate and modulation phase sensitivity vary across layers in auditory cortex (reviewed by Linden and Schreiner, 2003). Depth selectivity could potentially provide another way for the processing of different sound features to be carried out in parallel.

4.4.8 Summary

There is some evidence for a differentiation of function across auditory cortical areas in the present study, but the distributions across cortical fields do not fit easily into a model of dual cortical processing streams for sound localization and identification. A more striking finding in this study is that very few neurons are selective for one perceptual dimension (pitch, timbre or azimuth) whilst remaining invariant to the others. Instead, most neurons are sensitive to multiple stimulus parameters. Non-linear codes of stimulus feature combinations can also be found, particularly for combinations of pitch and timbre, but these non-linear representations of conjunctions of stimulus features are less common in higher cortical areas. Although such combination sensitivity will ultimately limit information transmission, this form of processing could facilitate the formation of auditory objects early in auditory cortex.

It is possible that greater differences in stimulus sensitivity within the auditory cortex would have been observed in awake, behaving animals, particularly as the tuning characteristics of ferret cortical neurons can change in a task-dependent fashion (Fritz *et al.*, 2005a, 2003). It would be interesting to see how performing a task in a given perceptual dimension influenced the sensitivity of neurons throughout auditory cortex. This type of behavioural engagement with acoustic stimuli may be necessary to observe the neural spiking activity that can reconcile the clear functional division of sound processing observed in deactivation and human imaging studies with single cell electrophysiological results.

4.5 Acknowledgements

This work was carried out in partnership with Dr. Jennifer Bizley. Dr. Bernard Silverman consulted on statistical procedures. Prof. Andrew King and Dr. Jan Schnupp supervised the experiment, and Dr. Schnupp designed the VAS filters that were applied to the stimuli. A manuscript of this chapter has been published in the Journal of Neuroscience (Bizley *et al.*, 2009).

Chapter 5

Discrimination by Ferrets of the Direction of Pitch Changes

5.1 Introduction

To interpret a vocal call appropriately, animals must process a number of perceptual attributes of complex sounds, including loudness, timbre, and pitch. In the context of human speech, pitch perception is particularly pertinent to identifying the speaker and inferring their emotional state (Johnson, 1990), and is thought to play similarly important roles in vocal communication among non-human primates (Koda and Masataka, 2002; Kojima *et al.*, 2003), songbirds (Nelson, 1989), and even frogs (Capranica, 1966). But even though pitch is clearly a fundamental perceptual attribute of sound, few studies to date have examined how well animals can judge the pitch of complex stimuli.

The physiology underlying pitch perception for complex sounds is also incompletely understood. In the case of pure tones, which comprise a sine-wave at a single frequency, perceived pitch correlates directly with pure tone frequency, and different pure tones maximally stimulate different parts of the cochlea (Kim *et al.*, 1990; Kim and Parham, 1990). These widely appreciated and fundamental facts can easily lead to the tempting, but probably incorrect, conclusion that pitch perception is the result of a simple frequency discrimination problem, which can be easily solved through "place coding" in the tonotopically-organized ascending auditory pathway. However, most naturally occurring sounds are spectrally complex, i.e. they carry acoustic energy at large numbers of fre-

quencies. For many types of complex sounds, the relationship between frequency content and perceived pitch is anything but straight-forward and pitch cannot be readily derived from a tonotopic place code. Instead, pitch perception may rely more on "time domain" information carried by temporally-phase-locked neural discharges in the auditory nerve (Cedolin and Delgutte, 2005). Thus, the periodicity of a complex sound, as measured using autocorrelation functions, is highly predictive of the perceived pitch, and inter-spike interval distributions of neurons early in the auditory pathway accurately reflect this periodicity pitch cue (Cariani, 1999). If pitch information is indeed derived from a neural analysis of periodicity in the inter-spike intervals of phase-locked discharges, then this need not necessarily result in an anatomically-ordered, topographic representation analogous to tonotopic maps. Nevertheless, some studies have provided evidence in favour of periodotopic arrangements at the level of the gerbil inferior colliculus (Schreiner and Langner, 1988) and primary auditory cortex (Schulze *et al.*, 2002; Schulze and Langner, 1997a), which might serve as pitch maps. But others have failed to find any clear topographic arrangement of pitch preference in the auditory cortex of either ferrets (Nelken *et al.*, 2008) or primates (Bendor and Wang, 2005; Fishman *et al.*, 1998; Steinschneider *et al.*, 1998). Therefore, it is presently unclear whether pitch maps that represent the pitch of complex sounds through a place code exist in the auditory pathway of most mammalian species

There is at present no conclusive psychophysical evidence that non-human mammals can perceive pitch changes in complex sounds as occurring along an ordered, low-to-high scale. Recent studies of individuals with cochlear implants have emphasized that the ability to detect changes in the pitch of ongoing sounds does not necessarily imply that the listener can order these sounds along a pitch scale. While adults and children with cochlear implants are severely impaired on tasks that require discriminations of the direction of pitch changes (Fujita and Ito, 1999; Gfeller *et al.*, 2002; Vongpaisal *et al.*, 2006), children with cochlear implants have been shown to exhibit much finer pitch acuity on tasks that simply require them to report changes in the pitch of a sequence of complex sounds (Vongpaisal *et al.*, 2006).

Numerous studies have shown that diverse species of animals are sensitive to changes in the pitch of tones or more complex stimuli. Most commonly, this is tested using

a "go/no-go" task in which an animal is conditioned (with food or water rewards, or by shock avoidance) to make a response if, and only if, the pitch of ongoing stimuli changes. This paradigm has been used to examine pure tone frequency discrimination in a wide variety of species, including macaques (Pfingst, 1993; Sinnott *et al.*, 1985), chinchillas (Nelson and Kiester, 1978; Shofner, 2000), cats (Elliott *et al.*, 1960; Witte and Kipke, 2005), rats (Syka *et al.*, 1996; Talwar and Gerstein, 1998, 1999), mice (Ehret, 1975), guinea pigs (Heffner *et al.*, 1971), budgerigars (Dooling and Saunders, 1975), and electric fish (Marvit and Crawford, 2000). However, as noted above, pure tone frequency discrimination is not the same thing as pitch discrimination for complex sounds. In a few studies, these paradigms have also been used to demonstrate that animals can detect changes in the pitch of complex sounds, such as click trains (Marvit and Crawford, 2000) and harmonic tone complexes (Dooling *et al.*, 2002; Shofner, 2000). What is not known, however, is whether the animals were also able to tell the direction of the pitch change.

Psychophysical paradigms required to address this question are much more difficult to carry out. Only a handful of studies of this kind have been undertaken so far, and none of them have used complex sounds. Go/no-go tasks, in which animals are trained to respond to a particular direction of frequency change (i.e., frequency increases or decreases) within a sequence of tones, have been used to demonstrate that primates (Brosch *et al.*, 2004; D'Amato, 1988) and birds (Cynx, 1995; Page *et al.*, 1989) do have the capacity to make relative pure tone frequency judgments. However, as noted above, the neural mechanisms of pure tone frequency discrimination are likely to be quite different from those used when judging the pitch of a complex sound.

Investigations of pitch discrimination in humans tend to favour two-alternative forced choice (2AFC) designs, in which the subject must not only detect pitch changes, but also identify the change on each trial as a relative increase or decrease in pitch from a standard, reference value (Wier *et al.*, 1977). This type of task is not commonly used in animal psychoacoustics because it is considerably more difficult to train animals to perform 2AFC tasks than go/no-go tasks (Burdick, 1979). However, these differences in task design make it problematic to compare human and animal frequency or pitch difference limens directly, since frequency discrimination performance has been shown to vary across different training paradigms (Burdick, 1979; Nelson and Kiester, 1978;

Talwar and Gerstein, 1999).

In the present study, I trained ferrets to discriminate the direction of pitch change between pairs of successively-presented sounds on a positively conditioned, 2AFC discrimination task that is similar in design to that of previous human studies. On each trial, two sounds were presented in succession and the ferret had to indicate, by choosing a water spout either to the left or to the right, whether the second sound was higher or lower in pitch than the first. The anatomy and physiology of the ferret auditory system are well documented, and there are a number of observations that suggest that the ferret may be a highly suitable species in which to study the role of pitch cues in vocalization processing. Ferrets are highly sensitive to low-frequency pure tones (Kelly *et al.*, 1986a), and many of their vocalizations contain low-frequency energy and are strongly periodic, which gives them a clearly identifiable pitch. The responses of ferret auditory cortical neurons encode information about the pitch of artificial vowels (Chapter 4), and support the discrimination of human speech sounds (Mesgarani *et al.*, 2008). However, the ability of ferrets to discriminate the pitch of complex sounds has not been previously measured.

Here, I measured both artificial vowel pitch discrimination and pure tone frequency discrimination in ferrets, using an identical paradigm, so that difference limens and Weber fractions could be directly compared. The artificial vowel stimuli were sufficiently complex to exhibit key features that are found commonly in vocalization sounds used in human and animal communication, yet simple enough to be described by a small number of numerical parameters. The overall spectral distribution of the vowels is largely determined by its' formants and was similar across all pitches tested. Finally, I also measured the performance of naive human listeners for comparison, using a very similar paradigm.

5.2 Methods

5.2.1 Subjects

Four adult pigmented ferrets (one male) were trained in this study. Ferrets were housed either singly (males) or in groups of two or three (females), with free access to high-protein food pellets and water bottles. On the day before training, water bottles were removed from the ferret's home cages and they were replaced on the last day of a training run.

Training runs lasted for five days or less, with at least two days between each run. On training days, ferrets received drinking water as positive reinforcement while performing a sound discrimination task. Water consumption during training was measured, and was supplemented as wet food in home cages at the end of the day to ensure that each ferret received at least 70 ml of water per kilogram of body weight daily. Regular otoscopic and typanometry examinations were carried out to ensure that both ears of the animals were clean and healthy. All experimental procedures were approved by the local ethical review committee and were carried out under license from the UK Home Office, in accordance with the Animals (Scientific Procedures) Act 1986.

5.2.2 Training apparatus

Ferrets were trained to discriminate sounds in custom-built testing chambers constructed from double glazing units that incorporated a sound-insulating vacuum. The testing chambers were approximately 45 cm wide, 62 cm long, and 54 cm high (Figure 5.1). A plexiglass wall 12 cm from the back of the box separated the animal from the electronics and tubing of the apparatus. Three metallic water spouts were mounted on the plexiglass wall, one centrally positioned "start spout" and two "response spouts" positioned to the left and right. An aluminium 32x34 cm footplate covered the floor below the water spouts. When the ferret licked a water spout while standing on the footplate a small change in voltage between the steel spout and the aluminium plate resulted, allowing the experimenter to register the animal's licking responses using electronic circuitry, as described by Hayar *et al.* (2006). Sound stimuli as well as acoustic feedback signals were delivered via three Visaton FRS-8 speakers (VISATON GmbH and Co. KG, Haan, Germany) which were mounted above the spouts. Plexiglass partitions, 13 cm long and 15 cm high, were positioned between the central spout and each of the peripheral spouts to increase the "perceived cost" involved in the ferrets' response, as initial testing had indicated that ferrets were less likely to pay careful attention to the acoustical cues if hopping between response spouts required essentially zero time or effort.

The behavioral task, data acquisition, and stimulus generation were all automated using custom MATLAB scripts running on personal computers, which communicated with TDT RM1 real-time signal processors.

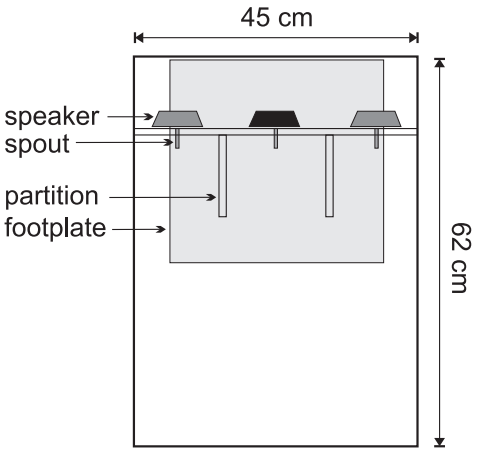


Figure 5.1: Schematic of the training apparatus, shown from above. Ferrets made behavioural responses and received water rewards from the three stainless steel water spouts while standing on the aluminium footplate. In the first training stage, sounds were presented from the centre speaker (black) as well as the two peripheral speakers (grey). In later training stages and during testing, sounds were presented from the central speaker only.

5.2.3 Training

Ferrets were trained on a 2AFC discrimination task, using drinking water as positive reinforcer. To assist learning, animals were trained in several stages of 2AFC tasks, and each ferret was advanced to the next stage when they reached a criterion of at least 85% correct on three consecutive sessions. The main differences between each of the five training stages and the final testing stage are highlighted in Table 5.1, and are described in detail below. Ferrets ran two training sessions daily and typically completed between 60 - 150 trials per session.

Training stage	Speaker positions	Stimulus type	Centre lick triggers	Num. target pitches
1	Centre and peripheral	Tones	Target	2
2	Centre	Tones	Target	2
3	Centre	Click trains	Target	2
4	Centre	Click trains	Reference and target	2
5	Centre	Artificial vowels	Reference and target	2
Test	Centre	Artificial vowels	Reference and target	30

Table 5.1: Summary of training stages

Pre-training was carried out in the testing chamber, during which ferrets learned to lick the spouts for a water reward. The animal was required to maintain contact with a spout for 1 - 1.2 seconds before receiving a water reward from that spout. During pre-training, ferrets were required to alternate between the central and peripheral spouts

in order to receive a water reward, but no sound stimuli were presented at this time. The amount of water used to reward a single response varied across animals, but for all animals the water reward presented from the peripheral response spouts (0.3 - 0.5 ml per trial) was larger than the water reinforcer presented at the central start spout (0.1 - 0.2 ml per trial).

5.2.3.1 Training stage 1

In the first training stage, the ferrets performed a pure tone frequency discrimination task, but with a lateralization cue to assist. At the start of each trial, a "reference signal" consisting of a continuously repeated pure tone (5 kHz, 100 ms duration, 150 ms inter-tone interval, 0.5 ms cosine ramped rise/fall) was presented from the central speaker until the ferret activated the central spout to start the trial (Figure 5.2A). This activated a small water reinforcer at the central spout, followed by a second "target" pulse of pure tones (300 ms duration, 150 ms inter-tone interval, 0.5 ms cosine ramped rise/fall) that differed in frequency from the initial reference tones. The target pulse continued to play until the ferret activated one of the peripheral spouts. If the target frequency was higher than the reference, the animal was required to move to the right response spout to receive a water reward. For frequencies lower than the reference the animal was required to move to the left. Responses at the central spout were ignored, and responses at the incorrect side spout resulted in a 500-ms broadband noise "punishment sound" at the central speaker, followed by a time out of 10 - 12 seconds. To make this task easy I provided an additional spatial cue during the initial training phases: the target tones were presented only from the speaker above the correct water spout for a given trial. Thus, if the target was higher in frequency than the reference, then the target sounds were presented from the right peripheral speaker and right spout responses were rewarded with water.

If an animal responded incorrectly on a given trial, the same stimuli were presented on the next trial. Such "correction trials" continued until the animal responded correctly. For all training stages, the pitch of the reference sound was fixed during a session, and the target sound varied between two pitches across trials - one pitch that was at least an octave higher than the reference and another that was at least an octave lower than the

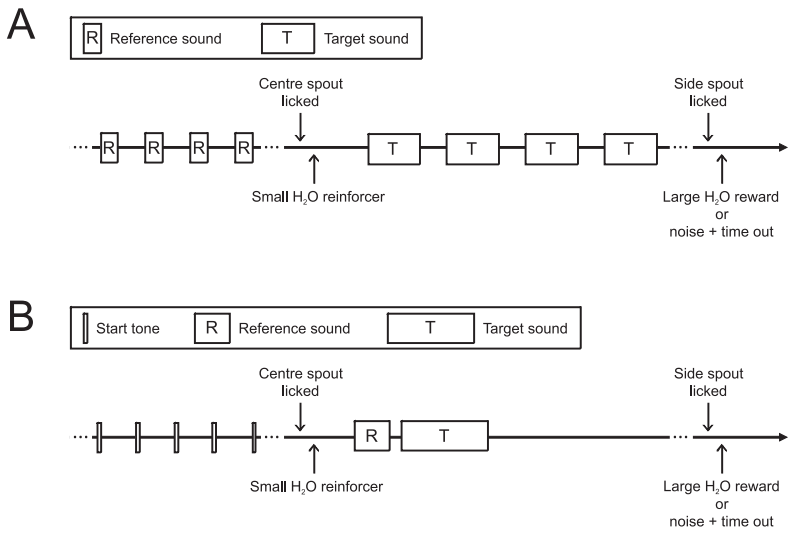


Figure 5.2: Discrimination trial schematics. (A) In the first 3 stages of training, a reference stimulus was repeated until the ferret initiated a trial by activating the centre spout. Then a target stimulus of a different frequency was presented repeatedly until the animal responded. The ferrets' task was to indicate whether the pitch of the target sound was higher or lower than the reference by activating the right or left peripheral spouts, respectively. In training stage 1 and 2 the stimuli were pure tones, and in stage 3 they were click trains. (B) In training stages 4 and 5 and in testing, a "ready signal" consisting of a repeating 5000 Hz tone pip was presented to indicate to the animal that a trial could be started by licking the centre "start" spout. The reference and target sound were then presented only once, in quick succession. As above, the ferrets' task was to indicate the direction of pitch change between the reference and target sounds by responding at the correct peripheral spout. The sounds to be discriminated were click trains in training stage 4 and artificial vowel sounds (formant-filtered click trains) in the final training and testing stages.

reference.

5.2.3.2 Training stage 2

Once animals learned to perform the above task, the localization cue was removed so that the ferret was left only with the frequency differences between reference and target sounds as an acoustic cue to the correct response. For this and all further tasks, all stimuli were presented from the middle speaker above the start spout only.

5.2.3.3 Training stage 3

Once criterion had been reached on the pure tone frequency discrimination task, the tones were replaced with click trains, where the click rate varied between the reference and target sounds. Thus, the reference and target sounds were both broadband, but differed in pitch.

5.2.3.4 Training stage 4

In stage 4 of training, the previous pitch discrimination task was modified so that the sounds to be discriminated were presented only once per trial, rather than being repeated until the animal made its response (Figure 5.2B). Before each trial, a series of tone pips (5 kHz tone, 20 ms duration, 0.5 ms rise/fall time, 200 ms inter-tone interval) was presented as a "ready signal" to let the ferret know that the central spout could now be triggered to start the trial. When the ferret triggered the central spout, a small water reward was administered from the spout on 10% of trials, chosen at random. On the remaining 90% of the trials, rewards were given only for correct responses to the test sounds. The test sounds consisted of two consecutive click trains, presented from the centre speaker: a 200-ms-long reference click train (5 ms rise/fall time) followed by a 50-ms-long inter-stimulus interval of silence, and then a 500-ms-long target click train (5 ms rise/fall time). The pitch of these stimuli was determined by the rate of clicks within the train. As before, in order to receive a reward the ferret was required to activate the right peripheral spout if the target was higher in pitch than the reference sound, and to activate the left peripheral spout if the target pitch was lower in pitch. Incorrect responses were negatively reinforced with a timeout of 10-12 s, and were followed by correction trials.

5.2.3.5 Training stage 5

In this stage, click train stimuli were then replaced with "artificial vowel" sounds that were created using custom MATLAB code, based on an algorithm adapted from Malcolm Slaney's Auditory Toolbox[1]. The vowel sounds were composed of click trains that were band-pass filtered to add "formants" centred at 430, 2132, 3070 and 4100 Hz (Figure 5.3).

[1] http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/

These formants correspond to the first four formants of the English vowel /i/ (as in "pill"). This switch from unfiltered click trains to artificial vowels did not cause any transient drop in the animals' performance, indicating that the animals were able to generalize the pitch task very rapidly across the two types of complex sounds. Across trials, the amplitudes of the reference and target vowels were varied independently. Each was set to one of ten possible values, spanning a range of approximately 15 dB SPL, chosen at random with a uniform distribution. Sound levels were calibrated using a B&K sound level meter and free-field, 1/4 inch microphone (Type 4191; Brüel and Kjær, Naerum, Denmark). The average sound intensity of the artificial vowels was approximately 80 dB SPL (±3 dB across different pitches). The ±7.5 dB random variation in sound level was introduced to ensure that ferrets were not inadvertently provided with relative level difference cues within particular spectral bands. Once animals had reached criterion on a given session of this task, the set of target pitches used were occasionally jittered (±150 Hz) across additional sessions. By randomizing levels and target pitches I encouraged the animals to follow a general pitch difference rule, rather than mapping absolute sound parameters of the training targets onto the left and right spouts.

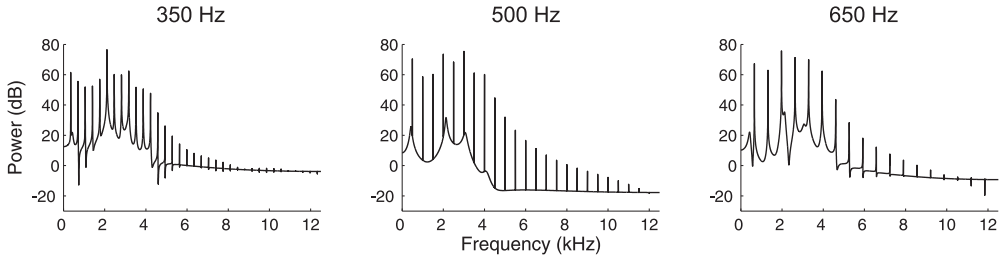


Figure 5.3: Effect of pitch on /i/ spectrum. Power spectra of the artificial vowel stimulus (/i/) used in this experiment (/*i*/), shown for pitches of 350 (left panel), 500 (centre panel) and 650 (right panel) Hz.

5.2.4 Pitch discrimination testing

Once ferrets performed at $\geq 85\%$ correct in at least three consecutive sessions of training stage 5, I switched the ferrets from training to testing. The testing task was very similar to stage 5 of training, except that 30 different target pitches were now typically presented

within a given session, rather than just two target pitches. Within each weekly testing "run", composed of ten individual testing sessions, the reference pitch was held constant. Each run was initiated with a session resembling the final stage of training, whereby only two target sounds with fixed pitches, one octave above and one octave below the reference pitch, were used. Animals were required to perform at $\geq 85\%$ on this task before progressing to the "variable target" condition, and most animals reached this criterion in one or two testing sessions. In the variable target condition, the pitch of the target sound was varied from trial-to-trial across 30 values. Fifteen of these targets were higher in pitch than the reference and 15 were lower, with a sampling density of 12 pitches per octave. Correction trials were still included in testing. To help ensure that the animals were still attending to the pitch of the sounds when the task was very difficult, if the target pitch was within 5/12 octaves of the reference, then the correction trial was presented at the most extreme value in the pitch range in the appropriate direction. Pitch discrimination testing was repeated across sessions using a constant reference pitch for typically 300-600 trials per reference. The animal was then restarted on stage 5 of training using a different reference pitch at the beginning of the next testing run (i.e. the following week). This procedure was repeated for a number of reference pitches between 200 - 1200 Hz.

5.2.5 Data analysis

Correction trials were excluded from the data analysis, as were trials on any testing session in which the animal scored less than 60% correct overall (i.e. across all target pitches). Each animal completed at least 300 trials for each reference pitch, although in many cases the number was closer to 1000. Test trials with the same reference pitch were pooled across all sessions in the analyses.

Plots of the proportion of trials on which the animal responded at the right spout as a function of the log of the target pitch were sigmoidal in shape and approximated a cumulative Gaussian distribution function. Therefore, psychometric curves were estimated from the data by fitting cumulative Gaussian distributions using probit generalized linear models. The pitch difference limens for each reference pitch were calculated from the fitted psychometric curves as half of the distance between the pitch at which the right

spout was chosen on 69.15% of trials, and the pitch at which the right spout was chosen on 30.85% of trials. A threshold of 69.15% was chosen because this level of performance on a 2 alternative discrimination task is equivalent to a *d*' of 1 (Wickens, 2002), making the present results more readily comparable to previous studies of frequency discrimination in non-human animals on go/no-go tasks. Weber fractions were also calculated as the ratio of the pitch difference limen divided by the value of the pitch of the reference sound.

Pearson correlations and ANOVAs are used throughout to test whether comparisons of performance measures were significantly above chance or whether performance differed significantly as a function of parameters such as stimulus type and reference pitch. A significance level (alpha) of 0.05 was used as a criterion for null hypothesis rejection, and where multiple comparisons were carried out across a number of subjects and/or reference values, significance levels were Bonferroni corrected as appropriate.

5.3 Results

5.3.1 Discrimination of the pitch of artificial vowels

All four ferrets reached criterion on the five training stages after 1-2 weeks of training per stage. The ferrets' discrimination of the pitch of artificial vowels was then measured using a number of reference pitches on a 2AFC discrimination task. The discrimination performance of each animal at each reference pitch was visualized by plotting the percentage of "higher" responses (i.e., responses at the right spout) as a function of the pitch of the target vowel. Figure 5.4 plots the performance of one ferret for trials on which the reference pitch was 700 Hz. As expected, the ferret responded correctly more often when the pitch distance between the reference and target vowels was larger. A cumulative Gaussian function was fitted to the raw data in order to obtain a psychometric function of pitch discrimination performance. Threshold estimates were then directly calculated from the psychometric function.

Figure 5.5 shows the psychometric functions obtained from all four ferrets for each reference pitch tested. The pitch acuity of ferrets is indicated by the steepness of the psychometric curves, and can also be expressed as the minimum pitch difference required

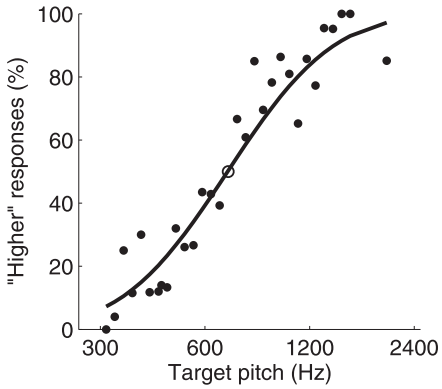


Figure 5.4: Ferret pitch psychometric. Performance of one ferret when the reference vowel had a pitch of 700 Hz. The percentage of right spout choices are plotted as a function of the pitch of the target sound (black dots), and the reference pitch is indicated by an open circle plotted at 50% choice probability. The fitted psychometric curve is also shown (black line).

for the animal to reach a given performance threshold (i.e. the pitch difference limen, here at 69.15% correct). In Figure 5.6A I investigated how the ferrets' pitch difference limen (i.e. the minimum detectable change in pitch) for vowels depends on the reference pitch. Pearson correlations showed that ferret pitch difference limens increased significantly with the pitch of the reference vowel ($r = 0.859$, $p < 0.001$). Weber fractions were derived by normalizing each pitch difference limen by the corresponding reference pitch (Figure 5.6B), and these normalized measures of discrimination acuity showed no overall across the range of reference pitches tested ($r = -0.010$, $p = 0.951$). Nevertheless, Weber fractions did differ significantly depending on reference pitches (1-way ANOVA, $F(12, 25) = 3.98$, $p = 0.002$). Post hoc tests indicated that the Weber fraction measured at a reference of 200 Hz was significantly higher than the Weber fractions for references of 300 - 417 Hz and 700 Hz (Tukey's HSD test, $p < 0.05$).

In Figure 5.5, some of the fitted psychometric curves do not pass through the reference pitch at the 50% choice probability point. These small shifts of the psychometric curve are partly attributable to statistical errors in the fit of the psychometric function, but also likely indicate a "response bias", i.e. when uncertain about the correct response, the ferret did not make "higher" (right) or "lower" (left) responses with exactly equal probability, but exhibited a small preference for either higher or lower responses. I quantified this bias in each psychometric curve as the distance, in Hz, between the 50% right choice probability point of the psychometric function and the reference pitch. In this calculation, higher response biases are represented as negative values and lower biases as positive. In Figure 5.7, the biases are plotted as a percentage of the reference pitch,

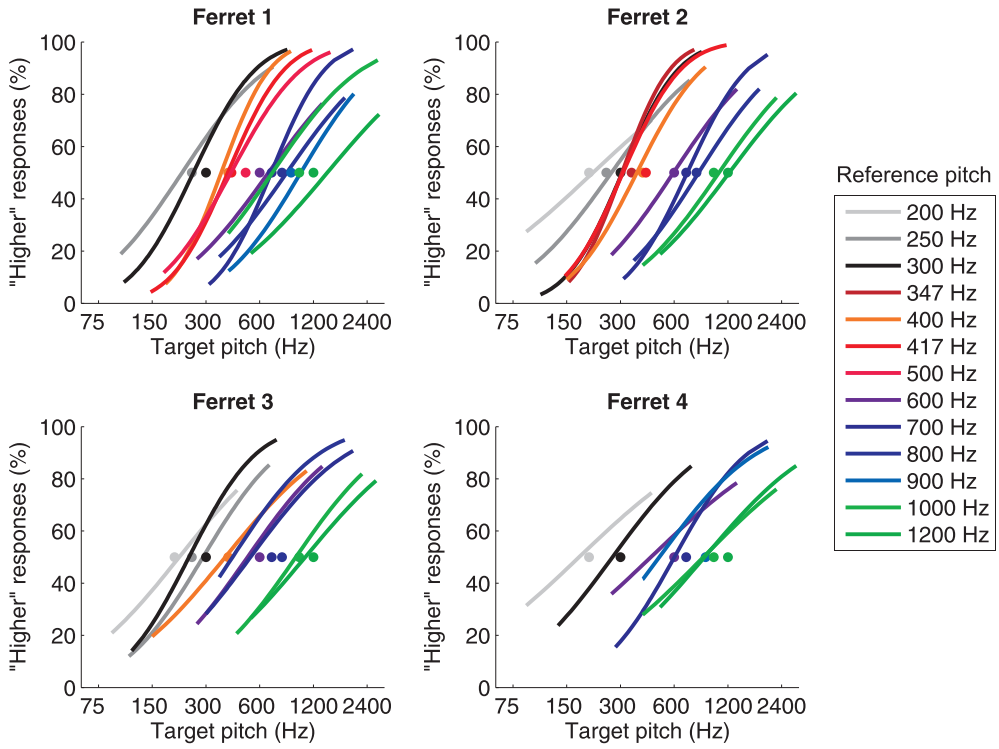


Figure 5.5: Psychometric curves for a complex pitch discrimination task. Psychometric curves describe the performance of four ferrets trained to discriminate the pitch of artificial vowel sounds. Each panel shows the psychometric functions for a single animal, and the colour of each psychometric curve corresponds to the pitch of the reference vowel. The reference pitch of each curve is also indicated by a colour-coded filled circle at 50% choice probability.

and these values are compared across ferrets and across reference pitches. Some animals were more likely to show response biases than others, and most of the significant biases on this task (see Figure 5.7 caption) were towards the right spout (i.e. higher pitch).

To ensure that the ferrets were making a general pitch difference judgment, the levels of the reference and target sounds were independently and randomly varied across trials. To confirm that the intensity of the target sounds did not affect discrimination performance, psychometric functions were fitted independently for target sounds at each intensity level for each reference pitch. A 2-way ANOVA was then carried out on the slopes of the fitted psychometric functions (expressed as percent choice probability per octave), using ferret identity and target intensity as predictor variables. No significant

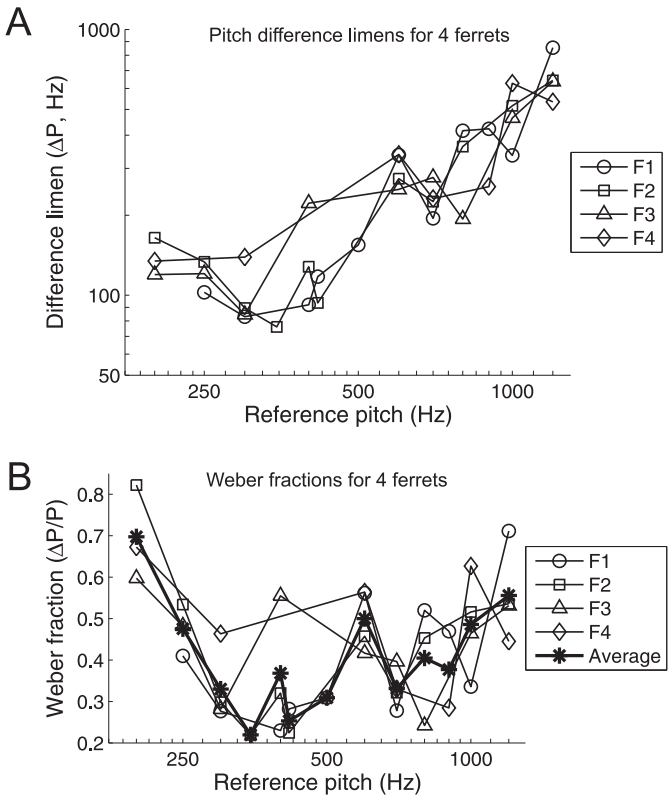


Figure 5.6: Ferrets' acuity in discriminating the pitch of two artificial vowels depends on the pitch of the reference vowel. (A) In the top panel, the pitch difference limens of four ferrets across a range of reference pitches are graphed. Data for each ferret are displayed with a unique symbol and connected with solid lines. (B) Weber fractions for four ferrets across a range of reference pitches, calculated using the frequency difference limens in (A), are graphed in the bottom panel. The average Weber fractions across all ferrets are shown by the asterisks.

effect of target sound intensity ($F(19, 436) = 1.09$, $p = 0.359$) was found, nor did the intensity significantly interact with the performance of individual ferrets ($F(57, 436) = 1.15$, $p = 0.221$). Furthermore, trial-to-trial response probabilities were not predicted by the intensity differences between the reference and target vowel (2-way ANOVA, $F(28, 569) = 1.25$, $p = 0.174$) or the interaction between this intensity difference and ferret identity ($F(84, 569) = 1.23$, $p = 0.090$). These results show that across the 15 dB range tested, the ferrets' performance on the pitch discrimination task did not depend on sound intensity. Moreover, because sound levels varied randomly from trial to trial and between reference and target sound, the ferrets could not have used any inadvertent

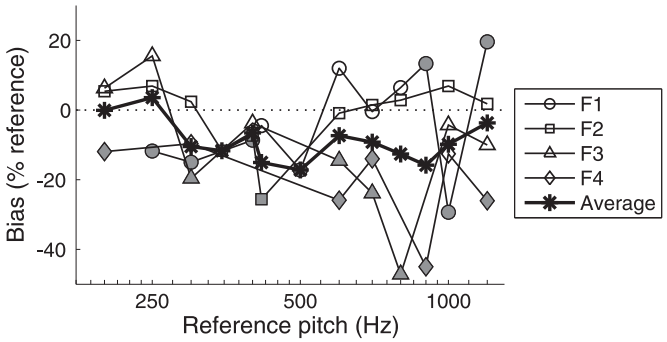


Figure 5.7: Bias of four ferrets on the vowel pitch discrimination task. Biases were calculated as the distances, in Hertz, between the 50% right choice probability obtained from the fitted psychometric function and the corresponding reference pitch, so that positive values indicate a bias towards the left ("lower") spout and negative values indicate a bias towards the right ("higher") spout. These biases are plotted here as a percentage of the reference pitch. Data for different animals are plotted with different symbols, and bias values which are significantly different from zero (99.87% confidence intervals; 0.05 alpha with Bonferroni correction) are shown in grey, non-significant values in white. The average percentage of bias across all four animals is plotted with asterisks. A dotted line is shown at a bias of zero, for reference.

intensity differences that might have remained between stimuli.

5.3.2 Discrimination of the pitch of tones

Once trained to discriminate the pitch of artificial vowels on the 2AFC task, the same four ferrets were retested using pure tones instead of artificial vowels, but on an otherwise identical paradigm. On each trial, a reference tone was presented (200 ms duration, 20 ms rise/fall time), followed by a silent inter-stimulus interval (50 ms) and a target tone (500 ms duration, 20 ms rise/fall time). Performance was again measured around a range of reference frequencies between 200 and 1200 Hz. For each new reference tone, the animals were trained to criterion performance using two fixed targets before more variable targets were introduced, as described in Section 5.2.4. All ferrets generalized from the vowel to tone discrimination task well, and reached criterion on the latter within the first 2 weeks of training.

Ferrets' psychometric curves for the tone version of the pitch discrimination task are shown in Figure 5.8, while Figure 5.9A shows the difference limens calculated from these psychometric curves, together with the artificial vowel pitch difference limens to facilitate comparison. The discrimination performances (i.e., Weber fractions) obtained with the vowel and tone versions of this task were clearly similar, and these values were not statistically different across reference pitches (2-way ANOVA, $F(1, 43) = 1.36$, $p = 0.250$). Just as with the vowel data, the pure tone difference limens again scaled with reference pitch ($r = 0.882$, $p < 0.001$), but when expressed as Weber fractions, discrimination performance did not show a systematic increase or decrease across the range of reference frequencies tested ($r = -0.042$, $p = 0.866$). Small but significant response biases were again observed in some cases, as shown in Figure 5.9B, and they again tended to favour "higher" (right) responses.

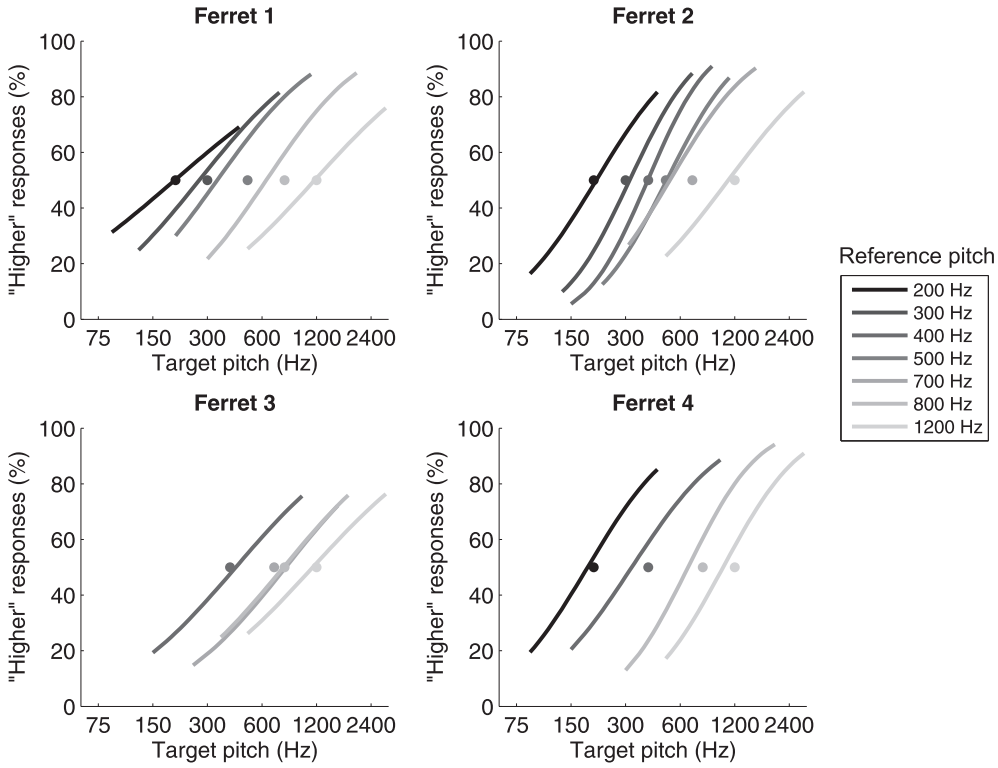


Figure 5.8: Psychometric curves for a simple pitch (i.e., frequency) discrimination task. Fitted psychometric curves show the performance of four ferrets on a tone frequency discrimination task. The reference frequency for each curve is coded in greyscale, and the value of this frequency is also indicated by the position of the filled greyscale circles along the x-axis.

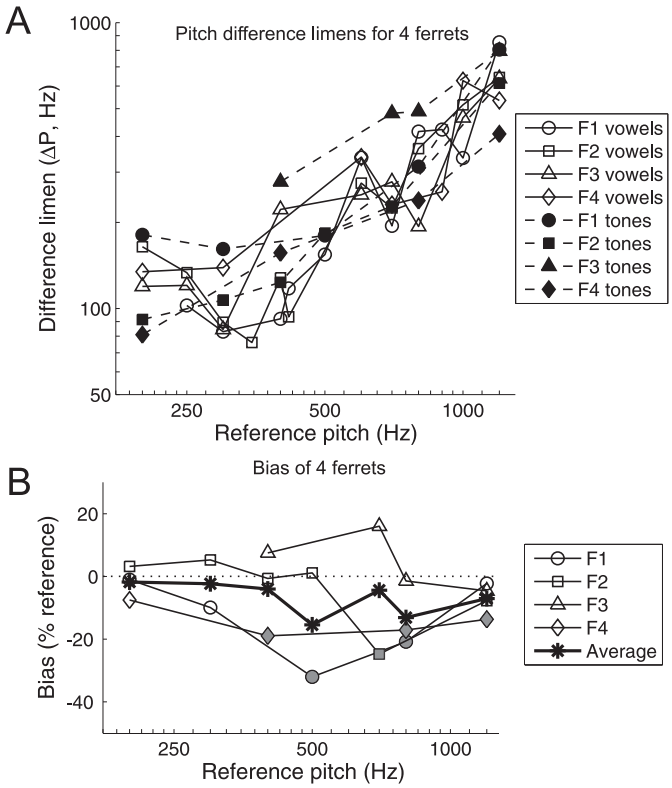


Figure 5.9: Performance of four ferrets on a 2AFC tone discrimination task. (A) Pitch difference limens of four ferrets across a range of reference pitches, measured with two stimulus types, are shown in the top panel. Data for each ferret are displayed with a unique symbol, as shown in the legend. Difference limens were measured on a common pitch discrimination task using artificial vowel sounds (open symbols, solid lines) and a pure tones (filled symbols, dashed lines). (B) The bottom panel shows the bias of four ferrets on the tone discrimination task. Bias calculations are as described for Figure 5.7. The percentage biases (normalized by the reference pitch) are plotted as a function of reference pitch. Data for different animals are plotted with different symbols, with statistically significant (99.87% confidence intervals; 0.05 alpha with Bonferroni correction) values shown in grey, and the average bias across animals plotted with asterisks. A dotted line is shown at a bias of zero, for reference.

5.3.3 Pitch discrimination performance of human subjects

Five adult humans (two male, ages 24 - 40 years) were tested on a very similar pitch discrimination task, using both vowels and tones as stimuli. Human psychophysical procedures were carried out under the guidelines of the Ethical Review Committee of the University of Oxford. I attempted to make the pitch discrimination task performed by human subjects as similar as possible to the task performed by the ferrets. The stimuli were presented from the same central speaker of the testing chamber in which ferrets were trained, and the sounds were presented with the same random intensity attenuations across trials as in the animals' task. The human subjects, being too large to fit in the ferret testing chamber, listened to the sounds through the opened lid, and initiated trials and responded by pressing keys on a keyboard positioned near the chamber. Upon initiation of each trial, reference and target sounds identical to those used with the ferrets were presented. However, as humans performed considerably better at these pitch discrimination tasks, the range of pitches sampled around each reference was adjusted based on pilot data (not shown), and was occasionally readjusted for individual subjects, based on their performance on the first 50-100 trials. Instead of water rewards, subjects received feedback on a computer monitor after each trial, and incorrect choices resulted in a time-out of 2-4 seconds. Each subject completed 250-350 trials on each of four reference pitches, in both artificial vowel and tone versions of the discrimination task. The reference pitch was constant within any one testing run, tone and vowel trials were presented in blocks, and the order of reference pitches presented was pseudo-randomized across subjects. Only one subject (H5) was musically trained, and none were given extensive training on the task prior to testing.

In Figure 5.10, the pitch difference limens of human subjects are plotted together with the difference limens of ferrets. The pitch acuity of humans was clearly and significantly better than that of ferrets on these discrimination tasks (3-way ANOVA, $F(1, 82) = 840.86$, $p < 0.001$). As in the ferret data, the performance of the human listeners in the two discrimination tasks did not differ significantly (2-way ANOVA, $F(1, 35) = 1.81$, $p = 0.188$), and the pitch difference limens of humans scaled with the reference pitch in both the vowel ($r = 0.738$, $p < 0.001$) and tone ($r = 0.780$, $p < 0.001$) versions of the task, while Weber fractions did not change systematically across the range of reference

frequencies tested ($r = -0.245$, $p = 0.297$ for vowels; $r = -0.044$, $p = 0.855$ for tones). A small but statistically significant bias was observed for 10 of the 40 subject-by-reference-pitch conditions ($p < 0.05/20$, data not shown). In the majority (8 of 10) of these cases, the bias was again towards "higher" responses.

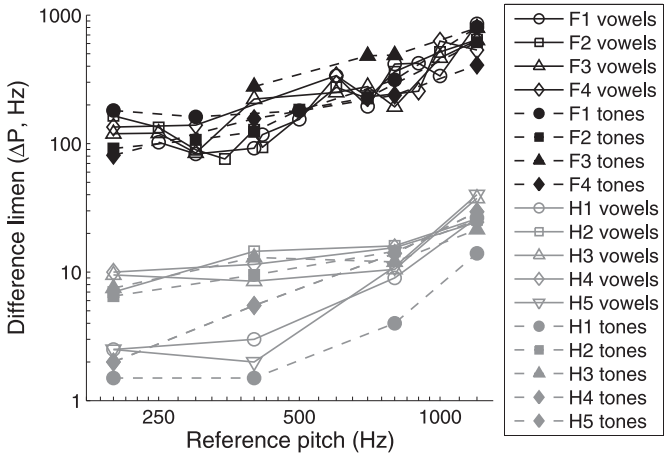


Figure 5.10: Pitch difference limens of four ferrets and five human subjects, measured using two stimulus types. Data for each ferret are displayed with black symbols connected by black lines, and human data are displayed with grey symbols connected by grey lines. Difference limens for each subject were measured using artificial vowels (open symbols, solid lines) and tones (filled symbols, dashed lines).

5.4 Discussion

In natural environments, pitch often correlates with the size of vibrating objects. For example, the fundamental frequency of a mechanical oscillator like a tuning fork is inversely proportional to its mass, and the fundamental frequency of resonant cavities similarly decreases with increasing volume. Consequently, the ability to order the pitches of complex sounds along an ordinal scale can be useful, as it provides information about the physical properties of the sound sources. Pitch ordering also plays an important role in vocal communication, as it often carries information, not just about the gender and size, but also the emotional state of the vocalizing individual. For example, excited individuals tend to have greater tension in the vocal folds, and therefore have higher pitched voices

than relaxed speakers. The ability to perceive the pitch of periodic, complex sounds along a continuous scale, from low to high, might facilitate estimating continuous valued properties of the sound source (heavy or light, large or small, relaxed or tense, empty or full). This would be a useful faculty for many species, but although there have been a number of previous investigations into pitch and frequency discrimination in mammals, none so far have asked how well animals can judge the direction of a pitch change in complex sounds. Here, I show that ferrets can distinguish increases from decreases in the pitch of artificial vowel sounds and that their pitch discrimination performance for this class of stimuli closely matches that in an equivalent pure tone frequency discrimination task, even though the neural processing necessary to carry out these two perceptual judgments may be quite different.

5.4.1 Pitch direction judgments in animals

The few previous studies that have examined animals' ability to judge the direction of pitch changes have all used simple tones as stimuli. These studies have often been motivated by the question of whether animals can identify the direction of pitch changes by comparing the "relative" pitches of several tones presented on a given trial, or, instead, by comparing the "absolute" pitch of each tone to an internal representation. For instance, Cynx and colleagues carried out psychophysical tasks in birds that required the subjects to respond to increases or decreases in frequency across sequences of tones (Cynx, 1995; Page *et al.*, 1989). While they showed that it is possible to train birds to make relative pitch judgments under carefully designed experimental conditions, these relative judgments do not appear to come easily. Their results suggest that both songbirds and non-songbirds prefer to use absolute pitch cues to solve frequency discrimination tasks. Other groups have also experienced difficulty in training rats (D'Amato, 1988) and non-human primates (Brosch *et al.*, 2004; D'Amato, 1988; Izumi, 2001) to respond to the direction of relative frequency change in tone sequences. The design of the present task encouraged subjects to respond to the relative change in pitch between the reference and target sound, which is the strategy that human listeners report using. However, since the reference sound remained constant throughout a given session, the ferrets may also have performed the task by responding to the absolute pitch of the target sound; that is,

by indicating whether the target is higher or lower in pitch than some internal reference. Nevertheless, whether ferrets are comparing the target sounds to the reference presented on each trial or an internal representation of a pitch scale, they are clearly ordering the pitches along a linear progression that correlates with the periodicity of the sound.

5.4.2 Frequency acuity of animals

In many ways, the present results indicate interesting parallels between humans and ferrets. For both species difference limens scaled with reference pitch in very similar ways and thresholds in the pure tone and the artificial vowel tasks were very similar. However, humans performed consistently and substantially better than ferrets. This was not entirely surprising, given that previous studies have already well established that the pure tone frequency discrimination performance of humans is superior to that of many other mammals across the reference pitch range tested here (Fay, 1988b). It has been proposed that the superior performance of humans in these tasks might be due to differences in basilar membrane mechanics or higher densities of ganglion cells in the human cochlea (Elliott *et al.*, 1960). While humans' improved sensitivity to low-frequency pure tones may facilitate pitch discrimination, ferrets are easily able to detect 200 - 1200 Hz tones presented at the levels used in the present study (Kelly *et al.*, 1986b). Furthermore, the fundamental frequencies of ferrets' vocalizations are within this frequency range (see Section 1.2), so it is reasonable to expect that they can perceive the pitches of artificial vowels.

Of course, when comparing behavioural results across different studies and species, one must be careful to consider not just sensory factors, but also cognitive ones. In particular, the design of the present pitch discrimination task required animals to report the direction of pitch change correctly, while in most other previous animal studies, the subjects were required only to report any detectable pitch change, regardless of whether they were able to tell in which direction the pitch had changed. The present design presents a more demanding task, and one might expect the animals' performance to be correspondingly lower. The pure tone Weber fractions of ferrets in this study were indeed considerably higher than those previously reported for other species (reviewed by Heffner *et al.*, 1971; Fay, 1988a,b; and Shofner, 2005). For example, on frequency discrimination

tasks that use a 500-Hz reference tone, Weber fractions have been measured to be 3.4% in chinchillas (Nelson and Kiester, 1978), 1.6% in guinea pigs (Heffner *et al.*, 1971), 1.4% in bushbabies (Heffner *et al.*, 1969b), 2.5% in treeshrews (Heffner *et al.*, 1969a), 3.6% in budgerigars (Dooling and Saunders, 1975) and 1.7% in cats (Elliott *et al.*, 1960). In comparison, the average Weber fraction measured for ferrets in my direction judgment task at this reference frequency was about tenfold greater, at 36.4%. Even across studies that use the same species and very similar tasks, Weber fractions for pitch discrimination can vary widely. The average Weber fractions of chinchillas for a 250-Hz reference tone were measured on go/no-go tasks to be 4.6% by Nelson and Kiester (1978) and 21.2% by Shofner (2000). Large differences in pure tone frequency discrimination performance are also sometimes present across individual members of the same species within the same study. For instance, in one study, a group of 9 male macaque monkeys tested with a 1000-Hz reference tone had individual Weber fractions that ranged from approximately 0.9 to 9.0% (Prosen *et al.*, 1990).

Even though the inherent variation in performance on very similar behavioural tasks can be considerable and is well documented, it is unlikely that this can account for the relatively elevated Weber fractions reported here for ferrets, which are instead most likely attributable to the considerably higher cognitive and/or sensory demands posed by a pitch change direction judgment relative to a mere pitch change detection. This would be consistent with the results of a study of pitch discrimination in children with cochlear implants, which found that these individuals showed more acute pitch discrimination for complex sounds on a pitch change detection task than on a task that required subjects to make pitch direction judgments (Vongpaisal *et al.*, 2006). The authors of this study suggest that while subjects may have used spectral cues associated with pitch changes to solve the former task, they are unable to order the same sounds (synthesized piano notes) along an ordered pitch scale on the latter task.

A 2AFC discrimination task involves a number of additional cognitive demands compared to a go/no-go task. In most go/no-go tasks, the animal simply has to make a particular response when it detects a stimulus change and refrain from making this response in the absence of a stimulus change. In a 2AFC task, the animal has to categorize changes in the signal as belonging to two different classes and map these onto two different

types of responses, which it has to learn. But there is a second, less intuitive, difficulty in carrying out 2AFC frequency discrimination tasks in animals. Previous studies have suggested that when performing go/no-go tasks animals tend to make response choices based on the "quality" of sounds (such as frequency or timbre), while in 2AFC tasks animals prefer to respond to the spatial location of the sound source (reviewed by Burdick, 1979). The reason for this task specificity, which appears to be much more pronounced in the auditory than the visual modality, is still unclear, but sound quality discriminations have been shown to be more difficult to train on 2AFC tasks than in go/no-go alternatives in a number of species, including guinea pigs (Upton, 1929), chinchillas (Burdick, 1980), cats (Elliott *et al.*, 1962), and monkeys (Elliott *et al.*, 1971). Of particular interest to the present discussion is a frequency discrimination study by Dobrzecka and Konorski (1968), described in Burdick (1979). They report that dogs more accurately discriminated the frequencies of pairs of tones in a go/no-go task than in a two-choice procedure (but see Neill and Harrison, 1987). This would result in higher frequency difference limens on the 2AFC version of the task than the corresponding go/no-go version.

Another difference between the present task design and that of previous studies is the number of stimuli presented on each trial. In my task, as in most human studies, the ferrets heard only one instance of the reference and target before making their response choice on each trial. In the typical go/no-go frequency discrimination task, a sequence of reference tones is presented and the signal for detection is a frequency alteration in 2 or more of the tones near the end of this sequence. Therefore, in these studies, the animals are provided with "multiple looks" at the tones to be discriminated in any one trial, whereas in my task a sensory decision is made after hearing the stimulus pair only once.

Comparisons of the present results with previous studies suggest that a traditional go/no-go discrimination task might be more useful for determining the limits of frequency acuity in animals. However, my design was necessary to investigate animals' ability to respond to the pitch of complex sounds along a low-to-high continuum - a question that has not been addressed in previous investigations. Ferrets have been trained on a number of sound localization (Bizley *et al.*, 2007b; Kacelnik *et al.*, 2006; Kelly and Kavanagh, 1994; Nodal *et al.*, 2008; Parsons *et al.*, 1999) and detection tasks (Fritz

et al., 2007b, 2003; Hine *et al.*, 1994; Kelly *et al.*, 1986b, 1996) in the past, but rarely in discrimination tasks. In a series of studies by Shamma and colleagues, ferrets were trained on an aversively reinforced, go/no-go task to discriminate between tones of two fixed frequencies (Fritz *et al.*, 2005a,b). These experiments have demonstrated that ferrets can make discriminations based on simple spectral cues, and that the response properties of primary auditory cortical neurons undergo systematic changes during this behavior. However, the present study is the first to measure the frequency acuity of ferrets. From a comparative psychophysics point of view, it will be interesting for future studies to measure the frequency difference limens of ferrets on a more traditional go/no-go, change detection task, in order to determine if and how these results might differ from those of the current 2AFC results.

5.4.3 Pitch discrimination for complex sounds

I observed that, for both ferrets and humans, thresholds obtained with the artificial vowel version of the task were not significantly different from those obtained with pure tone stimuli. The pitch acuity of humans and other species has previously been tested with a variety of stimulus types (reviewed by Shofner, 2005). Some of these studies found pitch discrimination for complex periodic sounds to be better than pure tone frequency discrimination, while others found it to be worse, differences that may be attributable to stimulus class. In chinchillas (Shofner, 2000) and humans (Henning and Grosberg, 1968; Moore *et al.*, 1984), discrimination of the pitch of harmonic tone complexes is reported to be more acute than that of pure tones. In contrast, discrimination of the repetition rate of iterated rippled noise stimuli is poorer than tone frequency discrimination in humans (Yost, 1978) and chinchillas (Shofner *et al.*, 2007), while these stimulus types yield similar pitch thresholds in goldfish (Fay *et al.*, 1983). Finally, discrimination of the modulation frequencies of sinusoidally amplitude modulated noise bursts is poorer than the discrimination of pure tone frequencies in macaque monkeys (Moody, 1994), chinchillas (Long and Clark, 1984), parakeets (Dooling and Searcy, 1981) and humans (Formby, 1985). Previous studies have measured pitch discrimination in human subjects using artificial vowel sounds (Flanagan and Saslow, 1958; Klatt, 1973; Smith *et al.*, 2006). The average Weber fractions reported by previous authors for the artificial vowel /i/

(reference frequency = 120 Hz) was 0.23% - 0.40%, which was only slightly better than the average Weber fraction of 0.44% measured for pure tones by Flanagan and Saslow (1958). This is broadly consistent with the results of the present study, in which both ferret and human subjects showed no statistically significant difference between pitch difference limens measured using artificial vowels and tones.

The human Weber fractions reported here are larger than those previously reported (Flanagan and Saslow, 1958; Klatt, 1973; Wier *et al.*, 1977). This is most likely due to the much smaller number of training and testing trials used in the present study. Human subjects in this experiment carried out approximately 1200 pitch discrimination trials, while listeners in Flanagan and Saslow's (1958) study were highly practiced, having carried out approximately 14,000 trials. Both human and animal frequency difference limens are known to improve with training beyond the first few hundred trials (Banai and Ahissar, 2004; Prosen *et al.*, 1990). Another potentially important difference between my study and previous investigations of human pitch discrimination is that my sounds were presented via a free-field speaker, whereas others have presented stimuli over headphones, which offer superior sound isolation. While it is therefore likely that I would have obtained lower human difference limens had I used highly trained subjects performing a headphone task, results obtained in that way would also be less representative of the capability of 'average' human listeners functioning in real world acoustic environments. Also, in this study the authors were primarily interested in measuring performance for two stimulus types across a range of reference pitches, in order to compare the resulting thresholds with those obtained in essentially the same way from the ferrets.

For both ferret and human listeners, pitch difference limens were poorer for reference sounds with higher periodicities - an effect that has also been demonstrated in previous studies of tone frequency discrimination in humans and non-human animals (Fay, 1988a). Figure 5.10 shows that the slopes of the difference limens across reference pitch were quite shallow for reference values between 200 - 500 Hz, but had a steeper positive slope thereafter. This effect is manifest in both the human and ferret data, though it is more pronounced in the latter. A similar trend across frequency difference limens has been observed in past studies of pure-tone frequency discrimination in humans (Moore, 1973; Rosenblith and Stevens, 1953; Wier *et al.*, 1977). While few non-human frequency

discrimination studies have sampled reference frequencies below 500 Hz, those that have show comparable trends in the frequency difference limens of pigeons (Sinnott *et al.*, 1980), bushbabies (Heffner *et al.*, 1969b), chinchillas (Nelson and Kiester, 1978) and cats (Elliott *et al.*, 1960). The physiological mechanisms underlying pitch discrimination remain poorly understood, but these similarities suggest that similar mechanisms may be at work across a wide variety of species.

5.4.4 Conclusions

This study shows that ferrets can make pitch direction judgments, and that their performance on this discrimination task is dependent on the pitch of the reference sound. The relationship between discrimination performance and reference periodicity is similar when measured with tones, where the periodicity information is mapped throughout the auditory system, and with complex sounds, where an anatomical periodicity map may not exist. In order to fully understand the significance of these similarities between simple frequency discrimination and periodicity discrimination, investigations of the neurophysiological processes that give rise to pitch perception will be essential. In the field of vision and somatosensory research, "neurometric" studies that have combined measurements of neural responses and behavior while animals perform a 2AFC discrimination task have successfully addressed these types of questions (de Lafuente and Romo, 2006; Liu and Newsome, 2005; Parker and Newsome, 1998). In order to carry out similar methodologies to study auditory perception, we require an animal model, like the ferret, that is suited to both psychophysical tasks and electrophysiological recordings. Neurophysiological recordings of responses to complex sounds that are carried out in ferrets as they perform pitch discrimination tasks are likely to be particularly useful in demonstrating the relationship between brain activity and pitch perception.

5.5 Acknowledgements

This work was carried out in partnership with Dr. Jennifer Bizley. Sheelah Hart-Schnupp and a host of undergraduate dissertation students helped run the ferret behaviour over the course of three years. Prof. Andrew King and Dr. Jan Schnupp supervised the

study. A manuscript of this chapter has been accepted for publication in the Journal of Acoustical Society of America.

Chapter 6

Behavioural and Physiological Measures of Ferrets' Pitch Judgments

6.1 Introduction

Although the timbre of a sound is defined to be the quality used to distinguish sound sources, the periodicity, and thus perceived pitch, of a vibrating object often scales with physical properties of that object, such as its' size or the density of its' cavity. This correlation makes pitch discrimination an additionally important tool for animals that need to identify auditory objects in their natural environment. For instance, animals can use pitch perception to determine the fullness of a coconut based on the sound it makes when it is tapped. Furthermore, pitch perception is essential to speech, music, and the complex communication behaviours of non-human animals (Capranica, 1966; Koda and Masataka, 2002; Kojima *et al.*, 2003; Nelson, 1989). Yet although processing the pitch of natural, periodic sounds is almost certainly a key function of the mammalian auditory system, it is unclear how the brain carries out this task.

An important development in the area of pitch processing was provided in a recent study by Bendor and Wang (2005). These authors identified a "pitch centre" in the auditory cortex of the marmoset, near the border of the low-frequency regions of area A1

and R. A collection of neurons in this small region of cortex were selective for a particular periodicity of sound, whether that periodicity was present as the frequency of a pure tone or the missing fundamental of a tone complex. These neurons might be described as having "pitch-selective" responses. This finding fits nicely with the results of previously existing functional magnetic resonance imaging (Griffiths, 2003; Griffiths *et al.*, 1998; Hall *et al.*, 2006; Patterson *et al.*, 2002; Penagos *et al.*, 2004) and electromagnetic (Gutschalk *et al.*, 2002; Krumbholz *et al.*, 2003) studies of the human brain, which have repeatedly shown that the activation of regions of Heschl's gyrus is modulated by the periodicity of sounds. These studies suggest that there is an anatomically and functionally distinct collection of neurons in primate auditory cortex that is selectively tuned to pitch cues and is invariant to other spectral and binaural properties of sounds. However, the evidence for a single, dedicated pitch centre in auditory cortex is not unambiguous. A recent study by Hall and colleagues has demonstrated that the pattern of activation elicited by two stimuli which have the same pitch can be non-overlapping (Hall and Plack, 2007). Nevertheless, if a specialized pitch centre does exist in auditory cortex, it would be reasonable to expect that neurons in this area might underlie pitch discrimination decisions, with neurons in other cortical areas performing other perceptual functions.

However, rather than looking for neurons that are selectively activated by the pitch of periodic sounds, another approach to studying the neurophysiological representations of pitch is to present sounds of different pitches, (which may also differ along other stimulus dimensions), and to then ask which neurons are modulated by the stimulus periodicity. That is, rather than looking for pitch-selective neurons, as in Bendor and Wang's (2005) study, one might look for neurons that are sensitive to pitch. I carried out such an experiment in anaesthetised ferrets, and showed that the low-frequency regions of primary auditory cortex and the posterior fields tended to contain neurons that were more sensitive, on average, to the pitch of artificial vowel sounds than neurons on more anterior cortical areas (Chapter 4). However, I also located neurons that were significantly sensitive to pitch across all 5 core and belt areas of ferret auditory cortex studied. Rather than showing selectivity for the pitch of periodic sounds, the responses of these distributed "pitch-sensitive" neurons were often also modulated by the timbre and azimuth of complex sounds. This leads to the question of functionality across areas. That

is, are neurons within a pitch-selective region necessary to explain animals' perception of pitch, or does the activity of pitch-sensitive neurons distributed throughout auditory cortex provide the information necessary for animals' to carry out pitch discrimination tasks?

To address this question, I use the approach of "neurometric" analysis. Newsome and colleges (Liu and Newsome, 2005; Shadlen and Newsome, 2001) have previously been demonstrated the power of this technique in investigating the neural correlates of motion perception in the visual system. Similarly, neurometric analyses have been central to Romo and colleagues' (Hernandez *et al.*, 2000; Luna *et al.*, 2005) elegant research which identifies the neural underpinnings of somatosensory discrimination. The approach relates neural activity to perception by comparing changes in firing rate within sensory neurons to perceptual decisions made by animals on 2-alternative forced choice (2AFC) behavioral experiments using the same stimuli. The rationale is that if the neural activity of interest provides the 'psychophysical signal' on which the animals' sensory discriminations are based, then the performance of an ideal observer of this neural activity should perform in a way that is analogous to the behaving animal. That is, the shape of the "neurometric curve" of the ideal observer's discrimination performance across a stimulus dimension should match the psychometric curve of the animal on a similar discrimination task. If this is shown to be the case, then the spiking activity of this neuron can be concluded to be a possible neural underpinning of the animals' perception of this stimulus dimension. While this correlation in discriminability does not, of course, demonstrate unequivocally that this neural activity directly *causes* the percept of interest, a lack of correlation in this type of analysis would be strong evidence against this neural activity directly giving rise to that percept.

In the present study, I aim to determine whether the activity of pitch-sensitive neurons across 5 areas of ferret auditory cortex is sufficient to account for ferrets' performance on pitch discrimination task. Ferrets were trained on a positively reinforced, pitch increase/decrease discrimination task using artificial vowel sounds as stimuli (as previously reported in Chapter 5). The same vowel sounds were then presented across a range of pitches to anaesthetised and passively listening ferrets, and neural responses were extracellularly recorded across 5 auditory cortical areas. Neurometric analyses of individual

neural units and populations of neurons were compared to behavioural results across the 5 fields, to determine which may provide a neural code for pitch perception in this species.

6.2 Methods

6.2.1 Stimuli

Sounds were generated using system 3 hardware from TDT and custom software written in MATLAB. The sound stimuli were presented to anaesthetised animals in an anechoic room through customized earphone drivers, and to unanaesthetised animals in the same anechoic room via a free-field speaker, as described in Section 2.4.

Artificial vowel stimuli were created using a adapted version of Malcolm Slaney's Auditory Toolbox[1]. In the physiological experiments, click trains (100 ms in duration) with a repetition rate corresponding to the desired fundamental frequency were passed through a cascade of four bandpass filters centred on 460, 1105, 2735 and 4052 Hz to impart spectral peaks corresponding to the formant frequencies of the vowel /i/. These artificial vowel stimuli were root-mean-square normalised to ensure that changes in pitch did not influence the overall sound pressure level. The artificial vowel /i/ was typically presented at 17 periodicities, or pitches, spaced approximately 1/5 octaves apart, across the range of 200 to 1838 Hz. The pitch range spanned 175 to 1805 Hz, in 1/5 octave steps, for one animal. In any given electrode penetration, the vowels were presented at the same intensity or set of intensities across all pitches ($\pm$ 3 dB). However, the intensity of the vowels presented in each recording session could range from 70-90 dB, depending on the response properties of the neurons in that electrode penetration. Vowels were presented at a rate of once per second, and the responses of cortical units were recorded for each sweep throughout a 1000 ms period beginning at the onset of the vowel. The artificial /i/ vowels presented in the behavioural experiment were highly similar to those of the physiological experiment. On each trial of the behavioural task, the ferret was presented with two vowels: a 200-ms-long "reference" /i/ (5 ms rise/fall time), followed by a 50-ms-long inter-stimulus interval of silence, and then a 500-ms-long "target" /i/

[1] http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/

(5 ms rise/fall time). The bandpass filters were centred at 430, 2132, 3070 and 4100 Hz, and the average sound intensity of the vowels was approximately 80 dB (± 3 dB across different pitches). The amplitudes of the reference and target vowels were varied independently from trial-to-trial across a 15 dB range to ensure that ferrets were not inadvertently provided with relative level difference cues.

Broadband noises (100 ms duration) and pure tones (100 ms duration; 200 - 24000 Hz) were also presented during the recording sessions. The responses to these sounds were used to identify neural units that were driven by sound, to classify their characteristic frequencies, and to determine the cortical area to which each neural unit belonged (Section 2.5).

6.2.2 Behavioural testing

Four adult pigmented ferrets were trained to discriminate the direction of pitch changes in pairs of artificial vowel sounds on a positively reinforced, psychophysical task. Methodological details of the testing and training involved in this task are described in detail elsewhere (Section 5) and so are only briefly discussed here. All experimental procedures were approved by the local ethical review committee and were carried out under license from the UK Home Office, in accordance with the Animals (Scientific Procedures) Act 1986.

The training apparatus for the psychophysical task (recall Figure 5.1) contained three stainless steel spouts that ferrets could lick to make behavioural responses and from which water rewards could be delivered. Prior to each trial, a repeating tone pip (5 kHz tone, 20 ms duration, 0.5 ms rise/fall time, 200 ms inter-tone interval) was presented as a "ready signal" to indicate to the ferret that the central "start" spout could now be triggered to start a new trial. When the ferret triggered the central spout, two vowel sounds were presented (as described in Section 6.2.1). If the target vowel was higher in pitch than the preceding reference vowel, the ferret was required to respond at the right peripheral spout to receive a water reward. If the target was lower in pitch than the reference vowel, responses at the left spout were rewarded. Responses at the incorrect peripheral spout resulted in a 500-ms broadband noise and a 10-12 sec time-out. Within a training session, the pitch of the reference was fixed, but the target pitch varied, so that psychophysical

curves of pitch discrimination could be determined.

6.2.3 Electrophysiological recordings

Extracellular recordings of the responses of neural units in the auditory cortices of six adult, anaesthetised ferrets were carried out according to the procedures outlined in Section 2.2. Neural units were carried out in five cortical regions (Bizley *et al.*, 2005; Kowalski *et al.*, 1995; Nelken *et al.*, 2004): primary auditory cortex (A1), the anterior auditory field (AAF), the posterior pseudosylvian and suprasylvian fields (PPF and PSF), and the anterior dorsal field (ADF) (recall Figure 1.3). In five out of six animals, recordings were made in all five cortical fields, and in the remaining animal responses were recorded in three fields. Extracellular recordings were also carried out in the primary auditory cortex of an adult, unanaesthetised, head-fixed ferret, as described in Section 2.3. Single neurons and multi-unit clusters were isolated in BrainWare (as described in Section 2.6) and all further data analyses were carried out in MATLAB.

Anaesthetised recordings were made with silicon Michigan probe electrodes. In three animals, electrodes with an 8 x 4 configuration (eight active sites on four parallel probes, with a vertical spacing of 150 μm) were used. In a small number of recordings in one of these animals, and in another experimental animal, electrodes with a 16 x 2 configuration (16 active sites spaced at 100 μm intervals on each of two probes) were used. In the final two animals, electrodes with 4 x 4 and 16 x 1 configurations were used (100-150 μm spacing of active sites on each probe). The electrodes were positioned so that they entered the cortex approximately orthogonal to the surface of the ectosylvian gyrus. A photographic record was made of each electrode penetration to allow later reconstruction of the location of each recording site relative to anatomical landmarks (i.e. surface blood vessels and sulcal patterns), to construct maps of auditory cortical activity.

6.2.4 Neurometric analysis of individual neural units

6.2.4.1 ROC method

I used neurometric analysis to examine whether the spike rates of neural units could support ferrets' pitch direction discrimination. Neurometric analyses approach this type of question by calculating how well an ideal observer of the neural activity in question

would perform on a relevant discrimination task - here, a pitch discrimination task. My analysis assumed a rate code for pitch, so each unit's response to a single presentation "sweep" of a vowel could be represented as a scaler - the sum of action potentials across a 75-ms-long response window. If a neuron responded reliably with a consistent spike rate to repeated presentations of a stimulus pitch, and also responded differentially across vowels of different pitches, then the neurometric algorithm would perform well on the pitch discrimination task.

The neurometric procedure is best explained using an example. The responses of one neural unit to vowels of different pitches are summarized in the histograms in Figure 6.1A. The top left panel shows two histograms of the neural unit's spike rate across repeated presentations of the vowel stimulus. The connected open circles show the spike counts when the artificial vowel when presented at a pitch of 919 Hz, and the connected closed circles show the responses when the pitch was 460 Hz. This plot shows that the neural unit generally fires more spikes in response to the lower-pitched vowel. In the four panels below, the histogram of responses to the 919 Hz vowel (open circles) are compared to vowels of increasing pitch (1/5 octave steps), from 528 to 801 Hz (closed circles). The pitch to which 919 Hz is compared is displayed indicated in the legends. Similarly, the five panels on the right side of Figure 6.1A display histograms of responses to vowels with pitches above 919 Hz, in 1/5 octave steps from top to bottom. The neural unit fires less spikes for vowels with pitches above 919 Hz. However, note that there is "noise" in the spike responses, so the overall correlation between spike rate and vowel pitch does not perfectly predict the spike count on each sweep.

I aimed to design a neurometric algorithm that was analogous to the pitch discrimination task on which I had previously trained ferrets. On each trial, the "observer" was presented with two of the units' spiking responses to artificial vowel sounds, and used these responses to decide if the second, "target", sound was higher or lower in pitch than the first, "reference", sound. For instance, interpreting the output of the low-pass neuron in Figure 6.1A, an observer might decide that the target was higher in pitch than a 919 Hz reference when it evoked less spikes than the response to the reference vowel. If both sounds resulted in the same number of spikes, the observer would be forced to make a guess, with a 50% probability of a "higher than reference" response. The results of

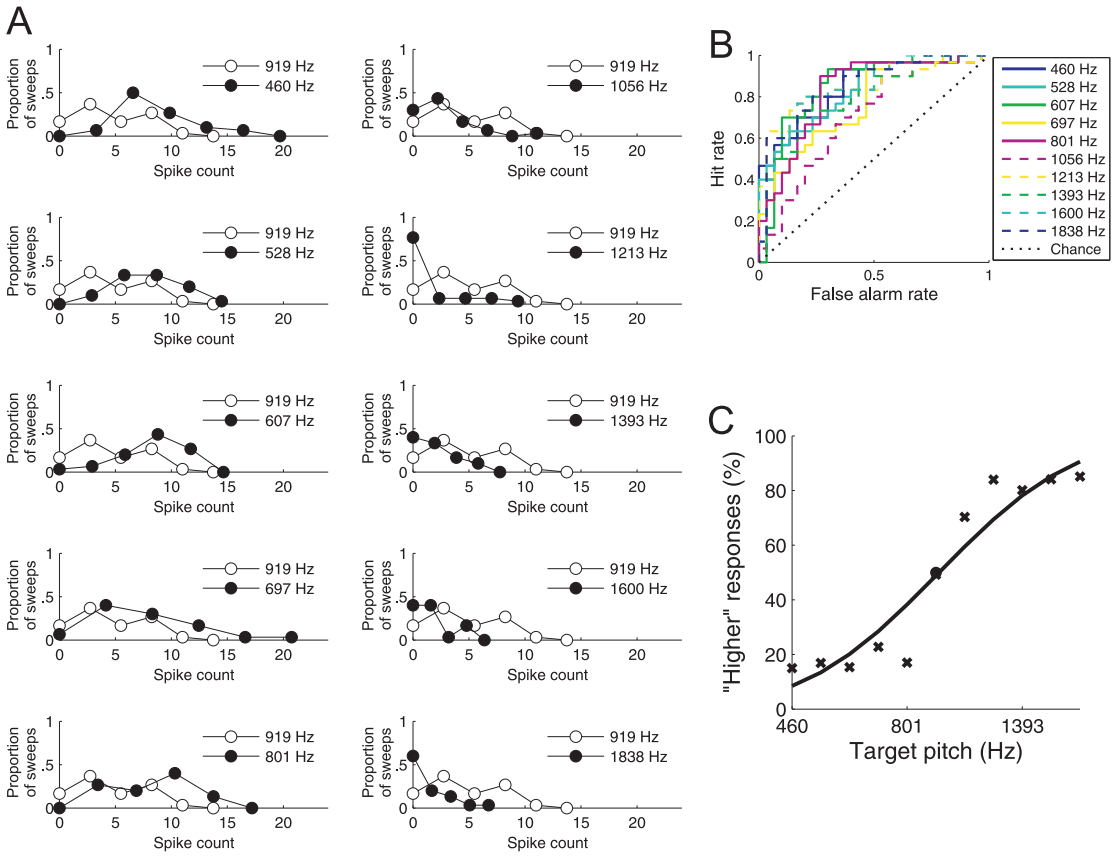


Figure 6.1: Constructing a pitch neurometric for one neural unit using ROC analysis. (A) Histograms show the distributions of spike counts of one neural unit across a 75 ms response period, in response to target (filled black circles) pitches and a 919-Hz reference (open grey circles). Each histogram shows data for a different target pitch (see legends). (B) Receiver Operating Characteristic (ROC) curves constructed from the distributions shown in A. ROC curves are shown across a range of target pitches (see legend). (C) A neurometric function was constructed by plotting the area under the ROC curves in B (black x's), and fitting a cumulative binomial distribution to these data points (black line). The reference pitch is indicated along the x-axis by the position of a black dot at 50% choice probability.

applying these decision rules to the data described in Figure 6.1A were derived using the ROC curves, which are illustrated in Figure 6.1B. Each of these curves was calculated using an algorithm that classifies each response within the spike rate distributions in Figure 6.1A as a response to the target or to the reference stimulus, based on whether it is above or below a spike rate threshold. The ROC curves were calculated separately for each target condition. For each target pitch condition (coloured lines), the observer's

proportion of "hits" (i.e. target stimuli that are correctly identified as a target) and "false alarms" (i.e. reference stimuli incorrectly identified as a target) can be calculated, but these proportions will depend on the value of the spike rate threshold (or, "response criterion") that the observer used. Therefore, the hits and false alarms were re-calculated for each possible response criterion, across the entire range of spike rates observed within the target pitch condition. The area under the ROC curve integrates the response probabilities across all possible spike thresholds, and so this value is free from criterion bias (Green and Swets, 1974). The area under the ROC curves is mathematically equivalent to performance of an ideal observer using the simple trial-by-trial spike rate decision rule described above. That is, if the area of the ROC curve for a given distribution of target and reference responses was .74, then an ideal observer of the neural units' spike rate responses could identify the pitch change between these 2 stimuli correctly on 74% of trials.

In Figure 6.1C, the discrimination performance derived from the ROC curves in Figure 6.1B is plotted as the probability of the observer reporting that the target is higher in pitch than the reference, across a range of target pitch conditions (black x's). A neurometric curve (black line) was fitted to these data to summarize discrimination performance across all conditions. The neurometric curve is a cumulative binomial distribution function, and was fitted to the discrimination performance data using the least squares fit of a probit link function. Neural units that support better pitch discrimination performance will produce neurometric curves with steeper slopes.

For each neural unit, neurometrics were calculated across the range of reference frequencies for which the responses to five lower and five higher target pitches were available. This meant that for the majority of units in this dataset I was able to calculate neurometric curves across 7 reference pitches (ranging from 400 - 919 Hz), and in one animal I was able to calculate them for 9 references (from 304 - 919 Hz). Neurometric analyses were also repeated for each unit using three response windows, since the duration of response varied across units. Response windows were 75, 150 and 300 ms in duration from stimulus onset.

6.2.4.2 Median reference response method

To examine if the performance afforded by responses of auditory cortical responses could be improved beyond that suggested by ROC-based neurometric analyses, neurometric curves were re-calculated using a constant reference criterion rather than the trial-by-trial responses to reference stimuli. The criterion was calculated as the median spike rate response across all presentations of the reference stimulus. Figure 6.2A shows histograms of the spike rate of one neural unit across presentations of a range of target pitches (closed circles). The target responses are the same as those presented in Figure 6.1A, but here they are compared to the units' average spike rate in response to the reference (919 Hz) pitch (dotted vertical line), rather than the distribution of responses to the reference pitch. Neurometric calculations were based on the difference between each response to target pitches and the reference criterion. If there was no difference between these values, the observer made a 50/50 guess as to whether the target was lower or higher in pitch than the reference. Otherwise, the difference was interpreted as indicating either a pitch increase or decrease between the reference and target vowels, depending on whether the unit was assumed to be positively or negatively monotonic for pitch. The neurometric performance was then plotted as the proportion of "higher target" decisions made within each target condition (Figure 6.2B, black x's). Neurometric curves (black line) were fitted as above. As with ROC-based neurometric calculations, neurometrics based on median reference responses were repeated using a range of reference pitches, and using 3 different response windows (75, 150 and 300 ms).

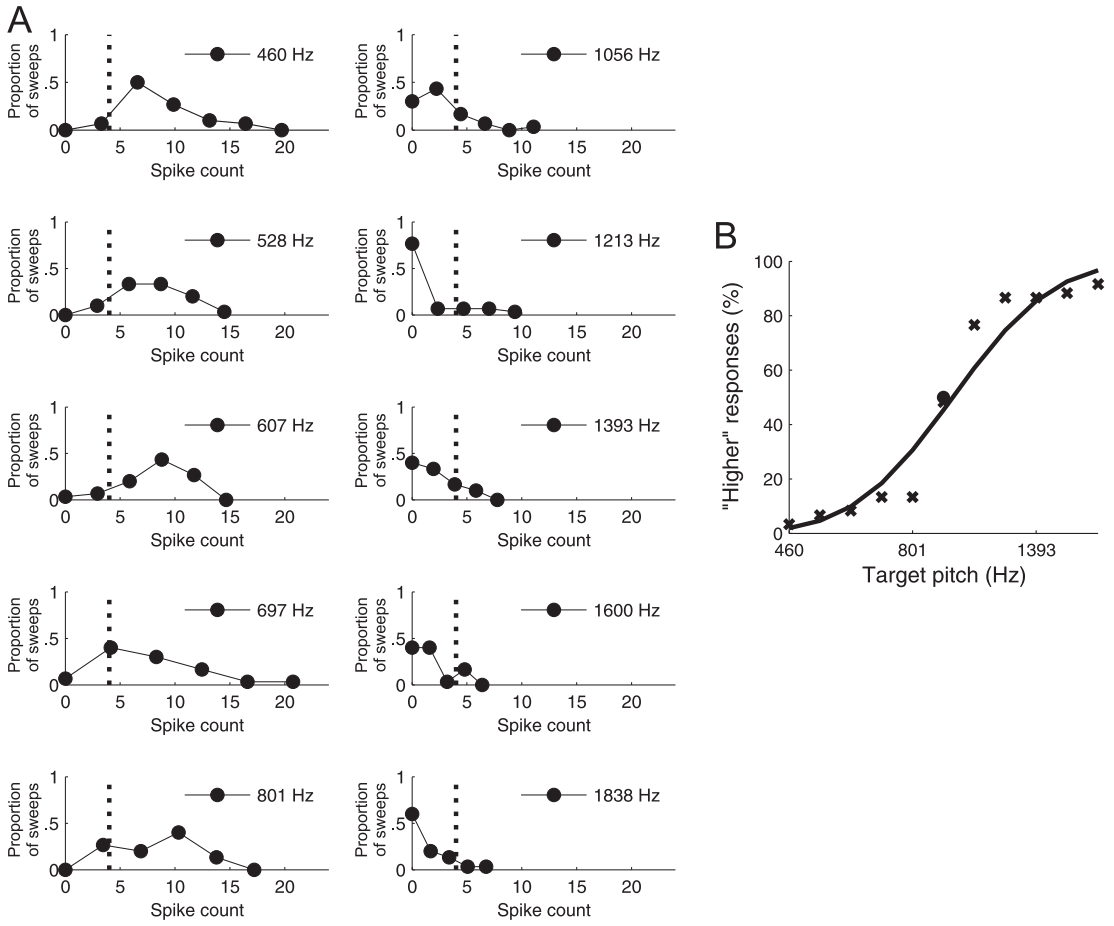


Figure 6.2: Constructing neurometrics based on the average spike rate response to the reference stimulus. (A) Histograms show the distributions of spike counts of one neural unit (the same shown in Figure 6.1) across a 75 ms response period, in response to the target (filled black circles) pitch. A vertical dotted line on each histogram indicates the median spike count in response to the 919-Hz reference pitch. Each histogram shows data for a different target pitch (see legends). (B) A neurometric function was constructed for the neural unit in A, based on whether the response to each target pitch was higher or lower in spike count than the average response to the reference. A neurometric curve (black line) has been fitted to these raw choice probabilities (black x's).

6.2.5 Neurometrics of neural populations

Multichannel electrodes were used to collect the physiological data, so in any given electrode penetration I recorded the simultaneous responses of several neural units that were separated by 100 - 150 μm. In addition, I sometimes recorded the activity of more than one neuron or group of neurons at a single electrode site, which could be separated based on their spike shapes. The number of units recorded simultaneously in any one electrode penetration varied between 1 - 64. The activity of at least three neural units were recorded in 47 electrode penetrations, and activity within these 47 populations of units were pooled and analysed in a population neurometric analysis.

To interpret the activity of neural units within a population, I wanted to design a neurometric algorithm that was as simple as possible so as to avoid unnecessary assumptions. I also wanted this algorithm to be as similar as possible to the median reference response method used for individual unit analysis, so that the results of these two analyses could be compared. For individual units, the observation on a single "trial" was the number of spikes in the unit's response to a vowel. Here, the observations were made up of the responses of many units. For each presentation i of a stimulus to neural unit j, the spike counts $r_{j,i}$ across all n neural units in the population were concatenated into an array, $\mathbf{R}_i$. The order of neural units within this array was arbitrary, but was fixed across all observations. The array representing the response of a population of n units to a single presentation of a stimulus was calculated as:

$$\mathbf{R}_i = [r_{1,i} \quad r_{2,i} \quad \dots \quad r_{n,i}]\,. \tag{6.1}$$

These representations of responses are visualized for one penetration across three pitches in Figure 6.3. Note that while the spike counts of some neurons in the pool are similar across pitches, others are more informative about the pitch of the vowel.

The responses to all presentation of a vowel at any one pitch can also be visualized as a collection of points in n-dimensional hyperspace. If the responses of neural units in the pool are informative about the pitch of vowels, then points representing the population's responses to different pitches will form separate clusters in this "response hyperspace". The population neurometric was based on this idea.

Like the individual unit analysis (Section 6.2.4.2), the population neurometric algo-

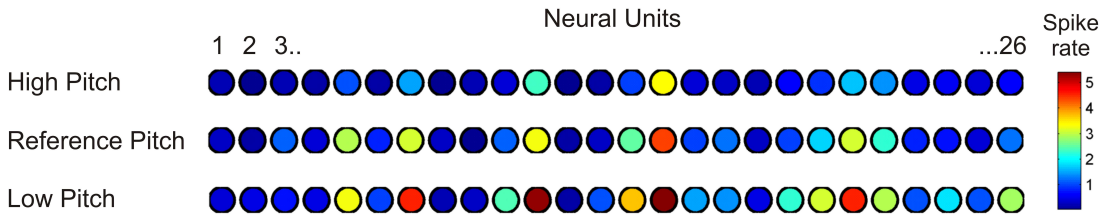


Figure 6.3: Schematic of algorithm for pooling responses across neural units in a population. The schematic illustrates how spike counts of several neural units simultaneously recorded in a single electrode penetration were pooled to prepare for population neurometric calculations. Each circle within a horizontal row represents one neural unit, with the colour of the circle corresponding to the number of spikes that the unit fired in response to a single stimulus presentation. The ordered array of average spike responses is shown for a vowel presented at the reference pitch (middle row), at the highest pitch (top row), and at the lowest pitch (bottom row) tested.

rithm consisted of two steps: first determining if the reference and targets sounds were noticeably different, and secondly labelling the direction of the pitch change between the reference and target. In the first step, I determined if the target response in question (i.e., the response to this particular presentation of the target pitch, $\mathbf{R}_i$) was more similar to responses to the reference pitch, or to other responses to that target. That is, I wanted to know if this pitch "sounded different" from the reference pitch. In practical terms, I answered this question by calculating the Euclidean distances, in response hyperspace, between the response to the target on this trial and (a) the mean response to the reference pitch, and (b) the mean response to other targets at this pitch. The mean response at a pitch k was calculated as:

$$\mathbf{S}_k = \sum_{i \in P_k} \frac{\mathbf{R}_i}{|P_k|}, \tag{6.2}$$

where P_k is the set of all stimulus presentations at a given pitch k, and $|P_k|$ is the size of P_k.

The Euclidean distance between a given response $\mathbf{R}_i$ and the average response to a particular pitch k is:

$$d_{i,k} = \sqrt{(\mathbf{R}_i - \mathbf{S}_k) \cdot (\mathbf{R}_i - \mathbf{S}_k)}. \tag{6.3}$$

If the response to the presentation of the target was more similar to responses to the reference pitch than to other targets of this type, it was considered to be "confused" with the reference, the observer could not detect a pitch change on which to base a pitch direction judgment. Thus, in the next step the observer made a 50/50 guess about the direction of pitch change. Alternatively, if the response on this trial was closer to other responses to the same target than to the average reference response, it was concluded that the observer could "hear" the difference between these pitches, and it then went on to make a pitch discrimination judgment based on the neural responses in step two. Note that if a neural population responded reliably and differently to the reference and target pitches, these responses would form two non-overlapping clusters in response space, and the observer would never guess, but would base all decisions on the neurons' responses.

To decide if the trial consisted of a pitch increase or decrease in step two of the population neurometric algorithm, the observer determined if the response to the target sound more closely resembled a "typical" high pitch response or low pitch response. In the individual unit neurometric algorithm, this question was asked by assuming, a priori, that the neuron was either positively or negatively monotonic for pitch, and then simply calculating whether the neuron spiked more or less times on the given trial than it usually spiked (on average) in response to the reference sound. Here, the population of neural units may include those that are monotonically increasing and decreasing for pitch. Therefore, it would be inappropriate to look the same effect of pitch spike rate across all neural units in the population. Instead, the response of each neural unit was interpreted independently, based on its own response characteristics, but again assuming that each neuron responded monotonically for pitch. Like the neurometrics of individual units, this analysis assumed that the spike rates of each neuron in the population changed monotonically across pitch. The average response to the highest and to the lowest pitch in the target range were used to represent the position of "high" and "low" templates in the neural response space. Thus, to decide whether the target on a given trial was higher or lower than the reference pitch, Euclidean distances were calculated between the response to this target and both the high and low templates. If the response was more similar to the high template, the observer decided that the target was higher than the reference. Otherwise, the observer decided that the target was lower than the reference.

In this manner, pitch direction judgments were carried out for each presentation of the target stimuli, and neurometric curves were fit to the resulting choice probabilities, as described for individual unit neurometrics above.

6.3 Results

6.3.1 Neural sensitivity to the pitch of artificial vowels

Neural responses to a computer-generated artificial vowel /i/ were recorded in six anaesthetised ferrets and one awake, passively listening ferret, across five auditory cortical fields. Single neurons and multi-neuron clusters were isolated based on spike shapes and spike auto-correlograms. Spike clusters which were composed of the activity of a single, well-isolated neuron and those which were likely made up of several neurons are treated indistinguishably in these analyses, and are thus referred to collectively as neural units. Raster plots and post-stimulus time histograms of the responses of each neural unit were inspected, and I identified 948 neural units that appeared to respond to the vowel stimulus. To confirm whether each neural unit was driven by the vowel, a t-test was carried out on the differences in spike count during the 7-200 ms period following vowel onset and the proceeding 807-1000 ms period of silence, across all responses of the neural unit. A modest alpha of 0.05 was chosen for this statistical test, so as to avoid discarding the responses of potentially interesting neural units early on in the analysis. Of all units tested, 77.64% (736 of 948) were found to be significantly driven by vowel stimuli (t-test, $p < 0.05$), and all of my further analyses were carried out on this subset of the data.

A careful inspection of the raster plots of neural units' responses indicated that many neurons fell into one of two classes: those that fired more spikes in response to higher pitches, and those that fired more in response to vowels that were lower in pitch. The responses of two neural units to the vowel stimuli are shown in Figure 6.4. In panel A of this figure, note that the unit did not respond selectively to a particular pitch, in the way that a frequency-tuned neuron would respond to pure tones across a range of frequencies. Instead, the responses to vowels were more often monotonic, demonstrating a increase and plateau in firing rate when the vowel was above (or, in panel B, below) a pitch threshold. These neural units may be described as acting as "high-pass" or "low-

pass" filters for vowel pitch. In order to determine quantitatively if the responses of each unit were modulated by the pitch of artificial vowels, I carried out ANOVAs on the spike responses. The pitch of vowels was used as a predictor variable for the sum of spikes elicited by a neuron during the 75 ms response window. If vowel pitch was found to be a significant predictor of spike count ($alpha = 0.05$), the neural unit's response was concluded to be sensitive to pitch. Of the 736 neural units that were driven by the artificial vowel /i/, the spike counts of 329 (44.7%) were modulated by the pitch of the vowel. These pitch-sensitive units were used in the neurometric analyses that follow.

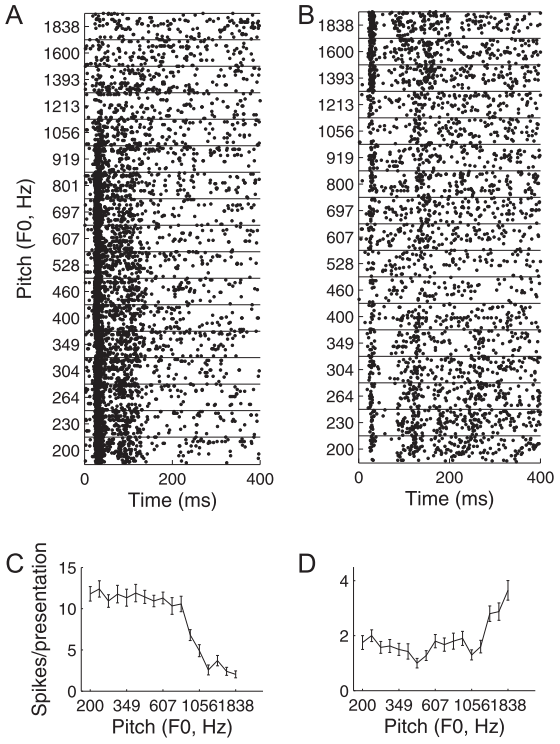


Figure 6.4: Raster plots of responses to vowels across pitches. (A) Action potentials (black dots) fired by a neural unit in auditory cortex in response to presentations of the artificial vowel /i/, across a range of fundamental frequencies (y-axis). Less spikes were fired in response to vowels that were higher in pitch, and so this is described as having "low-pass" pitch sensitivity (linear regression, $p < 0.05$). (B) Responses of a second neural unit in response to the same vowel sounds. This unit displays "high-pass" pitch sensitivity, as its' firing rate increases for higher pitched stimuli ($p < 0.05$). (C, D) The average spike counts of the neural units shown in A and B above, within the first 150 ms following stimulus onset. Spike counts are averaged across all presentations of the artificial vowel at each of 17 pitches.

To determine the number of units that showed monotonically increasing (high-pass) or decreasing (low-pass) pitch responses, linear regressions were carried out on the mean firing rates (during the first 75 ms following vowel onset) of each unit as a function of the vowel periodicity (Hz). The firing rates of 288 neural units (87.5% of pitch-modulated units) were shown to be monotonically related to pitch in this analysis ($p < 0.05$), with 39.58% of these (114/288) showing low-pass pitch sensitivity and 60.42% (174/288) showing high-pass pitch-sensitive responses. The remaining 41 non-monotonic pitch-sensitive units showed a variety of tuning properties. A small number of them had firing rates that responded selectively for a limited range of pitches (i.e. "band pass" pitch sensitivity), but more commonly these units showed more complex responses, with increased firing rates for several pitches dispersed throughout the pitch range. To illustrate, there is some evidence of this type of complex pitch sensitvity in the onset and offset response of the neural unit displayed in Figure 6.4B. Neural units from all three pitch-sensitivity classes were found throughout auditory cortex, although there were a tendency for low-pass units to cluster within low-frequency tonotopic regions, and similarly for high-pass units to cluster within regions of higher-frequency tuning (Figure 6.5A). The characteristic frequencies (CFs) of units was determined using methods described in Section 2.5, and the procedure for mapping neural units recorded from separate animals onto a "composite" auditory cortical image is described in Section 4.2.3. A Kruskal-Wallis test determined that the characteristic frequencies of units was different across the three pitch-senstivity classes ($p < 0.01$; Figure 6.5B). Post-hoc tests indicated that units with low-pass pitch sensitivity had lower CFs than those that were high-pass or non-monotonic for vowel pitch (Tukey's Honestly Significant Difference pairwise tests, $p < 0.01$). However, Figure 6.5B also shows the pitch sensitivity of an individual neural unit could not be predicted solely based on it's pure-tone frequency tuning, as there was considerable overlap in the CFs of low-pass and high-pass pitch units.

In Figure 6.6, the relation between pitch sensitivity and characteristic frequency is further illustrated. Figures 6.6A and B show the CFs of all units recorded in the six anaesthetised animals that were driven by the vowel stimuli, including those which were not significantly modulated by pitch, while Figures 6.6C and D show the CFs of the subset of these units that were significantly pitch-modulated. For some neurons, particularly those

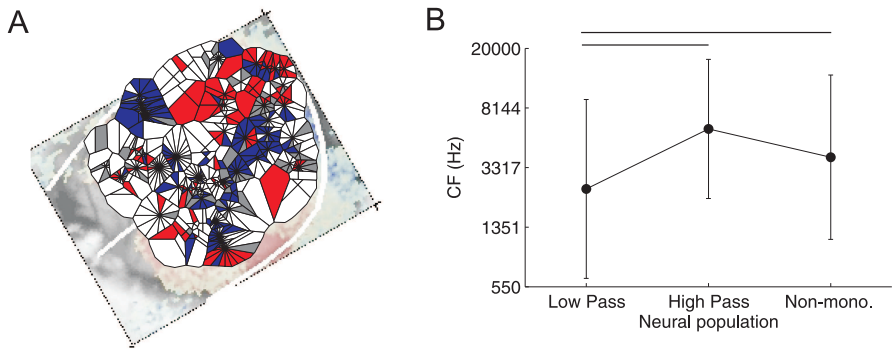


Figure 6.5: Cortical map and CFs of three classes of pitch-sensitive neural units. (A) Voronoi tessellation of the pitch sensitivity classification of all driven neural units recorded in six anaesthetised ferrets, compiled onto a map of auditory cortex. Each tile of the tessellation represents one neural unit, with the colour indicating whether the unit's response was unmodulated by pitch (white), high-pass modulated by pitch (red), low-pass modulated by pitch (blue), or modulated by pitch but not in a high- or low-pass manner (grey). Tiles representing units from a single penetration are arranged counter-clockwise according to depth, with the most shallow unit at the far right. (B) Mean characteristic frequencies of neural units in the three pitch sensitivity classes, with standard deviation bars. Lines above graph indicate significant differences between groups (Tukey HSD test, $p < 0.01$).

in ADF, a CF could not be determined because these cells were not frequency-selective. These units are included in Figures 6.6B and D, and are colour-coded white. These CF maps show that neurons which were modulated by the pitch of vowel sounds were not limited to those with low CFs, nor did all neurons with low CFs show pitch sensitivity. These maps also suggest that some cortical areas contained a larger proportion of pitch-sensitive neurons than others, and this is explored empirically in Figure 6.7. The core areas of auditory cortex - areas A1 and AAF - contained a larger percentage of pitch-modulated neural units than the more posterior fields (binomial distribution confidence intervals, $p < 0.01$).

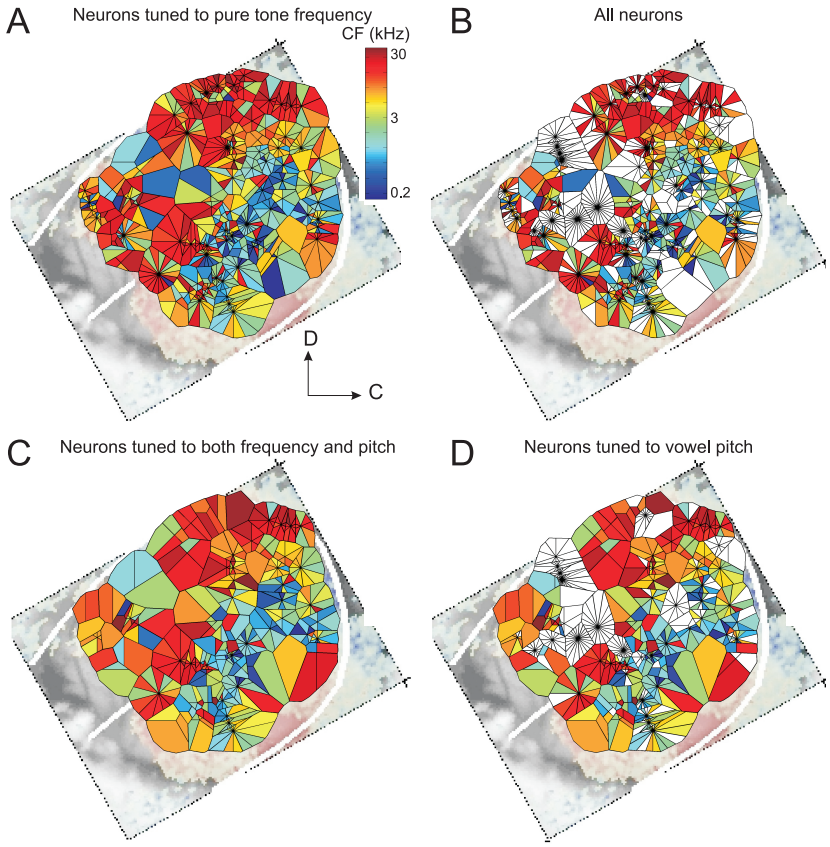


Figure 6.6: Maps of unit CFs across auditory cortex. Voronoi tessellations show the CFs of individual units recorded across the surface of auditory cortex. Data collected from six anaesthetised animals have been compiled onto one auditory cortex map. Tessellations were created as described for Figure 6.5, but here the colour of each tile indicates the CF of the unit (color scale shown applies to all four maps). The map in panel A includes all neural units from which a CF was determined. Map B includes all neural units in my dataset, including those that were not frequency-tuned (white tiles). The bottom two maps show the CFs of units that were frequency tuned and pitch modulated (panel C), and all those that were pitch modulated irrespective of frequency tuning (panel D).

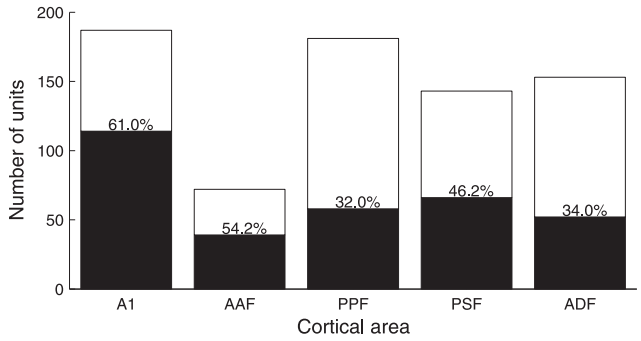


Figure 6.7: Proportion of pitch-modulated units across five cortical fields. The bar graphs indicate the total number of vowel-driven neural units identified in each cortical area (black+white bars), and the number of units within each area that were sensitive to the pitch of vowels (black bars). The percentage of pitch-sensitive neural units within each field is given within each corresponding bar.

6.3.2 Neural discrimination of pitch

6.3.2.1 Receiver Operating Characteristics (ROC) algorithm

Most neural units that were sensitive to the pitch of artificial vowels had spike rates that were either monotonically increasing (high-pass) or decreasing (low-pass) for vowel pitch. Therefore, my next analysis assumed that neurons in auditory cortex use spike rate to encode the pitch of vowels. Neurometric analysis based on ROC curves were carried out to estimate the pitch discrimination performance supported by the responses of each neural unit, as described in Section 6.2.4.1.

In my behavioural study, ferrets' performance was found to vary across reference pitches, and I expected that neurometric performance might also vary as a function of the reference pitch, since the position of the high and low-pass pitch filters varied across neural units. The neurometric curves of one unit, calculated for none different reference pitches (coloured lines), are shown in Figure 6.8A. The value of the references are indicated along the x-axis with coloured circles. The neurometric curves of this neural unit perform better when the reference is higher in pitch. For comparison, Figure 6.8B shows the set of neurometric curves across reference pitches for another neural unit. Here, the neurometric curves are steepest for the lowest reference pitches. Because the spiking response to the artificial vowels varied in duration across units, neurometric analyses for each unit were repeated using three different response windows: 75, 150 and 300 ms in duration from stimulus onset. The neurometric curves shown in Figure 6.8 and in most figures that follow were calculated using the response window that resulted in the steepest curves for each unit.

ROC-based neurometric curves were calculated for all pitch-sensitive neural units, across a range of reference pitches. The neurometrics of each unit assumed a monotonic spike rate code for pitch, in the appropriate pitch direction (i.e. the monotonically increasing or decreasing, depending on which resulted in neurometrics with positive slopes).

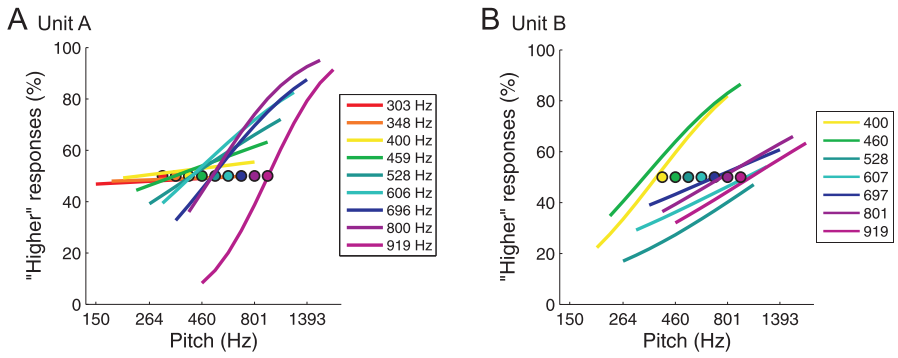


Figure 6.8: Neurometric curves across reference pitches. Each panel shows the ROC-based neurometric curves calculated for one neural unit, across a range of reference pitches (coloured lines). At each reference pitch, neurometrics were computed using three response durations (75, 150, and 300 ms) and the neurometric with the steepest slope is shown. The coloured circles at 50% choice probability indicate the reference pitch for the neurometric curve of corresponding colour. (A) Neurometrics for the neural unit described in Figures 6.4A and 6.1. (B) Neurometrics for another neural unit.

6.3.2.2 Median reference response neurometrics

Although most previous neurometric studies have favoured the criterion-bias-free approach of ROC analysis, the downstream neural networks that interpret the stimulus-evoked activity of auditory cortical neurons may not be subject to the limitations of this approach. In the neurometric analysis above, I have modeled an observer of spike rates in auditory cortical neurons which uses a neuron's response to one presentation of a target and reference vowel to decide whether that target was higher or lower in pitch than the reference. However, in my behavioural task the reference pitch is kept constant throughout a testing session, so the ferret does not have to depend on it's trial-to-trial representation of the reference vowel to make a decision about the relative pitches of the reference and target. If the neural representation of the reference is somewhat unreliable across trials, (as we have seen the spike rates of individual neural units to be), the ferret might perform better on this task comparing the trial-by-trial responses to target stimuli to a more stable, internal representation of the reference. One possible representation of the reference might be the average firing rate of the neuron in response to the reference, assuming the ferret has had previous exposure to the reference from which an average firing rate could be estimated. Training trials may have provided sufficient experience

for the animal to build up such a neural representation of the reference.

To test this idea, the neurometric performance of each unit was re-calculated using a modified, "median reference response" neurometric algorithm (see Section 6.2.4.2). The resulting neurometrics for the unit shown in Figures 6.2, 6.5A,C and 6.8A are shown in Figure 6.9, as calculated using the median reference response algorithm across a range of reference pitches.

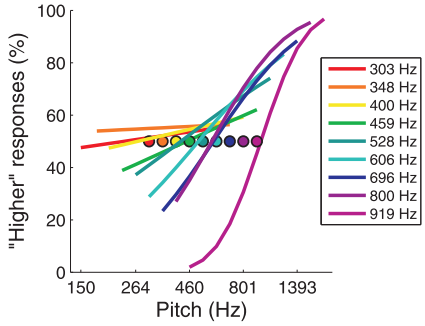


Figure 6.9: Median reference response neurometrics of one neural unit across a range of reference pitches. Neurometrics were calculated using a number of reference pitches (coloured lines) as in Figure 6.8, but here using the median reference response algorithm rather than ROC curves. At each reference pitch, neurometrics were computed using three response durations (75, 150, and 300 ms) and the neurometric with the steepest slope is shown. As in Figure 6.8, the neural unit is again shown to provide the best neurometric discrimination when the reference is higher in pitch. Coloured circles indicate the value of the reference pitch for each neurometric curve.

Figure 6.10 summarizes the results of the two neurometric algorithms. For each pitch-sensitive neural unit, ROC neurometrics were calculated across a range of reference pitches and three response windows and the neurometric curve with the steepest slope is plotted in Figure 6.10A. In the three panels, neurometric curves are plotted separately for high-pass (left), low-pass (centre) and non-monotonic (right) pitch-sensitive units. Similarly, the panels in Figure 6.10B show the best neurometric curve calculated for each unit using the median reference response algorithm. Using either technique, it is clear that while a small subset of pitch-sensitive neural units do well in supporting pitch direction judgments, many more units perform quite poorly on this task. I also observed that while most of the ROC-based neurometric curves are centred on chance performance

(50% "higher" responses), the neurometric algorithm that used the average response to the reference stimulus as a decision criterion are often biased towards one response choice. Thus, fixing a decision criterion can often introduce a response bias into the results of the neurometric discrimation.

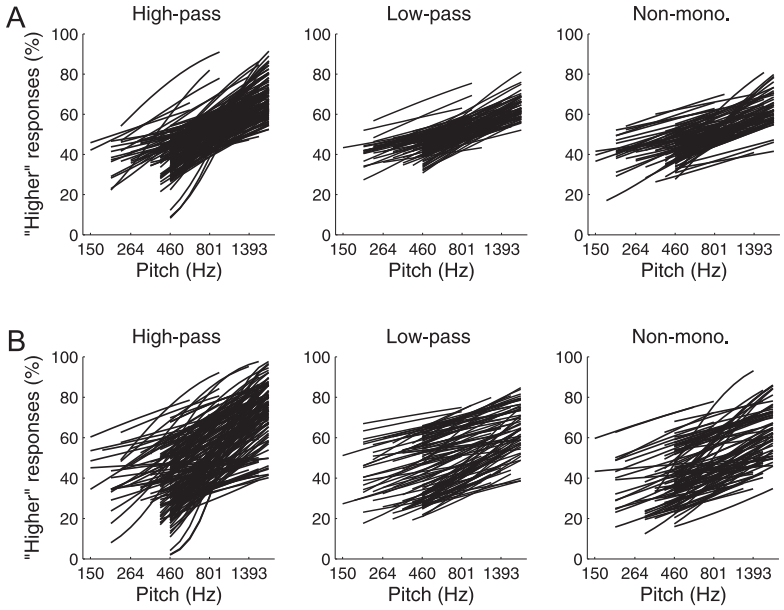


Figure 6.10: Pitch discrimination neurometrics for all neural units (n = 329). (A) Black lines show the best (i.e. steepest slope) fitted neurometric curves of each unit, across all three response windows and reference pitches. Neurometrics are plotted separately for units that were classified as high-pass (left panel), low-pass (centre panel), or non-monotonic for pitch (right panel). Neurometrics were calculated using ROC curves, as described in Figure 6.1. (B) Best neurometrics of all units across the three categories of pitch-sensitive neural units, as described for (A), but here neurometrics are calculated using the median response reference, as outlined in Figure 6.2.

On the other hand, the median reference response neurometrics often result in steeper neurometric curves than those calculated with ROC analyses (Figure 6.11). A paired t-test was carried out on the best slopes calculated for each unit using the ROC and median reference response neurometric algorithms. The result confirmed that slopes were usually steeper when calculated with a fixed criterion ($t = -7.75$, $p < 0.01$). Therefore, while using a fixed reference can bias pitch direction judgments, it also improves the discrimination performance of the most informative neural units.

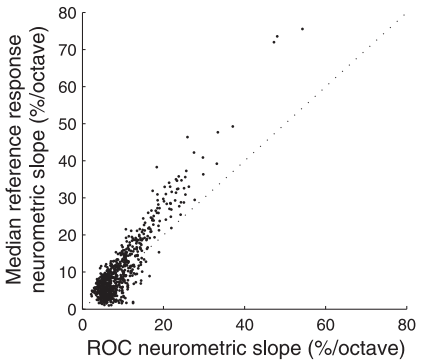


Figure 6.11: Comparison of neurometric slopes calculated using the ROC and median reference response algorithms. The scatterplot compares the best neurometric slopes for each unit (across all reference pitches and 3 response durations) when calculated using the ROC (x-axis) and median reference response (y-axis) neurometric algorithms. A dotted line of equality ($x = y$) is shown to aid comparison. Neurometrics slopes are generally better when calculated using the median spike rate as a reference, rather than the trial-by-trial spike rates (t-test, $p < 0.01$).

One might expect that the best pitch discrimination performance would result from using the average response to the reference stimulus as a decision criterion in a subset of neural units, and either ignoring the signal from other neural units or employing a more appropriate method of interpreting their spiking activity. But does the brain need to "tune out" the spikes of the less informative units in order to perform the pitch discrimination task well? Implications of pooling responses across neural units are explored later in this study, but first the fixed-criterion neurometrics of individual neural units are examined in further detail.

6.3.2.3 Neurometric discrimination of pitch across individual units

As mentioned briefly above, a neuron's neurometric pitch discrimination performance varied with the reference pitch used as well as the duration over which spikes were summed. Most neural units performed best at the highest reference pitch in the range tested (Figure 6.12A), and this result held within the subsets of high-pass and low-pass units (data not shown). The response window providing the best neurometric performance was more evenly distributed across neural units (Figure 6.12B).

In Figure 6.13A, the steepest neurometric curves, across all units and response windows, are plotted for a number of reference pitches. Since these neurometric curve were chosen based on their relative slopes, some show obvious response biases. At each reference pitch, there is at least one neural unit that provides information about relative pitch changes within their spike rate responses. However, the maximum neurometric

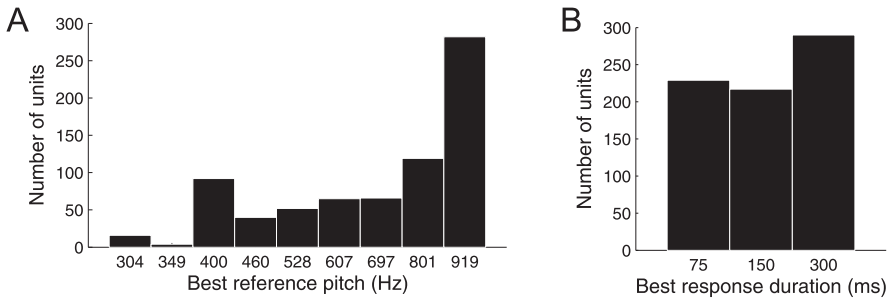


Figure 6.12: Best reference pitches and response durations for neurometrics. Histograms of the reference pitch (A) and response duration (B) resulting in the steepest neurometric function, across all neural units.

performance does differ across references in a way that is not easily predictable from the number of neurometrics optimized for each reference pitch in Figure 6.12A.

Nine neural units were recorded in the primary auditory cortex of an unanaesthetised, passively listening, head-fixed ferret. There was no significant difference in the best neurometric slopes (across references and response windows) of these units and that of units recorded in A1 of anaesthetised ferrets (t-test, $p = 0.138$).

If the ferret auditory cortices contain a devoted "pitch centre" analogous to that previously described in primate species (Bendor and Wang, 2006), one might expect neurons with the steepest neurometric slopes to cluster within that area. Neural units were sorted according to their best neurometric slopes (across all reference pitches and response durations tested), and the positions of the units giving the 25 steepest slopes are plotted onto the surface of auditory cortex in Figure 6.13B (coloured patches). The color scale indicates the rank order of neurometric slopes, with the best slopes shown in bright blue. The positions of all other neural units are also indicated for comparison (black dots). While the majority of the neural units supporting the best pitch discrimination performance tend to come from low-frequency tonotopic areas, these units are spread throughout the posterior and anterior banks. In addition, a collection of high-performing pitch neurometrics are based on the responses of neurons in the high-frequency tonotopic border of A1 and AAF.

The clustering of the steepest neurometric curves within low-frequency regions of

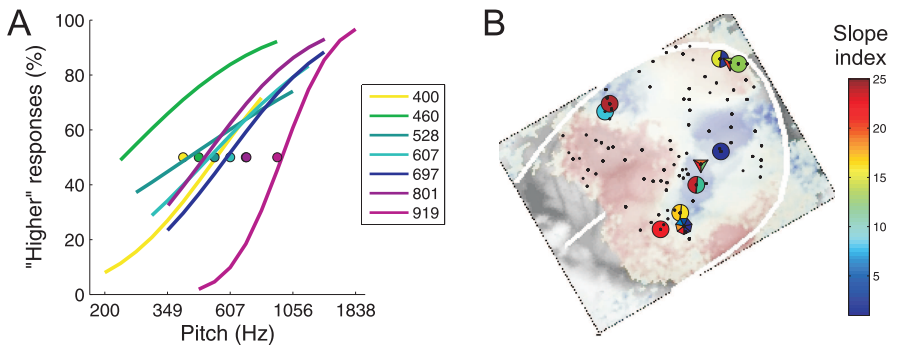


Figure 6.13: Best neurometrics across all units. (A) The best neurometric curve, across all neural units and response windows, is plotted for each reference pitch. The reference for each curve is indicated by the line colour (see legend) and by a coloured circle at 50% choice probability. (B) The position of the units resulting in the 25 steepest neurometric curves are mapped onto a composite auditory cortex map as coloured patches. The color scale of the auditory cortical map in the background shows the tonotopic organization of the cortex, as visualized using optical imaging of intrinsic signals (modification of figure originally reported in (Nelken *et al.*, 2004)). Patches that are subdivided represent multiple neural units from a common electrode penetration. Within these patches, units are represented counter-clockwise in order of increasing depth, with the most shallow unit in the far right section. The colour code represents the rank order of the neurometric slopes for these units, so that the best neurometric is bright blue (see colour bar). Black dots indicate the positions of all other neural units in my dataset.

auditory cortex would seem to suggest that units' CF was related to their neurometric pitch discrimination performance. However, there was no correlation between the CF and best neurometric slope of neural units (Figure 6.14A; Pearson correlation, $r = -0.11$, $p > 0.01$). There was also no significant correlation between CF and best neurometric slopes for neural units within each of the three classes of pitch-sensitive neurons (Figure 6.14B; $p > 0.01$). However, the three classes of units did differ in neurometric performance (Figure 6.15). There was a significant effect of pitch sensitivity class on slope (Kruskal-Wallis test, $\chi^2 = 68.58$, $p < 0.01$). Neural units with high-pass pitch sensitivity had better neurometric slopes than low-pass and non-monotonic units (Tukey's HSD test, $p < 0.01$), although there is clearly a large degree of overlap in the neurometric slopes of the three unit classes.

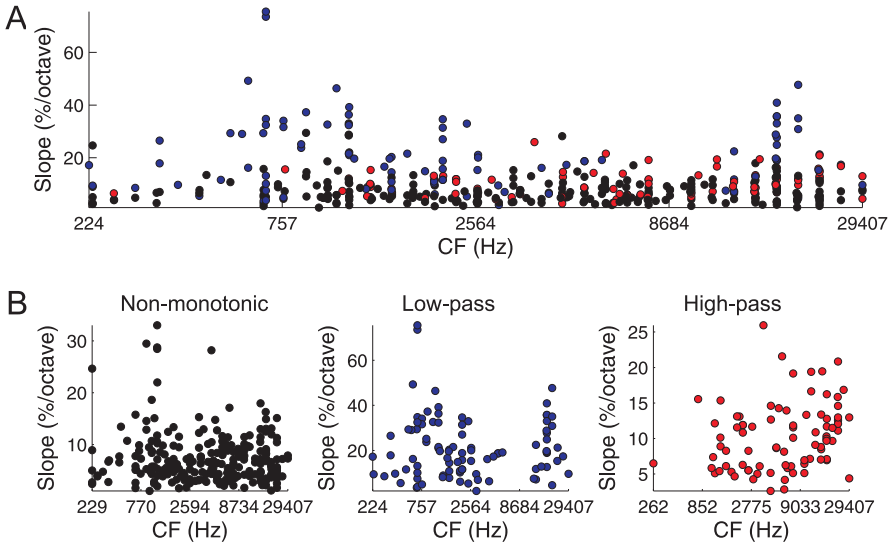


Figure 6.14: Neurometric slope as a function of the characteristic frequency of neural units. (A) Scatterplots of the CF and best neurometric slope (i.e. steepest slope across all reference pitches and response windows) for all neural units. The color of each dot indicates whether the unit was high-pass (red), low-pass (blue), or non-monotonically pitch-sensitive (black). (B) As described for A, but here scatterplots are drawn separately for non-monotonic (left panel), low-pass (centre panel), and high-pass (right panel) pitch sensitive units. There was no significant correlation between slope and CF within any of these group of neural units (Pearson correlation, $p > 0.01$).

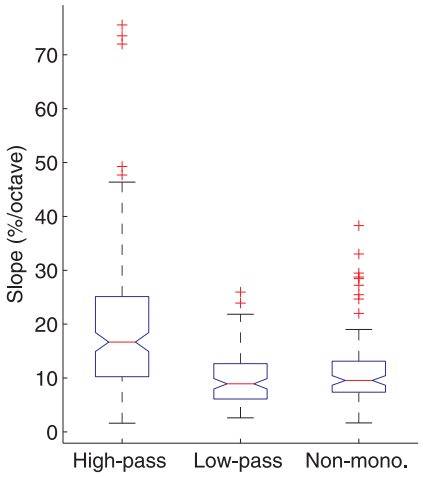


Figure 6.15: Neurometric slopes across high-pass, low-pass and non-monotonic pitch-sensitive units. The box plots describe the distributions of units' best neurometric slopes (across reference pitches and response durations) for neural units in each of the three pitch sensitivity classes. The notch in each box indicates the median value, and the lower and upper edges of the box indicate the lower and upper quartiles, respectively. The whiskers extend to 1.5 times the interquartile range of the sample, and outliers are marked by plus signs (+).

6.3.2.4 Pitch discrimination by ferrets and cortical neurons

The above results show that the spike rates of auditory cortical neurons can be used to discriminate pitch changes in artificial vowels. But how does this compare to ferrets' pitch judgments? If the spike rates of these neurons underlies ferrets' pitch discrimination judgments, the neurometric curves should perform at least as well as ferret psychometric curves on this task. The behavioural performance of four ferrets on the pitch direction discrimination task are summarized in Figure 6.16. Ferrets' pitch discrimination performance, measured as the slope of psychometric curves, was relatively constant across this range of reference pitches (Pearson correlation, $r = -0.312$, $p = 0.011$).

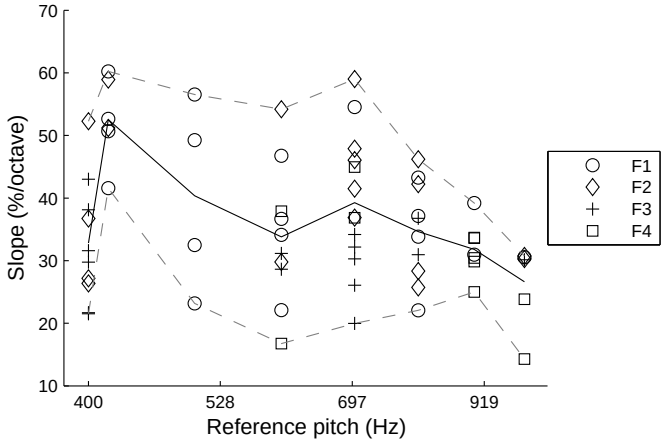


Figure 6.16: Pitch discrimination performance of four ferrets. The slopes of the psychometric curves describing the performance of 4 ferrets on a pitch direction discrimination task are shown. Data from each of the four animals are displayed with symbols (see legend). Each datapoint shows the performance of one ferret during one testing run (i.e. a week of testing). Pitch discrimination was tested using a number of different reference pitches (x-axis). The mean (solid line), minimum (lower dotted line) and maximum (upper dotted line) psychometric slopes are indicated.

In Figure 6.17A, the average slopes of ferrets' psychometric curves (black line) are compared to the neurometric slopes of cortical units (black dots). The dotted grey lines indicate the minimum and maximum behavioural performance measured at each reference pitch, so the complete range of behavioural slopes lies between these lines. While the neurometrics of the vast majority of neural units perform below the behavioural range, a small number of units perform as well as ferrets' averages on the pitch discrim-

ination task for reference pitches above 528 Hz. For the two highest reference pitches, a number of neural units outperform even the best behaving ferrets. In contrast, when the reference pitch is below 500 Hz, the neurometrics provide a very poor account of ferret's behaviour. Only one neural unit reaches the average behavioural slope value when a 400 Hz reference pitch is used, and none can support behavioural performance at references of 460 and 528 Hz. The "upper envelope principle" suggests that the limits of psychophysical performance are set by the most sensitive individual neurons in a sensory system (Mountcastle *et al.*, 1972). Said differently, an animal should discriminate stimuli no better than the best neural units that underlie this behaviour. This principle, as applied to my data, accounts for ferrets' pitch discrimination performance at high references, but not for low references.

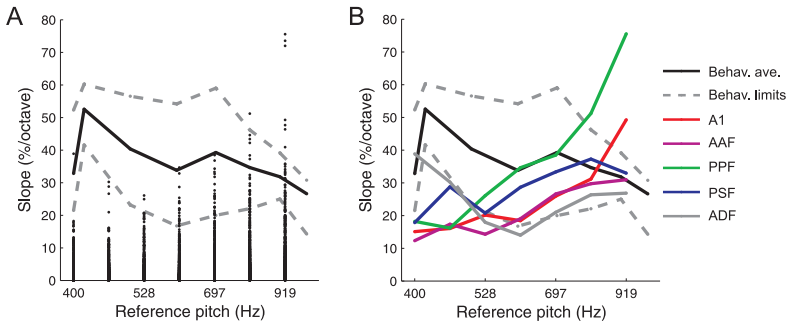


Figure 6.17: Pitch discrimination performance of ferrets and cortical units. (A) The slopes of the neurometric pitch discrimination curve of each neural unit was calculated across a range of reference pitches (black dots). The mean (black line), maximum (upper dotted grey line), and minimum (lower dotted grey line) psychometric slopes of ferrets on the pitch discrimination task are shown for comparison. (B) Coloured lines are the upper envelopes of neurometric performance across all neural units within a cortical area. Upper envelopes were calculated separately for neurons in area A1 (red line), AAF (purple line), PPF (green line), PSF (blue line), and ADF (grey line). The mean (black line), maximum (upper dotted grey line), and minimum (lower dotted grey line) psychometric slopes of ferrets on the pitch discrimination task are again shown for comparison.

The upper envelope of neurometric pitch discrimination performance is examined across cortical fields in Figure 6.17B. The upper envelopes of neural units in each of five auditory cortical fields is shown by the coloured lines, and is again superimposed upon the behavioural results. The upper envelopes are similar across areas A1 (red), AAF

(purple), and PSF (blue). These upper envelopes are generally within the behavioural range when the reference pitch is above 528 Hz, but are often below average ferret performance. For the highest reference pitch (919 Hz), the upper envelope of PSF units performs slightly better than the best ferret psychometric. The upper envelope of neural units in area ADF (grey) is also similar to these curves, except that ADF shows slightly improved discrimination performance for the lowest (400 Hz) reference pitch. Note that this improvement is present in the activity of only one ADF neural unit (Figure 6.17A). The upper envelope of units in PPF differs most dramatically from the other four fields, particularly for high reference pitches where the upper envelope outperforms the upper envelopes of other fields. Again, note that the dramatic increase in the upper envelope of are PPF at 919 Hz is accounted for by the outlying activity of a small number ($n = 3$) of neural units.

6.3.2.5 Neurometric discrimination based on populations of neural units

Figures 6.10 and 6.17A show that while the spike rates of some neural units can support the discrimination of pitch changes quite well, others do not, even though they are driven by the artificial vowel stimulus and are sensitive to changes in it's pitch. In the real-world situation, all of these neurons will be active simultaneously in auditory cortex, and decisions must be made based on some or all of these spiking patterns. While having two neurons that support discrimination of pitch changes may be better than having one, including the activity of less informative neurons in this decision process could impair performance. In this section, I explore how the brain might combine the activity of a collection of neurons that are simultaneously responding to the stimulus and use this "pooled" activity to solve the pitch change discrimination task.

The simultaneous activity of neural units recorded within each of 47 electrode penetrations were pooled (within a given electrode penetration only) and neurometrically analysed as described in Section 6.2.5. Each population contained between 3-64 neural units. The population neurometric algorithm was designed to be as similar as possible to the median reference response algorithm used to analyse the responses of individual units. A population's response to each stimulus was constructed as an ordered array of the spike count response of each neural unit in the pool. This response array was

compared across stimulus pitch conditions in order to compute pitch discrimination performance across a range of reference pitches (304 - 919 Hz). For each population, the reference pitch yielding the steepest neurometric curve was determined, and these 47 "best" neurometric curves are plotted in Figure 6.18. An informal comparison of these results to the neurometrics of individual neural units (Figure 6.10) suggests that the responses of populations of neural units usually resulted in sharper discrimination of the pitch of artificial vowels. The population neurometrics were also less prone to response bias than neurometrics of individual units.

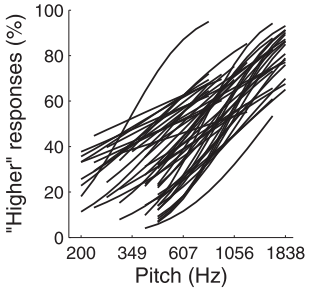


Figure 6.18: Best population neurometric curves of all penetrations. Black lines show the best (i.e. steepest slope) fitted neurometric curves of electrode penetration, across all reference pitches tested.

The results of individual unit and population neurometrics are further compared in Figure 6.19. The left panel displays a population neurometric curve (black line) along with the neurometric curves of each of the 26 neural unit within its' population (grey lines). The pitch of the reference is indicated by the open circle at 50% choice probability. These results show that discrimination based on the combined activity of these neurons was better than discrimination based on any one neural unit within the pool. In addition, although the median reference response neurometrics of most individual units within this population were biased towards either a "higher" or "lower" pitch change response, the performance of the population did not show a bias. This improvement in the performance of pooled the activity was a common finding across the majority of populations studied. Figures 6.19B and C show histograms of the best neurometric slopes of all neural units and all neural populations, respectively. The neurometrics of populations of units usually performed better than those of individual neural units ($t = 15.40$, $p < 0.01$).

Above, we saw that the neurometric performance of individual units on the pitch discrimination task often changed dramatically for references of different pitches. Would discrimination performance be more robust to changes in the reference pitch when it was

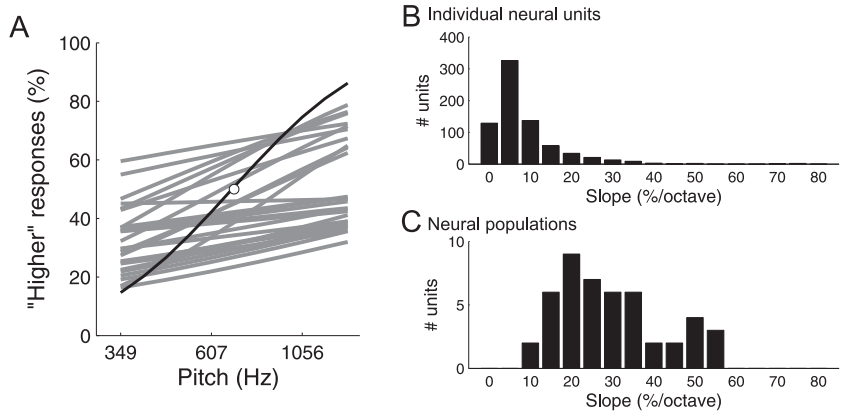


Figure 6.19: Comparison of individual unit and population neurometrics. (A) The population neurometric curve for a single, pooled penetration (black line), and the neurometric curves of each neural unit within that population (grey lines). The reference pitch is marked with an open circle. (B) Histogram of the best neurometric slope calculated for each individual neural unit. (C) Histogram of the best neurometric slope calculated for each population of simultaneously recorded neural units.

based on the activity of populations of neural units? The effects of reference pitch are explored for one neural population in Figure 6.20. The coloured lines are the neurometric curves for one neural population, calculated using a number of reference pitches between 304 - 919 Hz. In the interest of comparison, this particular population includes the neural unit investigated in Figure 6.9. The neurometric performance of this population is less dependent upon reference pitch than the neurometrics of the individual unit (Figure 6.9). These result, however, did not generalize across all populations of neurons. For each neural unit and each population, the variance of neurometric performance across all reference pitches was calculated as the standard deviation of neurometric slopes (across reference pitches), normalized by the mean of these neurometric slopes. A t-test showed that these values were the same for neural units and neural populations ($t = 0.71$, $p > 0.01$). Therefore, while neural populations supported better pitch discrimination performance, this performance was not particularly robust to reference changes.

In Figure 6.21A, the best population neurometric for each reference pitch are shown. More so than in the case of individual unit neurometrics (Figure 6.13A), a neural population that supports good pitch discrimination performance can be found for any reference pitch within the tested range. One would expect that penetrations providing the best

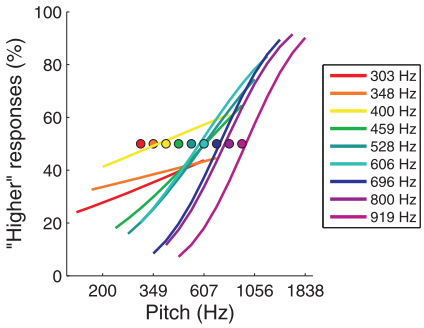


Figure 6.20: Population neurometrics of one neural population, calculated across a range of reference pitches. Population neurometrics were calculated across a range of reference pitches (colored lines, see legend), using the pooled, simultaneous responses of all neural units in a single electrode penetration. Coloured circles indicate the reference pitch for each neurometric curve.

population-based neurometric pitch discrimination performance would be localized in the same regions where the best individual unit neurometrics were found. Figure 6.21B, when compared to Figure 6.13B, confirms that this is indeed the case. Again, the most informative neural activity for the pitch discrimination task are found in the low frequency borders of cortical fields, but penetrations supporting good pitch discrimination performance are distributed across all five auditory cortical areas. There was no overall tendency for populations with larger numbers of neural units to produce steeper population neurometric slopes (data not shown; Pearson correlation, $r = 0.38$, $p > 0.01$).

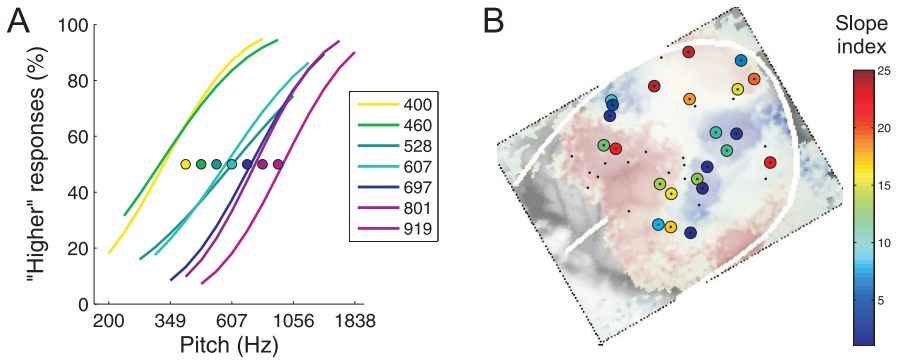


Figure 6.21: Best population neurometrics across all penetrations. (A) The best population neurometric curve is plotted for each reference pitch. The reference for each curve is indicated by the line colour (see legend) and is also indicated along the x-axis by a circle of corresponding colour at 50% choice probability. (B) The position of the electrode penetrations resulting in the 25 steepest population neurometric curves are mapped, as coloured circles, onto a composite map of auditory cortex (see Figure 6.13B caption for details). The cicle colour represents the rank order of the neurometric slopes for these populations, such that the best population neurometric is bright blue (see colour bar). Black dots indicate the positions of all electrode penetrates.

The spread of neurometric discrimination performance across auditory cortical areas is examined in more detail in Figure 6.22. The best neurometric performance of each neural unit is shown (colored circles), with units from a common penetration grouped along the x-axis. The pooled population neurometric for each set of units is also shown (black crosses) for comparison. With the exception of three outlying units in one PPF penetration, the distribution of population and individual neurometics appear evenly distributed across the five cortical fields. The three outlying neural units with particularly steep neurometric slopes were recorded in the same electrode penetration, in a low-frequency region of PPF. All three units showed low-pass pitch sensitivity, with a sharp decrease in their firing rates for pitches over 919 Hz. Therefore, their pitch neurometrics performance best with a 919 Hz reference. There was no direct relation between their pitch tuning and CFs - two of these units were tuned to pure tone frequencies of around 680 Hz, while the third was not frequency-tuned.

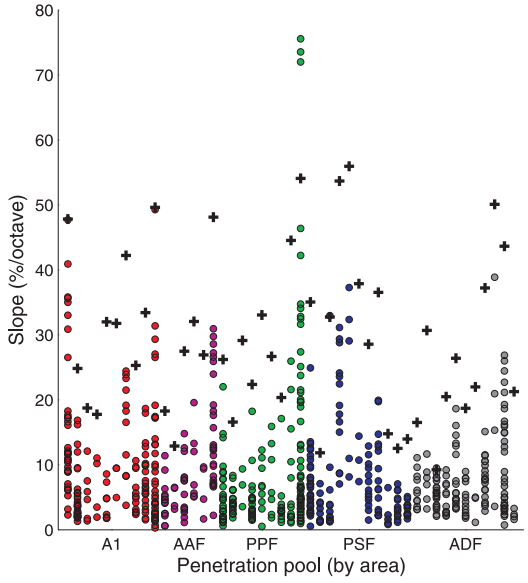


Figure 6.22: Population and individual unit neurometric slopes for each penetration. The slope of the population neurometric (black crosses) is given for each of 47 recording penetrations, and the slopes of the individual units in each penetration is also shown (coloured dots) at the same position along the x-axis. Penetrations are grouped (colour code) according to their cortical field.

Formal statistical tests failed to find a difference in the slopes of population neurometrics across cortical areas (Kruskal-Wallis test, $\chi^2 = 10.05$, $p = 0.04$). There was a modest but statistically significant effect of cortical area on the slope of individual unit neurometrics (Kruskal-Wallis test, $\chi^2 = 16.10$, $p = 0.0029$), which was due to the slopes of areas A1 and AAF being higher than the slopes of area PPF (Tukey's HSD test, $p < 0.05$). However, it is worth noting that this effect was not robust enough to reach

significance under an assumption of normality (ANOVA, $F = 1.59$, $p = 0.174$). In any case, the similarity of pitch sensitivity across cortical areas was more striking than the differences across these areas.

Figure 6.22 also allows a direct comparison of individual and pooled neurometric performance within penetrations. Generally, the population neurometric performed better than any one neural unit in the pool. In all penetrations, the population neurometric performed better than the average of all pooled units. Therefore, it is beneficial to base pitch direction judgments on the activity of multiple, simultaneously active units, even using my very simplistic pooling algorithm where the outputs of all neurons weight the perceptual decisions equally, and even if many of the pooled units have "noisy", poorly informative pitch responses.

6.3.2.6 Effects of neural correlations on population neurometrics

Given that pooling neural activity generally resulted in better pitch discrimination than the neurometrics of individual neural units, it is tempting to carry out a population neurometric analysis using all of the neural units in my sample to test the limits of neurometric performance. However, such an approach might not be valid since the spiking activity of simultaneously active neural units are correlated, and these correlations may have an important effect on the performance of the neural pool. A number of physiological conditions, such as anaesthetic levels, EEG rhythms or local spontaneous bursts of neural activity, may introduce correlated noise in the spiking response of several neurons within a penetration. The effect of correlations within this noise on the neurometric analysis is unclear a priori, but I can test for this effect within my data by removing noise correlations from the populations. This was done by randomly permuting the order of trials (within each stimulus condition) independently across each of the neural units in the population response array. The population neurometric of the "shuffled" population dataset was therefore based on the same neural signals as the original population analysis, but without the same trial-to-trial correlations in the signals of individual units. Because the performance of the shuffled population neurometric depends on the randomly determined order of trials, the shuffling procedure and subsequent population neurometric analysis was repeated 500 times at each reference pitch condition, for each

neural population.

The distribution of the neurometric slopes of shuffled populations across 500 repetitions was always approximately normal, so the average of these values was used to estimate the performance of the shuffled population on the pitch discrimination task. In Figure 6.23, the best population neurometric slopes (across all reference pitch conditions) calculated using the original, simultaneously recorded populations are compared to the best neurometric slopes for the shuffled populations. Overall, there was no difference in the slopes of these two datasets (t-test, $t = -1.89$, $p = 0.065$), although a small number of neural populations clearly performed better when correlations in the responses of their neural units were removed (e.g. the 3 outliers in Figure 6.23).

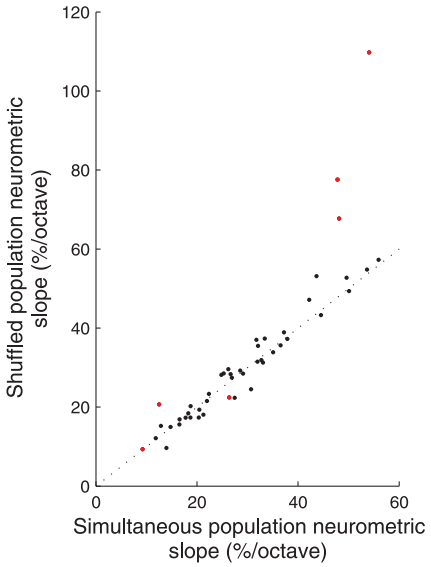


Figure 6.23: Neurometric slopes of simultaneous and shuffled neural populations. The best neurometric slope (across all reference pithes) for each population of simultaneous recordings (x-axis) is plotted against the average best neurometric slope calculated for "shuffled" activity from the same populations. Data points in red indicate penetrations whose shuffled population neurometric performed significantly better than the simultaneous population neurometrics for at least one reference pitch condition ($p = 0.01/329$).

For each population ($n = 47$) and each reference pitch condition ($n = 7$), I tested whether the neurometric performance of the simultaneous population differed from that of the shuffled populations. The neurometric slope calculated for the simultaneous dataset of a given population was tested against confidence intervals ($p = 0.01$, Bonferroni corrected for 329 comparisons) of the distribution of the 500 shuffled neurometric slopes. The shuffled data performed better on the pitch discrimination task for six populations, at several of the higher reference pitches. These six populations are indicated in red in Figure 6.23. No population performed more poorly in the population neurometric analysis when correlations were removed by shuffling. Since correlations in the activity

of neurons limits the neurometric performance of pools of these neurons, particularly for those pools that discriminate pitches well, pooling neural activity across separate recording sessions would almost certainly result in an artificially inflated estimate of the discrimination performance of populations of cortical units.

6.3.2.7 Pitch discrimination by ferrets and cortical populations

We have seen that pooling populations of simultaneously active neural units improves neurometric pitch discrimination performance (Figure 6.22), but does this improved neurometric performance help explain ferrets' behavioural discrimination of the pitch of artificial vowels? Figure 6.24 compares ferrets' pitch discrimination performance to the neurometric performance of populations of cortical units. Figure 6.24A shows that across the entire range of reference pitches, there is at least one population of neural units that performs within the behavioural range on the pitch discrimination task. As in the individual unit results (Figure 6.17A), the majority of population neurometrics perform below the range of psychometric slopes, and there are several neurometrics that outperform ferrets when the reference pitch is high (> 700 Hz). However, the upper envelope of population neurometric slopes more closely approximates behavioural performance than the upper envelope of individual neurometrics.

The upper envelopes of population neurometrics were remarkably similar across the five auditory cortical fields studied (Figure 6.24B). They all showed a general improvement in neurometric slope for higher reference pitches, except for field ADF, where neurometrics were best for the lowest references pitches (< 500 Hz). There was an overall agreement between the upper limits of population neurometrics and psychometric slopes, but the trends across reference pitches were different for the behavioural and physiological results. While ferrets' performance declined for reference between 500 - 1000 Hz, the upper envelopes of population neurometrics generally rose across this range (with the exception of ADF). In addition, the population neurometrics performed below ferrets' best psychometric curves when the reference had a pitch of less than ca. 600 Hz.

Because correlations among spiking neurons can effect the neurometric performance of pooled responses in a way that varies across different neuronal populations, I avoided pooling the responses of neurons recorded across electrode penetrations in order to esti-

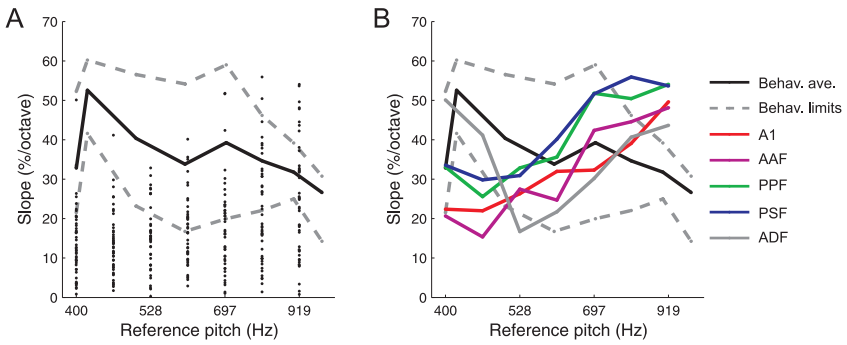


Figure 6.24: Pitch discrimination performance of ferrets and populations of cortical units. (A) The slopes of the neurometric pitch discrimination curve of each simultaneously recorded neural population was calculated across a range of reference pitches (black dots). The mean (black line), maximum (upper dotted grey line), and minimum (lower dotted grey line) psychometric slopes of ferrets on the pitch discrimination task are shown for comparison. (B) Coloured lines are the upper envelopes of population neurometric performance across all neural populations within a cortical area. Upper envelopes were calculated separately for area A1 (red line), AAF (purple line), PPF (green line), PSF (blue line), and ADF (grey line). Psychometric slopes of ferrets on the pitch discrimination task are again shown for comparison.

mate the neurometric performance of larger populations. However, I can further examine the neurometric performance of simultaneously active neurons by calculating pitch discrimination performance using subpopulations of neural units within an electrode penetration. While all the units in my analyses were responsive to artificial vowels and sensitive to their pitch, this does not in itself imply that they are all a part of the cortical circuits that a ferret would utilize to make pitch direction judgments. Therefore, population neurometrics were carried out on randomly chosen combinations of neural units, using a range of subpopulation sizes. For instance, if there were ten neurons recorded in a given electrode penetration, I calculated the population neurometrics based on all possible combinations of pairs of these ten units, then triplets of these units, and so on, up to groups of nine units, and finally the complete population neurometric. The number of possible subpopulations for a population of size p and subpopulation of size n is given by:

$$\frac{p!}{(p-n)!n!}, \tag{6.4}$$

which can become very large for modest population and subpopulation sizes. Therefore,

due to computing constraints, if there were more than 500 possible combinations of a subpopulation, I limited the analyses to 500 randomly chosen subpopulations.

These above analyses were carried out using a reference pitch of 697 Hz, and are summarized in Figure 6.25. The average neurometric performance (across all possible combinations of neural units) generally improved when a larger number of neural units were included in the subpopulation (Figure 6.25A). In many cases, the performance of subpopulations asymptote to the performance granted by the entire population of neural units. However, there were some populations for which average neurometric slopes were steepest when fewer units contributed to pitch discrimination decisions. This finding reminds us that depending on the response properties of the neural units that are being combined, basing perceptual decisions on larger numbers of responsive units will not always improve discrimination performance. Nevertheless, choosing the right combination of neural units on which to base pitch judgments usually resulted in better pitch discrimination performance than blindly basing decisions on all pitch-sensitive units available (Figure 6.25B).

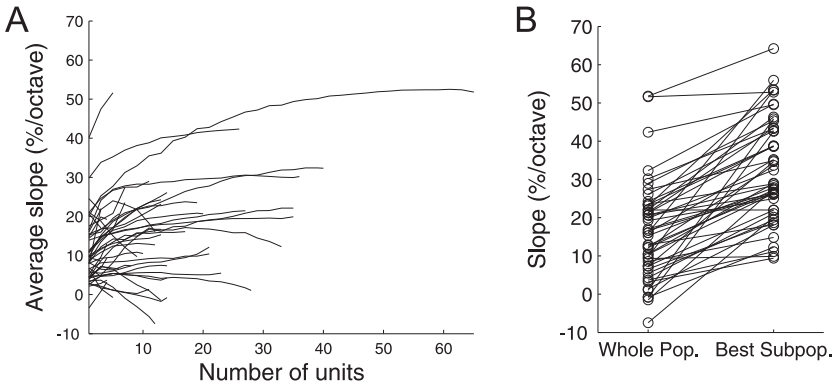


Figure 6.25: Neurometrics of subpopulations of simultaneously active neural units. (A) Average neurometric slope of subpopulations of neural units as a function of the number of neural units in the subpopulation. Each black line shows the results for 1 of the 47 neural populations ($reference = 697Hz$). The average was computed across all possible combinations of neural units where feasible (i.e. < 500 possible combinations of units), or across 500 randomly selected subpopulations otherwise. (B) Distribution of neurometric slopes calculated using the entire neural populations recorded in each of 47 electrode penetrations (left), and distribution of the steepest neurometric slope calculated across neural subpopulations in each electrode penetration (right).

Might the population neurometrics of these optimal subgroups of units help reconcile my physiological model of pitch discrimination with ferrets' behavioural results? The answer is not so simple. Figure 6.26A shows the best subpopulation neurometrics of each neural population ($n = 47$), calculated across a range of reference pitches, as in Figure 6.24A. When the best combination of neural units within each penetration was used to estimate neurometric curves, the performances across populations improved, and so the neurometric slopes overlapped more with the behavioural performance range. This provides stronger evidence of the activity of these neurons providing a sufficient physiological signal on which behavioural decisions could have been made. The upper envelopes of neurometric performance emphasize that the neurometric performance of optimal subpopulations of neural units was again best for the highest reference pitches in this range, which is in contrast to behavioural trends 6.26B. Nevertheless, at low reference pitches (< 600 Hz), many more neurometric slopes performed within the behavioural range when optimal subgroups of units were used to make pitch judgments than when individual units or entire populations were used (compare to Figure 6.17 and 6.24).

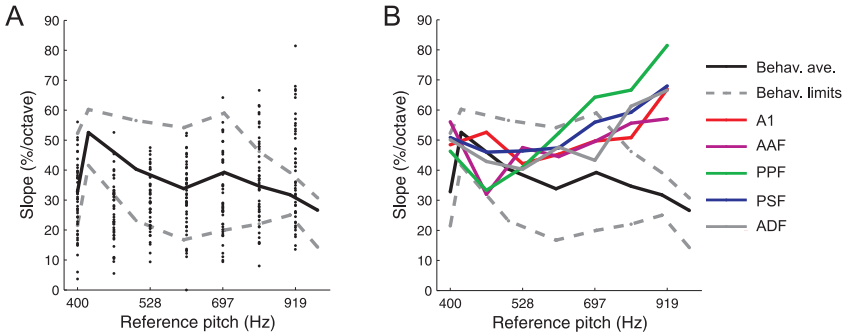


Figure 6.26: Pitch discrimination performance of ferrets and optimal subpopulations of cortical units. (A) Neurometric slopes calculated using the best possible subgroup of units from each electrode penetration population, across a range of reference pitches (black dots). The mean (black line), maximum (upper dotted grey line), and minimum (lower dotted grey line) psychometric slopes of ferrets on the pitch discrimination task are shown for comparison. (B) Coloured lines are the upper envelopes of the best neurometric slopes for subpopulations within each cortical area. Upper envelopes were calculated separately for area A1 (red line), AAF (purple line), PPF (green line), PSF (blue line), and ADF (grey line). Psychometric slopes of ferrets on the pitch discrimination task are shown (black and grey lines, as in A).

6.3.2.8 Examination of pitch encoding strategies

All the neurometric analyses above assumed that the pitch of artificial vowel sounds was encoded in the number of spikes fired by neurons. However, information about the pitch of stimuli might also be encoded in some aspects of the neural response, and so an alternative interpretation of the spiking responses might provide an even better prediction of ferrets' behaviour. Here, I compare the population neurometrics based on the number of spikes fired within a 75 ms response window (as above) to three other possible spiking codes.

The first is a temporal code, calculated as the latency, in ms, between the first spike fired in the pool following onset of the stimulus and the next spike fired by each other neural unit in the pool. If a given neuron did not fire a spike within the 75 ms response window, it was assigned the maximum latency of 75 ms. The array of first-spike latencies formed response arrays for each stimulus presentation, and the population neurometric analysis was carried out on these response arrays as before.

The second alternative neurometric analysis used a binary code to represent whether each neuron fired spikes or not on each trial. Neurons that fired at least one spike in the 75-ms response window were assigned a "1" in the response array, and those that did not spike were assigned a "0". The arrays of ones and zeros across neural units made up the response array for the population neurometric analysis. The information used in this analysis is essentially a compression of the information used in the spike-rate-based analysis.

Finally, I used Principal Components Analysis (PCA) to capture temporal patterns in the spike responses. PCA is a standard mathematical transformation for reducing the dimensionality of a dataset. It achieves this by projecting the dataset on axes along which the data vary most, and then expressing the data as principal components along these smaller number of dimensions. A convenient aspect of PCA is that the principal components it generates are in order of the amount of variance in the data that they capture, so that the first principal component accounts for the most variance that can be possibly described in any one projection of the data. Therefore, by expressing a neural unit's response to a stimulus presentation as its' value along the first principal component for that unit (instead of, say, the rate of spikes), I can express the most

informative temporal patterns of the response with a simple scaler. For each neural unit, PCA was carried out on the collection of 75-ms-long response to all stimulus presentations (binned in 1-ms segments), and the first principal component was used to build response arrays for the population neurometric analysis. To remove the effects of pitch on spike rate, each response was normalized by the average response to that stimulus category before the PCA, so that the population neurometric analysis that followed was based only on information in the temporal patterns of spikes.

Population neurometrics were carried out using first spike latencies, binary spiking responses, and first principal components for the pools of neurons in all 47 electrode penetrations, and across a range of reference pitches. The spike rate population neurometric displayed in the left panel is identical to the population neurometric discussed above, but the results are replicated here to facilitate comparison. The results of the three new population neurometric analyses are compared to the results of the spike-rate-based population neurometrics in Figures 6.27. In Figure 6.27A, the best neurometrics of each population across all reference pitches are shown in four panels. The results of the four neurometric algorithms look strikingly similar overall, despite the fact that the information contributing to the pitch direction decisions is different in each case. Figure 6.27B compares the best population neurometric curves calculated at each reference pitch, across the four neural codes. Again, despite some subtle differences in these results, overall, there is little advantage in using one of these codes over the others. Finally, Figure 6.27C compares the best slopes (across all references and response windows) of the spike-rate-based neurometric curves for each population to the best slope obtained with each of the other 3 neural encoding strategies. The slopes (dots) in each of these plots are centred around the line of equality (dashed line; $x = y$), since none of the three new neural codes produced neurometric slopes that differed from the results of the spike-rate-based population neurometrics (t-tests; $p > 0.01$).

In Figure 6.28, ferrets' behavioural pitch discrimination performance is compared to the neurometric performance of neural populations in each cortical area, interpreted using each of the four encoding assumptions. As before, the results of the four population neurometric analyses are similar, but here there are differences worth noting. The spike rate analysis performs better than all others when the reference is low in pitch, partic-

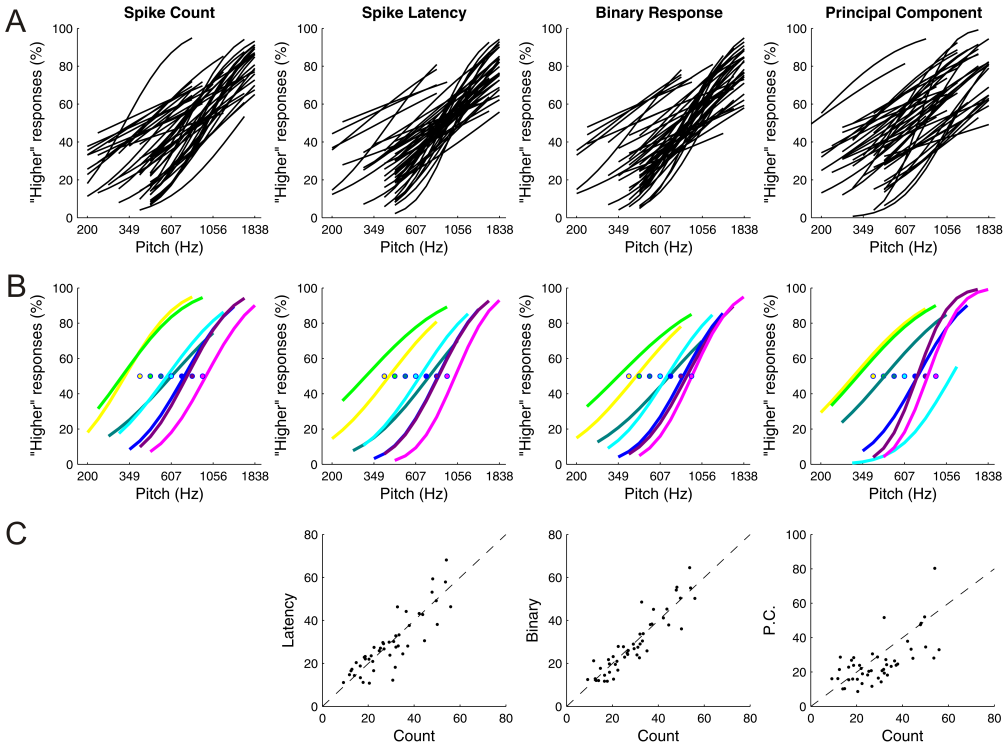


Figure 6.27: Population neurometrics based on spike counts, first spike latencies, binary spike codes, and first principal components. (A) Best population neurometric curves of all penetrations, calculated using either a spike rate, first spike latency, binary response, or principal component code (left to right, respectively). (B) In each panel, the best population neurometric curve calculated using one of the four spiking codes (see titles, above) is plotted for each reference pitch. The reference for each curve is indicated by the line colour (see legend, right) and is also indicated along the x-axis by a circle of corresponding colour at 50% choice probability. (C) Scatterplots compare the best population neurometric slopes (across all reference pitches) calculated using the spike-rate-based neurometric for each pentetration to the best slope calculated using each of the three other spiking codes.

ularly for populations in area ADF (grey line). Population neurometrics based on the other three neural codes are unable to account for behavioural performance when the reference is below 500 Hz. On the other hand, the spike latency and binary response codes sometimes perform better than spike rate neurometrics at the higher reference pitches ($>$ 600 Hz). The upper envelopes of neural populations in areas AAF (purple line) and PPF (green line) perform especially well when a first spike latency or binary spike/no-spike code are used. The principal component code performs far better than

the other three codes for neural populations in area PPF when the reference is high in pitch. However, even the relatively impaired spike-rate-based population neurometrics are more than sufficient to support psychophysical slopes for high reference pitch conditions. Despite the changes in the upper envelopes of the four types of population neurometrics, neither of the three new encoding analyses produced neurometric slopes that were significantly different from the spike-rate-based population neurometrics for populations in any of the five cortical areas (t-tests, $p > 0.05$, Bonferroni corrected for 18 comparisons).

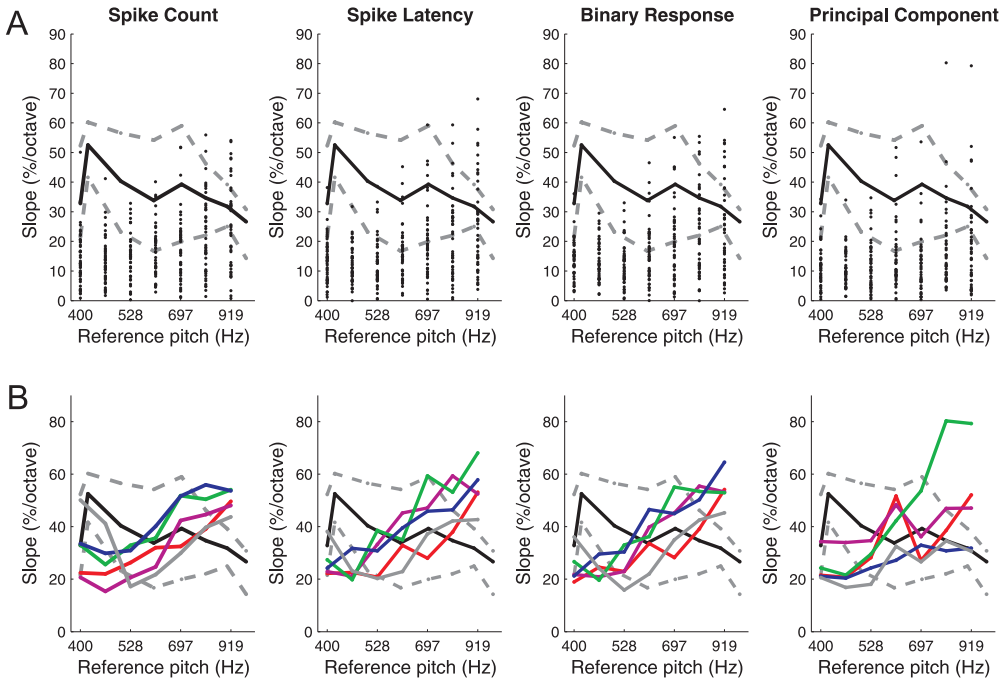


Figure 6.28: Pitch discrimination of ferrets compared to four population neurometric codes. Psychometric slopes are compared to population neurometric slopes calculated assuming a spike rate (left panel), spike latency (centre panel), binary spike response (right panel), and first principal component code. In each panel, the coloured lines are the upper envelopes of population neurometric performance across all neural populations within a cortical area. Upper envelopes were calculated separately for area A1 (red line), AAF (purple line), PPF (green line), PSF (blue line), and ADF (grey line). The mean (black line), maximum (upper dotted grey line), and minimum (lower dotted grey line) psychometric slopes of ferrets' performance on the pitch discrimination task are shown for comparison.

Since some neural responses to the artificial vowels lasted for longer than 75 ms, I

examined the performance of the four population neurometrics over two extended response windows: 150 ms and 300 ms (data not shown). The neurometrics based on response latencies remained relatively unchanged, since most responses to a stimulus in these cortical areas occur more rapidly than 75 ms post stimulus onset. The spike rate and binary spike/no-spike response codes also performed relatively similarly across the three response windows, although the upper envelopes for areas PPF and PSF were slightly impaired at longer response durations. The spike rate neurometrics in area ADF performed notably better with responses calculated over 75-ms durations than for those over longer periods. The principal component neurometrics showed a different trend. Here, the upper envelopes for neural populations in area PPF and PSF performed better for response windows of 150 and 300 ms than for the shorter duration. This suggests that while the spike rate or presence of a spike response can be most accurately calculated over quite short time periods (75 ms) in order to decode stimulus pitch, some neurons in the posterior bank of auditory cortex have temporal patterns that carry information about pitch for periods over 150 after stimulus onset.

6.4 Discussion

The purpose of this study was to determine if the pitch discrimination performance of ferrets could be supported by the spiking responses of neurons throughout ferret auditory cortex, or if these perceptual decisions require the activity of an anatomically segregated subpopulation of pitch-selective neurons. I found that neurons with spike rate responses that are monotonically responsive to the pitch of artificial vowels are distributed throughout auditory cortex. These neural units were found in each of the two core and three belt auditory cortical areas studied, and were present in both low-frequency and high-frequency tonotopic regions. This finding is consistent with Nelken *et al.* (2008), which used optical imaging of intrinsic signals in ferret auditory cortex to study periodicity encoding across these five auditory cortical fields. The authors found widespread neural modulation to periodicity pitch, and different distributions of pitch preferences depending on the type of periodic stimulus used. Here, I have further demonstrated that the activity of simultaneously recorded populations of pitch-modulated neurons throughout auditory cortex are sufficient to account for ferrets' average pitch discrimination per-

formance. Neurometric discrimination of pitch changes performed slightly worse than ferrets' best behavioural psychometrics when the reference pitch was low (*ca.* $450 - 530$ Hz), but note that even my quite extensive physiological recordings included just 64 population samples and there was much variance in neurometric performance across these populations. My failure to find neural populations with neurometric slopes equivalent to the limit of ferrets' low pitch discrimination performance may simply have been because my sample was too small.

6.4.1 Distributed correlates of pitch discrimination in auditory cortex

The vocalizations of ferrets have characteristic periodicities with syllabic pitch-shifts (see Section 1.2), so pitch perception is likely to be important for communication in this species, as it is known to be in other animals (Capranica, 1966; Koda and Masataka, 2002; Kojima *et al.*, 2003; Nelson, 1989). But how does the brain represent the pitch of these complex, natural sounds? Although the periodic features of sounds that are necessary to extract pitch are represented within the firing rates and firing autocorrelations of neurons as early in the auditory system as the auditory nerve (Cariani and Delgutte, 1996a,b; Cedolin and Delgutte, 2005), lesion (Warrier and Zatorre, 2004; Whitfield, 1980; Zatorre, 1988) and human imaging (Griffiths, 2003; Griffiths *et al.*, 1998; Hall *et al.*, 2006; Patterson *et al.*, 2002; Penagos *et al.*, 2004) studies suggest that pitch perception requires the activity of neurons in auditory cortex. In humans, Heschl's gyrus is preferentially activated during pitch discrimination tasks, and the marmoset anatomical analog of this area has been shown to contain cells that are uniquely tuned to the fundamental frequency of a variety of complex, harmonic sounds (Bendor and Wang, 2005). It is therefore likely, at least in primates, that the auditory cortex contains an area that is specialized to process the pitch of sounds. However, I have previously shown that neurons throughout the auditory cortex are sensitive to the fundamental frequency of complex sounds (Section 4), so these neurons outside the proposed pitch centre may also play a role in pitch perception. The present results suggest that a specialized pitch centre may not be necessary for ferrets to judge pitch changes in vocalization sounds, especially when the fundamental frequencies of the sounds are above about 600 Hz.

These findings do not dispute the existence of a dedicated pitch centre in primates,

or even in ferrets, nor did I address such questions of pitch selectivity in the present study. A specialized pitch centre may very well exist in ferrets to process pitch cues that are difficult to extract. For instance, specialized pitch neurons might be necessary to extract pitch from signals with highly unresolved harmonics, missing fundamentals, or from periodic signals in noisy environments. It is also possible that further pitch processing is necessary to improve ferrets' perception of pitches below 500 Hz, where the neural units in this study performed the pitch discrimination task relatively poorly.

Clearly, further investigation is needed in this area in order to determine the functional roles of pitch-sensitive and pitch-selective neurons in auditory cortex. Ultimately, these questions will require techniques that are designed to study functional organization, such as selective neural inactivation during behavioural tasks. In order to understand how neurons process pitch, it will also be essential for future investigations to study the neurophysiological responses of single neurons in animals as they perform behavioural tasks. In the present study, I found no notable differences between the pitch sensitivity, and thus the neurometric pitch discrimination performance, of primary auditory cortical neurons recorded in the awake, passively listening animal and the primary auditory cortex of anaesthetised animals. However, it is possible that the functional selectivity of cortical areas can only be observed at its' best while an animal actively engages in the relevant task. That is, the pitch tuning properties of cortical neurons may change when the animal makes use of them to solve a task. This type of behavioural effect on cortical tuning has already been demonstrated for ferrets performing frequency detection, frequency discrimination and temporal gap detection tasks (Fritz *et al.*, 2005a,b, 2007b, 2003), so it is likely to apply to pitch tuning as well.

6.4.2 The benefit of using a stable, internal pitch reference

I found that the neurometrics of cortical neurons performed better when a constant reference (namely, the neurons' average response to the reference stimulus across trials) was used to make pitch decisions, rather than the neurons' response to a reference pitch on each trial. This result is not entirely surprising, since the average reference will not incorporate noise from trial-to-trial variability in the neurons' response. But is it likely that ferrets would be able to use an internal representation of the pitch of the

reference sound to make judgments about the pitch of sounds presented on other trials? Behavioural trials were blocked by reference pitch condition, so it was entirely possible for animals to build up an internal representation (i.e. a sensory memory) of the reference sound. They could then determine whether each target sound presented during the test sessions was "high" or "low" based on whether it was above or below the pitch of this trained reference pitch. Although the ability to form memories of absolute pitches is rare among human listeners, behavioural research has shown that birds (Brosch *et al.*, 2004; D'Amato, 1988), rats (D'Amato, 1988) and non-human primates (Cynx, 1995; Page *et al.*, 1989) actually prefer to solve pitch discrimination tasks using an internal, absolute pitch scale rather than comparing the pitch of several sounds presented sequentially on one trial. Therefore, it is likely that ferrets also possess this capacity and could have been applying this strategy to solve my pitch discrimination task.

6.4.3 Pitch direction judgment as a 2-step process

My neurometric algorithms made pitch discrimination decisions using a 2-step process. First, the observer determined if it could detect a difference between the reference and target pitches. In the median reference response neurometric based on the responses of a single neural unit, this process involved determining if the average number of spikes fired in response to the reference was equal to number fired in response to the target. In the population neurometric algorithm, this decision was made by determining if the response to the target was more similar to the average response to targets of this type than the average response to the reference. Only if the observer determined that it could detect a pitch difference did it go on to try to classify the direction of pitch change based on neural responses. Otherwise, it made a random guess. The initial detection decision improved the performance of the population neurometrics (data not shown) by preventing the observer from basing its' decision on uninformative neural responses, which can lead to large biases in decision-making.

Ferrets may also utilize this 2-step tactic when solving the pitch discrimination task. That is, when the pitches of the target and reference are easily distinguishable, they may put in an honest effort to determine the direction of pitch changes. However, when the pitches are too similar, the direction of the pitch change may become more ambiguous,

and the ferret might elect to make a guess on this trial. This seems, to me, an obvious strategy to adopt since one cannot classify a difference that one has not perceived. I have informally observed that detecting a change in the pitch of my stimuli is easier than identifying the direction of that pitch change. In addition, behavioural studies of individuals with cochlear implants have demonstrated that their pitch change detection thresholds are much better than their thresholds for discriminating the directions of pitch changes (Fujita and Ito, 1999; Gfeller *et al.*, 2002; Vongpaisal *et al.*, 2006). This hypothesis is testable by studying ferrets' neural responses during the pitch discrimination task. If ferrets do use my proposed 2-step pitch direction discrimination strategy, there should be a significant drop in trial-to-trial correlations of their behavioural responses and the responses of neurons that underlie this process during conditions when the stimuli are closer together in pitch than some threshold value. This effect may also manifest as a step-like increase in the response times of human subjects on a pitch discrimination task when the pitches are separated in pitch beyond a threshold value, when the second processing step is implemented.

6.4.4 Frequency and periodicity tuning

The perceived pitch of pure tones correlates directly with the frequency of the single sine wave that composes them, so extracting the pitch of tones is a very straight-forward task for the tonotopically organized auditory nervous system. The tonotonicity provides a place code for frequency, so that the pitch of these simple sounds can be determined by locating the area of maximal stimulation along the cochlear, inferior colliculus, or auditory cortex. However, the vast majority of periodic sounds that a ferret would encounter in its natural environment are made up of many sine waves of different frequency, and so the pitch of these sounds cannot be arrived at using a simple place code along the tonotopic axis. Nevertheless, pitch perception is often conceptualized as a spectral decomposition problem. In contrast, my results suggest that the relations between frequency and pitch tuning are anything but straight-forward at the level of auditory cortex.

Although there was a general trend for neurons that supported the best pitch discrimination to be located in low-frequency tonotopic areas, there were exceptions to this

finding. Some neurons that were highly informative about pitch were located in high-frequency or non-tonotopic regions, and many neurons in low-frequency regions of cortex were informative on the pitch discrimination task. Furthermore, the neurometric performance of units was not correlated with their frequency-tuning. In fact, even the shape of typical tuning curves differed dramatically for periodicity pitch and pure tone frequency cues. While neurons in areas A1, AAF, PPF and PSF typically respond maximally to pure tones around one frequency, and are less responsive to pure tones above and below this characteristic frequency, these cells responded to artificial vowels of different pitches in a more high-pass or low-pass manner. That is, high-pass units responded maximally for artificial vowels that were above a certain pitch threshold, and low-pass units responded maximally for vowels below a pitch threshold. These results suggest that the stimulus parameters that determine auditory cortical neurons' responses to complex sounds at different pitches are not simply by-products of their pure-tone frequency tuning.

6.4.5 Neurometric performance of populations of neurons

I found that neurometric discrimination based on arrays of responses of populations of neural units (namely, 3 - 64) generally performed as well neurometrics based on the best unit in the pool or, more commonly, better than any one unit in the pool. The beneficial effect of pooling information across several neural units did not require complicated mechanisms to weight the inputs of more and less informative neural units, but was evident when neural responses were pooled in a simplistic, unweighted array. However, neurometric performance could be improved further by calculating neurometrics based on an optimal combination of neural units from a population of simultaneously recorded units. Previous studies have shown that when the inputs of cortical units are weighted using optimization calculations or artificial neural networks, the neurometric performance of their pooled activity can be substantially improved (Britten *et al.*, 1992b; Furukawa *et al.*, 2000; Purushothaman and Bradley, 2005). I fully expect that modification of the relative weightings of inputs of auditory cortical neurons (i.e. Hebbian learning) within the cortical circuits that underlie pitch discrimination play a key role in determining ferrets' performance on this task. Thus, even better pitch discrimination thresholds might be found for the neural populations examined here if this approach

were incorporated into neurometric analyses. However, I have shown that assumptions about the relative contributions of neural units to pitch discrimination decisions are not necessary to account for average ferret behavioural thresholds on this task. Rather, the raw spike counts of small (< 100) numbers of simultaneously active cortical neurons can support these pitch judgments.

My results also suggest that while basing pitch direction judgments on larger numbers of neurons can often improve discrimination performance, this improvement tends to asymptote at a upper performance limit. This asymptotic behaviour is a predictable consequence of pooling sources of information that are not completely independent. Indeed, I have confirmed empirically that naturally-occurring correlations among the spike rates of simultaneously active neurons can limit neurometric performance when these correlated responses are pooled. This is consistent with previous models of cortical neurons, which used inter-neuronal correlation terms to reconcile discrimination performance of animals and populations of neurons (Britten *et al.*, 1992b; Purushothaman and Bradley, 2005).

6.4.6 Neural codes for pitch

Most of the analyses in this study assumed that pitch was encoded in the spike rates of neurons, since the majority of neural units had spike rates that were monotonically increasing or decreasing for pitch. But since the responses of these units was often temporally rich, I also considered three other possible encoding strategies within populations of neurons. The latency of auditory cortical neurons' first spike following stimulus onset has been previously shown to encode information about sound source locations (Nelken *et al.*, 2005), so I reasoned that this response feature might carry information about the pitch of sounds as well. I calculated the first spike latencies relative to the first spike in the population following stimulus onset, rather than stimulus onset itself. Not only is this latency more biologically plausible computation, since the cortex does not have a record of stimulus onset, but it has also been shown to be more informative (Chase and Young, 2007). I also examined the pitch discrimination that resulted from observing the most predominant temporal pattern in spike responses, or from simply observing whether each neuron fired or not in response to vowels of different pitches.

I predicted that some of these response features would be informative about stimulus pitch, while others would be uninformative. In contrast, I found that all four encoding strategies produced neurometric slopes that differed only subtly for most neural populations. Yet there were two neural populations recorded in PPF that supported much better pitch discrimination with the temporal patterns of its' responses than with any of the other three response codes. While there were no overall effects of encoding strategy on neurometric performance in any of the five cortical areas studies (t-tests, $p > 0.05$, Bonferroni corrected for 18 comparisons), the best slopes of individual populations of neurons sometimes differed by as much as 100% for different response codes. Therefore, pitch may be encoded using different strategies across populations of neurons within a common auditory cortical area.

Luna *et al.* (2005) similarly found that the vibrations of tactile stimuli could be discriminated based on five different features of somatosensory cortical responses: the rate of spikes, the absolute spike count, the average inter-spike time interval, the number of bursts of spikes, and the rate of spikes within bursts. These authors could only determine which of these neural codes related directly to monkeys' frequency discrimination (namely, the spike count) by examining trial-to-trial correlations between neural responses to stimuli and animals' behaviour. In this manner, recording neural responses to stimuli while an animal performs a psychophysical task provides a very powerful investigative approach for linking neural activity to perception. Studies such as these will be one of the essential steps in the near future for determining how the auditory cortex encodes the pitch of complex sounds.

6.5 Acknowledgements

This study has been carried out in partnership with Dr. Jennifer Bizley. Dr. Victoria Bajo and Dr. Fernando Nodal supervised the surgery and recovery of the animal for awake electrophysiology, and Johannes Dahmen assisted in carrying out electrophysiological recordings in the awake animal. Prof. Andrew King and Dr. Jan Schnupp supervised my work on this study.

Chapter 7

General Discussion

7.1 Representations of multiple vocalization features in cortical neurons

This thesis has shown how neurons throughout the auditory cortex work in parallel to encode several features used by animals to interpret vocal calls. The temporal envelopes of vocalizations are represented in the temporal spiking patterns of A1 neurons. The resolution of these temporal spiking patterns reflects the perceptual acuity of human listeners, and the reliability of this code is statistically similar to listeners' choice probabilities. This suggests that listeners rely on this code to distinguish the temporal envelopes of vocal calls. Here, representations of the temporal envelopes of natural sounds were only examined in A1, but it would be interesting to see how these sounds are represented in higher cortical areas. For example, how do the width of temporal integration windows, which seem to be on the order of 10 - 20 ms at the level of A1, change in higher cortical fields? A very recent study suggests that temporal integration windows widen from A1 to more rostral regions of the marmoset core auditory cortex, as temporal spiking representations are converted into rate codes (Wang *et al.*, 2008). This is consisten with human imaging studies that show evidence of wider temporal integration windows in higher cortical areas (Boemio *et al.*, 2005). Electrophysiological studies of responses to temporally varying stimuli should be carried out in the ferret to investigate how computational conversions of this type take place from the auditory core to belt regions.

While the temporal order of vocalization events are represented in the temporal pat-

tern of A1 spikes, the pitch, timbre and spatial location of steady-state vocal calls are represented in the onsets of cortical responses. Pitch and timbre, more so than azimuth, are also encoded by spike rates sustained throughout these sounds and those that occur in response to sound offsets. The existence of multiple codes in auditory cortex allows a single neuron to represent several distinct stimulus features by utilizing a number of these encoding strategies. For instance, a neuron could represent the temporal envelope of a sound with a synchronized temporal spiking pattern, the azimuthal position of the sound in its onset response, and the pitch of the sound in the sustained portion of its response. Alternatively, or in addition to this multi-code strategy, a neuron can also represent a number of stimulus features by scaling its response according to linear or non-linear conjunctions of features.

I have seen evidence of both these tactics in the responses of auditory cortical neurons. In fact, cortical neurons that are modulated by either the pitch, timbre or azimuthal location of vocalizations alone are quite rare. The interdependence of neural receptive fields will be missed by electrophysiological studies that vary only one stimulus feature at a time. I examined neural responses as a function of three stimulus attributes, which meant that it was only feasible to sample the stimulus space sparsely. Varying each stimulus parameter across only four values resulted in a set of 64 stimuli, and so it took nearly an hour to present 40 repetitions of this stimulus set within one electrode presentation. This practical limitation can be overcome by studying the interaction of only two stimulus features, which would allow the investigator to sample stimulus features twice as densely in the same amount of time. This would provide a more detailed examination of receptive fields, and is likely to be necessary to uncover the more abstruse combination effects and regional specificities.

In considering the results of Chapter 4, it is important to appreciate the difference between neurons that are linearly sensitive to combinations of features and those that showed non-linear sensitivites to feature combinations. While the former responds to both stimulus parameters in a graded fashion, the latter acts more like a “combination detector”. Its response to a particular conjunction of two stimulus features is not simply a sum of its average response to each of these features. Neurons in the core auditory cortical regions were more likely than those in the cortical belt to respond uniquely to

particular combinations of pitch and timbre cues. This finding is consistent with the view that the primary cortical fields are responsible for identifying auditory objects through the integration of stimulus features. If these non-linearities underlie auditory object identification, we might expect them to be dependent on the statistics of an animals' acoustical environment, and this should be tested in future studies. Specifically, do non-linear A1 neurons act as filters for pitch and timbre combinations that commonly occur in the animals' acoustic environment?

7.2 Identifying the neural basis of what we hear

It has also been demonstrated here that ferrets can be trained to discriminate changes in the periodicity of vocalization sounds along an ordered pitch scale. Their performance on this task is closely matched by the upper limit of discriminability of neural responses throughout auditory cortex. Neurons in low-frequency regions of tonotopic cortical fields are particularly informative for pitch tasks. A small subset of cortical cells encode pitch using temporally complex spiking patterns, but most encode pitch information in a way that can be captured by the spike count over a single, appropriately chosen response window or even by the absense or presence of spikes within a single time window. Dr. Bizley, Dr. Schnupp, Prof. King and I are currently undertaking similar behavioural and physiological studies of vowel timbre discrimination in ferrets. In addition, we are training ferrets on a go/no-go version of the pitch discrimination task in order to determine if this results in better behavioural pitch thresholds than those reported here, and if there is a similar improvement in the discriminability of neural responses when neurometric algorithms are modelled after a go/no-go pitch task.

I have shown that neurometric comparisons of behavioural and neural data can be revealing about the potential relevance of neural responses to perception, and my comparisons of anaesthetised and unanaesthetised animals show that the effects of anaesthesia on these types of research questions is often minimal. However, I fully expect that there might be large differences between the auditory cortical activity of an animal that is awake and passively listening to sounds, and one that is actively listening to the same sounds in order to perform a psychophysical task. In a series of studies, one resarch group have illustrated the remarkable task-dependent plasticity of ferret A1 (Fritz *et al.*,

2007a, 2005b, 2007b, 2003), and these effects are also now beginning to be explored in the rat (Hromàdka and Zador, 2007). These studies show that the receptive field properties of auditory cortical neurons often undergo dramatic short-term changes when a ferret is performing a psychophysical task, and can revert back to its' original receptive field within hours after the task ends. The nature of the receptive field change is dependent on the type of task the animal is performing. Therefore, the cortical responses to sounds measured while an animal is at rest may underestimate the neurometric performance of neurons in the listening animal. Clearly, to relate neural activity to perception it will be essential to carry out neurometric studies like those presented here in the awake, behaving animal.

We are currently undertaking these types of experiments in animals that are performing the pitch discrimination task. Neurometric performance measured during the task will be compared to that measured in the same animals while they passively listen to the same stimuli. These types of experiments allow us to measure trial-to-trial correlations between behavioural choices and physiological responses, which provides a powerful tool for identifying neurons that directly influence animals' perceptual decisions. Eventually, the go/no-go task may prove especially useful for studying animals' responses while they perform a behavioural task in which the stimuli can change along several perceptual dimensions. That is, since animals merely detect the presence of a stimulus change on these tasks (rather than labelling stimuli with left/right responses), I can alter the nature of that change across trials and examine the effects of different perceptual problems on the receptive field properties of neurons.

7.3 Functional organization in auditory cortex

My results showed that perceptual attributes of vocal calls are represented redundantly across auditory cortical fields. Subtle differences were noted in neurons' relative sensitivity to pitch, timbre and azimuthal cues across cortical areas and depths, but neurons that encode each of these cues were found throughout the entire auditory cortex. How does this result fit with the striking specificities across cortical areas that have been identified in cortical deactivation (Lomber and Malhotra, 2008; Lomber *et al.*, 2007) and brain imaging studies (Alain *et al.*, 2001; Barrett and Hall, 2006; Maeder *et al.*, 2001;

Warren and Griffiths, 2003)? Again, a key breakthrough may rely on recording neural responses in these cortical fields while animals are actively listening to the sounds. The anaesthetised and/or head-restrained animals investigated in single cell recording studies are undoubtedly paying far less attention to the repetitive, experimental sounds than animals performing psychophysical tasks (Lomber and Malhotra, 2008; Lomber *et al.*, 2007) or even human listeners that take part in brain imaging studies. The degree of functional divisions across cortical areas may only become apparent in the presence of the activity that arrises in these pathways when the animal "uses" its cortex to listen carefully to sounds. A limitation of cortical deactivation studies is that they disrupt activity not just in the deactivated region, but in all other areas along the pathway that depend on the output of this region to perform properly. In addition, the fibers of passage in the deactivated region are also compromised. Therefore, careful studies of the effects of deactivation of one brain region on the activity of others need to be carried out in order to interpret the results of these reversible lesioning experiments properly. These types of studies are already underway, and have demonstrated profound influences of cortical processing on the tuning properties of neurons in the inferior colliculus (Nakamoto *et al.*, 2008).

Yet even when recording in awake, behaving animals, comparisons of cellular recording to human imaging studies will be non-trivial. This is due in part to the difficulties of across-species comparisons, but also to the differences in the types of activity measured when using each of these techniques. Magnetic Resonance Imaging measures the blood oxygenation level, rather than neural responses directly. A correlation has been demonstrated between this hemodynamic response and local field potentials, suggesting that fMRI results are strongly dependent on dendritic activity (Goense and Logothetis, 2008; Logothetis *et al.*, 2001). Thus, an experimental effect could reflect local processing in the region of interest via synaptic connections. Alternatively, fMRI activity in a region could reflect the activity of synaptic inputs from a projecting neural structure where the process of interest is taking place. Local field potential measurements can usually be easily collected during extracellular recording experiments, and the interpretation of these signals in addition to local neural responses may assist the comparison of fMRI results to neural response properties. Along these lines, we are currently measuring local

field potentials during extracellular recording experiments, so that comparisons of local dendritic and cellular spiking activity can be undertaken.

The differences in cortical specificity suggested by single cell recordings and human imaging studies might also be a consequence of the time scales of the underlying processes measured by these techniques. Blood oxygenation level dependent signals are measured over seconds and they reflect the mean rates of membrane potential flucuations within populations of cells, but are insensitive to fine temporal modulations in spike firing patterns (Goense and Logothetis, 2008; Logothetis *et al.*, 2001). My own work, and that of others, has shown that temporal aspects of neural spike responses can carry significant information about the temporal envelopes (Chapter 3; Schnupp *et al.*, 2006), pitch (Chapter 4; Chapter 6; Steinschneider *et al.*, 1998), timbre (Chapter 4), and spatial location (Chapter 4; Furukawa and Middlebrooks, 2002; Nelken *et al.*, 2005) of sounds. Thus, fMRI measures only a subset of the neural representations of vocalizations, and it is possible that areas which fail to show pitch modulation in fMRI studies do in fact have temporally-delimited or temporally-based aspects of their spike responses that are highly modulated by the periodicity of sounds.

7.4 Using more than one neuron to solve a task

Finally, my experiments have shown that when the simultaneous responses of several cortical neurons are pooled, their combined activity often support much better discrimination of the temporal envelopes and periodicities of vocalizations than that of any one neuron alone. This is the case even when the activity of many, or even all, of the neurons in the pooled population are uninformative about these stimulus features when interpreted independently. A key refinement in future studies will be to examine the perceptual acuity supported by more elaborate pooling mechanisms, where the inputs of individual neurons are weighted using statistical optimization techniques or artificial neural networks. Dynamic causal modelling, used previously in imaging studies (Kumar *et al.*, 2007), offers a further promising approach for determining the connectivity between spiking neurons recorded in separate brain regions.

7.5 Conclusions

Auditory cortical neurons represent multiple features of vocal calls in both the timing and number of spikes in their responses. These spike responses are noisy across repeated presentations of the same stimulus, and the reliability of these noisy responses is similar to the choice probabilities of listeners if the proper code is used to "read" the neural response. Physiological and behavioural data can be compared using these types of neurometric analyses to identify which neural representations of sounds might underlie perception, and which are not sufficiently informative. Future studies that examine neural responses in awake, behaving ferrets will be essential to achieving our goal of identifying the neural correlates of sound perception.

Appendix A

Code for "Flipping" Sounds

The following script, when run in MATLAB, cuts the signal 's', sampled at a rate of 'fs' hertz, into 'wdwLen' ms-wide, overlapping segments. The amplitude of each segment is multiplied by a cosine window, flipped and reassembled. This was used to generate the flipped vocalization stimuli in Chapter 3.

```
function out = granularCosFlip(s,wdwLen,fs);

%  make sure s is row vector
if size(s,1)==1,
    s=s';
end;

if wdwLen < 0,
    error('wdwLen must be positive');
end;

slen=length(s);
wdwStep=round(fs*(wdwLen)/2000);
wdwLen=2*wdwStep; % ensures that wdwLen is even
wdw=cosWin(wdwStep);
out=zeros(size(s));
out(1:wdwLen)=flipud(s(1:wdwLen).*wdw);
from=wdwStep; to=from+wdwLen-1;
```

```
while to < slen,
    out(from:to)=out(from:to)+flipud(s(from:to).*wdw);
    from=from+wdwStep; to=to+wdwStep;
end;

function x=cosWin(N);  % returns an 2*N point cosine window
t=0:pi/(N-1):pi;
x=(1-cos(t'))/2;
x=[x; flipud(x)];
```

Appendix B

Further Examples of Cortical Responses to Artificial Vowels

Chapter 4 described the responses of auditory cortical neurons to artificial vowel sounds that varied in pitch, timbre and spatial location around the azimuth. Raster plots (Figure 4.2) and Post-Stimulus Time Histograms (PSTHs) (Figure 4.4) summarizing the responses of two neural units to 64 artificial vowel stimuli were shown to illustrate how spiking responses can vary interdependently as a function of these three stimulus parameters. Here, further examples of the responses of neural units are presented in the same manner, in order to give the reader a better appreciation of the diversity of responses across cortical units. The responses of one neural unit in each cortical field are shown. These examples also illustrate common response properties across units. That is, most units were sensitive to more than one stimulus parameter, their response patterns could change at various times throughout the duration of the response, and interactions between stimulus features were not uncommon.

In each of the figures that follow, the responses of a neural unit in the indicated cortical field (see figure captions) are described. Three raster plots of responses will be shown in the top panels, and two PSTH arrays describing the same responses are shown immediately below. The raster plots show action potentials (black dots) fired by the neural unit in response to 64 artificial vowel stimuli. In each of the three raster plots, the complete dataset are shown, but these are organized according to stimulus azimuth (Az,

in degrees) in the left panel, pitch (F0, in Hz) in the middle panel, and timbre (arranged by first formant frequency) in the right panel. Negative azimuth values indicate sounds designed to be lateralized to the ipsilateral (left) side of the animal.

The PSTHs in each figure illustrate the main effects and 2-way interactions of timbre, pitch and azimuth on the responses of the neuron. The black lines trace the mean firing rate (in Hz) of the neural unit. The first four rows and columns of the array of PSTHs show the responses to 16 combinations of 2 stimulus attributes, averaged across the third (for example, all pitch and timbre combinations, averaged across all azimuth conditions). The cells to the right of each row and at the bottom of each column show "main effects". The bottom right PSTH shows the grand average response (purple line), across all 64 stimuli. The color scale underlay in each "main effect" PSTH represents the difference between that stimulus condition and the grand average PSTH, with blue indicating a relative decrease in firing rate, and red showing a relative increase. The color scale underlay in the 16 "interaction" PSTHs show non-linear combination sensitivity - that is, the difference between the response to this stimulus combination and the response predicted from a linear combination of the two main effects.

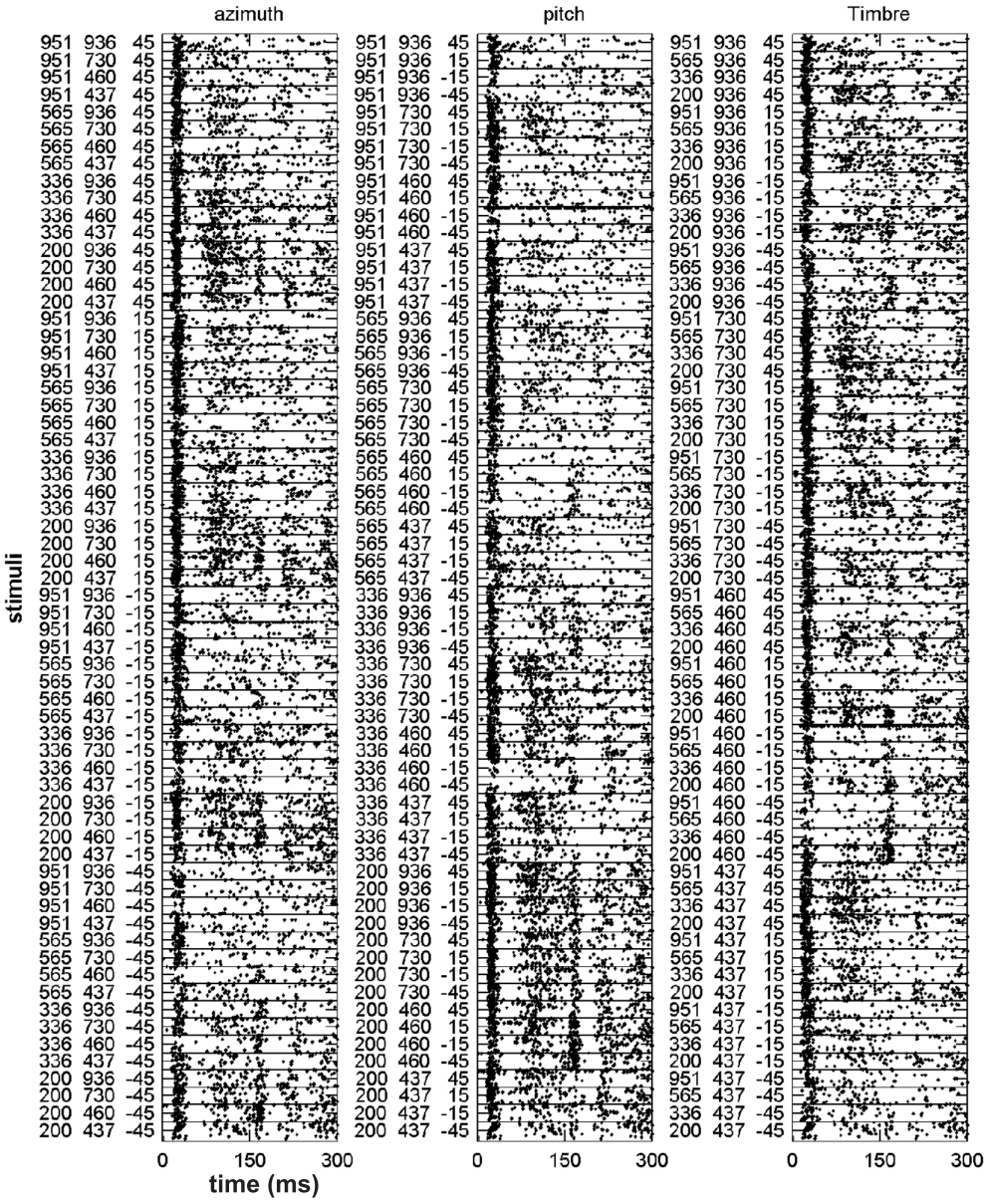


Figure B.1: Raster plot of the responses of a neural unit in ferret A1 to the presentation of 64 artificial vowel sounds.

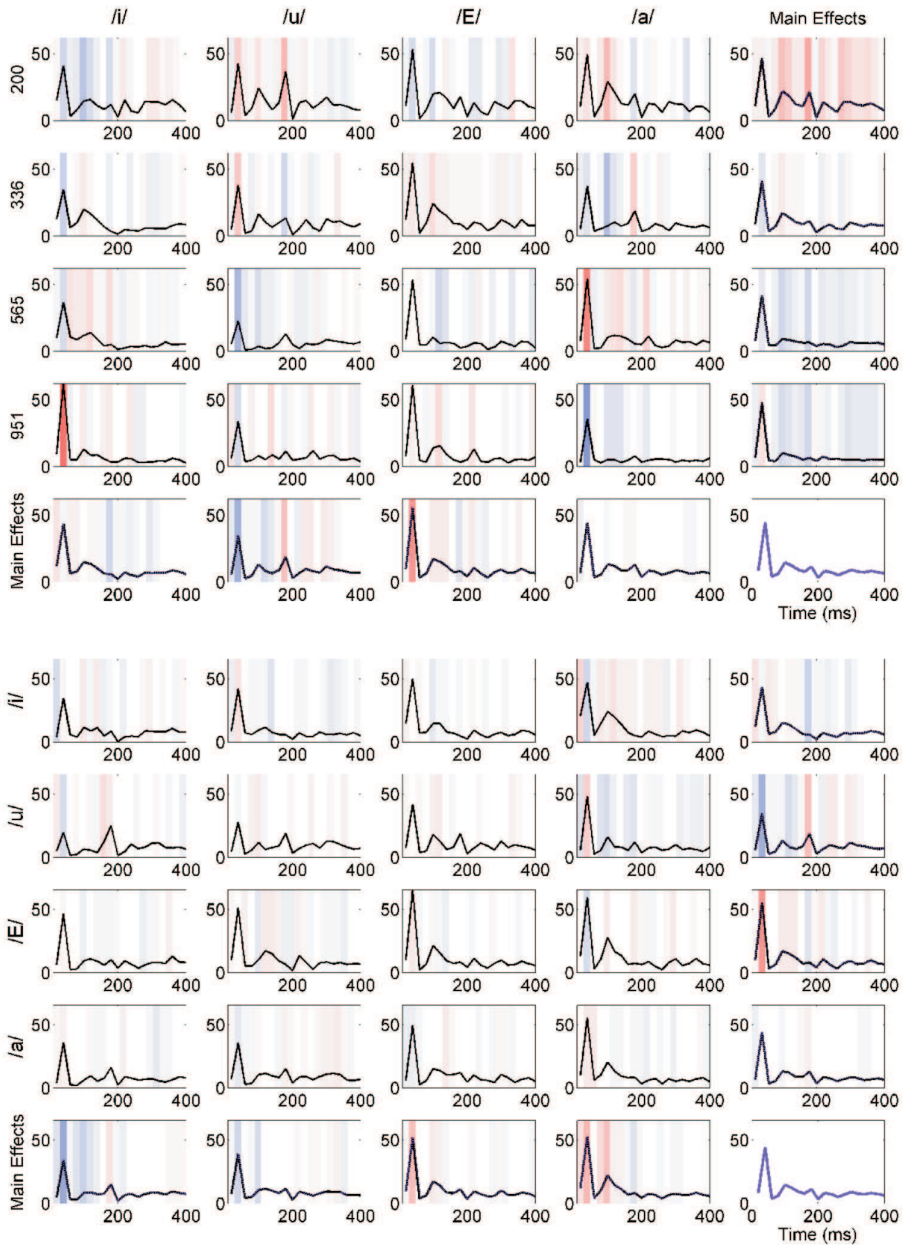


Figure B.2: PSTH descriptions of the responses of a neural unit in ferret A1 to the presentation of 64 artificial vowel sounds.

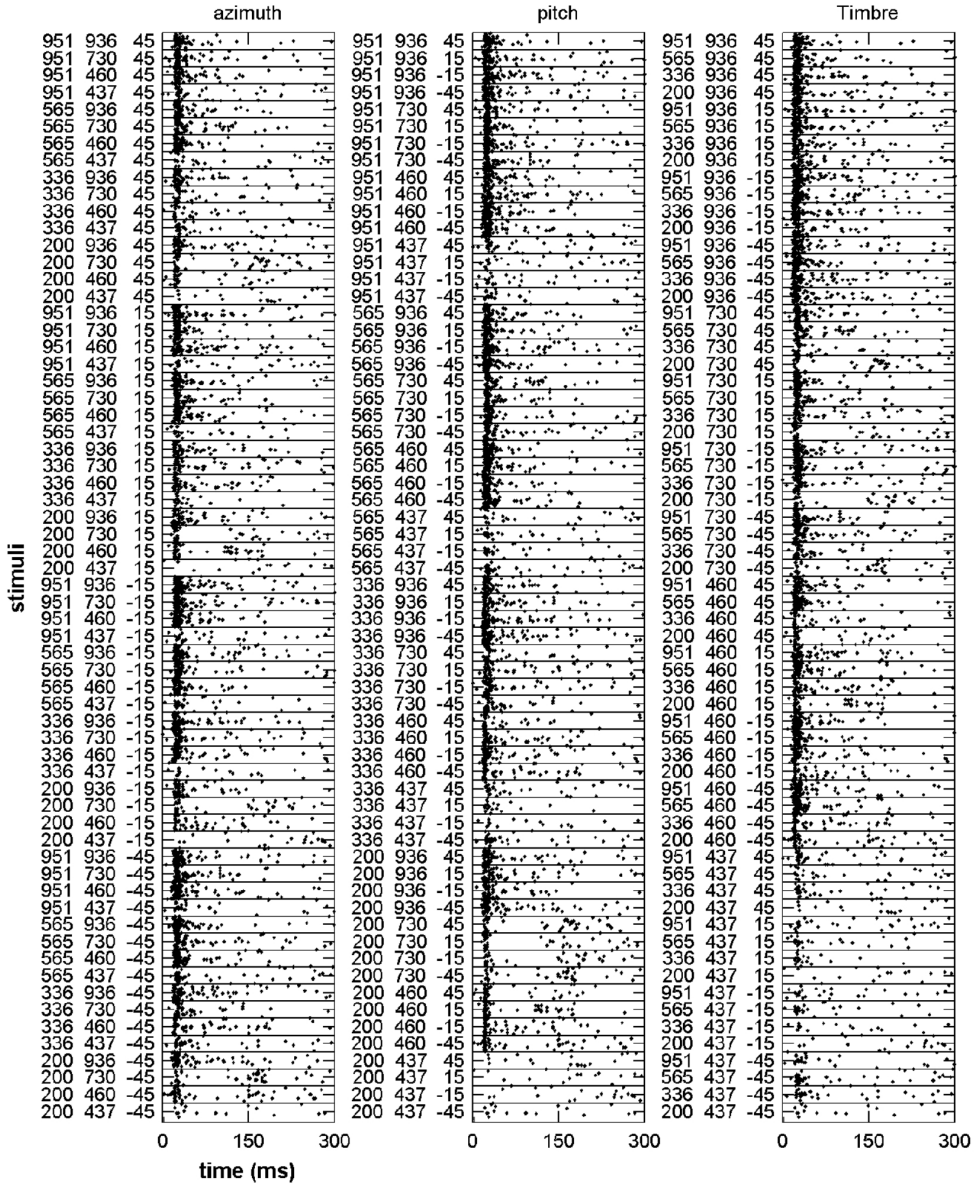


Figure B.3: Raster plot of the responses of a neural unit in ferret AAF to the presentation of 64 artificial vowel sounds.

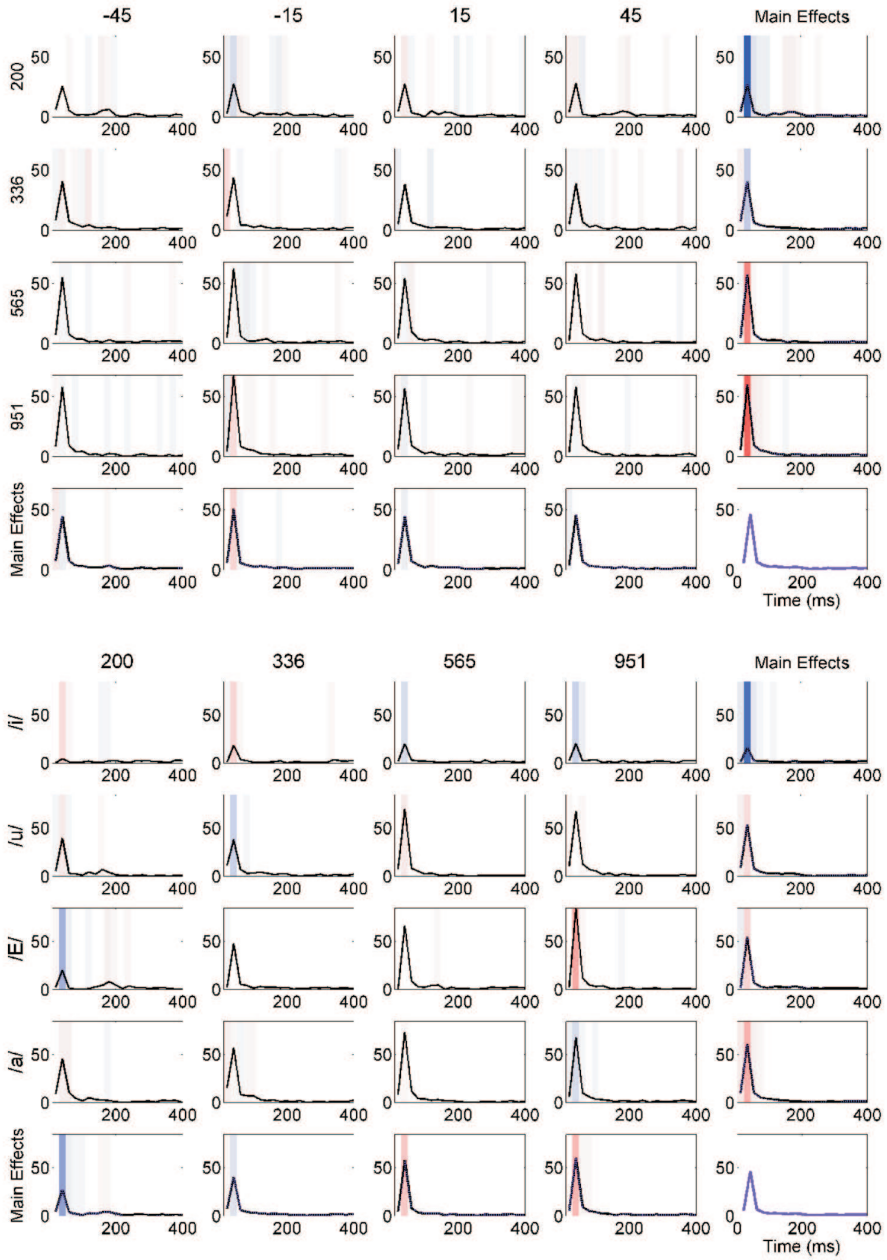


Figure B.4: PSTH descriptions of the responses of a neural unit in ferret AAF to the presentation of 64 artificial vowel sounds.

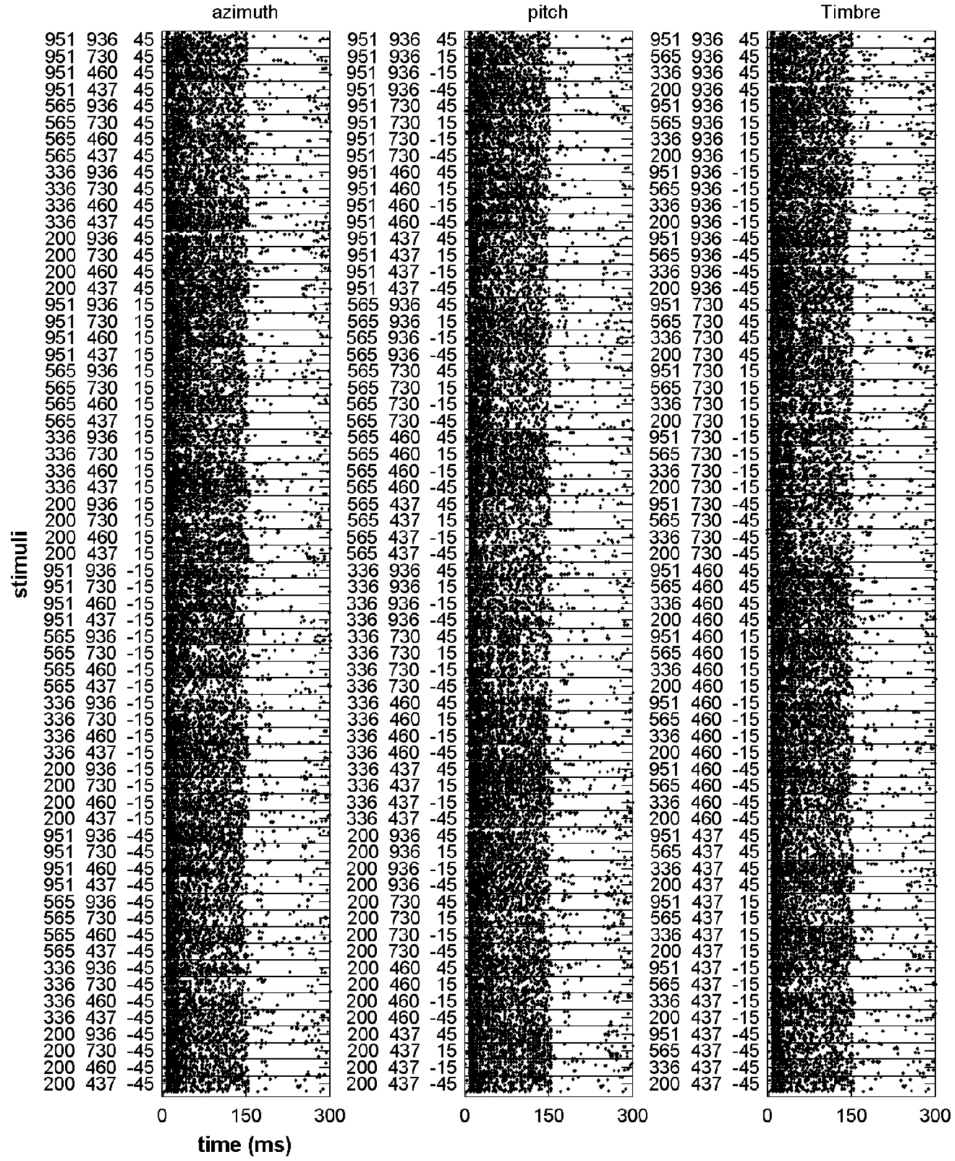


Figure B.5: Raster plot of the responses of a neural unit in ferret PPF to the presentation of 64 artificial vowel sounds.

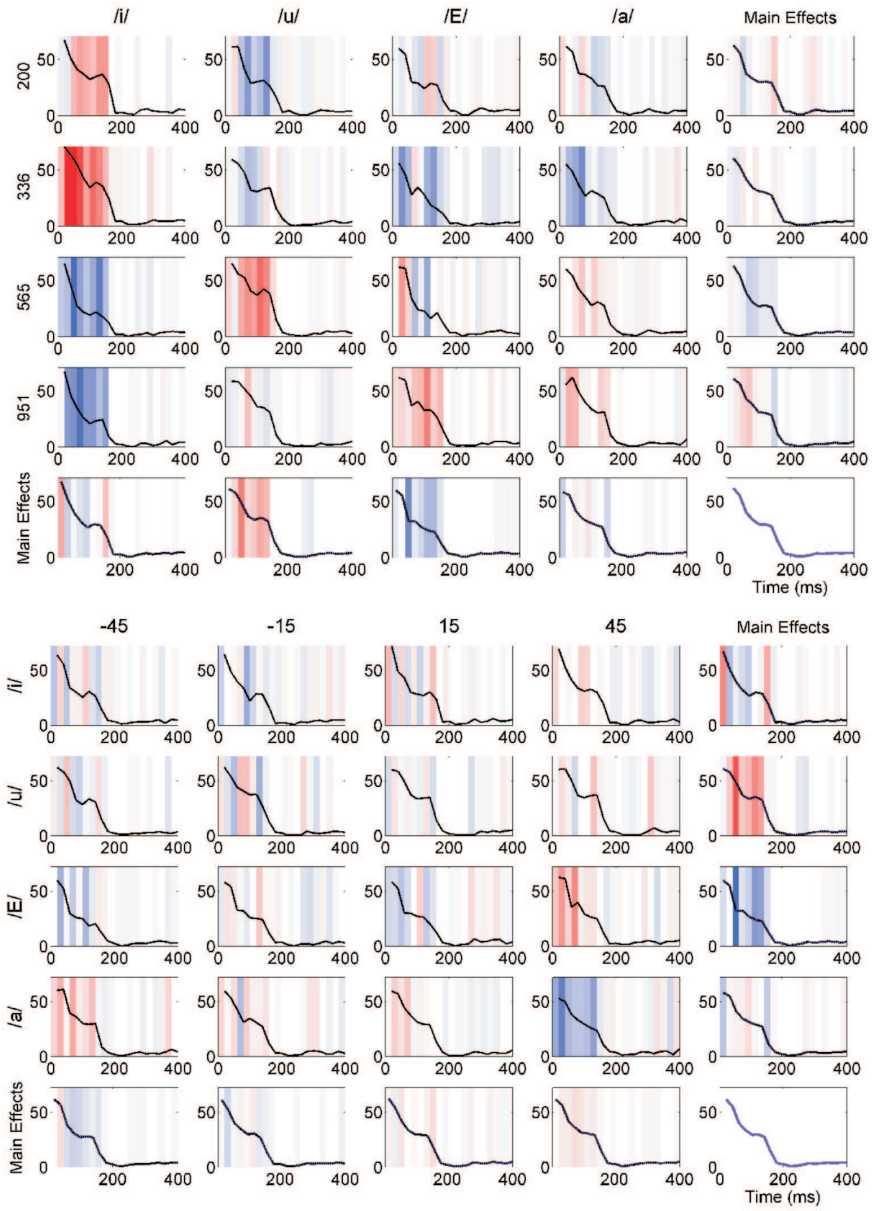


Figure B.6: PSTH descriptions of the responses of a neural unit in ferret PPF to the presentation of 64 artificial vowel sounds.

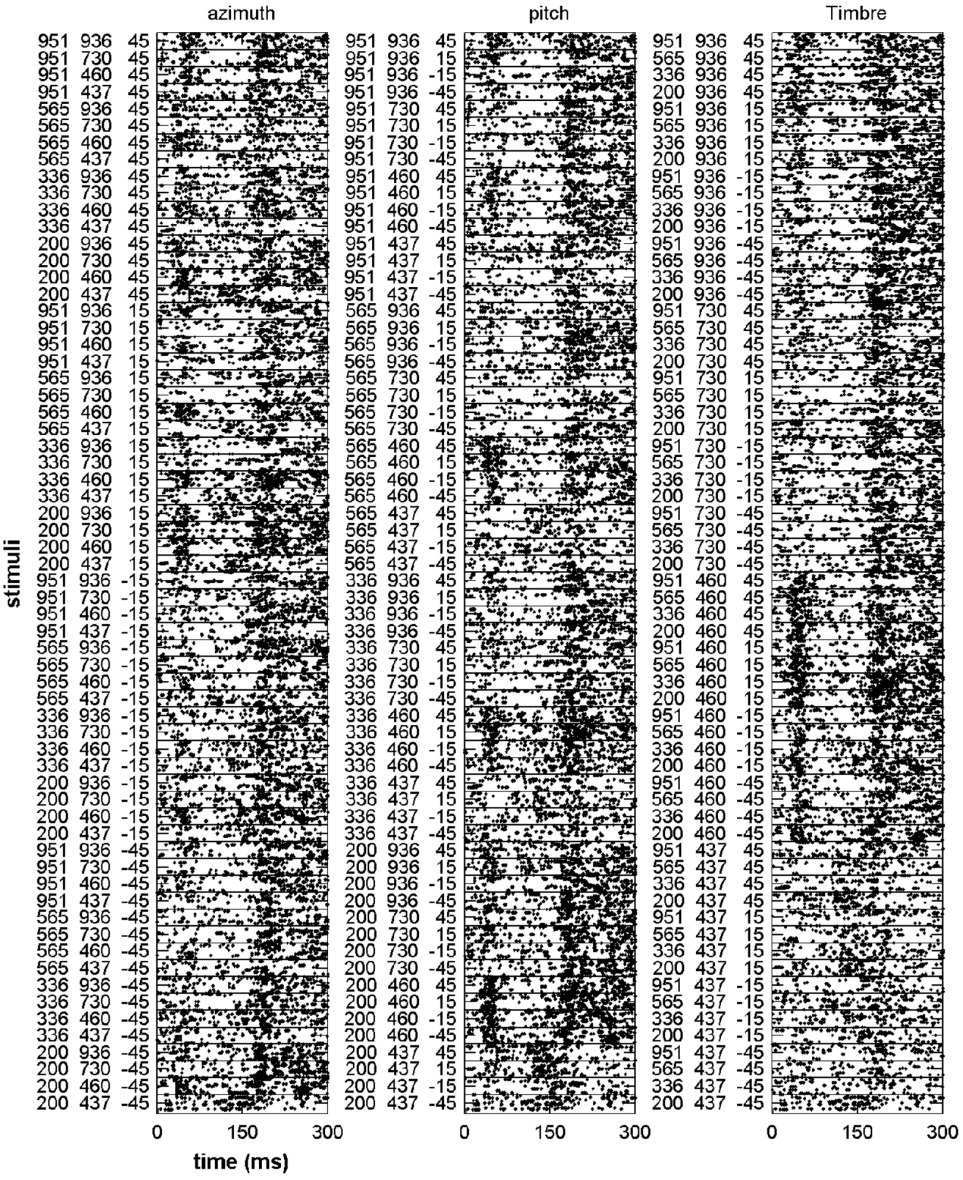


Figure B.7: Raster plot of the responses of a neural unit in ferret PSF to the presentation of 64 artificial vowel sounds.

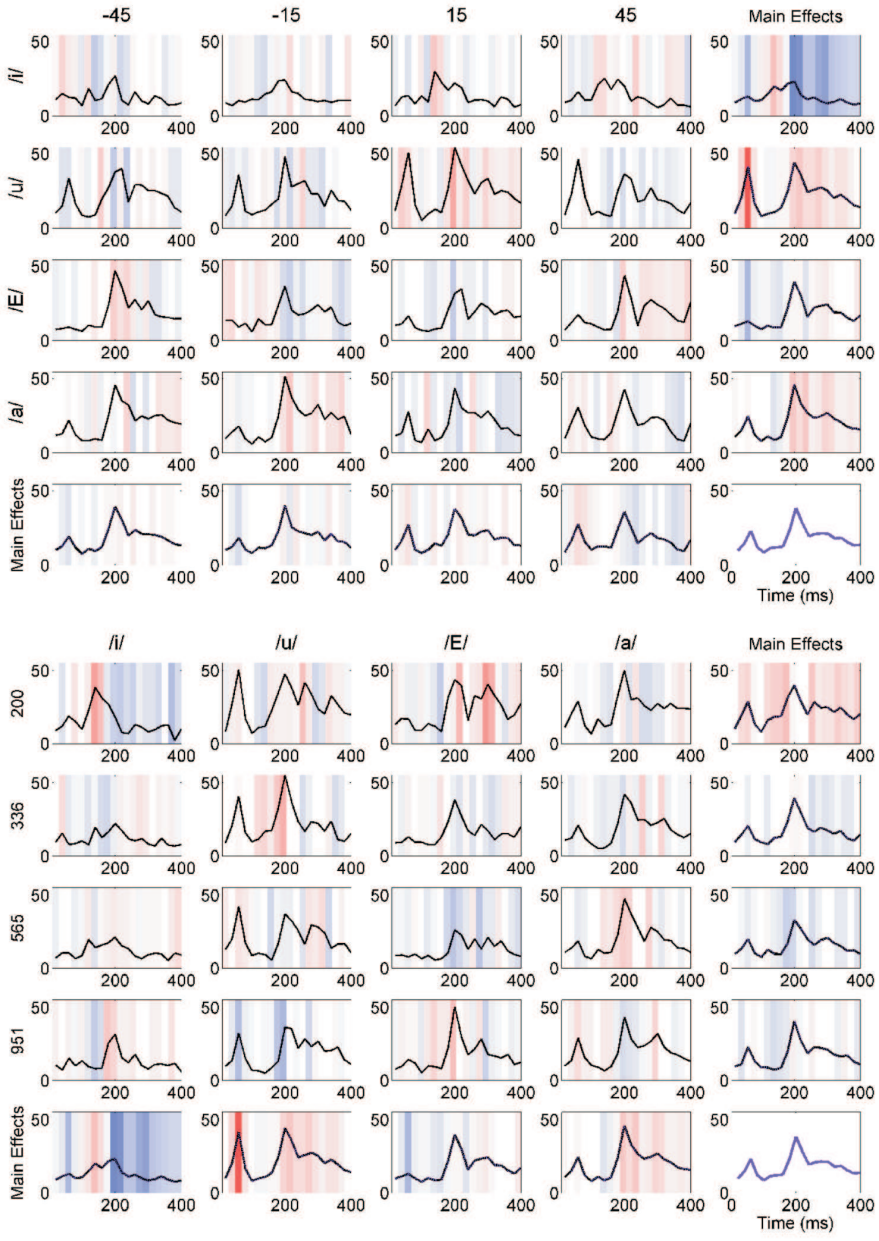


Figure B.8: PSTH descriptions of the responses of a neural unit in ferret PSF to the presentation of 64 artificial vowel sounds.

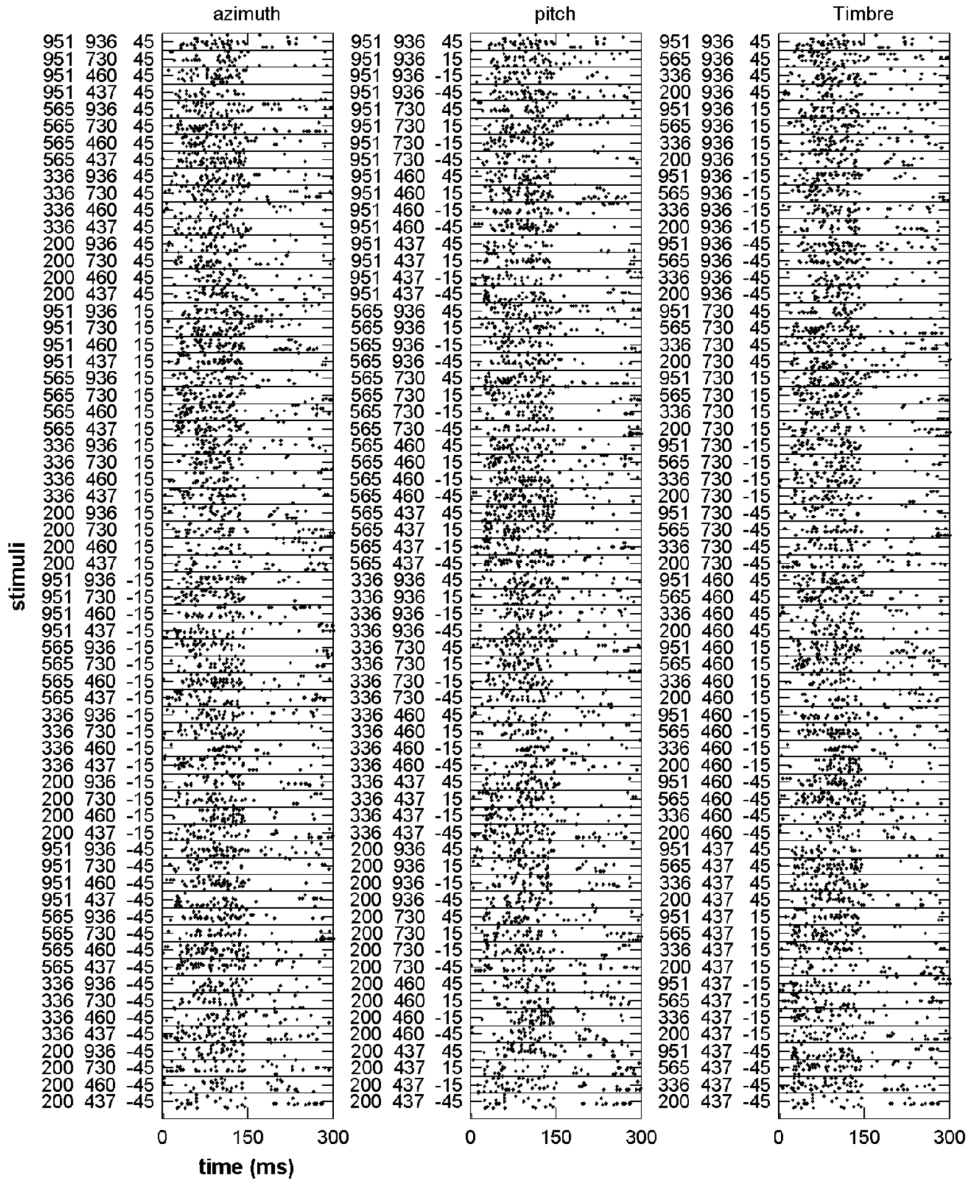


Figure B.9: Raster plot of the responses of a neural unit in ferret ADF to the presentation of 64 artificial vowel sounds.

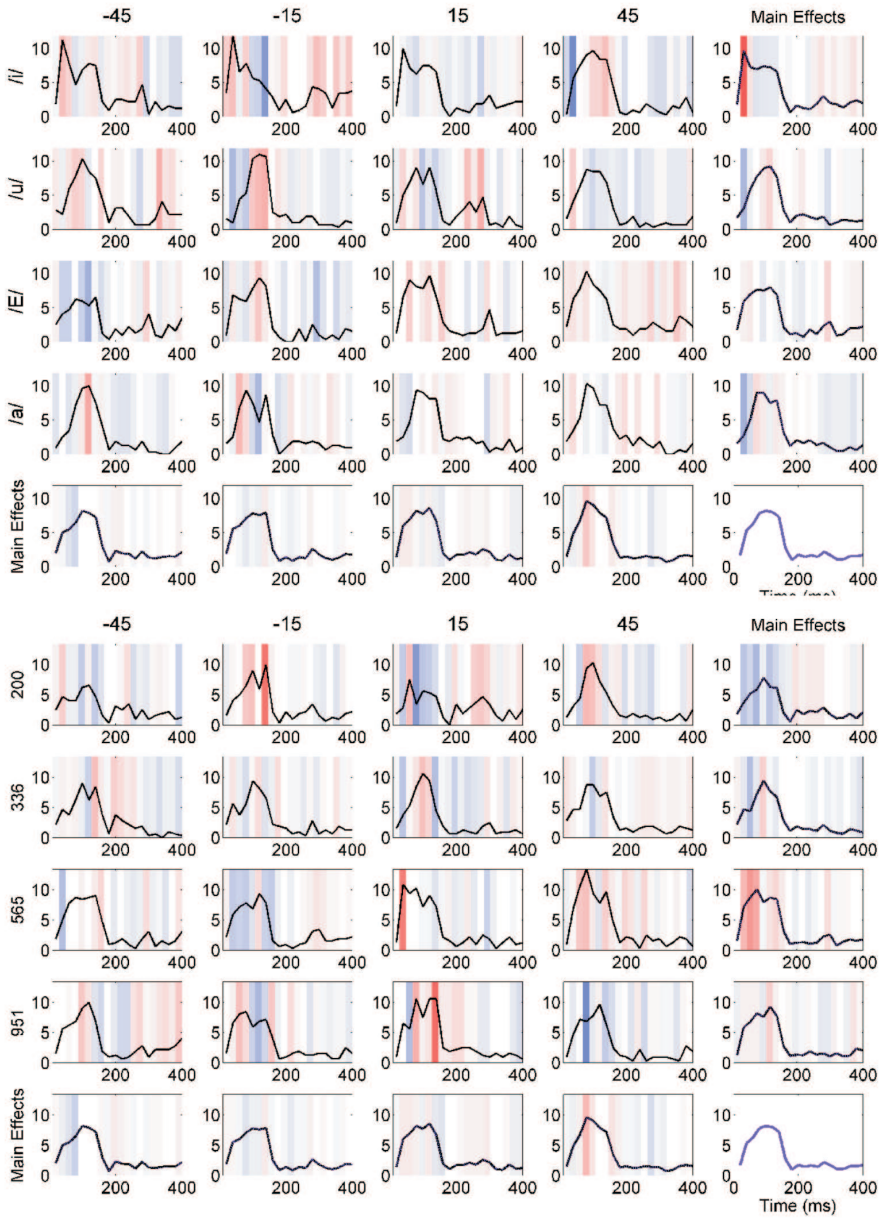


Figure B.10: PSTH descriptions of the responses of a neural unit in ferret ADF to the presentation of 64 artificial vowel sounds.

References

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F. H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., and Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proc Natl Acad Sci U S A*, 103(39):14608–13.

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc Natl Acad Sci U S A*, 98(21):12301–6.

Anderson, L. A., Malmierca, M. S., Wallace, M. N., and Palmer, A. R. (2006). Evidence for a direct, short latency projection from the dorsal cochlear nucleus to the auditory thalamus in the guinea pig. *Eur J Neurosci*, 24(2):491–8.

ANSI (1994). *American National Standards Acoustical Terminology*. American National Standards Institute, New York, NY.

Bajo, V. M., Nodal, F. R., Bizley, J. K., Moore, D. R., and King, A. J. (2007). The ferret auditory cortex: descending projections to the inferior colliculus. *Cereb Cortex*, 17(2):475–91.

Banai, K. and Ahissar, M. (2004). Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiol Neurootol*, 9(6):328–40.

Barlow, H. B. (1972). Single units and sensation: a neuron doctrine for perceptual psychology? *Perception*, 1(4):371–94.

Barrett, D. J. and Hall, D. A. (2006). Response preferences for "what" and "where" in human non-primary auditory cortex. *Neuroimage*, 32(2):968–77.

Belin, P. and Zatorre, R. J. (2000). 'What', 'where' and 'how' in auditory cortex. *Nat Neurosci*, 3(10):965–6.

Bendor, D. and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, 436(7054):1161–5.

Bendor, D. and Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Curr Opin Neurobiol*, 16(4):391–9.

Bendor, D. A. and Wang, X. (2007). Differential neural coding of acoustic flutter within primate auditory cortex. *Nat Neurosci*, 10:763Ű71.

Bizley, J. K. and King, A. J. (2008). Visual-auditory spatial processing in auditory cortical neurons. *Brain Res*, 1242:24–36.

Bizley, J. K., Nodal, F. R., Bajo, V. M., Nelken, I., and King, A. J. (2007a). Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cereb Cortex*, 17(9):2172–89.

Bizley, J. K., Nodal, F. R., Nelken, I., and King, A. J. (2005). Functional organization of ferret auditory cortex. *Cereb Cortex*, 15(10):1637–53.

Bizley, J. K., Nodal, F. R., Parsons, C. H., and King, A. J. (2007b). Role of auditory cortex in sound localization in the midsagittal plane. *J Neurophysiol*, 98(3):1763–74.

Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J., and Schnupp, J. W. H. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *J Neurosci*, 29(7):2064–75.

Blackburn, C. C. and Sacks, M. B. (1990). The represenation of the steady-state vowel sound /*e*/ in the discharge patterns of cat anteroventral cochlear nucleus neurons. *J Neurophysiol*, 63:1191–211.

Boehnke, S. E. and Phillips, D. P. (2005). The relation between auditory temporal interval processing and sequential stream segregation examined with stimulus laterality differences. *Percept Psychophys*, 67(6):1088–10.

Boemio, A., Fromm, S., Braun, A., and Poeppel, D. (2005). Heirarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci*, 8(3):389–95.

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference. *Nature*, 417(6888):543–7.

Brawer, J. R., Morest, D. K., and Kane, E. C. (1974). The neuronal architecture of the cochlear nucleus of the cat. *J Comp Neurol*, 155:251–300.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press, Cambridge, MA.

Britten, K. H., Newsome, W. T., and Saunders, R. C. (1992a). Effects of inferotemporal cortex lesions on form-from-motion discrimination in monkeys. *Exp Brain Res*, 88(2):292–302.

Britten, K. H., Shadlen, M. N., Newsome, W. T., and Movshon, J. A. (1992b). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J Neurosci*, 12(12):4745–65.

Brosch, M., Selezneva, E., Bucks, C., and Scheich, H. (2004). Macaque monkeys discriminate pitch relationships. *Cognition*, 91(3):259–72.

Brugge, J. F. and Merzenich, M. M. (1973). Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J Neurophysiol*, 36(6):1138–58.

Brugge, J. F., Reale, R. A., and Hind, J. E. (1996). The structure of spatial receptive fields of neurons in primary auditory cortex of the cat. *J Neurosci*, 16(14):4420–37.

Burdick, C. K. (1979). The effect of behavioural paradigm on auditory discrimination learning: A literature review. *J Aud Res*, 19(1):59–82.

Burdick, C. K. (1980). Auditory discrimination learning by the chinchilla: comparison of go/no go and two-choice procedures. *J Aud Res*, 20(1):1–29.

Burdick, C. K. and Miller, J. D. (1975). Speech perception by the chinchilla: discrimination of sustained /a/ and /i/. *J Acoust Soc Am*, 58(2):415–27.

Campbell, R. A., Schnupp, J. W. H., Shial, A., and King, A. J. (2006). Binaural-level functions in ferret auditory cortex: evidence for a continuous distribution of response properties. *J Neurophysiol*, 95(6):3742–55.

Capranica, R. R. (1966). Vocal response of the bullfrog to natural and synthetic mating calls. *J Acoust Soc Am*, 40:1131–9.

Cariani, P. (1999). Temporal coding of periodicity pitch in the auditory system: an overview. *Neural Plast*, 6(4):147–72.

Cariani, P. A. and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones i. pitch and pitch salience. *J Neurophysiol*, 76:1698–716.

Cariani, P. A. and Delgutte, B. (1996b). Neural correlates of the pitch of complex tones ii. pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. *J Neurophysiol*, 76:1717–34.

Cedolin, L. and Delgutte, B. (2005). Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *J Neurophysiol*, 94(1):347–62.

Chase, S. M. and Young, E. D. (2005). Limited segregation of different types of sound localization information among classes of units in the inferior colliculus. *J Neurosci*, 25(33):7575–85.

Chase, S. M. and Young, E. D. (2007). First-spike latency information in single neurons increases when referenced to population onset. *Proc Natl Acad Sci U S A*, 104(12):5175–80.

Cheung, S. W., Bedenbaugh, P. H., Nagarajan, S. S., and Schreiner, C. E. (2001). Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. *J Neurophysiol*, 85(4):1732–49.

Clugnet, M. C., LeDoux, J. E., and Morrison, S. F. (1990). Unit responses evoked in the amygdala and striatum by electrical stimulation of the medual geniculate body. *J Neurosci*, 10(4):1055–61.

Cynx, J. (1995). Similarities in absolute and relative pitch perception in songbirds (starling and zebra finch) and a nonsongbird (pigeon). *J Comp Psychol*, 109:261–7.

Dahmen, J. C. and King, A. J. (2007). Learning to hear: plasticity of auditory cortical processing. *Curr Opin Neurobiol*, 17(4):456–64.

D'Amato, M. R. (1988). A search for tonal pattern perception in cebus monkey: why monkeys can't hum a tune. *Music Perception*, 5:452–80.

de Lafuente, V. and Romo, R. (2006). Neural correlate of subjective sensory experience gradually builds up across cortical areas. *Proc Natl Acad Sci U S A*, 103(39):14266–71.

Depireux, D. A., Simon, J. Z., Klein, D. J., and Shamma, S. A. (2001). Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *J Neurophysiol*, 85(3):1220–34.

DeWeese, M. R., Wehr, M., and Zador, A. M. (2003). Binary spiking in auditory cortex. *J Neurosci*, 23(21):7940–9.

Dobrzecka, C. and Konorski, J. (1968). Qualitative versus directional cues in differential conditioning. IV. Left leg - right leg differentiation to nondirectional cues. *Acta Biol Exper (Warsaw)*, 28:61–9.

Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (2002). Auditory temporal resolution in birds: discrimination of harmonic complexes. *J Acoust Soc Am*, 112(2):748–59.

Dooling, R. J. and Saunders, J. C. (1975). Hearing in the parakeet (*Melopsittacus undulatus*): absolute thresholds, critical ratios, frequency difference limens, and vocalizations. *J Comp Physiol Psychol*, 88(1):1–20.

Dooling, R. J. and Searcy, M. H. (1981). Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*). *J Comp Physiol*, 143(3):383–8.

Efron, R. (1963). Temporal perception, aphasia and deja vu. *Brain*, 86:403–24.

Eggermont, J. J. (1995). Representation of a voice onset time continuum in primary auditory cortex of the cat. *J Acoust Soc Am*, 98(2 Pt 1):911–20.

Ehret, G. (1975). Frequency and intensity difference limens and nonlinearities in the ear of the housemouse (*Mus. musculus*). *J Comp Physiol*, 102:321–36.

Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., and Shamma, S. A. (2004). Dynamics of precise spike timing in primary auditory cortex. *J Neurosci*, 24(5):1159–72.

Elliott, D. N., Frazier, L., and Riach, W. (1962). A tracking procedure for determining the cat's frequency discrimination. *J Exp Anal Behav*, 5:323–8.

Elliott, D. N., Frazier, L. A., and Haydon, R. C. (1971). Relational and absolute cues in auditory discrimination by monkeys. *Percept Psychophys*, 10:278–82.

Elliott, D. N., Stein, L., and Harrison, M. J. (1960). Determination of absolute intensity thresholds and frequency difference thresholds in cats. *J Acoust Soc Am*, 32(3):380–84.

Eramudugolla, R., McAnally, K. I., Martin, R. L., Irvine, D. R., and Mattingley, J. B. (2008). The role of spatial location in auditory search. *Hear Res*, 238(1-2):139–46.

Farb, C. R. and LeDoux, J. E. (1997). NMDA and AMPA receptors in the lateral nucleus of the amygdala are postsynaptic to auditory thalamic afferents. *Synapse*, 27(2):106–21.

Farley, S. D., Lehner, P. N., Clark, T., and Trost, C. (1987). Vocalizations of the Siberian ferret (*Mustela eversmanni* and comparisons with other mustelids). *J Mamm*, 68(2):413–416.

Fay, R. R. (1988a). Comparative psychoacoustics. *Hear Res*, 34(3):295–305.

Fay, R. R. (1988b). *Hearing in Vertebrates: A Psychophysics Databook*. Hill-Fay Associates, Winnetka, Il.

Fay, R. R. and Coombs, S. L. (1992). Psychometric functions for level discrimination and the effects of signal duration in the goldfish (*Carassius auratus*): psychophysics and neurophysiology. *J Acoust Soc Am*, 92(1):189–201.

Fay, R. R., Yost, W. A., and Coombs, S. (1983). Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system. *Hear Res*, 12(1):31–55.

Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (1998). Pitch vs. spectral encoding of harmonic complex tones in primary auditory cortex of the awake monkey. *Brain Res*, 786(1-2):18–30.

Flanagan, J. L. and Saslow, M. G. (1958). Pitch discrimination for synthetic vowels. *J Acoust Soc Am*, 30(5):435–42.

Formby, C. (1985). Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners. *J Acoust Soc Am*, 78(1):70–7.

Fritz, J. B., Elhilali, M., David, S. V., and Shamma, S. A. (2007a). Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1? *Hear Res*, 229(1-2):186–203.

Fritz, J. B., Elhilali, M., and Shamma, S. A. (2005a). Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hear Res*, 206(1-2):159–76.

Fritz, J. B., Elhilali, M., and Shamma, S. A. (2005b). Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *J Neurosci*, 25(33):7623–35.

Fritz, J. B., Elhilali, M., and Shamma, S. A. (2007b). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J Neurophysiol*, 98(4):2337–46.

Fritz, J. B., Shamma, S. A., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci*, 6(11):1216–23.

Fujita, S. and Ito, J. (1999). Ann otol rhin laryngol. *J Neurosci*, 20(3):1216–28.

Fuller, B. F., Horii, Y., and Conner, D. A. (1992). Validity and reliability of nonverbal voice measures as indicators of stressor-provoked anxiety. *Res Nurs Health*, 15(5):379–89.

Furukawa, S. and Middlebrooks, J. C. (2002). Cortical representation of auditory space: information-bearing features of spike patterns. *J Neurophysiol*, 87(4):1749–62.

Furukawa, S., Xu, L., and Middlebrooks, J. C. (2000). Coding of sound-source location by ensembles of cortical neurons. *J Neurosci*, 20(3):1216–28.

Garcia-Lazaro, J. A., Ahmed, B., and Schnupp, J. W. (2006). Tuning to natural stimulus dynamics in primary auditory cortex. *Curr Biol*, 16(3):264–71.

Garde, M. M. and Cowey, A. (2000). "Deaf hearing": unacknowledged detection of auditory stimuli in a patient with cerebral deafness. *Cortex*, 36(1):71–80.

Gehr, D. D., Komiya, H., and Eggermont, J. J. (2000). Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. *Hear Res*, 150(1-2):27–42.

Gelfer, M. and Mikos, V. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J Voice*, 19(4):544–54.

Gfeller, K. E., Turner, C., Woodworth, G., Mehr, M., Fearn, R., Knutson, J., Witt, S., and Stordahl, J. (2002). Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults. *Coch Implants Int*, 3:29–53.

Ghazanfar, A. A., Smith-Rohrberg, D., and Hauser, M. D. (2001). The role of temporal cues in rhesus monkey vocal recognition: orienting asymmetries to reversed calls. *Brain Behav Evol*, 58(3):163–72.

Goense, J. B. and Logothetis, N. K. (2008). Neurophysiology of the bold fmri signal in awake monkeys. *Curr Biol*, 18(9):631–40.

Green, D. M. and Swets, J. A. (1974). *Signal detection theory and psychophysics*. Kreiger, New York.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J Acoust Soc Am*, 61(5):1270–7.

Griffiths, T. D. (2003). Functional imaging of pitch analysis. *Ann N Y Acad Sci*, 999:40–9.

Griffiths, T. D. (2005). Functional imaging of pitch perception. In Plack, C., Oxenham, A., Fay, R. R., and Popper, A., editors, *Pitch: Neual Coding and Perception*, Springer

Handbook of Auditory Research, pages 147–168. Springer Science+Business Media, Inc., New York.

Griffiths, T. D. (2008). Sensory systems: auditory action streams? *Curr Biol*, 18(9):R387–8.

Griffiths, T. D., Buchel, C., Frackowiak, R. S., and Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nat Neurosci*, 1(5):422–7.

Gulyas, B., Heywood, C. A., Popplewell, D. A., Roland, P. E., and Cowey, A. (1994). Visual form discriminaiton from color or motion cues: functional anatomy by positron emission tomography. *Proc Natl Acad Sci U S A*, 91(21):9965–9.

Gutschalk, A., Patterson, R. D., Rupp, A., Uppenkamp, S., and Scherg, M. (2002). Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. *Neuroimage*, 15(1):207–16.

Hall, D. A., Edmondson-Jones, A. M., and Fridriksson, J. (2006). Periodicity and frequency coding in human auditory cortex. *Eur J Neurosci*, 24(12):3601–10.

Hall, D. A. and Plack, C. J. (2007). The human 'pitch center' responds differently to iterated noise and Huggins pitch. *Neuroreport*, 18(4):323–7.

Hall, J. L. and Goldstein, M. H. J. (1968). Representation of binaural stimuli by single units in primary auditory cortex of unanesthetized cats. *J Acoust Soc Am*, 43(3):456–61.

Hanley, J. A. and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36.

Hayar, A., Bryant, J. L., Boughter, J. D., and Heck, D. H. (2006). A low-cost solution to measure mouse licking in an electrophysiological setup with a standard analog-to-digital converter. *J Neurosci Methods*, 153(2):203–7.

Heffner, H. E. (1997). The role of macaque auditory cortex in sound localization. *Acta Otolaryngol Suppl*, 532:22–7.

Heffner, H. E. and Heffner, R. S. (1986). Effect of unilateral and bilateral auditory cortex lesions on the discrimination of vocalizations by Japanese macaques. *J Neurophysiol*, 56(3):683–701.

Heffner, H. E. and Masterton, B. (1975). Contribution of auditory cortex to sound localization in the monkey (*Macaca mulatta*). *J Neurophysiol*, 38(6):1340–58.

Heffner, H. E., Ravizza, R. J., and Masterton, B. (1969a). Hearing in primate mammals III: Tree shrew (*Tupaia glis*). *J Aud Res*, 9:12–18.

Heffner, H. E., Ravizza, R. J., and Masterton, B. (1969b). Hearing in primitive mammals IV: Bushbaby (*Galago senegalensis*). *J Aud Res*, 9:19–23.

Heffner, R. S., Heffner, H., and Masterton, B. (1971). Behavioral measurements of absolute and frequency-difference thresholds in guinea pig. *J Acoust Soc Am*, 49(6):1888–95.

Heil, P., Rajan, R., and Irvine, D. R. (1992). Sensitivity of neurons in cat primary auditory cortex to tones and frequency-modulated stimuli. II: Organization of response properties along the 'isofrequency' dimension. *Hear Res*, 63(1-2):135–56.

Heil, P., Rajan, R., and Irvine, D. R. (1994). Topographic representation of tone intensity along the isofrequency axis of cat primary auditory cortex. *Hear Res*, 76:188–202.

Helmholtz, H. L. F. (1954). *On the Sensations of Tone*. Dover Publications, Mineola, New York.

Henning, G. B. and Grosberg, S. L. (1968). Effect of harmonic components on frequency discrimination. *J Acoust Soc Am*, 44(5):1386–9.

Hernandez, A., Zainos, A., and Romo, R. (2000). Neuronal correlates of sensory discrimination in the somatosensory cortex. *Proc Natl Acad Sci U S A*, 97(11):6191–6.

Hill, K. G., Stange, G., and Mo, J. (1989). Temporal synchronization in the primary auditory response in the pigeon. *Hear Res*, 39(1-2):63–73.

Hine, J. E., Martin, R. L., and Moore, D. R. (1994). Free-field binaural unmasking in ferrets. *Behav Neurosci*, 108(1):196–205.

Hirsh, I. J. (1959). Auditory perception of temporal order. *J Acoust Soc Am*, 31:759–767.

Hromàdka, T. and Zador, A. M. (2007). Toward the mechanisms of auditory attention. *Hear Res*, 229(1-2):180–5.

Imig, T. J. and Adrian, H. O. (1977). Binaural columns in the primary field (A1) of cat auditory cortex. *Brain Res*, 138:241–57.

Irvine, D. R. and Gago, F. (1990). Binaural interaction in high-frequency neurons in inferior colliculus of the cat: effects of variations in sound pressure level on sensitivity to interaural intensity differences. *J Neurophysiol*, 63:570–91.

Izumi, A. (2001). Relative pitch perception in Japanese monkeys (*Macaca fuscata*). *J Comp Psychol*, 115:127–31.

Jäncke, L., Mirzazade, S., and Shah, N. J. (1999). Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci Lett*, 266(2):125–8.

Javel, E. and Viemeister, N. F. (2000). Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination. *J Acoust Soc Am*, 107(2):908–21.

Johnson, D. (1980). The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones. *J Acoust Soc Am*, 68:1115–22.

Johnson, K. (1990). The role of perceived speaker identity in f0 normalization of vowels. *J Acoust Soc Am*, 88(2):642–54.

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitudemodulated sounds. *Physiol Rev*, 84:541–77.

Kaas, J. H. and Hackett, T. A. (1999). 'What' and 'where' processing in auditory cortex. *Nat Neurosci*, 2(12):1045–7.

Kaas, J. H. and Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A*, 97(22):11793–9.

Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (2006). Training-induced plasticity of auditory localization in adult mammals. *PLoS Biol*, 4(4):e71.

Kavanagh, G. L. and Kelly, J. B. (1987). Contributions of auditory cortex to sound localisation in the ferret (*Mustela putorius*). *J Neurophysiol*, 57:1746–66.

Keeling, M. D., Calhoun, B. M., Krüger, K., Polley, D., and Schreiner, C. E. (2008). Spectral integration plasticity in cat auditory cortex induced by perceptual training. *Exp Brain Res*, 184(4):493–509.

Kelly, J. B. and Judge, P. W. (1994). Binaural organization of primary auditory cortex in the ferret (*Mustela putorius*). *J Neurophysiol*, 71:904–13.

Kelly, J. B., Judge, P. W., and Phillips, D. P. (1986a). Representation of the cochlea in primary auditory cortex of the ferret (*Mustela putorius*). *Hear Res*, 24(2):111–5.

Kelly, J. B. and Kavanagh, G. L. (1994). Sound localization after unilateral lesions of inferior colliculus in the ferret (*Mustela putorius*). *J Neurophysiol*, 71(3):1078–87.

Kelly, J. B., Kavanagh, G. L., and Dalton, J. C. (1986b). Hearing in the ferret (*Mustela putorius*): thresholds for pure tone detection. *Hear Res*, 24(3):269–75.

Kelly, J. B., Rooney, B. J., and Phillips, D. P. (1996). Effects of bilateral auditory cortical lesions on gap-detection thresholds in the ferret (*Mustela putorius*). *Behav Neurosci*, 110(3):542–50.

Kim, D. O., Chang, S. O., and Sirianni, J. G. (1990). A population study of auditory nerve fibers in unanaesthetized decerebrate cats: responses to pure tones. *J Acoust Soc Am*, 87:1648–55.

Kim, D. O. and Parham, K. (1990). Auditory nerve spatial encoding of high frequency pure tones: population response profiles derived from d' measure associated associated with nearby places along the cochlea. *Hear Res*, 52:167–80.

King, A. J., Schnupp, J. W. H., and Doubell, T. P. (2001). The shape of ears to come: dynamic coding of auditory space. *Trends Cog Sci*, 5:261–70.

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception. *J Acoust Soc Am*, 53(1):8–16.

Koda, H. and Masataka, N. (2002). A pattern of common acoustic modification by human mothers to gain attention of a child and by macaques of others in their group. *Psychol Rep*, 91(2):421–2.

Kojima, S., Izumi, A., and Ceugniet, M. (2003). Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates*, 44(3):225–30.

Kowalski, N., Versnel, H., and Shamma, S. A. (1995). Comparison of responses in the anterior and primary auditory fields of the ferret cortex. *J Neurophysiol*, 73(4):1513–23.

Krug, K., Cumming, B. G., and Parker, A. J. (2004). Comparing perceptual signals of single V5/MT neurons in two binocular depth tasks. *J Neurophysiol*, 92(3):1586–96.

Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., and Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cereb Cortex*, 13(7):765–72.

Kudoh, M., Nakayama, Y., Hishida, R., and Shibuki, K. (2006). Requirement of the auditory association cortex for discrimination of vowel-like sounds in rats. *Neuroreport*, 17(17):1761–6.

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., and Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Comput Biol*, 3(6):e100.

Langner, G., Albert, M., and Briede, T. (2002). Temporal and spatial coding of periodicity information in the inferior colliculus of the awake chinchilla (*Chinchilla laniger*). *Hear Res*, 168:110–30.

Langner, G., Sams, M., Heil, P., and Schulze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *J Comp Physiol [A]*, 181(6):665–76.

Langner, G. and Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. i. neuronal mechanisms. *J Neurophysiol*, 60:1799–822.

Levitin, D. J. (1999). Memory for musical attributes. In Cook, P. R., editor, *Music, Cognition and Computerized Sound: An Introduction to Psychoacoustics*, pages 209–227. MIT Press, Cambridge, MA.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol Rev*, 74:431–61.

Linden, J. F., Liu, R. C., Sahani, M., Schreiner, C. E., and Merzenich, M. M. (2003). Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *J Neurophysiol*, 90(4):2660–75.

Linden, J. F. and Schreiner, C. E. (2003). Columnar transformations in auditory cortex? a comparison to visual and somatosensory cortices. *Cereb Cortex*, 13(1):83–9.

Liu, J. and Newsome, W. T. (2005). Correlation between speed perception and neural activity in the middle temporal visual area. *J Neurosci*, 25(3):711–22.

Liu, L. F., Palmer, A. R., and Wallace, M. N. (2006). Phase-locked responses to pure tones in the inferior colliculus. *J Neurophysiol*, 95(3):1926–35.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412(6843):150–7.

Lomber, S. G. and Malhotra, S. (2008). Double dissociation of 'what' and 'where' processing in auditory cortex. *Nat Neurosci*, 11(5):609–16.

Lomber, S. G., Malhotra, S., and Hall, A. J. (2007). Functional specialization in nonprimary auditory cortex of the cat: areal and laminar contributions to sound localization. *Hear Res*, 229(1-2):31–45.

Long, G. R. and Clark, W. W. (1984). Detection of frequency and rate modulation by the chinchilla. *J Acoust Soc Am*, 75(4):1184–90.

Lu, T., Liang, L., and Wang, X. (2001a). Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. *J Neurophysiol*, 85(6):2364–80.

Lu, T., Liang, L., and Wang, X. (2001b). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat Neurosci*, 4(11):1131–8.

Luna, R., Hernandez, A., Brody, C. D., and Romo, R. (2005). Neural codes for perceptual discrimination in primary somatosensory cortex. *Nat Neurosci*, 8(9):1210–9.

Lütkenhöner, B. (2003). Single-dipole analyses of the n100m are not suitable for characterizing the cortical representation of pitch. *Audiol Neurootol*, 8(94):222–33.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage*, 14(4):802–16.

Malhotra, S., Hall, A. J., and Lomber, S. G. (2004). Cortical control of sound localization in the cat: unilateral cooling deactivation of 19 cerebral areas. *J Neurophysiol*, 92(3):1625–43.

Malhotra, S. and Lomber, S. G. (2007). Sound localization during homotopic and heterotopic bilateral cooling deactivation of primary and nonprimary auditory cortical areas in the cat. *J Neurophysiol*, 97(1):26–43.

Malhotra, S., Stecker, G. C., Middlebrooks, J. C., and Lomber, S. G. (2008). Sound localization deficits during reversible deactivation of primary auditory cortex and/or the dorsal zone. *J Neurophysiol*, 99(4):1628–42.

Malmierca, M. S., Merchan, M. A., Henkel, C. K., and Oliver, D. L. (2002). Direct projections from cochlear nuclear complex to auditory thalamus in the rat. *J Neurosci*, 22(24):10891–7.

Manger, P. R., Engler, G., Moll, C. K., and Engel, A. K. (2005). The anterior ectosylvian visual area of the ferret: a homologue for an enigmatic visual cortical area of the cat? *Eur J Neurosci*, 22(3):706–14.

Marvit, P. and Crawford, J. D. (2000). Auditory discrimination in a sound-producing electric fish (*Pollimyrus*): tone frequency and click-rate difference detection. *J Acoust Soc Am*, 108(4):1819–25.

May, B. J. and Sachs, M. B. (1992). Dynamic range of neural rate responses in the ventral cochlear nucleus of awake cats. *J Neurophysiol*, 68:1589–602.

McAlpine, D., Jiang, D., and Palmer, A. R. (2001). A neural code for low-frequency sound localization in mammals. *Nat Neurosci*, 4(4):396–401.

Mellot, J. G., Van der Gucht, E., Lee, C. C., Larue, D. T., Winer, J. A., and Lomber, S. G. (2005). Subdividing cat primary and non-primary auditory areas in the cerebrum with neurofilament proteins expressing SMI-32. In *Assoc Res Otolaryngol Abs*, page 348.

Merchant, H., Zainos, A., Hernandez, A., Salinas, E., and Romo, R. (1997). Functional properties of primate putamen neurons during the categorization of tactile stimuli. *J Neurophysiol*, 77(3):1132–54.

Merzenich, M. M., Knight, P., and Roth, G. L. (1975). Representation of the cochlea within the primary auditory cortex in the cat. *J Neurophysiol*, 38:231–49.

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2008). Phoneme representation and classification in primary auditory cortex. *J Acoust Soc Am*, 123(2):899–909.

Mickey, B. J. and Mddlebrooks, J. C. (2003). Representation of auditory space by cortical neurons in awake cats. *J Neurosci*, 23:8649–63.

Middlebrooks, J. (1999). Cortical represenations of auditory space. In Gazzaniga, M. S., editor, *The New Cognitive Neurosciences*, pages 425–36. MIT Press, Cambridge M.A., 2 edition.

Middlebrooks, J. C. (2002). Auditory space processing: here, there or everywhere? *Nat Neurosci*, 5(9):824–6.

Middlebrooks, J. C., Dykes, R. W., and Merzenich, M. M. (1980). Binaural respons-especific bands in primary auditory cortex (ai) of the cat: topographical organization orthogonal to isofrequency contours. *Brain Res*, 181:31–48.

Middlebrooks, J. C. and Pettigrew, J. D. (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *J Neurosci*, 1(1):107–20.

Middlebrooks, J. C., Xu, L., Eddins, A. C., and Green, D. M. (1998). Codes for sound-source location in nontonotopic auditory cortex. *J Neurophysiol*, 80(2):863–81.

Miller, L. M., Escabi, M. A., Read, H. L., and Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J Neurophysiol*, 87:516–27.

Mishkin, M. and Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav Brain Res*, 6(1):57–77.

Moody, D. B. (1994). Detection and discrimination of amplitude-modulated signals by macaque monkeys. *J Acoust Soc Am*, 95(6):3499–510.

Moore, B. C. (1973). Frequency difference limens for short-duration tones. *J Acoust Soc Am*, 54(3):610–9.

Moore, B. C., Glasberg, B. R., and Shailer, M. J. (1984). Frequency and intensity difference limens for harmonics within complex tones. *J Acoust Soc Am*, 75(2):550–61.

Moore, G. P., Perkel, D. H., and Segundo, J. P. (1966). Statistical analysis and functional interpretation of neuronal spike data. *Annu Rev Physiol*, 28:493–522.

Mountcastle, V. B., LaMotte, R. H., and Carli, G. (1972). Detection thresholds for stimuli in humans and monkeys: Comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand. *J Neurophysiol*, 35:122–36.

Mrsic-Flogel, T. D., King, A. J., Jenison, R. L., and Schnupp, J. W. (2001). Listening through different ears alters spatial response fields in ferret primary auditory cortex. *J Neurophysiol*, 86(2):1043–6.

Mrsic-Flogel, T. D., King, A. J., and Schnupp, J. W. (2005). Encoding of virtual acoustic space stimuli by neurons in ferret primary auditory cortex. *J Neurophysiol*, 93(6):3489–503.

Musicant, A. D., Chan, J. C., and Hind, J. E. (1990). Direction-dependent spectral properties of cat external ear: new data and cross-species comparisons. *J Acoust Soc Am*, 87(2):757–81.

Nakamoto, K. T., Jones, S. J., and Palmer, A. R. (2008). Descending projections from auditory cortex modulate sensitivity in the midbrain to cues for spatial position. *J Neurophysiol*, 99(5):2347–56.

Narayan, R., Grana, G., and Sen, K. (2006). Distinct time scales in cortical discrimination of natural sounds in songbirds. *J Neurophysiol*, 96(1):252–8.

Neill, J. C. and Harrison, J. M. (1987). Auditory discrimination: the Konorski quality-location effect. *J Exp Anal Beh*, 48:81–95.

Nelken, I. and Bar Yosef, O. (2008). Neurons and objects: the case of auditory cortex. *Front Neurosci*, 2(1):107–113.

Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., King, A. J., and Schnupp, J. W. (2008). Responses of auditory cortex to complex stimuli: functional organization revealed using intrinsic optical signals. *J Neurophysiol*, 99(4):1928–41.

Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., Schnupp, J. W., and King, A. J. (2004). Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals. *J Neurophysiol*, 92(4):2574–88.

Nelken, I., Chechik, G., Mrsic-Flogel, T. D., King, A. J., and Schnupp, J. W. (2005). Encoding stimulus information by spike numbers and mean response time in primary auditory cortex. *J Comput Neurosci*, 19(2):199–221.

Nelken, I., Fishbach, A., Las, L., Ulanovsky, N., and Farkas, D. (2003). Primary auditory cortex of cats: feature detection or something else? *Biol Cybern*, 89:397–406.

Nelson, D. A. (1989). Song frequency as a cue for recognition of species and individuals in the field sparrow (*Spizella pusilla*). *J Comp Psychol*, 103(2):171–6.

Nelson, D. A. and Kiester, T. E. (1978). Frequency discrimination in the chinchilla. *J Acoust Soc Am*, 64(1):114–26.

Nodal, F. R., Bajo, V. M., Parsons, C. H., Schnupp, J. W. H., and King, A. J. (2008). Sound localization behavior in ferrets: Comparison of acoustic orientation and approach-to-target responses. *Neurosci*, 154(1):397–408.

O'Connor, K. N. and Sutter, M. L. (2000). Global spectral and location effects in auditory perceptual grouping. *J Neurosci*, 12(2):342–54.

Page, S. C., Hulse, S. H., and Cynx, J. (1989). Relative pitch perception in the European starling (*Sturnus vulgaris*): further evidence for an elusive phenomenon. *J Exp Psychol Anim Behav Process*, 15(2):137–46.

Pallas, S. L. and Sur, M. (1993). Visual projections induced into the auditory pathway of ferrets: II. Corticocortical connections of primary auditory cortex. *J Comp Neurol*, 337:317–33.

Palmer, A. R. and Russell, I. J. (1986). Phase locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells. *Hear Res*, 24:1–15.

Pantev, C., Hoke, M., Lütkenhöner, B., and Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: pitch versus frequency representation. *Science*, 246(4929):486–8.

Park, T. J. (1998). IID sensitivity differs between two principal centers in the interaural intensity difference pathway: the lso and the ic. *J Neurophysiol*, 79(5):2416–31.

Parker, A. J., Krug, K., and Cumming, B. G. (2002). Neuronal activity and its links with the perception of multi-stable figures. *Philos Trans R Soc Lond B Biol Sci*, 357(1424):1053–62.

Parker, A. J. and Newsome, W. T. (1998). Sense and the single neuron: probing the physiology of perception. *Annu Rev Neurosci*, 21:227–77.

Parsons, C. H., Lanyon, R. G., Schnupp, J. W., and King, A. J. (1999). Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets. *J Neurophysiol*, 82(5):2294–309.

Pastore, R. E. (1983). Temporal order judgment of auditory stimulus offset. *Percept Psychophys*, 33(1):54–62.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36(4):767–76.

Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *J Neurosci*, 24(30):6810–5.

Peterson, G. E. and Barney, H. L. (1952). Control methods used in the study of vowels. *J Acoust Soc Am*, 24:175–84.

Pfeiffer, R. R. (1966). Classification of the response patterns of spike discharges for units in the cochlear nucleus: Tone burst stimulation. *Exp Brain Res*, 1:220–35.

Pfingst, B. E. (1993). Comparison of spectral and nonspectral frequency difference limens for human and nonhuman primates. *J Acoust Soc Am*, 93(4 Pt 1):2124–9.

Phillips, D. P. (1999). Auditory gap detection, perceptual channels, and temporal resolution in speech perception. *J Am Acad Audiol*, 10(6):343–54.

Phillips, D. P. and Hall, S. E. (1990). Response timing constraints on the cortical representation of sound time structure. *J Acoust Soc Am*, 88:1403–11.

Phillips, D. P. and Irvine, D. R. F. (1981). Responses of single neurons in physiologically defined area ai of cat cerebral cortex: sensitivity to interaural intensity differences. *Hearing Res*, 4:299–307.

Phillips, D. P., Judge, P. W., and Kelly, J. B. (1988). Primary auditory cortex in the ferret (mustela putorius): neural response properties and topographic organization. *Brain Res*, 443(1-2):281–94.

Prosen, C. A., Moody, D. B., Sommers, M. S., and Stebbins, W. C. (1990). Frequency discrimination in the monkey. *J Acoust Soc Am*, 88(5):2152–8.

Purushothaman, G. and Bradley, D. C. (2005). Neural population code for fine perceptual decisions in area MT. *Nat Neurosci*, 8(1):99–106.

Ramsay, A. M. and Meredith, M. A. (2004). Multiple sensory afferents to ferret pseudosylvian sulcal cortex. *Neuroreport*, 15(3):461–5.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Curr Opin Neurobio*, 8:516–521.

Rauschecker, J. P. and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci U S A*, 97(22):11800–6.

Rauschecker, J. P., Tian, B., Pons, T., and Mishkin, M. (1997). Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol*, 382(1):89–103.

Recanzone, G. H. (2000). Spatial processing in the auditory cortex of the macaque monkey. *Proc Natl Acad Sci U S A*, 97(22):11829–35.

Recanzone, G. H., Guard, D. C., Phan, M. L., and Su, T. K. (2000). Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. *J Neurophysiol*, 83(5):2723–39.

Recanzone, G. H., Schreiner, C. E., Sutter, M. L., Beitel, R. E., and Merzenich, M. M. (1999). Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. *J Comp Neurol*, 415:460–81.

Reiss, L. A. and Young, E. D. (2005). Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus. *J Neurosci*, 25(14):3680–91.

Reissland, N., Shepherd, J., and Herrera, E. (2003). The pitch of maternal voice: a comparison of mothers suffering from depressed mood and non-depressed mothers reading books to their infants. *J Child Psychol Psychiatry*, 44(2):255–61.

Relkin, E. M. and Pelli, D. G. (1987). Probe tone thresholds in the auditory nerve measured by two-interval forced-choice procedures. *J Acoust Soc Am*, 82(5):1679–91.

Romanski, L. M., Tian, B., Fritz, J. B., Mishkin, M., Goldman-Rakic, P. S., and Rauschecker, J. P. (1999). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci*, 2(12):1131–6.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philos Trans R Soc Lond B Biol Sci*, 336(1278):367–73.

Rosenblith, W. A. and Stevens, K. N. (1953). On the DL for frequency. *J Acoust Soc Am*, 25(5):980–5.

Saberi, K. and Perrott, D. R. (1999). Cognitive restoration of reversed speech. *Nature*, 398(6730):760.

Samson, S., Zatorre, R. J., and Ramsay, J. O. (2002). Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling. *Brain*, 125(Pt 3):511–23.

Sanes, D. H. (1990). An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive. *J Neurosci*, 10(11):3494–506.

Schnupp, J. W., Booth, J., and King, A. J. (2003). Modeling individual differences in ferret external ear transfer functions. *J Acoust Soc Am*, 113(4 Pt 1):2021–30.

Schnupp, J. W., Mrsic-Flogel, T. D., and King, A. J. (2001). Linear processing of spatial cues in primary auditory cortex. *Nature*, 414(6860):200–4.

Schnupp, J. W. H., Hall, T. M., Kokelaar, R. F., and Ahmed, B. (2006). Plasticity of temporal pattern codes for vocalisation stimuli in primary auditory cortex. *J Neurosci*, 26(18):4785–95.

Schreiner, C. E. and Calhoun, B. M. (1994). Spectral envelope coding in cat primary auditory cortex: properties of ripple transfer functions. *Aud Neurosci*, 1:39–61.

Schreiner, C. E. and Langner, G. (1988). Periodicity coding in the inferior colliculus of the cat. II. topographical organization. *J Neurophysiol*, 60:1823–40.

Schreiner, C. E. and Mendelson, J. R. (1990). Functional topography of cat primary auditory cortex: distribution of integrated excitation. *J Neurophysiol*, 64(5):1442–59.

Schulze, H., Hess, A., Ohl, F. W., and Scheich, H. (2002). Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil. *Eur J Neurosci*, 15(6):1077–84.

Schulze, H. and Langner, G. (1997a). Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): two different coding strategies for pitch and rhythm? *J Comp Physiol [A]*, 181(6):651–63.

Schulze, H. and Langner, G. (1997b). Representation of periodicity pitch in the primary auditory cortex of the Mongolian gerbil. *Acta Otolaryngol Suppl*, 532:89–95.

Shadlen, A. and Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *J Neurophysiol*, 86(4):1916–36.

Shamma, S. A. (1999). Physiological basis of timbre perception. In Gazzaniga, M. S., editor, *The New Cognitive Neurosciences*, pages 411–23. MIT Press, Cambridge M.A., 2 edition.

Shamma, S. A., Fleshman, J. W., Wiser, P. R., and Versnel, H. (1993). Organization of response areas in ferret primary auditory cortex. *J Neurophysiol*, 69(2):367–83.

Shamma, S. A. and Versnel, H. (1995). Ripple analysis in ferret primary auditory cortex. ii. prediction of unit responses to arbitrary spectral profiles. *Aud Neurosci*, 1:255–70.

Sharma, A., Marsh, C. M., and Dorman, M. F. (2000). Relationship between N1 evoked potential morphology and the perception of voicing. *J Acoust Soc Am*, 108(6):3030–5.

Shofner, W. P. (2000). Comparison of frequency discrimination thresholds for complex and single tones in chinchillas. *Hear Res*, 149(1-2):106–14.

Shofner, W. P. (2005). Comparative aspects of pitch perception. In Plack, C., Oxenham, A., Fay, R. R., and Popper, A., editors, *Pitch: Neual Coding and Perception*, Springer Handbook of Auditory Research, pages 56–98. Springer Science+Business Media, Inc., New York.

Shofner, W. P., Yost, W. A., and Whitmer, W. M. (2007). Pitch perception in chinchillas (*Chinchilla laniger*): stimulus generalization using rippled noise. *J Comp Psychol*, 121(4):428–39.

Sinnott, J. M., Petersen, M. R., and Hopp, S. L. (1985). Frequency and intensity discrimination in humans and monkeys. *J Acoust Soc Am*, 78(6):1977–85.

Sinnott, J. M., Sachs, M. B., and Hienz, R. D. (1980). Aspects of frequency discrimination in passerine birds and pigeons. *J Comp Physiol Psychol*, 94(3):401–15.

Smith, A. L., Parsons, C. H., Lanyon, R. G., Bizley, J. K., Akerman, C. J., Baker, G. E., Dempster, A. C., Thompson, I. D., and King, A. J. (2004). An investigation of the role of auditory cortex in sound localization using muscimol-releasing Elvax. *Eur J Neurosci*, 19(11):3059–72.

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (2006). The processing and perception of size information in speech sounds. *J Acoust Soc Am*, 117(1):305–318.

Sommers, M. S., Moody, D. B., Prosen, C. A., and Stebbins, W. C. (1992). Formant frequency discrimination by Japanese macaques (*Macaca fuscata*). *J Acoust Soc Am*, 91(6):3499–510.

Stecker, G. C., Harrington, I. A., Macpherson, E. A., and Middlebrooks, J. C. (2005a). Spatial sensitivity in the dorsal zone (area DZ) of cat auditory cortex. *J Neurophysiol*, 94(2):1267–80.

Stecker, G. C., Harrington, I. A., and Middlebrooks, J. C. (2005b). Location coding by opponent neural populations in the auditory cortex. *PLoS Biol*, 3(3):e78.

Stecker, G. C., Mickey, B. J., Macpherson, E. A., and Middlebrooks, J. C. (2003). Spatial sensitivity in field paf of cat auditory cortex. *J Neurophysiol*, 89(6):2889–903.

Stecker, G. C. and Middlebrooks, J. C. (2003). Distributed coding of sound locations in the auditory cortex. *Biol Cybern*, 89(5):341–9.

Steinschneider, M., Fishman, Y. I., and Arezzo, J. C. (2003). Representation of the voice onset time (VOT) speech parameter in population responses within primary auditory cortex of the awake monkey. *J Acoust Soc Am*, 114(1):307–21.

Steinschneider, M., Reser, D. H., Fishman, Y. I., Schroeder, C. E., and Arezzo (1998). Click train encoding in primary auditory cortex of the awake monkey: evidence for two mechanisms subserving pitch perception. *J Acoust Soc Am*, 104(5):2935–55.

Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., and Howard, M. A., r. (2005). Intracortical responses in human and monkey primary auditory

cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cereb Cortex*, 15(2):170–86.

Stevens, A. A. and Weaver, K. (2005). Auditory perceptual consolidation in early-onset blindness. *Neuropsychologia*, 43(13):1901–10.

Sumner, C. J., Palmer, A. R., and Moore, D. R. (2008). The need for a cool head: reversible inactivation reveals functional segregation in auditory cortex. *Nat Neurosci*, 11(5):530–1.

Syka, J., Rybalko, N., Brozek, G., and Jilek, M. (1996). Auditory frequency and intensity discrimination in pigmented rats. *Hear Res*, 100(1-2):107–13.

Talwar, S. K. and Gerstein, G. L. (1998). Auditory frequency discrimination in the white rat. *Hear Res*, 126(1-2):135–50.

Talwar, S. K. and Gerstein, G. L. (1999). A signal detection analysis of auditory-frequency discrimination in the rat. *J Acoust Soc Am*, 105(3):1784–800.

Theunissen, F. E., Sen, K., and Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci*, 20(6):2315–31.

Tian, B., Reser, D., Durham, A., Kustov, A., and Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science*, 292(5515):290–3.

Tolhurst, D. J., Movshon, J. A., and Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res*, 23(8):775–85.

Tomlinson, R. W. and Schwarz, D. W. (1988). Perception of the missing fundamental in nonhuman primates. *J Acoust Soc Am*, 84(2):560–5.

Trussell, L. O. (1999). Synaptic mechanisms for coding timing in auditory neurons. *Annu Rev Physiol*, 61:477–96.

Trussell, L. O. (2002). Modulation of transmitter release at giant synapses of the auditory system. *Curr Opin Neurobiol*, 12(4):400–4.

Ungerleider, L. G. and Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol*, 4(2):157–65.

Upton, M. (1929). The auditory sensitivity of guinea pigs. *Am J Psychol*, 41:412–21.

Versnel, H., Mossop, J. E., Mrsic-Flogel, T. D., Ahmed, B., and Moore, D. R. (2002). Optical imaging of intrinsic signals in ferret auditory cortex: responses to narrowband sound stimuli. *J Neurophysiol*, 88(3):1545–58.

Versnel, H. and Shamma, S. A. (1998). Spectral-ripple representation of steady-state vowels in primary auditory cortex. *J Acoust Soc Am*, 103(5):2502–14.

Vongpaisal, T., Trehub, S. E., and Schellenberg, E. G. (2006). Song recognition by children and adolescents with cochlear implants. *J Speech Lang Hear Res*, 49:1091–1103.

Walker, K. M., Ahmed, B., and Schnupp, J. W. (2008). Linking cortical spike pattern codes to auditory perception. *J Cogn Neurosci*, 20(1):135–52.

Wallace, M. N., Roeda, D., and Harper, M. S. (1997). Deoxyglucose uptake in the ferret auditory cortex. *Exp Brain Res*, 117(3):488–500.

Wallace, M. N., Rutkowski, R. G., and Palmer, A. R. (2005). Responses to the purr call in three areas of the guinea pig auditory cortex. *Neuroreport*, 16(18):2001–5.

Wallace, M. N., Shackleton, T. M., and Palmer, A. R. (2002). Phase-locked responses to pure tones in the primary auditory cortex. *Hear Res*, 172(1-2):160–71.

Wang, L., Narayan, R., Graña, G., Shamir, M., and Sen, K. (2007). Cortical discrimination of complex natural stimuli: can single neurons match behavior? *J Neurosci*, 27(3):582–9.

Wang, X. (2000). On cortical coding of vocal communication sounds in primates. *Proc Natl Acad Sci U S A*, 97(22):11843–9.

Wang, X. and Kadia, S. C. (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *J Neurophysiol*, 86(5):2616–20.

Wang, X., Lu, T., Bendor, D., and Bartlett, E. (2008). Neural coding of temporal information in auditory thalamus and cortex. *Neuroscience*, 154(1):294–303.

Wang, X., Merzenich, M. M., Beitel, R., and Schreiner, C. E. (1995). Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *J Neurophysiol*, 74(6):2685–706.

Warren, J. D. and Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *J Neurosci*, 23(13):5799–804.

Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci U S A*, 100(17):10038–42.

Warrier, C. M. and Zatorre, R. J. (2004). Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. *Brain*, 127:1616–25.

Wehr, M. and Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442–6.

Whitfield, I. C. (1980). Auditory cortex and the pitch of complex tones. *J Acoust Soc Am*, 67(2):644–7.

Wickens, T. (2002). *Elementary Signal Detection Theory*. Oxford University Press, New York, NY.

Wier, C. C., Jesteadt, W., and Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *J Acoust Soc Am*, 61(1):178–84.

Winer, J. A. (1992). The functional architecture of the medial geniculate body and the primary auditory cortex. In Gazzaniga, M. S., editor, *The mammalian auditory pathway: neuroanatomy*, volume 1 of *Springer Handbook of Auditory Research*, pages 222–409. Springer-Verlag, New York N.Y.

Winer, J. A., Larue, D. T., Diehl, J. J., and Hefti, B. J. (1998). Auditory cortical projections to the cat inferior colliculus. *J Comp Neurol*, 400:147–74.

Winer, J. A. and Lee, C. C. (2007). The distributed auditory cortex. *Hear Res*, 229:3–13.

Winer, J. A., Miller, L. M., Lee, C. C., and Schreiner, C. E. (2005). Auditory thalamocortical transformation: structure and functions. *Trends Neurosci*, 28:255–63.

Winslow, R. and Sacks, M. B. (1988). Single tone intensity discrimination based on auditory nerve fiber responses in backgrounds of quiet, noise and with stimulation of the crossed olivocochlear bundle. *Hear Res*, 35:165–90.

Winter, I. M. (2005). The neurophysiology of pitch. In Plack, C., Oxenham, A., Fay, R. R., and Popper, A., editors, *Pitch: Neual Coding and Perception*, Springer Handbook of Auditory Research, pages 99–146. Springer Science+Business Media, Inc., New York.

Winter, I. M. and Palmer, A. R. (1990). Temporal responses of primarylike anteroventral cochlear nucleus units to the steady-state vowel /i/. *J Acoust Soc Am*, 88(3):1437–41.

Winter, I. M., Wiegrebe, L., and Patterson, R. D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea pig. *J Physiol*, 537:553–66.

Winter, P. and Funkenstein, H. H. (1973). The effect of species-specific vocalization on the discharge of auditory cortical cells in the awake squirrel monkey (*Saimiri sciureus*). *Exp Brain Res*, 18(5):489–504.

Witte, R. S. and Kipke, D. R. (2005). Enhanced contrast sensitivity in auditory cortex as cats learn to discriminate sound frequencies. *Brain Res Cogn Brain Res*, 23(2-3):171–84.

Yost, W. A. (1978). Pitch and pitch discrimination of broadband signals with rippled power spectra. *J Acoust Soc Am*, 63(4):1166–75.

Young, E. D. and Brownell, W. E. (1976). Responses to tones and noise of single cells in the dorsal cochlear nucleus of unanesthesized cats. *J Neurophysiol*, 39:282–300.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *J Acoust Soc Am*, 84:566–72.

Zatorre, R. J., Bouffard, M., Ahad, P., and Belin, P. (2002). Where is "where" in the human auditory cortex? *Nat Neurosci*, 5(9):905–9.