Physiological Organization of Auditory Cortex in Awake Macaque

by

Brian H. Scott

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Center for Neural Science
New York University
September, 2004

_____

Malcolm N. Semple, Ph.D

Dissertation Advisor

© Brian H. Scott

All Rights Reserved, 2004

For my parents, who got me here,

and for Kate, who got me through.

Acknowledgements

This work would not have been possible without the supportive and collaborative environment of the Semple lab and the Center for Neural Science at NYU. I am fortunate to have been associated with a group of such intellect, humor, and creativity. I thank my advisor, Dr. Malcolm Semple, for countless constructive and enlightening conversations (many about science). Mal, I owe you 3000 cups of coffee[1]. The data in this thesis represent a joint effort with my fellow graduate student, and primary collaborator, Brian Malone. Many of the ideas presented here evolved out of group discussion; I share credit for the good ones, and take full responsibility for the bad.

I also thank the members of my advisory committee, Mike Hawken and Dan Sanes, my readers Xiaoqin Wang and Souheil Inati, and everyone associated with the auditory journal club. Nicole Rust, Marianna Yanike and Maria Ter-Mikaelian all rotated through the lab and contributed to behavioral training and physiology. The anatomical work is credited to Anita Disney, and the magnetic resonance images are courtesy of Souheil Inati.

---

[1] Conservatively estimated by: 2 cups/day * 5 days/week * 50 weeks/year * 6 years

Abstract

The auditory cortex is essential to the functional analysis of sound, yet our current model of how this cortex is organized in the primate brain is based primarily upon anatomical structure. Most of what is known about the physiology of the auditory cortical fields derives from species with brains quite different from those of human. In addition, most data have been acquired under the influence of anesthesia, which strongly suppresses neural responses. These limitations are elided by recording single neurons in the awake, behaviorally trained rhesus macaque. Characterizations of how individual neurons are tuned to various parameters of sound are used to evaluate the predominant model of auditory cortex, in which two distinct fields of the "core" auditory cortex receive input in parallel, which is then relayed to the next level of the processing hierarchy. The primary and rostral fields of the core differ in their spectral and temporal response properties, such that the rostral field carries a strong but temporally asynchronous representation of the acoustic stimulus over a limited frequency range. Stimuli that modulate in sound amplitude or frequency reveal that neurons of the rostral field do not synchronize their discharges to the modulations of the signal as precisely as neurons in the primary field, suggesting that the two populations may perform analyses of the auditory scene at different time scales. When tested with stimuli that mimic acoustic motion by modulating the relative phase of a tone at the two ears, neurons

at the border of the two fields show strong tuning to this directional cue. The tuning of individual cells does not appear sufficient to account for the perceptual acuity of the macaque in a phase discrimination task, but these cells encode phase over a wider range than has been observed under anesthesia. Fields surrounding the core are hypothesized to subserve specific functions in the identification or localization of sound, but both fields of core auditory cortex appear suited to spectrotemporal analysis as well as representation of spatial location.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **VI**

ABSTRACT ..... **VII**

LIST OF FIGURES ..... **XVI**

LIST OF TABLES ..... **XVII**

LIST OF ABBREVIATIONS ..... **XVIII**

CHAPTER 1 – INTRODUCTION **1**

1.1 History *1*

1.2 Parallel Processing *3*

1.4 Is "what/where" a reasonable distinction? *6*

1.5 Organization of Primate Auditory Cortex *8*

1.5.2 Thalamocortical connections: 11

1.5.3 Basic physiological properties: 13

1.6 Physiological Evidence for Segregation of Function within Auditory Cortex *17*

1.6.1 Spatial Tuning 18

1.6.2 "Spectrotemporal" Sensitivity 20

1.7 Chapters *22*

CHAPTER 2 – GENERAL METHODS, PSYCHOPHYSICS AND PHYSIOLOGICAL EFFECTS OF BEHAVIOR 25

2.1 Introduction.....25

2.2 Methods: Subjects and Psychophysical Testing Procedure.....30

2.3 Comparing behaving and passive physiology.....36

2.4 Results I: Psychophysical discrimination of ILD and IPD.....38

2.5 Results II: Effects of behavioral context on neural responses to IPD 44

2.6 Discussion 62

2.6.1 Effects of behavioral state – comparison with previous studies 62

2.6.2 Effects of Training 67

2.6.3 Behavior and training beyond AI.....68

CHAPTER 3 – PHYSIOLOGICAL ORGANIZATION OF THE CORE 70

3.1 Introduction.....70

3.2 Anatomical methods and verification of recording sites.....74

3.2.1 Histology results 77

3.2.3 MRI 86

3.3 Physiological methods and measurement of basic response properties.....89

3.3.1 Basic response properties of auditory cortex 90

3.4 Tonotopic organization of the core and delineation of cortical fields.....98

3.4.1 AI vs. R – tonotopic reversal at the low-frequency border.....98

3.4.2 Core vs. Belt – tuning to pure-tone frequency 103

3.5 Comparing basic response properties in AI and R *109*

3.5.1 Discharge rates 111

3.5.2 Distributions of best frequencies 113

3.5.3 Minimum first-spike latency 116

3.5.4 Rate-level functions 117

3.5.5 Binaural properties 121

3.6 An alternative approach: mapping responses to standard stimuli across the cortex *122*

3.6.1 Measuring the spread of activation 123

3.6.2 A fractured, level-dependent representation of frequency 125

3.6.3 Restricted sensitvity to IPD and broad activation by band-passed noise 128

3.7 Discussion *131*

3.7.1 Response properties and organization of the core fields 131

3.7.2 Tonotopic organization in awake cortex 132

3.7.3 Binaural organization 134

3.7.4 Hemispheric asymmetries 136

3.7.5 Conclusions 137

CHAPTER 4: TEMPORAL ENVELOPE PROCESSING: AMPLITUDE AND FREQUENCY MODULATION 138

4.1 Introduction *138*

4.1.1 Relevance to natural stimuli 138

4.1.2 Representation of dynamic stimulus features 139

4.2 Methods 143

4.2.1 Stimulus protocols 143

4.2.2 Data analysis 146

4.3 Results I: SAM 147

4.3.1 General response properties 147

4.3.2 Coding in single neurons: dependence of MPH shape on SAM parameters 153

4.4 Results II: SFM 173

4.4.1 Frequency modulation and the receptive field 173

4.4.2 Coding in single neurons: dependence of MPH shape on SFM parameters 175

4.5 Comparing SAM and SFM 189

4.6 Comparing AI and R 194

4.7 Discussion 200

4.7.1 Envelope coding across modulation frequency 200

4.7.2 Coding of real-world envelopes 202

4.7.3 Mechanisms of envelope sensitivity 203

4.7.4 Comparison with previous studies 204

4.7.5 Implications for perception 208

4.7.6 Implications for cortical organization 211

CHAPTER 5: BINAURAL BEATS: TEMPORAL AND SPATIAL PROCESSING.....214

5.1 Introduction.....214

5.2 Background 216

5.3 Methods.....218

5.3.1 Subjects, recording procedure, and stimuli 218

5.3.2 Statistical Analysis of synchrony 220

5.3.3 Recording Sites 221

5.4 General Results 222

5.5 Tuning to IPD I: Effects of binaural beat frequency.....228

5.5.1 Effects of beat frequency on vector strength.....231

5.5.2 Effects of beat frequency on mean phase.....234

5.5.3 Effects of beat frequency on discharge rate 239

5.6 Tuning to IPD II: Effects of center frequency.....241

5.6.1 The phase/frequency response area.....242

5.6.2 Linear phase/frequency functions: Characteristic Delay and Characteristic Phase 245

5.6.3 ITD across carrier frequency: coding by peaks vs. slopes 251

5.7 Directional sensitivity and behavioral relevance 252

5.7.1 Effect of beat direction on discharge rate and synchrony 254

5.7.2 A d′ analysis of peaks vs. slopes of IPD tuning functions.....258

5.7.3 Neurometric analysis.....262

5.8 Discussion.....*265*

5.8.1 Binaural beat frequency.....266

5.8.2 Center frequency and mean phase.....270

5.8.3 Neural sensitivity to auditory motion.....272

5.8.4 Comparing neural and psychophysical discriminability.....274

5.8.5 Organization of auditory cortex.....277

CHAPTER 6: CONCLUSIONS.....280

6.1 Summary.....*280*

6.1.1 Behavior.....280

6.1.2 Physiological organization of the core.....282

6.1.3 Temporal and spectral modulation.....283

6.1.4 Interaural phase modulation.....285

6.2 Integration.....*286*

6.2.1 Temporal response properties between the core fields.....286

6.2.2 Rate and temporal coding by single neurons.....292

6.2.3 Cortical maps: gradients and clusters.....295

6.3 Implications.....*298*

6.3.1 Homology to auditory cortex of other species.....298

6.3.2 Functional implications of cortical organization.....301

6.3.3 Models of cortical function.....303

6.3.4 Relevance to human hearing 305

REFERENCES 307

List of Figures

Figure	Page
1.1	9
1.2	12
2.1	33
2.2	39
2.3	42
2.4	45
2.5	47
2.6	49
2.7	53
2.8	55
2.9	57
2.10	59
2.11	66
3.1	72
3.2	78
3.3	81
3.4	84
3.5	87
3.6	93
3.7	99
3.8	101
3.9	104
3.10	107
3.11	112
3.12	114
3.13	119
3.14	126
3.15	129

Figure	Page
4.1	150
4.2	154
4.3	156
4.4	158
4.5	161
4.6	163
4.7	166
4.8	170
4.9	176
4.10	181
4.11	183
4.12	186
4.13	191
4.14	196
4.15	198
5.1	225
5.2	227
5.3	229
5.4	233
5.5	236
5.6	237
5.7	240
5.8	243
5.9	247
5.10	249
5.11	253
5.12	255
5.13	260
5.14	264
6.1	288
6.2	290

List of Tables

Table	Page
Table 3.1	110
Table 4.1	148
Table 4.2	174
Table 4.3	197
Table 5.1	223

List of Abbreviations

AI	primary auditory cortex (core)
AL	anterolateral field (belt)
BF	best frequency
BMF	best modulation frequency
CD	characteristic delay
CM	caudo-medial field (belt)
CP	characteristic phase
EE	excitatory/excitatory (contra/ipsi binaural class)
EI	excitatory/inhibitory
EO	excitatory/no effect
fMRI	functional magnetic resonance imaging
FTF	frequency tuning function
IC	inferior colliculus (auditory midbrain)
ILD	interaural level difference
IPD	interaural phase difference
ITD	interaural time difference
L	lateral field (belt)
MGN	medial geniculate nucleus (auditory thalamus)
MGv	MGN, ventral division
MGd	MGN, dorsal division
MPH	modulation period histogram
MTF	modulation transfer function
rMTF	rate MTF
tMTF	temporal (synchrony) MTF
pMTF	phase MTF
MUA	multi-unit activity
PFC	prefrontal cortex
PSTH	peri-stimulus time histogram
PTC	pure-tone control
R	rostral field (core)
RLF	rate-level function
ROC	receiver operating characteristic
RT	rostro-temporal field
SAM	sinusoidal amplitude modulation
SFM	sinusoidal frequency modulation
SPL	sound pressure level (re: 20 μPa)
STG	superior temporal gyrus
STP	superior temporal plane
V1	primary visual cortex
VS	vector strength

Chapter 1 – Introduction

1.1 History

The basic function of any sensory system is to extract the behaviorally relevant features of a stimulus in the environment, so that this representation may serve as the basis for perception, decision and action. The conceptual framework in which we study a system is based on what those relevant features are, and how the system may be optimized to encode them. The auditory cortex sits at the top of an intricate hierarchy of neural processing centers that encodes, analyzes, filters, and relays a representation of sound impinging on the inner ear, and sends outputs to cortical and subcortical structures controlling decision and movement.

Situated at the intersection of stimulus and response, the auditory cortex has been a subject of investigation for over a century, but early lesion experiments in animals produced conflicting indications of what function the auditory cortex serves. Lesion experiments use the behavior of the entire organism to reveal the contribution of one component of a complex system, which can fog their interpretation. Ferrier, an early proponent of cerebral localization of function, used cortical lesions to demonstrate apparent deafness following destruction of the temporal lobes in monkeys (Ferrier 1876). Later experiments suggested that bilateral ablation of the auditory cortex need not abolish simple discriminations of frequency or intensity,

and that any apparent deficit was dependent on the species and method of behavioral testing employed (summarized in Whitfield 1984). Lesions of auditory cortex in primates – but not carnivores – cause a profound but partially recoverable hearing loss, and a threefold increase in frequency discrimination threshold (reviewed in Heffner and Heffner 2003). In experiments involving sound localization, animals trained on a simple avoidance response can discriminate the individual cues to sound laterality (discussed below) after removal of the cortex, yet cannot localize sounds in space when required to approach their source (Heffner and Masterton 1975), a finding confirmed by reversible cortical deactivation (Malhotra et al. 2004). Many simple aspects of a sound are analyzed in the brainstem, and their cortical representation may no be exclusively sensory, but tailored for integration with spatial orientation and motor response sytems.

Penfield and Perot (1963) produced compelling evidence of hierarchical representation in the human auditory cortex by mapping the sensations and utterances elicited by stimulation of the temporal lobe in awake human epilepsy patients. While stimulation of the central auditory areas within the superior temporal gyrus (STG) generated basic sensations like buzzing or tones, activation of tissue on the more lateral, exposed gyrus could evoke far more complex experiences, such as familiar voices or music. In addition to confirming the location of the auditory brain, these findings suggested a hierarchical organization

in which central auditory cortex processed basic elements of the sound stimulus, while surrounding tissue dealt with more complex aspects of sound.

Physiological recordings in animals have revealed the finer organization of the cortical fields. One consistent finding has been the organizational principle of tonotopy, a topographical gradient of optimal stimulus frequency across auditory structures, that is conserved to the level of the cortex. This reflects the organization of the peripheral sense organ, the cochlea, and is analagous to organization by retinal postion in the visual system. The original study by Woolsey and Walzl (1942) used stimulation of the spiral lamina to demonstrate cochleotopy directly via evoked-potential recordings across the cortex of the cat, and subsequent studies have confirmed the arrangement using tonal stimulation in a variety of species (e.g. Merzenich and Brugge 1973, Reale and Imig 1980). The search for additional ogranizational principles has proven difficult, but the auditory brain seems to echo other facets of the general evolutionary design of neural systems: the incoming signal is relayed along multiple pathways that may be specialized for different aspects of analysis.

1.2 Parallel Processing

The auditory pathway bifurcates in the brainstem to innervate multiple components of the cochlear nuclei, and this anatomical scheme of parallel pathways is maintained as the signal is relayed toward cortex. Assigning functions to these

anatomical streams will require extensive behavioral and physiological study, but putative roles have been assigned on the basis of similar architecture in other sensory systems. The visual system of primates is also described in terms of parallel neural pathways, one dedicated to the processing of motion, another concerned with form, which diverge into functionally disparate cortical streams: the ventral stream of cortical fields dedicated to object recognition, and the dorsal stream processing spatial location and movement (Ungerleider and Mishkin, 1982; but c.f. Goodale and Milner [1992] who see these as pathways for "perception and action"). This theory was advanced after decades of physiological studies of the primate retina, visual thalamus and cortex, but physiology in the primate auditory cortex is less complete. The idea of "what" and "where" streams has been transposed to the auditory system, in which information concerning the spatial location of a sound source may be processed in parallel with the spectral and temporal features essential to recognition (Rauschecker and Tian 2000). There is ample evidence for parallel processing in the auditory system (and, one may argue, the entire brain), but how well supported is the theory of what vs. where, and how applicable are the principles of visual cortical organization to the auditory cortex?

1.3 Auditory Cortical Organization in Humans

What we know of the cortical anatomy of hearing in humans derives primarily from two sources: clinical studies of patients with localized brain damage, and functional

neuroimaging of cerebral blood flow during auditory tasks. Damage to the auditory areas of the temporal or parietal lobe can lead to a range of deficits, some specific to speech comprehension or recognition of environmental sounds, others selectively affecting the ability to identify the origin of a sound in space (Clarke et al. 2000, Yamada et al. 1996). The apparent dissociation of recognition and localization would support the notion that these abilities are served by separate cortical systems, while the retention of basic hearing ability (e.g., detecting tones) suggests that these paths diverge beyond the level of simple sensory processing. In cases where the deficit is spatial, some investigators have studied the ability of patients to hear the binaural cues used in sound localization, interaural level (ILD) and time (ITD) differences. (A sound originating on the left must pass around the head to reach the right ear, so it will be lower in intesity, and later to arrive.) Yamada et al. (1996) found that bilateral injury to the temporal lobes (including the primary auditory cortex and rostral areas) abolished the ability to use ITD, but caused only a doubling of the threshold for detecting ILD, suggesting that the auditory cortex uses both spatial cues, but is crucial for the perception of time delay.

Functional magnetic resonance imaging (fMRI) has been used to confirm Penfields's observation in reverse, showing activation of the central auditory cortex by tones and activation of lateral temporal lobe areas by more complex sounds

(Wessinger et al. 2001). The technique also has been applied to the issue of spatial specialization in the temporal cortex and beyond. Griffiths et al. (2000) used auditory motion (simulated by dynamic ILD and ITD) to activate parietal cortex and regions of the prefrontal cortex involved in movement planning; but these areas can also be linked, by the same technique, to non-spatial tasks (Cohen and Wessinger 1999). Positron Emission Tomography (PET) sacrifices temporal and spatial resolution but allows auditory stimulation in the free-field (fMRI produces a loud background noise and a magnetic field that precludes the use of loudspeakers). Zattore and colleagues have demonstrated that the posterior end of auditory cortex is involved in disambiguating overlapping sound sources (Zattore 2001), and have shown a preferential response to temporal variation in central auditory cortex and spectral variation in surrounding tissue. Provocative as these results may be, imaging techniques are limited to detecting metabolic demand in a region of tissue, and can not discern separate populations of cells in the same area that may be active during different tasks or stimulus conditions.

1.4 Is "what/where" a reasonable distinction?

The physiological evidence from animal studies for the dual streams hypothesis in audition, discussed below, is primarily a demonstration of spatial tuning. Those areas of the cortex that are not spatial are considered, almost by default, "spectrotemporal": selective for acoustic signals with particular spectral and

temporal characteristics. But all acoustic signals can be characterized this way, and spatial processing is intertwined with spectral and temporal analysis throughout the subcortical auditory system. The interaural differences caused by the sound shadow of the head are frequency-dependent, so from the moment sound is transduced at the cochlea, all subsequent processing of ILD and ITD in the brainstem and midbrain occurs in frequency-specific channels (reviewed in Cohen and Knudsen 1999). In addition, the filter properties of the external ear allow monaural localization using spectral cues (Kulkarni and Colburn, 1998). At lower levels of the auditory system, the division between spatial and spectrotemporal processing may be a specious distinction. But after the process of acoustic analysis, where sound is broken down into its component parts, may come a process of synthesis, in which frequency specificity breaks down and the representation of sound becomes more object-oriented. Yost (1991) refers to this as sound-source detection, a process dependent on spectral and temporal cues and analogous to the transformation of visual information in the progression from V1 to inferotemporal cortex (Tanaka 1996). At the cortical level, divergent parallel streams for the processing of spatial and non-spatial attributes seem plausible, and this bifurcation may begin with the projections from AI to the surrounding fields of the auditory cortex. But until we understand the basic functional organization of the primate auditory system, this hypothesis cannot be tested.

To understand how physiological response properties of cells in the auditory system evolve as a signal is relayed from one cortical field to the next, we must look not in human brains, but in a lower primate. The macaque monkey is a commonly used species for studying the anatomy and physiology of a brain quite similar to our own (Hackett et al. 2001), and can be trained on a variety of tasks like those used to measure psychophysical abilities in human subjects. The next section will review what is known – and, more importantly, unknown – about the functional organization of the macaque auditory system, focusing on the cortex, and drawing on the more extensive cat literature when necessary.

1.5 Organization of Primate Auditory Cortex

Several schemes for subdividing the auditory cortex have been proposed over the past century, but the current prevailing model involves a central auditory "core" region buried within the lateral sulcus on the superior temporal plane (STP), including the primary auditory cortex (AI) and one or two contiguous, rostral fields (Hackett et al. 1998). Surrounding this core region is a "belt" of tissue receiving mixed inputs from both the thalamus and from the core. This layout is conserved to some degree across macaques, chimpanzees and humans (Hackett et al. 2001), and is schematized in Figure 1.1 (adapted from Hackett et al. 1998, Wallace et al. 2002, Semple and Scott 2003).

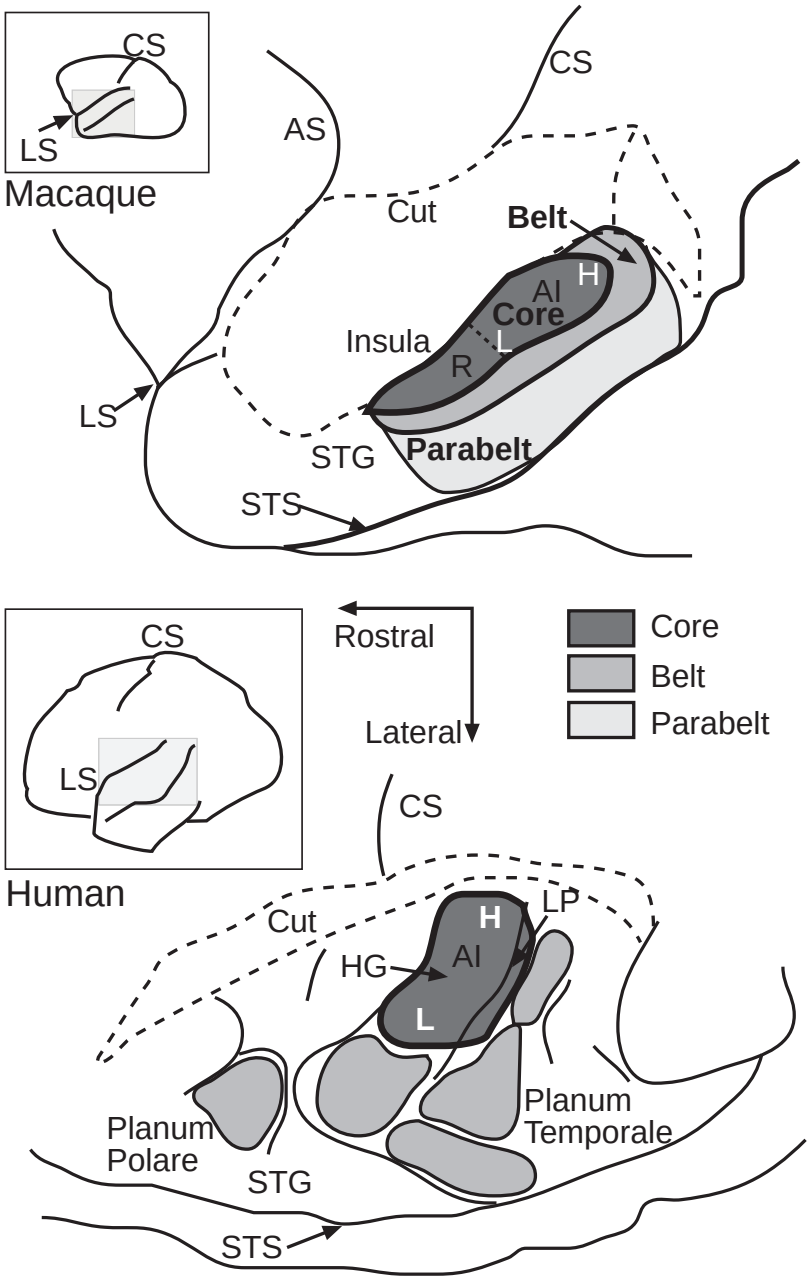


Figure 1.1 Homologous models of auditory cortical organization in macaque monkey (top) and human (bottom). Darkest shading indicates the "core" fields, including primary auditory cortex (AI) in both species, and a rostral field R in monkey. AI is tonotopically organized from high frequencies (H) to low (L), and in monkey AI adjoins R at a common low-frequency border. The core field(s) receive direct input from auditory thalamus and relay that information to the next level in the hierarchy, the belt (medium shading). In monkey, a third "parabelt" stage receives input from the belt (lightest shading). STG: superior temporal gyrus. STS: superior temporal sulcus. HG: Heschl's gyrus. CS: central sulcus. LS: lateral sulcus. AS: arcuate sulcus.

Felleman and Van Essen (1991) propose a set of four criteria by which proposed subdivisions may be validated as distinct cortical fields: architectonics, connectivity, topography, and function. These criteria will guide the following discussion of (1.5.1) the subdivisions of macaque auditory cortex as revealed by various staining techniques, (1.5.2) the thalamocortical connections to these regions, (1.5.3) physiological properties (especially tonotopic order) and (1.5.4) corticocortical connections.

1.5.1 Subdivisions of the auditory cortex:

Architectonically, the auditory core is differentiated from the surrounding temporal tissue by a granular architecture (*koniocortex*) characteristic of primary sensory cortices (Pandya and Yeterian 1985). Two groups (Jones et al. 1995, Hackett et al. 1998) have visualized the subdivisions of the auditory cortex using a battery of tissue staining methods, the most revealing being immunoreactivity for the calcium-binding protein parvalbumin, associated with the presence of a sub-population of interneurons containing the inhibitory neurotransmiter GABA(Hendry and Jones, 1991; Morino-Wannier et al. 1992). Parvalbumin staining formed a gradient, densest in the core areas AI and R and tapering rostrally to include a rostrotemporal field, RT, in the Hackett et al. (1998) study. This novel field was particularly apparent in *Macaca mulatta*, the species used here. A less dense band of immunoreactivity surrounded the core almost completely,

demarcating the auditory belt. A third, sparse region of staining lay anterior and lateral to the belt, called the parabelt. A possible implication of this parvalbumin distribution is that local GABAergic inhibition plays an important role in cortical sound processing at the input layers of core auditory cortex, and that local inhibitory circuitry differs between core and belt. Examination of the connectivity between these areas (discussed below) led Hackett et al. (1998) to extend prior models into a more extensive map of the macaque auditory cortex (Figure 1.2).

1.5.2 Thalamocortical connections:

Anatomical descriptions may only suggest how these different cortical areas, with their architectonic differences, function in the analysis of sound. To elaborate (and, by the same token, complicate) the picture, one must know whence these regions receive their input, and to where they send output. The medial geniculate nucleus (MGN) of the thalamus is an obligatory relay for afferent signals reaching the auditory cortex, and although it has been characterized in the monkey (e.g. Preuss and Muller-Preuss 1990), it has been more extensively studied in the cat (see Clarey, Barone and Imig 1992 for review). The ventral subdivision of the MGN, the MGv, is the primary recipient of auditory input from the inferior colliculus in the midbrain. Its cells are arranged tonotopically and respond to tones with short latency and sharp frequency tuning. The MGv stains darkly for parvalbumin, and projects to the similarly-stained middle layers of the auditory core via the thick

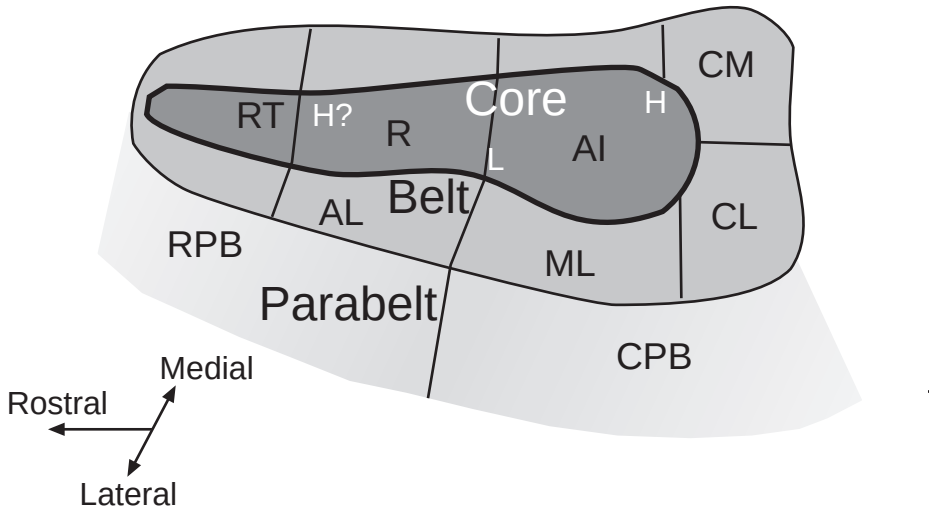


Figure 1.2 The current, anatomically derived model of auditory cortex organization in macaque.The core is subdivided into at least two tonotopic fields, AI and R, plus a hypothesized rostrotemporal field RT. Surrounding the core is a belt of secondary fields receiving the bulk of their input from the core (AL: antero-lateral, ML: middle lateral, CL: caudo-lateral, CM: caudo-medial). The parabelt, in turn, receives its input in serial from the belt (RPB: rostral parabelt, CPB: caudal parabelt). Adapted from Hackett et al. (1998).

fibers of the internal capsule (Molinari et al. 1995, Hashikawa et al. 1995). This predicts that the core regions might share the tonotopic, sharply tuned, short-latency responses of the MGv neurons, as has been confirmed by physiology (Merzenich and Brugge 1973, Kosaki et al. 1997; discussed below). The dorsal subdivision, MGd, also receives auditory input but responds at a longer latency, and less selectively, than the MGv.The MGd shows only moderate parvalbumin staining, and projects in parallel to the auditory belt regions via thin fibers. The MGd has two further subdivisions, anterior and posterior, that differ in their immunoreactivity and projections: the anterior projects to the caudal belt, whereas cells of the posterior division project to the rostral belt and parabelt. A component of the dual-streams hypothesis is that the rostral and caudal auditory fields reflect the division of labor between spatial and spectrotemporal tasks (Rauschecker and Tian 2000). The parallel inputs from different thalamic subdivisions establish an anatomical basis for physiological differences between these fields, but there has been little physiological mapping of core auditory cortex in the monkey.

1.5.3 Basic physiological properties:

Merzenich and Brugge (1973) offered a parcellation scheme based not only on architecture, but on the distribution of physiological response properties across the anesthetized macaque auditory cortex. They performed series of electrode penetrations from the insula (medially) to the STG (laterally), revealing a tonotopic

gradient, with high frequencies represented at the caudomedial end and low frequencies at the rostrolateral end of AI. These cells responded vigorously to a pure tone, and subsequent histology showed this field was coextensive with koniocortex. The lateral field L, a portion of the surrounding belt, displayed a tonotopy parallel to that of the primary field. Tonotopy in the caudo-medial region was vague, and BFs difficult to determine. Few responses were obtained in the medial belt, or the far rostral cortex. Morel et al. (1993) used short tones, clicks and speech to map the responses of multi-units across the core and belt regions. Their findings echoed those of Merzenich and Brugge (1973), and would be echoed by several subsequent papers (e.g. Kosaki et al. 1997): short latency, sharply tuned pure tone responses in the core, surrounded laterally and medially by a belt of less responsive, broadly tuned cells that responded poorly to pure tones.

Throughout their penetrations, Merzenich and Brugge (1973) found cells were most active in a region about 500um thick, corresponding to cortical layers III-V. This was likely a product of the anesthesia used, which suppresses responses in the superficial layers more strongly than in the middle layers, the site of thalamic input. This is especially true of barbiturate anesthesia (Shaw, 1986; Wang et al. 1987), commonly used in studies of cat auditory cortex. Thus many of these experiments were recording primarily thalamic responses as relayed to the cortex, and may not reflect cortical processing itself. Anesthesia also may explain the low spike rates

that plague many physiological studies in auditory cortex, since this region of the brain seems more vulnerable to barbiturate-induced depression than the visual cortex or lower auditory areas (Funkenstein and Winter 1973, Brugge and Merzenich 1973). Understanding cortical processing in the macaque calls for the use of awake animals. One such study, by Recanzone et al. (2000a), found a very rough tonotopic order, suggesting that the cleaner progression in anesthetized animals may be a representation of thalamic input patterns.

Tonotopic order in AI is to be expected, since most lower auditory structures have a similar organization (Brugge 1992), and the primary visual and somatosensory cortices also map the receptor surface (Felleman and Van Essen 1991, Mountcastle 1957). Though Merzenich and Brugge (1973) and later studies cite a reversal in the tonotopic gradient between AI and R, and hypothesize another reversal between R and RT (Hackett et al. 1998), the evidence that the frequency progression truly reverses is weak. At the rostral end of AI, the tonotopic gradient breaks down as BFs became higher, (Merzenich and Brugge 1973, Morel et al. 1993, Kosaki et al. 1997) but very few rostral cells have been recorded in any study.

1.5.4 corticocortical connections

The final criterion for distinguishing separate cortical fields is the pattern of interconnections between adjacent areas. Specific, reciprocal connections between

core, belt and parabelt have been established by tract tracing methods, and indicate that each area of the core connects primarily to the adjacent division of the belt, and vice versa (Hackett et al. 1998). Jones et al. (1995) describe three distinct streams of projections emanating from the core to the lateral, anteromedial and posteromedial belt, and suggest that these might correspond to the three pathways from auditory cortex to prefrontal cortex described earlier by Pandya and Yeterian (1985). Romanski et al. (1999a,b) confirmed and expanded the anatomical evidence for multiple pathways from temporal to prefrontal cortex by using electrophysiological recording to guide the injection of tracers into the lateral belt. The caudal belt has strong reciprocal connections with the dorsolateral prefrontal cortex (PFC), an area that has been shown to respond to acoustic stimulation during a localization task in an alert monkey (Vaadia et al. 1986), and is also implicated in visual spatial processing (Wilson et al. 1993). The rostral belt communicates with the rostral and ventral PFC, apparently non-spatial regions (Wilson et al. 1993). These connections, though not entirely exclusive, are interpreted as forming the basis for the “what” and “where” pathways, with rostral auditory cortex serving as the beginning of the spectrotemporal recognition pathway, and caudal auditory cortex beginning the spatial processing stream. Physiolgical correlates within the core have yet to be found.

1.6 Physiological Evidence for Segregation of Function within Auditory Cortex

The anatomy above provides a framework in which to evaluate the contributions of cells in different areas of the cortex to the problem of acoustic analysis and behavior. Investigations of the cortex generally assume that physiological response properties will be organized, or mapped, across the cortical surface by some set of parameters. Kaas has posited that the cortex is arranged into topographic maps to keep neurons with similar response properties in proximity, facilitating the formation of local circuits. These maps, he argues, are fundamental to sensory processing and ubiquitous in the brain (Kaas 1997). Whereas the visual cortex shows mapping of ocular dominance and orientation within the broader retinotopic layout (Hubel and Weisel, 1962), moving beyond tonotopy in the auditory system has been difficult.

Alternating bands of cells driven by sound in both ears, or driven by one ear and inhibited by the other, have been suggested to run orthogonal to the frequency progression in AI of the anesthetized cat, but the organization is far from clear cut (Imig and Adrian 1977, Middlebrooks et al. 1980), and these binaural bands have yet to be observed in monkey cortex (Semple et al. 2000, Reser et al. 2000). Schreiner has attempted to map many response properties across cat AI, including tuning sharpness, latency, and threshold, proposing a model in which many commonly measured response parameters are arrayed on orthogonal axes across the

cortical surface (reviewed in Schreiner 1998); many parameters cannot be clearly mapped, however, or are represented in patches or clusters. The failure to find a consistent, orderly stimulus representation in AI may be related to the problem of finding the ideal stimulus: as King and Schnupp (1998) have pointed out, auditory researchers have yet to find a stimulus for AI as effective and revealing as oriented bars are in V1. Thus, comparisons between visual and auditory cortical processing may be premature.

1.6.1 Spatial Tuning

Spatial attributes may be the class of parameters most likely to reveal an organization in auditory cortex more interesting than the vestiges of subcortical cochleotopy. Many cells in anesthetized cat AI and adjacent, nontonotopic fields, are selective for the origin of a stimulus in space (Middlebrooks and Pettigrew 1981, Brugge et al. 1996, Xu et al. 1998), and a number of investigations have examined this phenomenon in the monkey. Brugge and Merzenich (1973) reported low-frequency neurons in AI tuned to ITD, a tuning that was invariant with sound level and generally peaked at a discharge rate much higher than could be elicited by sound at either ear alone. Multiple cells studied in one penetration showed tuning to similar ITDs, suggesting a possible columnar organization to binaural selectivity, but few examples were obtained. Very few cells were recorded outside of AI, and no subsequent single-unit studies have been performed in awake monkeys using a

closed-field stimulus delivery system, which permits the explicit manipulation of interaural cues to sound localization.

Studies using free-field, broadband stimuli (noise presented from external speakers) have shown a coarse tuning for stimulus location in core auditory cortex of the behaving monkey (Benson et al. 1981, Ahissar et al. 1992), but recent studies including the belt areas have proposed an enhanced tuning for spatial location in the caudomedial belt region, CM (Recanzone et al. 2000b). The physiology of the auditory belt has been studied less extensively than that of AI, largely because more complex receptive fields preclude the use of standard laboratory stimuli such as tones (e.g. Morel et al. 1993, Rauschecker et al. 1997). Thus, determining a systematic difference between belt fields has been attempted by using complex stimuli such as band-pass noise and recorded vocalizations (e.g. Rauschecker, Tian and Hauser 1995). Area CM receives input from thalamic nuclei that receive input from the inferior and superior colliculus. As the colliclui are known to be involved in spatial orientation in the visual and auditory domains (Middlebrooks 1988), one may predict that neurons in the caudal belt would be sensitive to the spatial attributes of an acoustic stimulus. Rauschecker et al. (1997) report broad frequency tuning and a preference for complex stimuli in CM, and suggest these neurons are broadly tuned in the spectral domain to pool across frequency channels during sound localization. In later experiments, Rauschecker et al. (1999) attempted to

demonstrate spatial tuning by presenting vocalizations from a horizontal array of speakers while recording from the lateral belt areas. The populations of neurons recorded showed a slight bias toward spatial selectivity in area CL. Recanzone et al. (2000b) found 80% of units in AI and CM are spatially sensitive in awake monkeys performing a localization task, with units in CM better predicting the psychophysical performance of the animal. That the caudal belt projects preferentially to spatial areas of the PFC (Romanski et al. 1999a,b) is further evidence that the caudal processing stream may be involved in localizing sounds within auditory space, and eventually directing gaze to the relevant stimulus.

1.6.2 "Spectrotemporal" Sensitivity

What, then, is processed by the rostral auditory cortex? Rauschecker suggests, as indicated by the finding of vocalization selectivity in area AL, that this may be the beginning of a pattern recognition system (Rauschecker et al. 1999). A number of predictions follow from this hypothesis concerning spectral and temporal receptive field properties in this pattern-selective ("what") stream: neurons may be highly sensitive to amplitude or frequency fluctuations, sequential stimuli, or harmonic content of complex stimuli like vocalizations. While sensitivity to each of these factors has been sought in primate auditory cortex (e.g. Bieser and Muller-Preuss 1996, Brosch et al 1999, Schwarz and Tomlinson 1990, Wang et al. 1995) a definite rostral/caudal dichotomy has yet to be established. One reason may be the

dearth of physiology from rostral core. Another reason may be the use of anesthetized animals, described above, which may make responses outside AI difficult to elicit.

A third problem may be conceptual rather than procedural: the idea of "what" as a unitary concept in the two-streams model that dominates contemporary thinking. As Yost (1991) points out, object identification can incorporate many cues, including location. Identification can also be concerned with content, such as words or phonemes, or quality, such as distinguishing the voices of different speakers – the former task can be accomplished from purely temporal cues (Shannon et al. 1995), the latter relies on spectral structure. A recent report (Ghazanfar 2003) identifies cells in the belt selective not for a particular type of monkey call (as demonstrated previously [Tian et al. 2001]), but for the vocal tract characteristics of individual callers. Should we then consider these cells to form a third "who" stream? Anatomical tracer studies may show a general trend toward parallel pathways, but always with some substrate for crosstalk, and the interaction of spatial and non-spatial information. Understanding the cortical representation of sound will require moving outside the idea of spatial tuning vs. selectivity for call tokens, and require the systematic application of simple stimuli that measure neural tuning to dynamics of space, frequency, and intensity across the core regions of the

unanesthetized macaque auditory cortex. The remainder of the dissertaion summarizes my experiments toward this goal.

1.7 Chapters

Chapter 2 presents the awake macaque model of psychophysical discrimination and cortical physiology, including the thresholds of 2 animals for ILD and IPD. The bulk of the physiological data to be presented was collected under passive listening conditions, when the animal was not discriminating the stimulus being presented. Here neural responses to IPD collected under the behaving and passive conditions are analyzed to evaluate the reliability of physiolgical data collected as the animal sits passively.

The current model of auditory cortex organization in macaque is well grounded in anatomy, but little physiological evidence exists outside of AI, and what does exist is predominantly from anesthetized animals. Few sites have been recorded rostral to AI in an awake primate, and much work in the belt has been performed using simple tones or noise presented in the free field. Chronic recording in the awake primate allows complete mapping of a broad cortical region in a single animal, so the frequency organization proposed by Hackett et al. (1998) can be investigated. Chapter 3 uses the standard tonotopic criterion to delineate the core fields and borders with the belt regions. These field boundaries can then be used to compare

responsiveness to other stimuli. In addition, the latency of responses across the cortex will provide an indication of how signals are processed by the different fields in time, and perhaps shed light on the functional kinetics of serial and parallel inputs to the cortical fields.

Chapter 4 presents responses of cells across the core regions to non-spatial temporal fluctuations, such as sinusoidal amplitude and frequency modulation (SAM and SFM), to allow an evaluation of the temporal fidelity of neurons across the cortex in response to rapidly changing stimuli. Amplitude and frequency fluctuations are common to most animal vocalizations and many environmental sounds, and understanding their representation may bridge the gap between responses to simple tones and more complex sounds.

Very few data concerning binaural properties of neurons in the core exist, making free-field demonstrations of spatial selectivity (Recanzone et al. 2000a) difficult to interpret. This sensitivity could be based on ILD and ITD, and/or monaural spectral cues induced by the pinna (Kulkarni and Colburn, 1998). An efficient probe for ITD-based spatial sensitivity in low-frequency cells is the binaural beat, a continuously changing interaural phase disparity (IPD) between the two ears (Spitzer and Semple 1991, Reale and Brugge 1990, Fitzpatrick et al. 1997). Chapter 5 shows that many low-frequency core neurons are sensitive to these modulations,

and the distribution of binaural beat responsiveness may provide a valuable landmark for future investigations of the auditory cortex. Cells that are unresponsive to beats, but tuned in the appropriate frequency range, may indicate the divergence of a non-spatial cortical pathway. Binaural beats follow a cyclical modulation like SAM and SFM, allowing direct comparison of how cells encode spatial and spectrotemporal modulation across the core.

Chapter 6 draws these results together to present a model of the core in which spatial and spectrotemporal information are multiplexed within the tonotopic arrangement of AI and R.

Chapter 2 – General Methods, Psychophysics and Physiological Effects of Behavior

2.1 Introduction

One goal of studying the auditory cortex in lower animals is to understand the mechanisms and limits of human hearing. Investigating the function of the auditory cortex at the level of single cells requires a model species with brain architecture and perceptual abilities comparable to our own. The rhesus macaque (*Macaca mulatta*), an old-world monkey widely used in neuroscience research, has a cortical structure similar to that of humans and can be trained to perform a wide range of perceptual tasks in the visual, somatosensory and auditory domains. Beyond choice of species, the design of an experimental preparation must take into account the effects of anesthesia, behavior, and attention on the responses of cortical cells. This chapter will establish the procedures by which the data in this dissertation were acquired, and compare physiological observations in awake, trained macaques under performing and passive behavioral conditions.

Most physiological data in the auditory cortex have been collected from animals under the influence of anesthesia, which offers the advantage of extended, stable recordings from single cells at the cost of depressed responses. Under a strong anesthetic like pentobarbitol, pure tones generally elicit only one to three spikes per

trial in cat AI (Brosch and Schreiner 2000), and monkey AI responds similarly even under light anesthesia (ketamine/xylazine: Brosch *et al*. 1999). Responses in the awake animal are stronger, less prone to habituation, and may be found through a greater depth of cortical tissue. Because anesthetic tends to suppress responses in the supra- and infragranular layers more strongly than in the thalamic input layers, studies of the anesthetized cortex emphasize thalamic inputs over cortical responses or outputs. Those cells that do respond are heavily influenced by the anesthetic used: in rat AI, pentobarbitol was found to reduce spontaneous activity to 21% of its awake level, transform sustained response profiles into onset responses, prolong the response latency and, in some cells, abolish responses entirely (Gaese and Oswald, 2001). Isofluroane, a common surgical inhalant, can be even more disruptive (Cheung *et al*. 2001), and ketamine with or without pentobarbitol exerts a range of effects on cortical and thalamic discharge properties (Zurita *et al*, 1994). Thus recordings from anesthetized AI reflect a cortex in an unnatural pharmacological state, and offer a distorted representation of how cortical circuits may function in the waking brain.[1]

---

[1] Opiate anesthetics such as fentanyl citrate have been used with great success in studies of primate visual cortex, but this technique has not been adopted by the auditory community, in which the cat remained the dominant animal model until relatively recently. References to problems with "anesthesia" refer to the more suppressive agents typically employed in auditory research.

A more veridical sample of how cortical neurons respond to sound can be obtained from the unanesthetized preparation, but this presents a range of tradeoffs related to behavioral control. Experimenters have generally adopted an approach that falls into one of three paradigms, which I will call awake-untrained, awake-behaving, and awake-passive.

Awake animals may move, compromising recording stability, and extensive training may be required before they will tolerate lengthy experimental sessions. Thus, at a minimum, animals in the awake-untrained paradigm must be acclimated to – and rewarded for – sitting still, even if they are not trained on a stimulus-related behavior. The efficacy of this method may vary with the species involved. It has been used with success in rabbits (Fitzpatrick 1997), and in the marmoset, a small, new-world primate (Lu et al. 2001a,b). In rhesus macaques, this method has succeeded only for short-term chronic recordings: Brugge and Merzenich (1973) conditioned animals for >1 month, and recorded for 3-10 days, requiring the pooling of data across many animals. A contemporary study reported weak and inconsistent responses in AI of untrained macaques, compared to monkeys that had been trained on an acoustic task (Miller *et al*. 1972), leading subsequent investigations in awake macaque to incorporate a behavioral element.

Controlling for behavioral state and attention requires an additional level of training. To understand the contribution of individual neurons to a particular sensory discrimination, it is advantageous to use a preparation from which physiological recordings can be obtained as the animal actively performs the discrimination under study. This awake-behaving approach can be limiting to general physiological investigation, however, because an animal can only be trained on a small number of tasks. This restricts the range of stimuli that can be presented, and thus is most appropriate for experiments in cortical regions where basic physiological tuning properties are understood. For example, the middle temporal area is known to respond to visual motion, and a series of careful studies has linked individual neural activity to the perceived direction of randomly moving dots (Britten *et al*. 1996). The basic physiology of AI is less well-understood, making the design of such an experiment premature in the absence of an ideal stimulus. We have opted for a compromise between these two paradigms, which I call awake-passive: the monkeys were trained for an awake-behaving physiology task that exploited a suspected tuning property of AI neurons, but cells were also studied with a battery of stimuli as the animal sat passively. This provided behavioral control over the animal and allowed the application of a broad variety of stimuli in AI and surrounding fields.

The task on which we chose to train the animals was binaural lateralization of sound, based on interaural level differences (ILD) and interaural phase differences (IPD). Both of these cues are used by the auditory system to identify the origin of a sound source along the azimuth. Shadowing effects of the head and outer ear reduce the intensity of a sound at the ear farther from the sound source, and the additional travel time required for the sound wave to reach the farther ear creates a delay in both onset time and ongoing phase. These binaural cues are dependent on the frequency content of the stimulus, and this frequency-specific filtering by the outer ear imposes spectral modifications on the incoming sound that serve as a monaural cue to sound location (humans: Kulkarni and Colburn 1998; macaques: Spezio *et al*. 2000). Neural sensitivity to ILD and IPD are established in the lateral and medial nuclei of the superior olivary complex, respectively, and this sensitivity is relayed and transformed as it ascends to the cortex (Goldberg and Brown 1968, Imig and Adrian 1977, Fitzpatrick *et al.* 1997). Given that AI has been shown to be necessary for orienting to sounds in space (Heffner and Heffner 1990), these spatial cues were good candidate stimuli for probing the psychophysical limits of the macaque as well as the neural tuning properties of AI cells.

All physiological data presented will be from single cells, unless otherwise stated. Many prior studies examined the auditory cortex by recording multi-unit activity (MUA), which can be easier technically than isolating single units, and has proven

effective in mapping best frequency in the middle layers of anesthetized AI (e.g. Merzenich and Brugge 1973). We were able to record stable single cells through the cortical depth, so our data yield would not have been improved by this method. In addition, single-unit analysis is not clouded by the assumption that nearby cells respond homogeneously. Phillips and Irvine (1979) issued a caveat on this issue after the initial finding of aural dominance bands in cat AI by Imig and Adrian (1977). Although cells encountered in a given penetration shared a common excitatory input from the contralateral ear, the input from the ipsilateral ear could vary considerably, a property that would be missed by recording the summed activity of the local population.

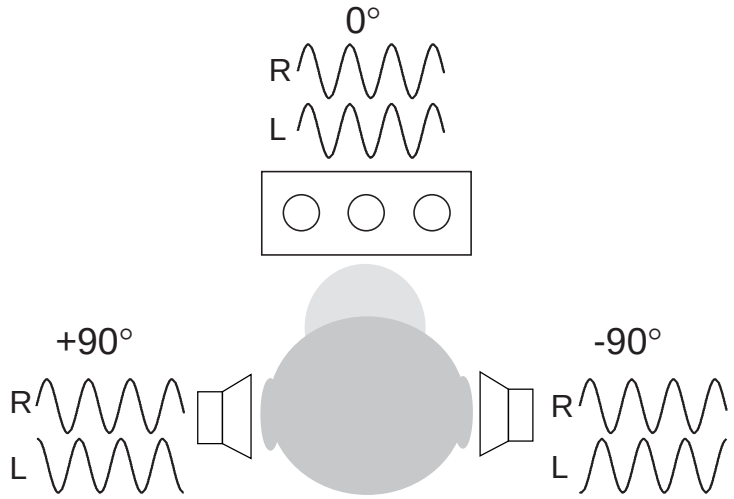
2.2 Methods: Subjects and Psychophysical Testing Procedure

Two male rhesus monkeys (*Macaca mulatta*, 6-10 kg over course of experiments, designated animals X and Z) were trained to discriminate IPD. All procedures were in accordance with the Society for Neuroscience guiding principles on the care and use of animals, and were approved by the IACUC of New York University. In daily sessions, animals sat in a Plexiglas chair (Christ, MD) in a double-walled anechoic room (IAC, Bronx NY) with their heads immobilized by a surgically implanted head-holder. Speakers coupled to ear inserts were positioned just within the opening of the ear canal for delivery of effectively closed-field stimuli, though the ear was not sealed from outside environmental (or self-produced) sounds.

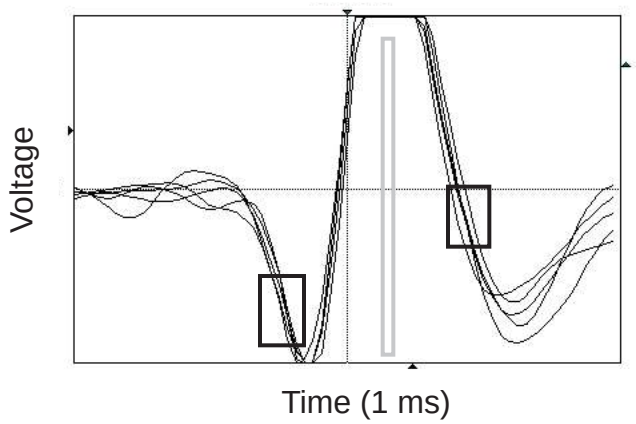
Thresholds for ILD and IPD discrimination were tested using the method of constant stimuli and a two-alternative, forced-choice (2AFC) procedure. A water spout and 3-button response box were placed in front of the animal, which initiated a trial by pressing and holding the center button for the duration of the stimulus (plus an additional 300 ms wait time), then pressed one of the flanking buttons on the left or right to indicate the laterality of the sound. Correct responses were rewarded by the opening of a valve allowing water to flow from the spout for a variable length of time (50 ms, doubling with each sequential correct response up to 800 ms). The valve also could be opened by the experimenter to provide supplemental liquid rewards. The first cue used to train the monkeys on lateralization was ILD, initially presented as an 80 or 90 dB tone to one ear (in blocks at first, then selected randomly). Once the monkeys performed reliably, the SPL in the other ear was incremented until threshold was determined. Tones were 2 s in duration, and carrier frequency was held constant within blocks (0.5-10 kHz tested). Once this task was acquired, a "dynamic" variant was introduced in which the tone came on at 0 dB ILD (midline) for 1 s, shifted by a linear ramp of intensity at the left or right ear, and remained at this second steady-state ILD for the remainder of the 2s tone. Thresholds were not determined by this method, it served primarily as a training exercise for the generalization of the task to IPD.

Once both animals acquired the lateralization task for ILD (which took many months to achieve), they were trained to perform the same task on the basis of IPD, which produces a comparable percept of laterality within the head. Figure 2.1A is a schematic depiction of the stimulus and testing setup. Under the "dynamic" condition each tone was initially diotic (IPD = 0°, SPL = 80 dB) for 500 ms, then linearly ramped at the left ear to a phase lead or lag (90° during training, decremented to 15, 10, 7.5, 5, and 2.5° for threshold determination) at 180°/s, and held constant at this IPD for the remainder of the 1 s pip (417-486 ms, depending on depth of the phase shift). These "dynamic" trials were interleaved randomly with "static" trials, in which the IPD was constant (at the same values used in dynamic trials) throughout the 1 s pip. The left and right tones were gated on simultaneously with a 10 ms cosine-squared ramp, so onset delay was not a cue to lateralization. Carrier frequency was held constant within blocks, but varied to cover the full range over which the animals were able and willing to perform the task (approximately 0.25-2.5 kHz, see Results I). All stimuli were selected and presented using the MALab system (Kaiser Instruments, see below), which also recorded behavioral responses (left, right, or no response). Percent correct at each ILD or IPD was folded across the postive and negative directions, plotted and fit with a spline curve (λ=10), from which the threshold was determined as the IPD at 76% correct (corresponding to d´=1 in a 2AFC task [Green and Swets 1966]).

A


0°
R
L
+90°
R
L
-90°
R
L

B


Voltage
Time (1 ms)

C

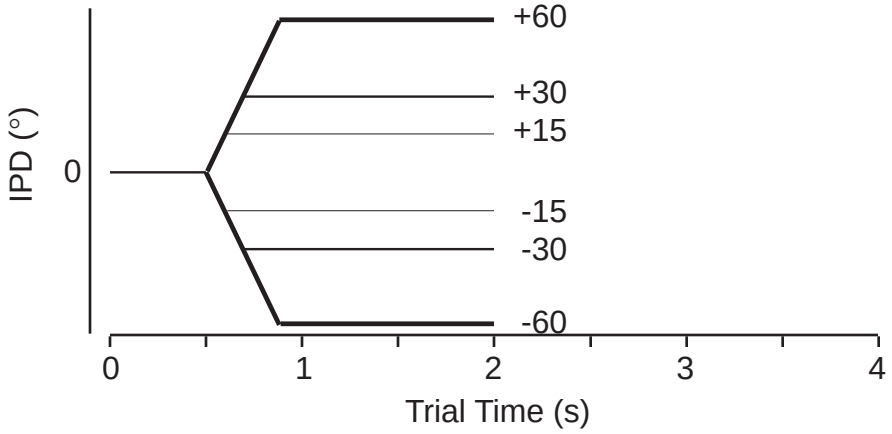

IPD (°)
0
+60
+30
+15
-15
-30
-60
0
1
2
3
4
Trial Time (s)

Figure 2.1 General methods for psychophysical testing, discrimination of single-unit spike waveforms, and comparison of neural activity under behaving and passive listening. **A:** Schematic depiction of testing setup: monkey faces a 3-button response box as signals are deliverd dichotically. A tone at 0° IPD (top) appears to be centered in the head; a phase lead at the left (-90°) or right (+90°) generates the percept of a sound lateralilzed to that side. **B:** Voltage/time plot of 5 spike waveforms from a single unit in AI, as viewed online in real time using MALab software. Discriminated spikes are required to pass through the voltage/time windows in black, and excluded if they pass though the window in gray. Abscissa is 1 ms, ordinate is arbitrary voltage (depends on analog amplification and filtering). **C:** Dynamic IPD stimuli used during behaving vs. passive comparison. IPD was initally 0° for 500 ms, then linearly ramped 15°, 30° or 60° in the positive or negative direction. Note these were employed only for the behavioral test; smaller steps closer to midline were used for threshold determination.

Physiology and Stimulus Generation:

For physiological recording, the animal was positioned as for psychophysics (above), and a stepping microdrive (Cal Tech) was placed on the recording cylinder under aseptic conditions to allow a vertical approach to the superior temporal plane (Pfingst and O'Connor, 1980). The stage of the microdrive was positioned in X-Y coordinates (medial/lateral and rostral/caudal axes). A guide tube was used to pierce the dura mater and allow vertical penetration of the parietal cortex by resin-coated tungsten microelectrodes (FHC, 10-12 MΩ). The signal was amplified (variable gain), filtered (typically 0.3-10 kHz), and fed into the MALab event processor interfaced with a Macintosh computer. Spikes from single neurons were discriminated by setting individual voltage/time windows, and waveforms were monitored for consistency throughout the recording (Figure 2.1B); spike times were logged to a precision of 1 μs. Stimulus generation was also controlled by the MALab software, which drives a dedicated hardware synthesizer that feeds the left and right signals to a Stax amplifier and, in turn, to the electrostatic speakers (Stax Lambda in Custom Sound Systems housings) coupled to custom-made ear inserts. Each channel was calibrated for level (dB, re: 20 μPa) and phase across frequency at the beginning of every session using a probe microphone (Bruel and Kæjer) positioned within the ear canal.

The electrode passed through the parietal cortex, where the somatotopic map provides a useful landmark: AI often lies ventral to somatosensory cortex corresponding to the contralateral paw (also observed by Brugge and Merzenich [1973]). Entry into auditory cortex was typically preceded by a gap in physiological activity believed to correspond to passage through the lateral sulcus. At this point, the animal stopped practicing his discrimination task and sat in a passive but alert state as tones or noise were employed as search stimuli. The depth of the dorsal and ventral limits of auditory responsiveness, and the depths of all units encountered, were recorded for comparison with histological analysis of laminar position within the cortex. Animals were monitored on closed-circuit video to ensure that they remained comfortable and awake during the experiments.

2.3 Comparing behaving and passive physiology

When a responsive cell was isolated, threshold and best frequency were determined using pure tones (See Chapter 3 – methods specific to particular analyses will be presented at the beginning of the appropriate chapter). The prevalence of nonmonotonic tuning to SPL, and a general dearth of clear tuning to ILD, suggested that the better task for studying the effect of behavior was IPD lateralization. Phase sensitivity was initially assessed by binaural beats, an ongoing modulation of interaural phase produced by presenting two tones of slightly different frequency to the two ears (See Chapter 5 for a full discussion). Cells that

synchronized their discharge to the resultant beat were candidates for recording during psychophysical discrimination.

The parameters of the task were similar to the dynamic IPD threshold discrimination described above, with certain modifications: trials were lengthened to 4 s, and pips to 2 s: a 500 ms 0° steady-state, a ramp of ±15°, ±30° or ±60°, and a second steady-state at that IPD for the remainder of the 2s (See Figure 2.1C). Tones were 70 dB SPL at the optimal carrier frequency for the cell, as measured by spike rate and synchrony in response to the binaural beat. The inter-stimulus interval was a minimum of 2 s, but was variable because each trial was initiated by the animal. The randomized phase excursions were chosen to be greater in magnitude to increase the chance that the stimulus would fall somewhere in the cell's preferred range of IPD – the phases used were standard across the recorded population, since animals were only trained to work around the midline and the stimulus could not be tailored to each cell. After a sufficient number of trials was run with satisfactory performance (15-25 trials per point, >90% correct), the task was stopped and the animal was allowed to sit until he stopped attempting to initiate trials. Then the equivalent number of stimuli was presented in blocks as the animal sat passively (stimulus order: +15, -15, +60, -60, -30, +30).

2.4 Results I: Psychophysical discrimination of ILD and IPD

After extensive training, both animals learned to lateralize ILD across a range of carrier frequencies. At 1 kHz, discrepancies of more than 5 dB (80 dB vs. <75 dB) were sufficient to allow performance of 95% correct or better, and performance degraded as the ILD diminished (Figure 2.2A). Plots like this were used to determine threshold (see Methods, Subjects and Psychophysical Testing Procedure, above). When performance at 5 dB ILD is compared across carrier frequency, the data are noisy and without any clear trend, though suprathreshold performance is demonstrated across 5 octaves (Figure 2.2B). At those frequencies where performance was tested down to threshold level, most thresholds fell between 2 and 3 dB (Figure 2.2C). Monkey Z was tested at ILDs relative to 90 dB at most frequencies, whereas X was trained relative to 80 dB, but this produced no difference in their measured thresholds at 1 and 2 kHz.

A prior study measured ILD discrimination in rhesus monkeys at 1 kHz, relative to a 40 dB standard, and found thresholds in the same range (1.8-3.7 dB), suggesting that threshold as measured in dB is relatively invariant with overall sound level (Wegener 1974). (Wegener's method also differed from ours in that his monkeys could use up to ten stimulus repetitions before responding, whereas ours had only a single presentation.) Overlaid on Figure 2.2C are thresholds obtained in *Macaca nemestrina* by Houben and Gourevitch (1979), which are consistently higher and

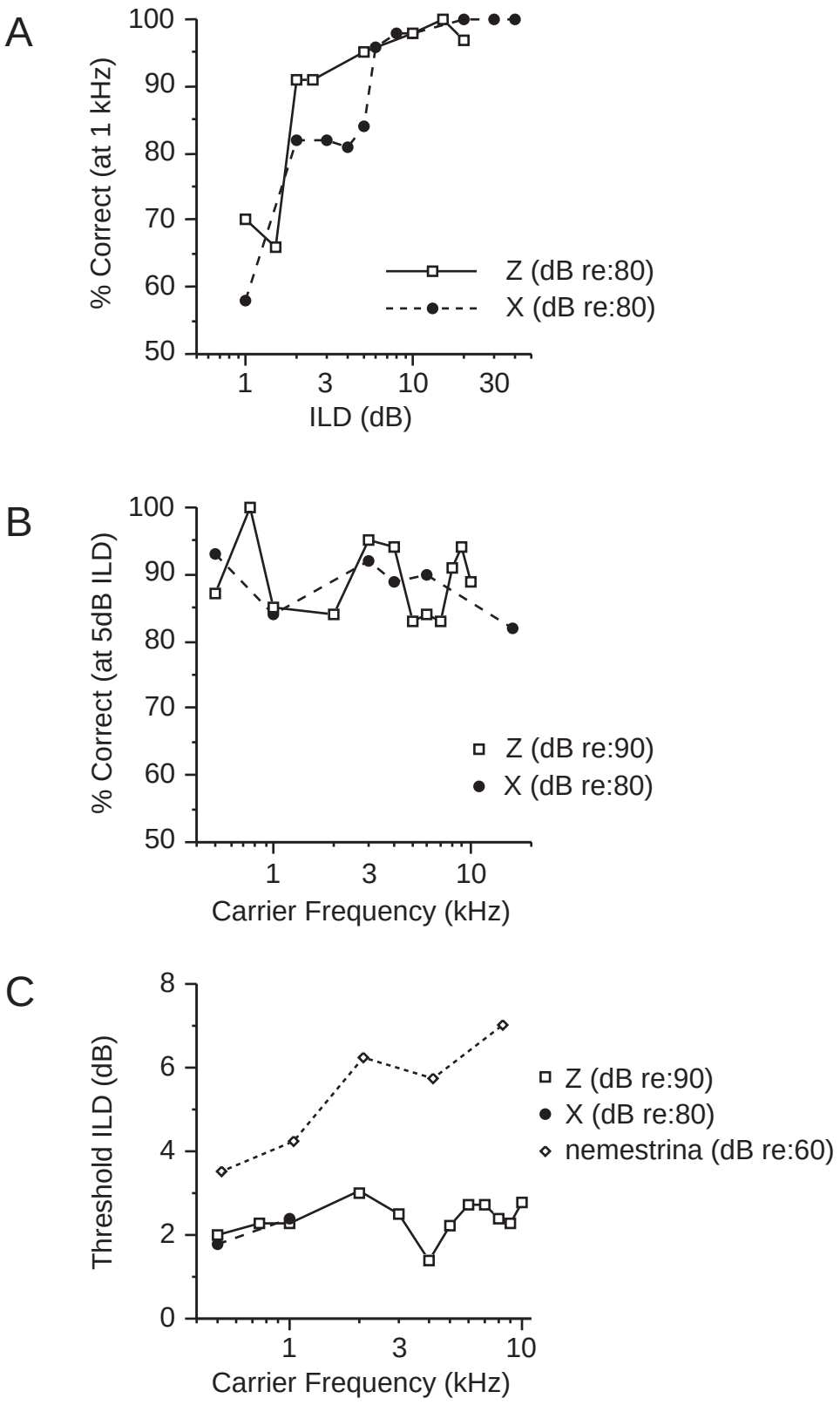
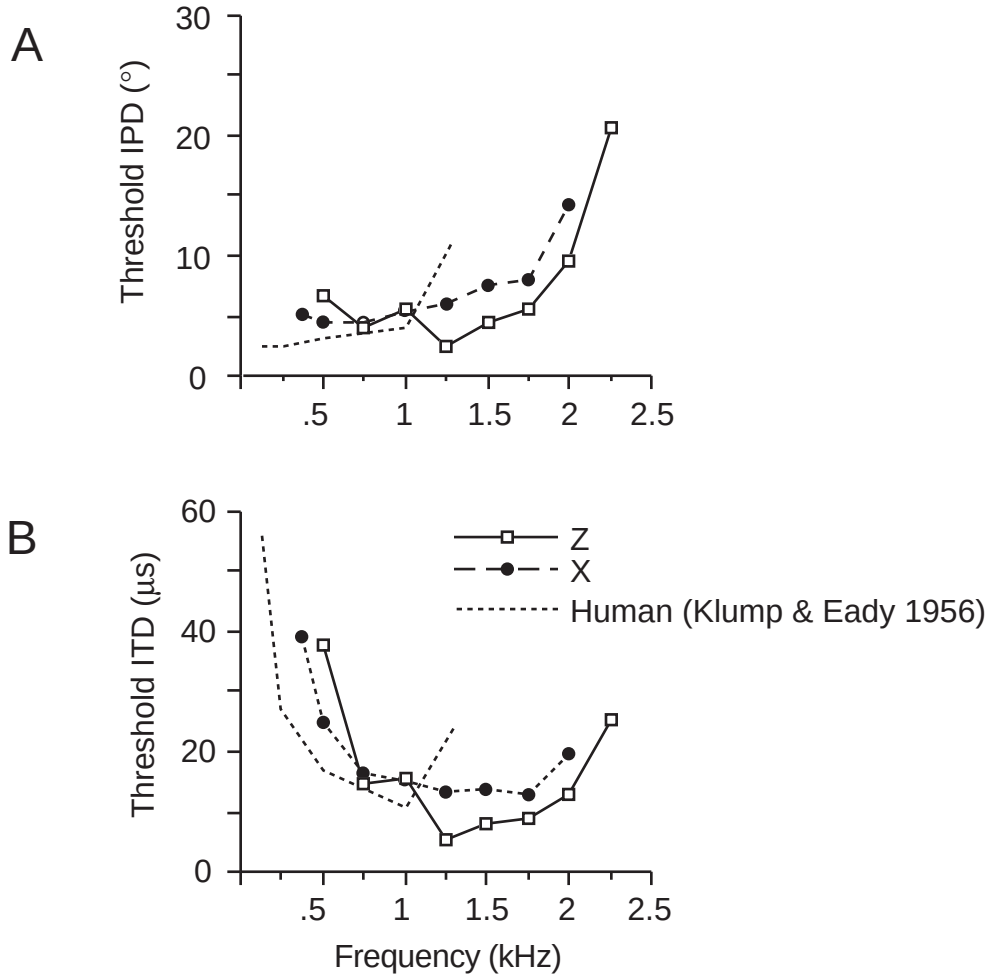

A
% Correct (at 1 kHz)
100
90
80
70
60
50
1
3
10
30
ILD (dB)
Z (dB re:80)
X (dB re:80)
B
% Correct (at 5dB ILD)
100
90
80
70
60
50
1
3
10
Carrier Frequency (kHz)
Z (dB re:90)
X (dB re:80)
C
Threshold ILD (dB)
8
6
4
2
0
1
3
10
Carrier Frequency (kHz)
Z (dB re:90)
X (dB re:80)
nemestrina (dB re:60)

Figure 2.2 Psychophysical results of static ILD lateralization. **A:** Percent correct lateralization across ILD at 1 kHz for both monkeys; louder ear was 80 dB. Spline fits of curves like these were used to determine ILD threshold at 76% correct. **B:** No clear effect of carrier frequency on discrimination of a 5 dB ILD, 0.5-16 kHz (monkey Z, louder ear was 90 dB; monkey X, louder ear was 80 dB). **C:** ILD thresholds across carrier frequency (threshold was only determined at two frequencies for animal X). Open diamonds are the results of Houben and Gourevitch (1979) in *Macaca nemestrina*.

show an upward trend with carrier frequency. By comparison, thresholds measured in humans range from 0.5-1 dB from 125 Hz to 8 kHz (Mills 1960, reproduced in Houben and Gourevitch 1979), and generally are insensitive to absolute intensity (Grantham 1995). The lower thresholds in humans and in our animals, and their relative invariance across carrier frequency, are most likely an effect of training and motivation. Measured ILDs for rhesus monkeys are smaller at lower frequencies (<4 kHz) where the shadowing effects of the head are weaker (Spezio *et al*. 2000). Thus a necessity for sharper sensitivity to ILD is a possible explanation for the frequency dependence in the prior data, but ILD is thought to be of less ethological importance in the low-frequency range (< 2 kHz) where IPD cues may be used.

Both animals subsequently were trained to discriminate static and dynamic IPD (see Methods) from 0.5 to 2 kHz, consistent with the only prior study to examine IPD discrimination by macaques across frequency (Houben and Gourevitch 1979). Animal X could perform down to .375 kHz, and Z up to 2.25 kHz – outside this range the animals required extremely large phase excursions to approach threshold, or refused to work altogether. Thresholds under the static and dynamic conditions were comparable for Z, but X performed better under the dynamic condition. Given that Z was more extensively trained, it is possible X could have improved with time, or that he relied on the presence of a standard comparison and/or dynamic sweep as a cue to lateralization. At low carrier frequencies, 76% correct directional


A
Threshold IPD (°)
30
20
10
0
.5
1
1.5
2
2.5
B
Threshold ITD (µs)
60
40
20
0
Z
X
Human (Klump & Eady 1956)
Frequency (kHz)

discrimination can be achieved with IPD excursions as small as 5°, and threshold increases with increasing carrier frequency for both animals (Figure 2.3A, dynamic IPD data shown). This frequency dependence is to be expected, since a fixed phase difference at higher frequencies represents a progressively shorter time delay[1]. When the same thresholds are re-plotted in terms of ITD (Figure 2.3B), thresholds for both animals reach a minimum around 1.25 – 1.75 kHz.

Houben and Gourevitch (1979) found a minimum threshold in a comparable frequency range, though their thresholds were at least twice as high overall. They also observed elevated thresholds (in ITD terms) at very low frequencies, speculating that this could be due to greater temporal jitter in phase-locked responses to very long-period tones in the auditory nerve, as suggested by Licklider (1950) and shown to be tenable by Rose *et al*. (1967). Brown *et al*. (1978) report thresholds for lateralization of pure tones in the free field that, when converted into IPD, show a fairly constant threshold around 11° IPD from 250-1000 Hz. Our thresholds, though lower, are also fairly constant in IPD up to 1 kHz. Overlaid on both plots are comparable data from human observers (Klump and Eady 1956),

---

[1] For example, a fixed IPD of 30° is a fraction (1/12) of the sound period, which is the inverse of frequency; if frequency doubles, the time delay (ITD) caused by that IPD halves. IPD refers to an ongoing phase difference in a periodic signal like a pure tone; ITD may also refer to a disparity in onset time – thus a click may have an ITD, but no definable IPD. For this stimulus, there is no onset disparity, just an ongoing phase difference expressed in μs instead of degrees of phase, so as to be comparable across carrier frequency.

showing that macaques are capable of fine discriminations in IPD at a level nearly comparable to humans at low frequencies, and superior to humans at carriers above 1 kHz. (The minimum ITD for monkey Z at 1.25 kHz is only 6 μs, lower than humans at any frequency.) This frequency span for behavioral acumen overlaps the distribution of BFs for the population of IPD-sensitive neurons (see Chapter 5). Sensitivity for IPD in human listeners appears to peak around 90 dB SPL (Zwislocki and Feldman 1956), suggesting the level used here (80 dB) may be nearly optimal for threshold determination in macaques. Robust cortical tuning to phase in low-frequency AI and R identified IPD discrimination as an appropriate behavior for the comparison of physiological data collected under behaving and passive listening conditions.

2.5 Results II: Effects of behavioral context on neural responses to IPD

The most common effect of behavioral state on neural responses to IPD is exemplified by the cell in Figure 2.4: consistent tuning to IPD under both conditions, with an overall increase in discharge rate during behavior. Spike time rasters on the left were collected as the animal performed the randomized IPD lateralization task (see Methods), those on the right were collected in blocks after the animal was taken off the behavioral task and left to sit passively. The cell gives a sustained response to the onset of the tone at 0° IPD, followed by a decrease or increase in firing rate in response to modulation in the negative or positive

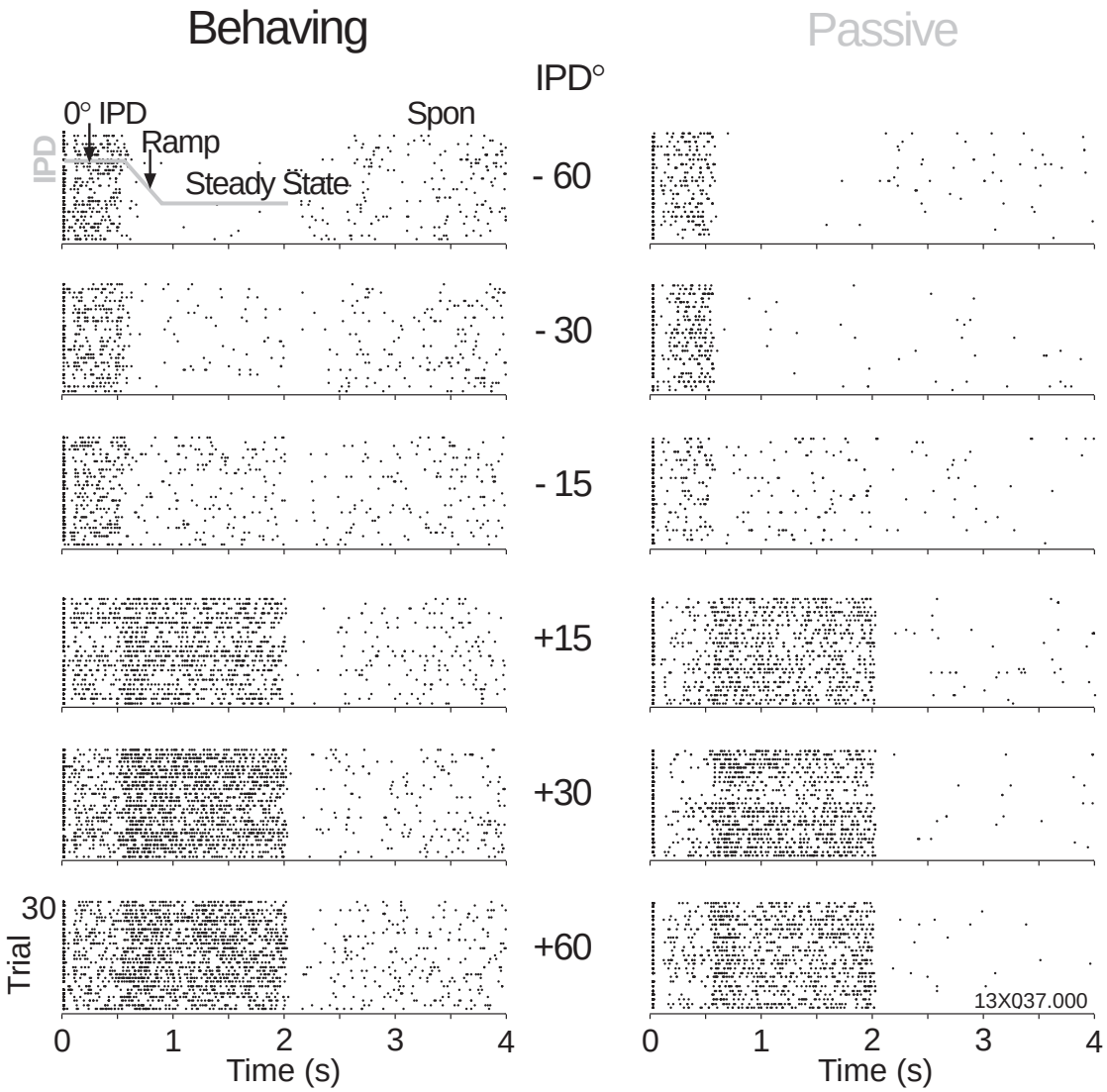
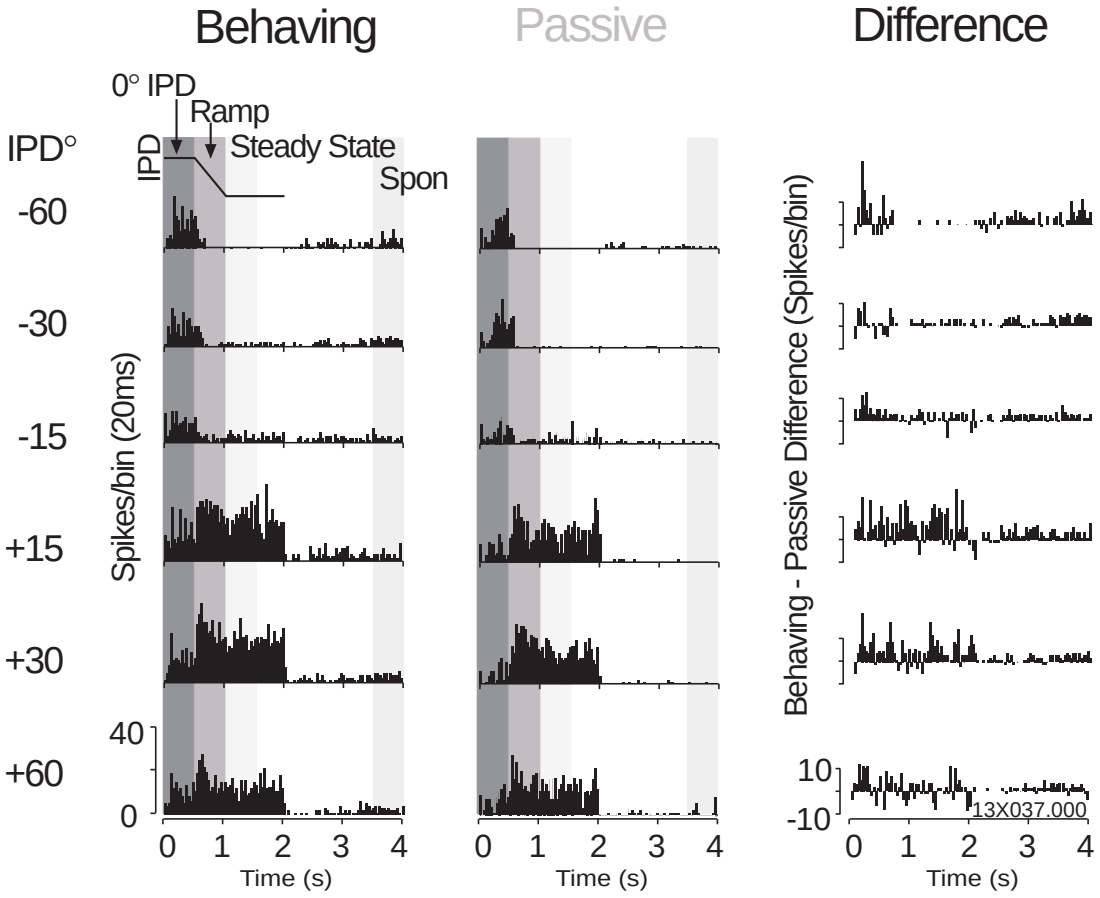


Figure 2.4 Spike rasters recorded in response to dynamic IPD under behaving and passive conditions. Behaving data are in the left-hand panels, passive on the right, arrayed from negative/ipsilateral phase shifts (top) to positive/contralateral phase shifts (bottom). Icon in gray over upper left raster shows time course of the stimulus.

direction, respectively (for this cell, positive phase shifts simulate motion to the contralateral side, relative to the recorded hemisphere). The icon overlaid on the uppermost behaving panel shows the time course of the stimulus (0-2 s). The rasters collected under the passive condition, at right, show the same stimulus effects with fewer overall spikes, particularly noticeable in the spontaneous discharge after the cessation of the stimulus (2-4 s).

I will refer to the "spontaneous rate" measured during part of this interval, but these spikes may not be truly stimulus-independent for two reasons. First, immediately after the response to the stimulus, an aftereffect is evident as a gap in activity after the strong sustained discharge at +60°, possible post-excitatory suppression. Second, in the behaving case, the animal may begin to initiate his response as soon as 300 ms after the sound turns off, after which he will press a button and (on >90% of trials) drink his reward, both of which may produce sounds. I attempted to minimize the stimulus effect by measuring spontaneous rate in the final 500 ms window of the trial (3.5-4 s), but this does not counteract the heightened possibility of self-produced sounds during behavior.

Spike rates were measured in four 500 ms epochs of the trial, demarcated in the peri-stimulus time histograms (PSTHs) of Figure 2.5 (same data as previous figure): the first 500 ms at 0° IPD, the second 500 ms which contains the linear


Behaving
Passive
Difference
0° IPD
Ramp
Steady State
Spon
IPD
IPD°
-60
-30
-15
+15
+30
+60
Spikes/bin (20ms)
40
0
Behaving - Passive Difference (Spikes/bin)
10
-10
13X037.000
0 1 2 3 4
Time (s)

ramp of IPD, the third 500 ms as a sample of the steady-state response to the shifted IPD, and the last 500 ms of the trial as a measure of spontaneous discharge rate. The right-hand set of histograms plots the difference in bin values between the two conditions, such that positive bin values indicate a higher spike count during behavior. Mean firing rates and standard deviations are plotted across IPD for each epoch in Figure 2.6A (behaving in black, passive in gray). The first panel shows the responses to the 0° IPD onset, as a control for overall excitability, which remains constant during the behaving trials at a rate higher than that seen during passive trials. The exceptions, at –30° and –60° passive, are the result of a stimulus artifact: these trials were collected in blocks, and in each case the onset response follows a strongly suppressive steady-state response. Despite a 2 s off time between each stimulus, the onset response is stronger in these trials than in those where the steady-state response is excitatory, suggesting an adaptive mechanism operating over a relatively long time scale (the effect of which may be averaged out when data are collected in a randomized order). The second and third panels of 2.6A show tuning to IPD measured during the dynamic and steady-state epochs of the stimulus, with a clear tuning to the contralateral side peaking at +30°. The excitatory response is augmented under the behaving condition; because the response to ipsilateral IPDs is suppressed down to zero, this effectively enhances the dynamic range of the cell's firing rate. The fourth panel shows that spontaneous

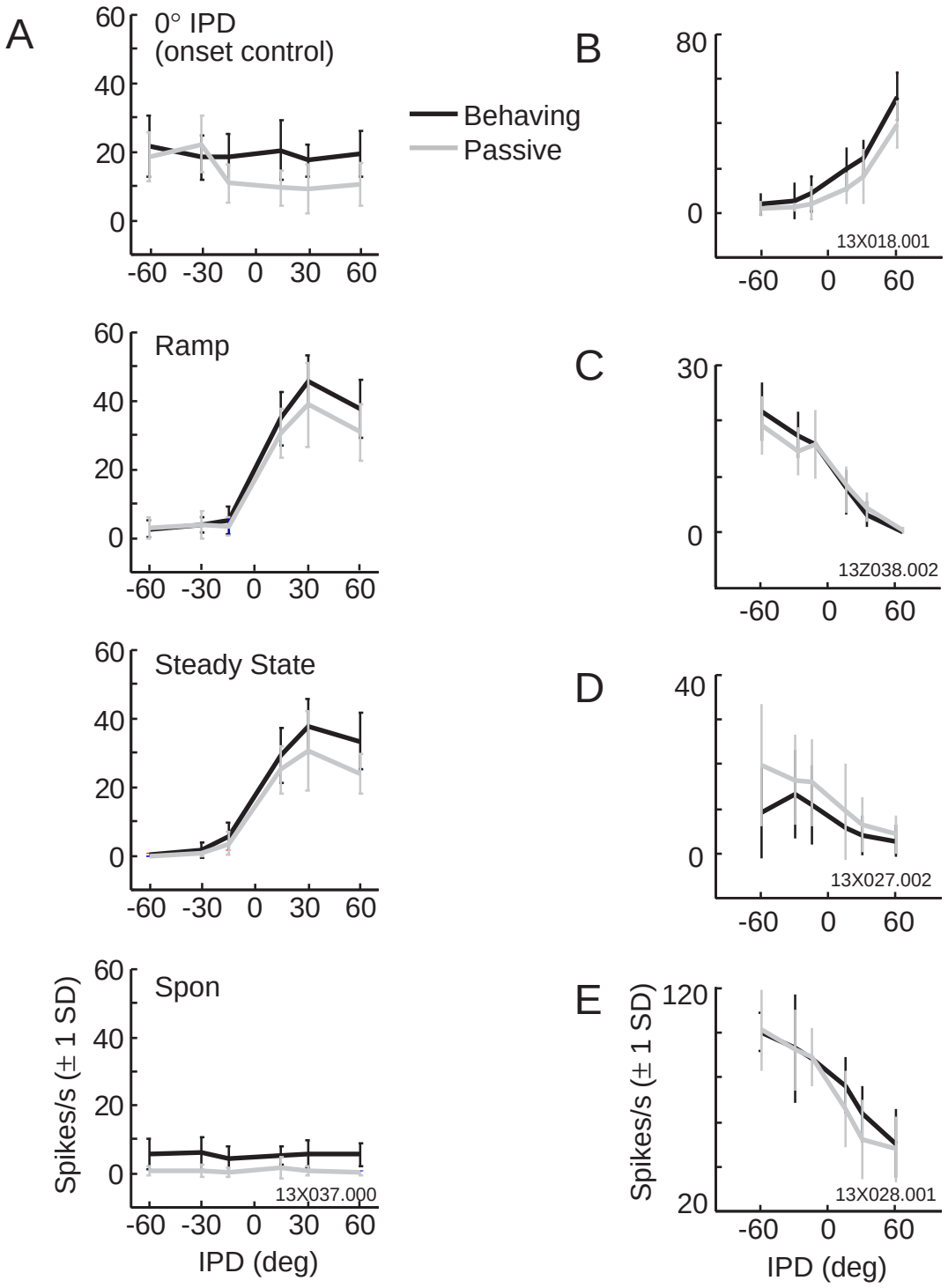

A
0° IPD
(onset control)
Ramp
Steady State
Spon
13X037.000
Behaving
Passive
Spikes/s (± 1 SD)
IPD (deg)
B
13X018.001
C
13Z038.002
D
13X027.002
E
13X028.001
Spikes/s (± 1 SD)
IPD (deg)

Figure 2.6 Firing rate tuning to IPD under behaving and passive conditions. **A:** Mean firing rates (error bars ± 1 standard deviation) during the four stimulus epochs defined in Figure 2.5, behaving (black) and passive (gray) conditions. **B-E:** IPD tuning functions during the steady-state for four example units.

activity is reliably constant across stimulus conditions, but consistently affected by behavioral condition. This was a common finding in the population.

Other examples of behavioral context affecting – or not affecting – responses in the steady-state epoch are arrayed on the right side of Figure 2.6. As in the example cell already discussed, contralateral tuning with a slight but consistent elevation in discharge rate during behavior was the most common finding (panel B). Some cells were tuned to ipsilateral phase excursions (panel C) and/or showed higher firing rates in the passive state (both seen in panel D, though these phenotypes did not covary). Other cells, like that in panel E, showed no reliable effect of behavioral context. Whatever the changes in response amplitude, behavioral state was never found to shift the *tuning* of the cell to IPD.

Tuning functions were analyzed in 31 single units, chosen for their sensitivity to IPD when tested with binaural beats, stable isolation during recording, and a sufficient number of behaving trials with adequate performance from the animal. Subsequent assignment of the cells to cortical fields (see Chapter 3) revealed 25 of the cells to be from AI, and 6 from R. Tuning and performance effects were comparable in the populations of cells recorded from each field, and each animal (13 from Z, 18 from X) so they were pooled in this analysis. Tuning to IPD contralateral to the recorded hemisphere was prevalent (22/31 cells), with some

tuned to ipsilateral IPD (7/31). (One unit had a response minimum at 0° with symmetrical responses about the midline, and another displayed no clear selectivity for IPD during the behavioral test despite responding to binaural beats.) Of the 31 cells, 23 showed some enhancement of the driven response during behavior, and 23 showed higher spontaneous discharge. These populations were largely overlapping, except for two units that showed one effect but not the other. Only 5 units showed higher driven rates in the passive state, and three showed elevated spontaneous rates. Three cells showed no effect of behavior.

To examine the trend in the population, firing rate pairs for all IPDs presented under the two conditions are plotted in Figure 2.7, for each epoch of the trial. (Including early data with only two IPDs tested per cell, there are 42 cells pooled: 12 contributing 2 points each, 30 contributing 6 points each.) Each epoch of the response shows a significant shift toward higher spike rates during behavior (paired *t*-test, $p<.01$). Because both driven response and spontaneous rate are elevated, the behavioral advantage of these extra spikes toward the lateralization task is not clear: the gain seems to affect both signal and noise. A mechanism that is adaptive, in the evolutionary sense, would enable discrimination by enhancing the contrast in firing rates between different IPDs.

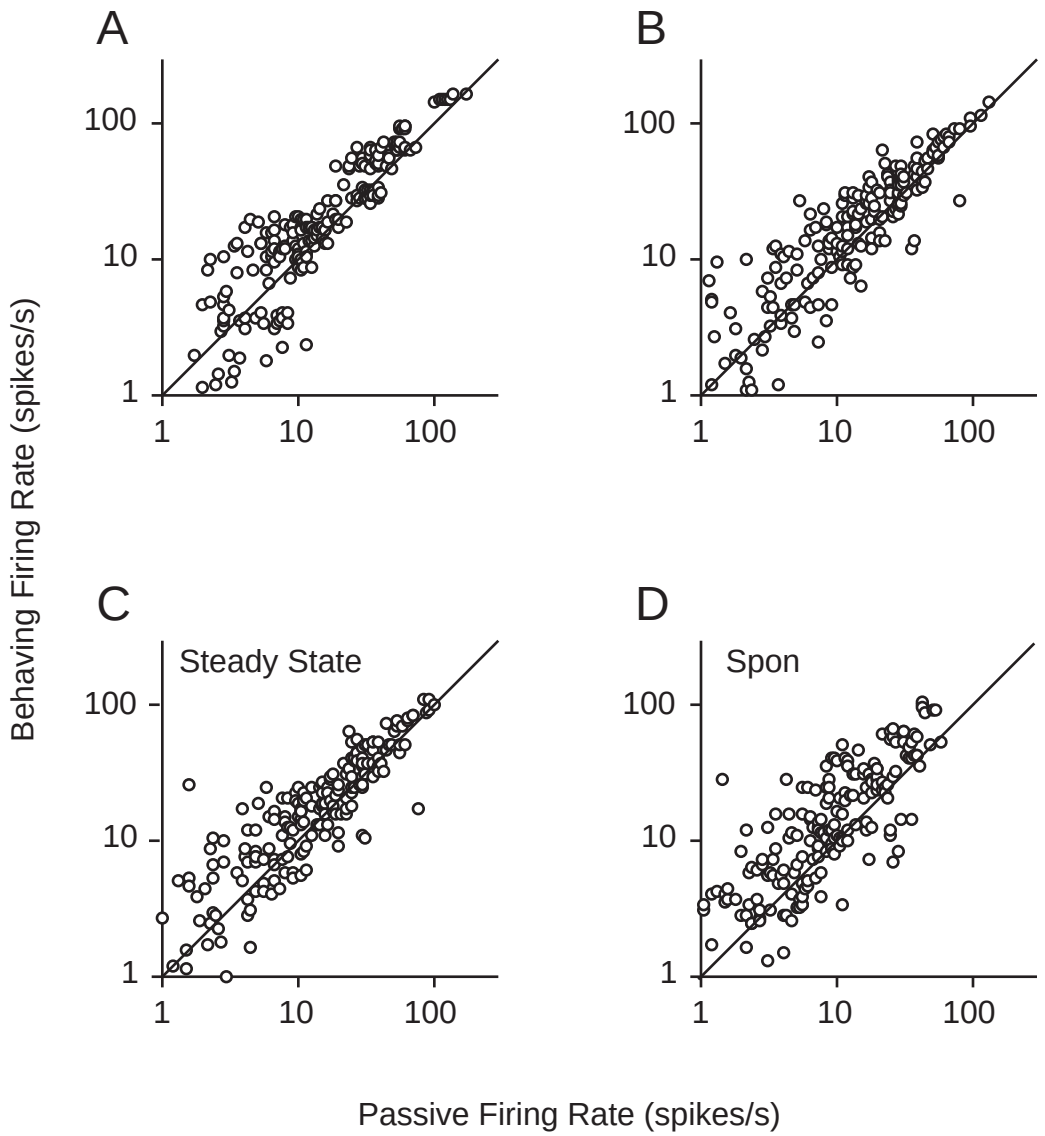
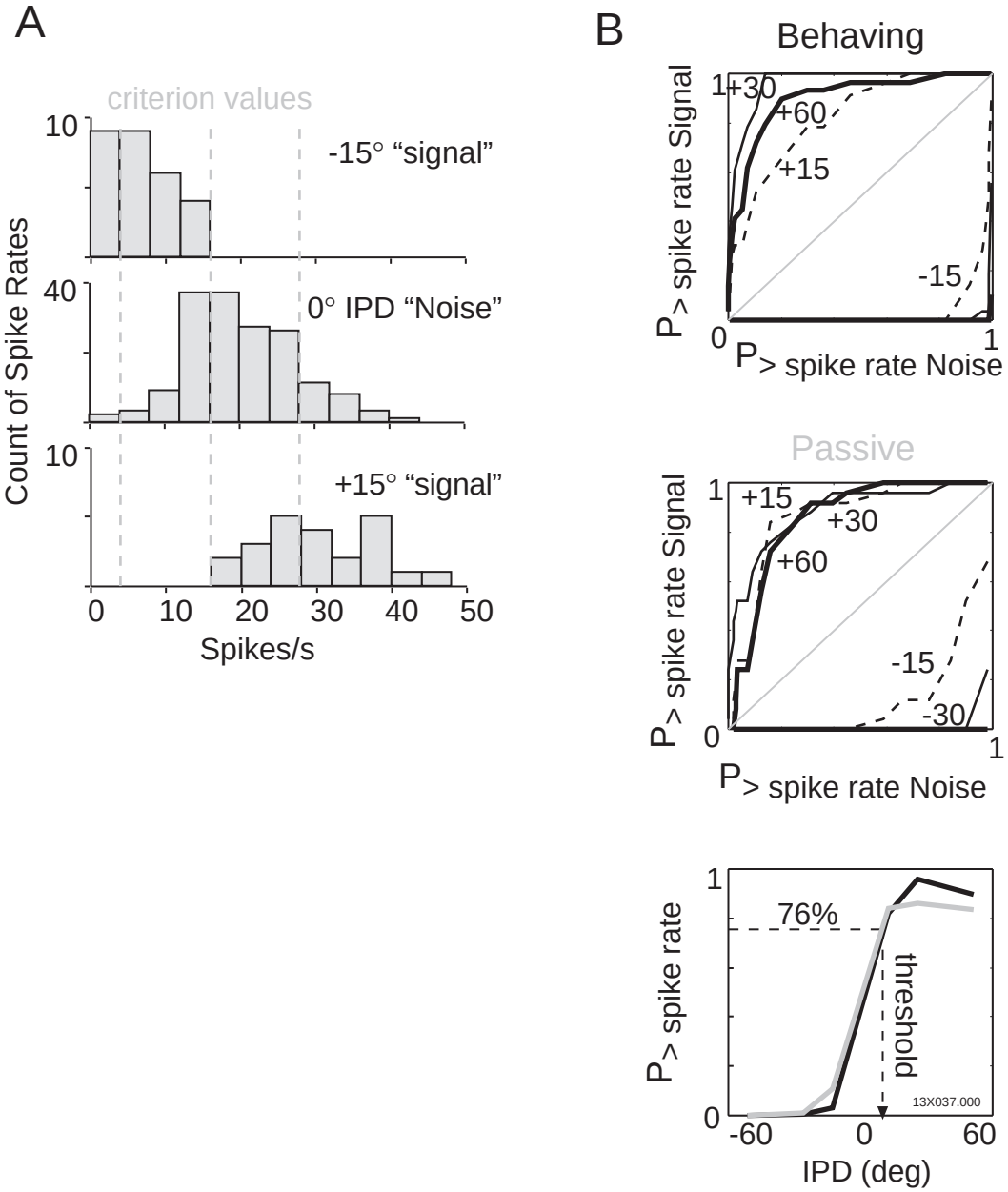


Figure 2.7 Comparison of firing rates under behaving and passive condtions for the full population of equivalent stimuli (each cell contributes mulitple points). Points above the line of unity indicate an elevated firing rate during behavior. All four epochs show a significant shift by a paired *t*-test, strongest for the spontaneous rate (see text).

To assess whether the gain associated with performance modifies the discharge of single units in a manner advantageous to the behavioral task, a Receiver Operating Characteristic (ROC) analysis was performed. This form of "neurometric" analysis compares the distributions of spike rates in response to a signal (e.g. a shift in IPD) to the distribution of rates associated with "noise" (Britten *et al*. 1996, Skottun *et al*. 2001). (The technique is based on engineering principles used to analyze psychometric functions by Green and Swets [1966].) In this case, all firing rates in response to the 0° IPD onset epoch were pooled as the noise distribution, for comparison to the distributions of rates elicited by the 6 IPD values. For instance, an ROC function for +15° is constructed by comparing the distribution of firing rates in response to 15° and 0° at a series of criterion values spanning the full range of firing rates (Figure 2.8A; same unit shown in Figures 2.4, 2.5, 2.6A). At each criterion, the probability that a spike rate drawn from the signal distribution would exceed the criterion is plotted against the probability that a spike rate drawn from the noise distribution would exceed the criterion. If the signal distribution is higher than noise, these points will trace a curve above the diagonal (2.8B, top panel, dashed line "+15"), whereas a distribution of rates lower than the noise traces a curve below the diagonal (same panel," –15"). This comparison was made for each IPD value under the behaving and passive conditions (2.8B, top and middle panels respectively). Integrating the area under each curve yields the probability that a given signal (IPD) will elicit a firing rate greater than the noise (0° IPD). These sets


A
criterion values
10
-15° "signal"
40
0° IPD "Noise"
10
+15° "signal"
Count of Spike Rates
0
10
20
30
40
50
Spikes/s
B
Behaving
1
+30
+60
+15
-15
0
1
P> spike rate Signal
P> spike rate Noise
Passive
1
+15
+30
+60
-15
-30
0
1
P> spike rate Signal
P> spike rate Noise
1
76%
threshold
0
-60
0
60
P> spike rate
IPD (deg)
13X037.000

of points define neurometric functions under each condition, seen in the lower panel of Figure 2.8. The two curves largely overlap, with a slightly steeper slope under the behaving condition suggesting slightly enhanced discriminability based on a comparison of firing rates in the first two epochs of the response. (It should be noted that we have no proof this is the cue the animal, or the brain, uses – e.g., the steady state response could also be compared to the spontaneous rate over a given window to yield comparable functions.)

The effects of behavior on ROC measurements across the sampled population showed the same variability seen in the spike rate effects, but the two did not necessarily correspond. Example functions in Figure 2.9 illustrate cases where the effects of behavior on ROC imitate the effects on spike rate. The first panel (Figure 2.9A) replots the data from 2.6B as neurometric functions, showing a behavioral advantage by either metric; 2.9B shows a very slight behavioral advantage in an ipsilaterally-tuned unit (same cell as 2.6C). The higher spike rates under passive conditions in 2.6D translate into a higher neurometric function in 2.9C. The curves in figure 2.9D are from units in which firing rates under the two conditions were highly similar (data not shown), resulting in similar neurometric functions. This continuity between firing rate and neurometrics did not hold for the population: of the 23 cells showing elevated discharge rates during behavior, only 11 showed enhanced discriminability as measured by ROC. Ten of these units had neurometric

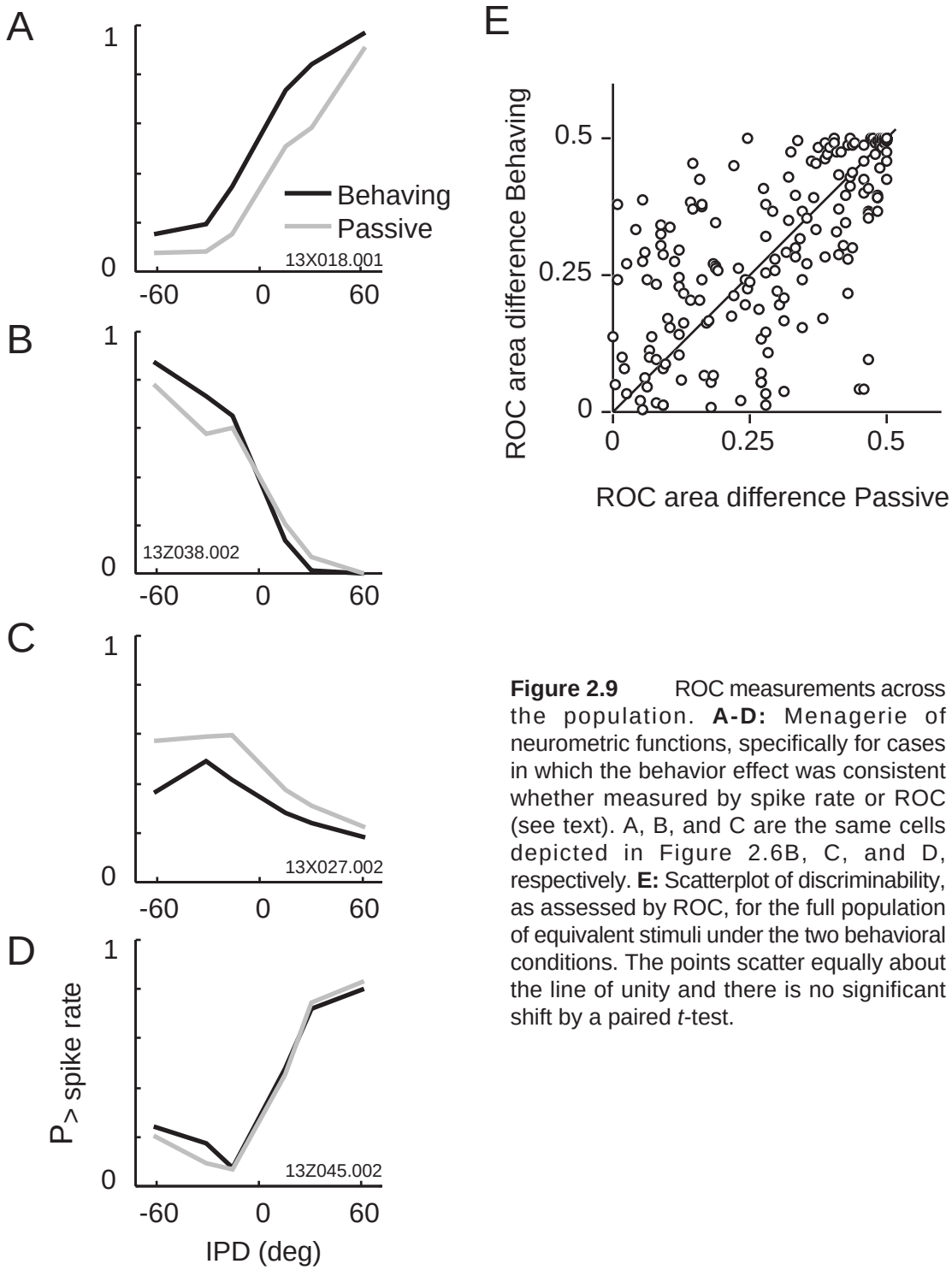


Figure 2.9 ROC measurements across the population. **A-D:** Menagerie of neurometric functions, specifically for cases in which the behavior effect was consistent whether measured by spike rate or ROC (see text). A, B, and C are the same cells depicted in Figure 2.6B, C, and D, respectively. **E:** Scatterplot of discriminability, as assessed by ROC, for the full population of equivalent stimuli under the two behavioral conditions. The points scatter equally about the line of unity and there is no significant shift by a paired *t*-test.

functions that were higher under the passive condition, and in two cases the curves were indistinguishable. To examine the effect of behavioral condition on the population overall, the area between each ROC curve and the diagonal was taken as a metric of discriminability for each measurement. (The area between the curve and the diagonal represents discriminability via an increase or decrease in firing rate, whereas measuring the total area under the curve presupposes that "signal" is always carried as an increase in spike rate.) When these areas are plotted for all 42 cells (Figure 2.9E), there is no evident bias toward enhanced discriminability under either behaving or passive conditions (paired *t*-test is insignificant). Thus the observed elevation in spike rates does not seem to translate into enhanced discriminability by a neurometric analysis.

Control experiments were carried out on some cells to verify the time course and generality of the effects observed. To assess the frequency specificity of the observed effects, a few cells were tested that were outside the frequency range for IPD sensitivity. Among 6 cells tested this way, 2 showed elevated discharge rates during behavior, one responded more strongly during passive listening, and 3 showed no effect (rate-tuning and neurometric functions were flat across IPD in all cases.) Figure 2.10A (top panel) shows spontaneous firing rates for a cell insensitive to IPD, but from a penetration surrounded by binaural beat-sensitive tissue. Spontaneous rate is not consistently different between the two conditions,

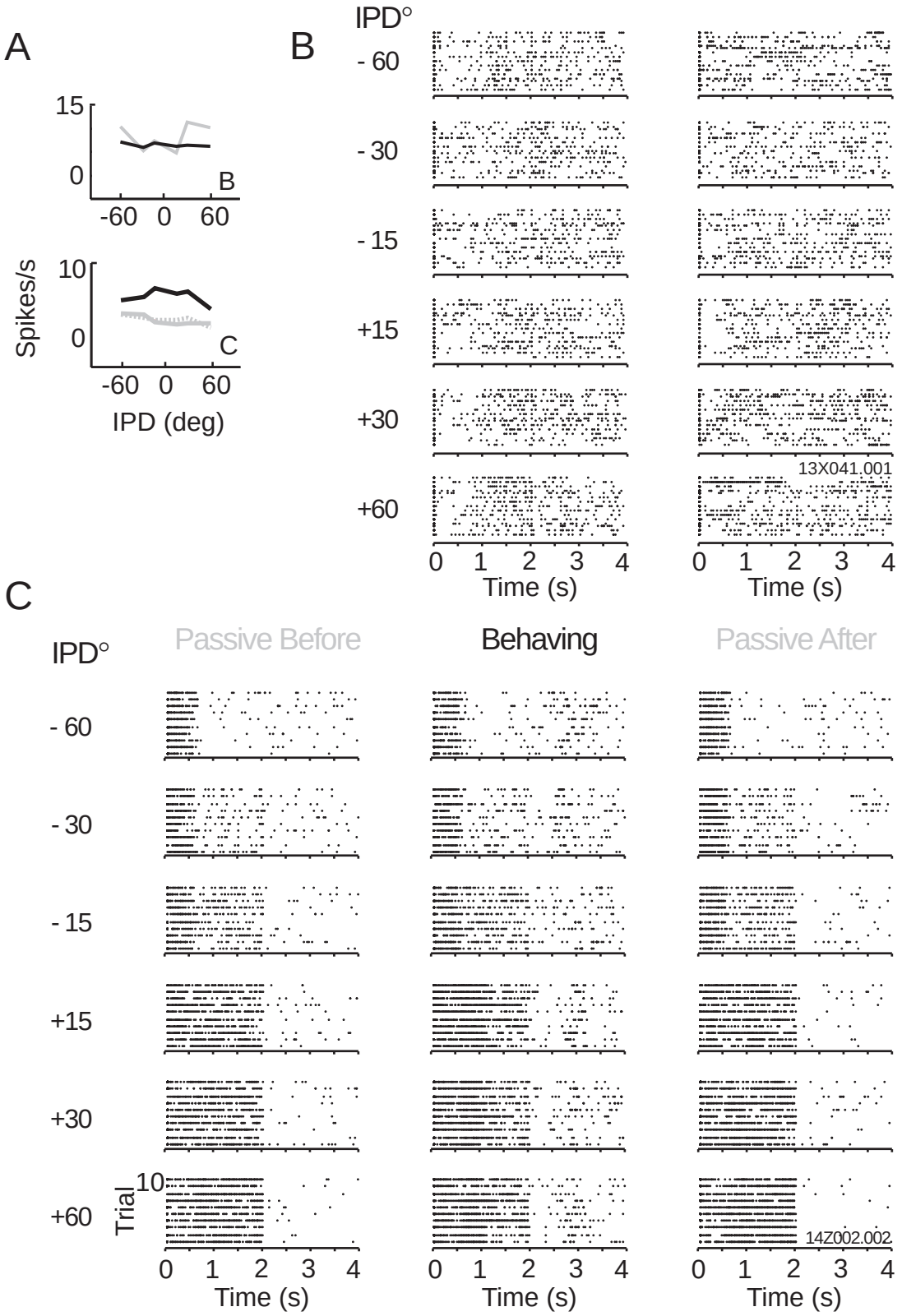

A
15
0
B
-60 0 60
10
0
C
Spikes/s
-60 0 60
IPD (deg)
B
IPD°
- 60
- 30
- 15
+15
+30
+60
13X041.001
0 1 2 3 4
Time (s)
0 1 2 3 4
Time (s)
C
IPD°
Passive Before
Behaving
Passive After
- 60
- 30
- 15
+15
+30
+60
Trial
10
14Z002.002
0 1 2 3 4
Time (s)
0 1 2 3 4
Time (s)
0 1 2 3 4
Time (s)

Figure 2.10 Controls for specificity of behavior effects in frequency and time. **A:** Spontanesous firing rates under behaving (black) and passive (gray) conditions for a unit unresponsive to tones at the carrier frequency used (upper panel), and for a responsive unit with passive data collected both before and after the behavior (dashed and solid gray, respectively; lower panel). **B:** Spike rasters for the poorly responsive cell shown in the upper panel of A; behaving on the left, passive on the right. Irregular bursting appears more common in the passive rasters, but rates are similar overall. **C:** Spike rasters collected before, during, and after behavior for the data shown in the lower panel of A. The elevation in spontaneous discharge rate during behavior, and its dissipation soon afterward, is visible.

but the error bars and the rasters (Figure 2.10 B) suggest more variability in firing rate during the passive trials. The fact that a cell insensitive to the stimulus could show an elevated discharge suggests that at least part of the effects described here may derive from a general arousal of auditory cortex during behavior, rather than a targeted enhancement of the cortical circuits analyzing the stimulus. A second factor that could bias the outcome is the order in which the data were acquired, since passive trials were always collected after the cessation of behavior. Two units were recorded with passive blocks before and after the behaving trials, and in both cases spike rates were elevated during behavior, whereas rates during both passive blocks were indistinguishable. The lower panel of 2.10A plots spontaneous rate during behavior (black line) and passive trials before and afterward (gray lines, dashed and solid respectively). The rasters from which these curves were generated are shown in 2.10C, where the temporary elevation of spontaneous rate (from 2-4 s) and consistency of IPD tuning (from 1-2 s) are evident. Taken together, these controls suggest that the gain observed in some cells may not be specific in tonotopic extent, but is confined in time to the behaving trials, and dissipates within a few minutes after performance stops.[1]

---

[1] A passive set of 25 trials lasts 100 s per stimulus, and in some cells a "residual" elevation of spontaneous discharge could be seen in response to the first (but only the first) stimulus. In these instances a second set was collected at that IPD.

2.6 Discussion

The "awake-passive" preparation is a valid compromise between recording in untrained animals, with a lack of behavioral control, and recording only in behaving animals, which restricts the investigator to those stimuli on which the animal has been trained to respond. In these experiments, both macaques acquired an auditory lateralization behavior that motivated their cooperation in the experiments, while the behavioral control gained by this training allowed extended recordings while the animal sat passively.

2.6.1 Effects of behavioral state – comparison with previous studies

Individual neurons in AI recorded during active discrimination and passive listening demonstrated comparable tuning to IPD, with a tendency toward higher discharge rates under the behaving condition. Similar results have been described previously in monkeys trained on auditory tasks, though these studies are part of a large literature that attributes response plasticity in auditory cortex to a wide range of overlapping phenomena: attention, performance, behavioral state, wakefulness, and stimulus predictability. Recording from awake, freely-moving cats, Hubel (1959) first described units in auditory cortex that responded only if the animals appeared to be "paying attention" to the stimulus, but the validity of these early observations is hampered by issues of sample size, behavioral control and recording location that would recur in later studies. In this case, behavior was not

controlled, and "attention" toward a stimulus was gauged by whether the cat was looking toward it; this could be an effect of attention, spatial tuning, position of the pinna, or multi-modal integration. The units recorded were outside AI, and accounted for only about 10% of the units studied. A tenable hypothesis was that this subset of cells was affected by the behavioral state of the animal, a hypothesis which would be tested in macaque monkeys by several groups.

Many of the effects described in this chapter were presaged by Miller *et al*. (1972), who explored the behavioral dependency of single-cell activity in the rhesus monkey auditory cortex. A short paper in which data from only two cells are shown, it argues from the standpoint that the "contribution of data obtained in acute studies to the understanding of neural functioning in awake, behaving animals is to some degree restricted." In awake, behaving animals trained on a simple reaction-time task, they find strong responses to repetitive tones, compared to similar but weaker responses in the passive state and a lack of consistently strong responses in awake but untrained monkeys. On the basis of this, many future studies would employ simple behaviors like the reaction-time task, intended not to explore the ability of macaques to do such tasks, but merely to engage the animal in listening (e.g. the "go/no-go" task employed by Recanzone *et al*. [2000]). Hocherman *et al*. (1976) compared unit activity during a selective-attention task with that collected under passive conditions, and found a mix of effects like that reported here. Effects

of reaction-time performance on single units throughout the auditory pathway are reported by Ryan *et al*. (1984), who found increased driven rates in some proportion of cells from the cochlear nucleus to the cortex, but strangely no effect on spontaneous discharge rates at the cortical level. They also show an example, and report a trend, of shorter and more reliable initial latencies in the cortex during behavior; latencies to tones collected under the passive condition (discussed in Chapter 3) were quite reliable in our sample.

Pfingst *et al*. (1977) expand the comparison of single-unit responses to include the awake-behaving, awake-passive, sleeping and anesthetized states. Not surprisingly, sleep and anesthesia (ketamine) could drastically alter responsiveness, echoing the finding of Brugge and Merzenich (1973) that driven and spontaneous activity were markedly reduced during sleep. Like Ryan *et al*. (1984), they report an enhanced driven response during behavior but no change in spontaneous rate (in 5 of the 6 cells studied). In monkeys trained to attend to the left or right ear, 14 of 77 cortical units showed an elevation in discharge during behavior relative to passive listening, but only 7 of 77 showed an effect of attention to the left versus the right ear (Benson and Heinz 1978). Our experiments did not allow performance and attention to be disambiguated, but it seems that "attention", as teased out by Benson and Heinz, affects a small proportion of cells. I cannot fully discount the possibility of "covert attention" (i.e. the animal mentally doing the task though not pressing

buttons), except by anecdotal observation: thirsty animals would sometimes continue to press buttons at the beginning of the passive blocks but gave up after a few trials, once they realized they were no longer in control of the stimulus and reward (these trials were discarded, and the passive block re-started after the cessation of button-pressing). The only disparity between prior studies and the data reported here is the elevation of spontaneous discharge during behavior in our sampled neurons. These papers do not specify how spontaneous rate was measured, but it is possible a different temporal window was used, one less prone to contamination from self-produced sounds during performance of the task.

In a prior study from this laboratory examining the effects of stimulus context, rather than behavioral context, on cortical tuning to IPD (Malone *et al*. 2002), the IPD presented prior to the tested stimulus evoked shifts in tuning averaging 60°. In many cells, the difference in firing rate dependent on the preceding stimulus was as great as that which defined the tuning curve itself. In one cell tested under the behaving vs. passive paradigm above, and with the trapezoidal IPD stimulus of Malone *et al*. (2002), stimulus context evoked a shift of 45° in its tuning curve (Figure 2.11, "origin" and "target"). By contrast, to align the passive and behaving tuning curves (which covered only the restricted range of –60° to 60° IPD) at their most disparate points would have required a shift of only 15°. Although I can only speculate as to the shapes of the complete tuning curves measured during behavior,

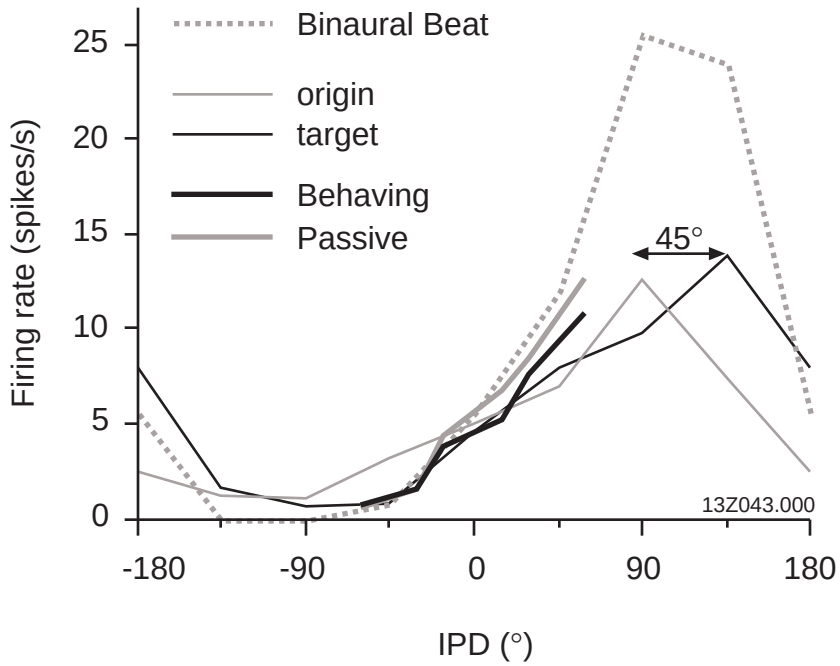


Figure 2.11 Behavior effects on tuning curves relative to the tuning shifts induced by stimulus context. Heavy dashed gray line: spike rate in response to a binaural beat. Tuning to IPD during passive listening is influenced by the preceeding stimulus – in this case, a positive ("origin", thin gray line) or negative ("target", thin black line) 90° phase shift (see Malone *et al.* 2002). The computed phase shift between these functions is 45°. Heavy black and gray lines: steady-state responses under behaving and passive conditions, respectively. Tuning from -60° to 60° is quite similar in all cases.

it appears that the effects of stimulus context (measured in the awake-passive state) may outweigh any effects of behavioral context.

2.6.2 Effects of Training

Along with anesthesia, sleepiness, or immediate behavioral performance, the cortical representation of sound may be altered by behavioral training itself. Miller *et al*. (1972) report that single units in the auditory cortex of awake but untrained animals showed unreliable responses to simple, repetitive stimuli like blocks of tones, and argue that two effects are at work in their data: performance condition and long-term training. Several studies have shown that certain types of training can distort individual AI receptive fields, and the cortical tonotopic maps which they build (Weinberger and Bakin 1998, Kilgard and Merzenich 1998, Fritz *et al*. 2003), but these studies differ from ours in several respects. Kilgard and Merzenich simulated behavioral reinforcement by direct stimulation of the nucleus basalis, which may be an exaggeration of natural reward processes; Fritz *et al* (2003) used aversive conditioning by shock avoidance, which presumably activates different modulatory circuitry from that involved in registering a liquid reward (LeDoux *et al*. 1988). The choice of training stimulus may also serve to distort cortical responses, if exposure to a single frequency is extensive (Beitel *et al*. 2003). Outside of our experiments, these animals were housed in a colony of macaques and exposed to a variety of environmental sounds and vocalizations, unlike animals

in plasticity studies that can be housed in soundproof boxes and exposed only to the conditioned stimulus. During experiments, our monkeys performed their tasks at a range of carrier frequencies that were varied within training sessions once the behavior was acquired, and by the time the behaving/passive comparisons were made, carrier frequency was chosen based on the best frequency for the cells encountered. In addition, the animals were given intermittent liquid rewards between passive recording runs if they remained still. Through training the animals may have come to associate the stimuli presented passively with their task and reward, which may explain the robust responses to simple stimuli found in AI in this report and Miller *et al*. (1972).

2.6.3 Behavior and training beyond AI

Effects of performance may not be of great consequence in determining the stimulus tuning of a primary sensory cortex like AI, but there is evidence that acoustic responses in cortical fields beyond AI may be more sensitive to these contingencies, and to training itself. Benson *et al*. (1981) compared responses to equivalent stimuli during two different behavioral tasks, localization and detection. Of cells showing a difference, none appear to be from AI on the basis of histology or response latency. In fields far rostral to AI on the STG, responses have been documented that depend on unpredictable reward delivery (they are not acoustic responses to the sound of drinking, because the cells do not respond when liquid is

delivered as a reward during performance) (Yin *et al*. 2003). Well outside of the auditory cortex, the lateral intraparietal area (LIP) can respond to acoustic stimuli, but only if the monkey has been trained to direct saccades to the location of the sound (Grunewald *et al*, 1999, Linden *et al*. 1999). Again, this cannot be considered a sensory response, as it is contingent on a trained motor output.

Within the primary auditory cortex, the awake-passive preparation appears to offer an unbiased sample of the cortical representation of acoustic stimuli, in the absence of anesthetic depression but free from the constraints and potential artifacts of awake-behaving physiology. Pfingst *et al*. (1977) conclude from their study that "this cortical tissue is more than a simple encoder and processor of auditory information", but subsequent studies – including this one – suggest that the effect of behavioral performance on AI responses in an awake, trained macaque is slight, especially when compared to cortical responses under anesthesia. While behavioral training may not be strictly necessary, it does offer the advantage of extended recordings from a more compliant animal under passive conditions, recordings that may be carried out in parallel with an awake-behaving experiment.

Chapter 3 – Physiological organization of the core

3.1 Introduction

At the foundation of the prevailing model of auditory cortical organization is the "core" auditory cortex, the first stage in the processing hierarchy and the exclusive recipient of auditory inputs from the MGv (Hashikawa et al. 1995, Molinari et al. 1995). As summarized in Chapter 1, the core and surrounding belt were defined on the basis of anatomy (Pandya and Yeterian 1984, Jones et al. 1995). Specifically, a dense, homogeneous immunoreactivity for parvalbumin in the thalamorecipient layers is coextensive with staining for acetylcholinesterase and cytochrome oxidase (Morel et al. 1993). Other anatomical studies have elucidated how subdivisions of the thalamus project selectively to the core and belt fields (Molinari et al. 1995, Hashikawa et al. 1995), and described the connectivity between core and belt (Hackett et al. 1998). Within this framework, we have only begun to investigate how anatomical structure relates to physiological function.

Tonotopic order within AI of the macaque monkey was revealed by Merzenich and Brugge (1973), who described the progression from high to low frequencies as their electrode was moved from caudo-medial to rostro-lateral. Transition into R (their RL) was marked by a reversal, as best frequencies reached their minimum and began to rise again further rostrally. Later studies would corroborate that finding

using the same technique: multi-unit mapping across the surface of the STG in the anesthetized macaque, determining best frequency audio-visually (Morel et al. 1993, Kosaki et al. 1997). Figure 3.1 compiles all published data on the tonotopic reversal between AI and R, aligned on a common rostral-caudal axis about the point of minimum frequency. In anesthetized macaque (panels A, B, and D), frequencies drop reliably as the electrode is moved rostrally in AI, spanning the full audible range of the monkey (Jackson et al. 1999). Beyond the point of reversal, there appears to be only a partial representation favoring low frequencies (panel A), or a general breakdown of tonotopic order (panels B and D). In the awake macaque, only Recanzone et al. (2000a) have produced maps of best frequency from isolated cells, revealing a less precise tonotopic order within AI, but confirming a reversal into R (Figure 3.1, panel C). The 17 rostral locations in this study (in the left hemisphere of one animal) comprise the only published single-unit data in R.

As Figure 3.1 illustrates, a central tenet of the current model – the tonotopic reversals delimiting the divisions of the core – is only partially supported. Rauschecker et al. (1997) confirm the tonotopic progression of AI and, in accordance with Merzenich and Brugge (1973), they suggest that low frequencies may be prevalent in R. Whereas the model of Hackett et al. (1998) depicts R as a core field spanning a low-frequency border with AI and a high-frequency border

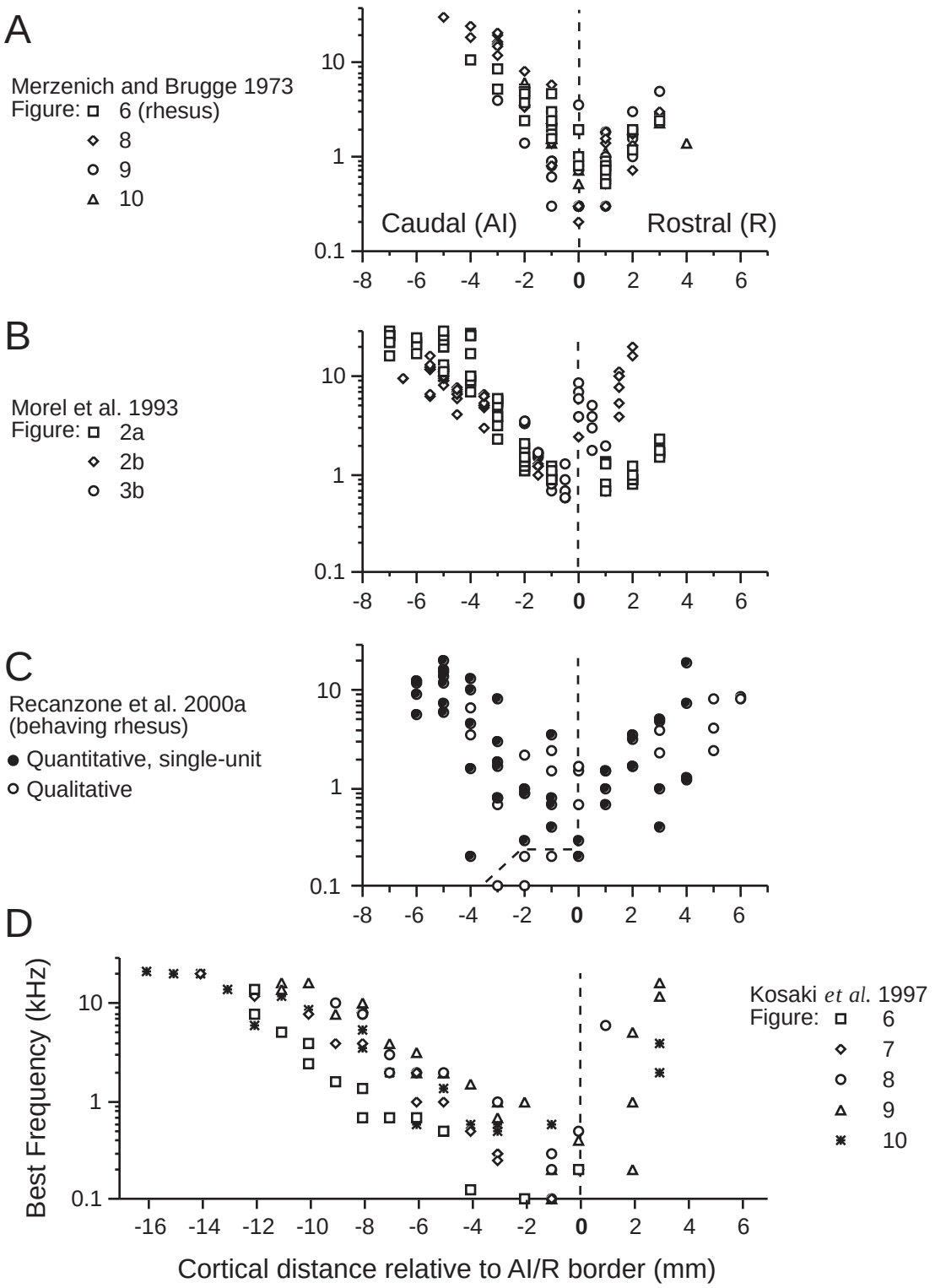

A
Merzenich and Brugge 1973
Figure: 6 (rhesus)
8
9
10
Caudal (AI)
Rostral (R)
B
Morel et al. 1993
Figure: 2a
2b
3b
C
Recanzone et al. 2000a
(behaving rhesus)
Quantitative, single-unit
Qualitative
D
Best Frequency (kHz)
Kosaki et al. 1997
Figure: 6
7
8
9
10
Cortical distance relative to AI/R border (mm)

Figure 3.1 Existing evidence for the reversal of tonotopy at the AI/R border. Panels **A**, **B**, and **D** are from anesthetized macaques (stumptail or Japanese); only A, open squares is from rhesus. From individual figures in each study, the rostro-caudal axis was drawn and each point assigned to a distance along that axis; all data were aligned to 0 mm at the border between AI and R if one was specified in the study; otherwise, the point of the frequency minimum was used (dashed line). Each panel represents one study, separate symbols designate the figures in that study from which the data are taken. Panel **C** is from awake behaving rhesus, and the only study using quantitative single-unit data (filled circles; open circles are multi-unit estimates of best frequency). Coordinates match those of the recording chamber, and the AI/R border was specified by the authors.

with the hypothesized field RT (Figure 1.2), there is little evidence that R, let alone RT, comprises a full representation of the monkey's hearing range. Another justification for placing AI and R at an equivalent level in the hierarchy is that each receives parallel input from the MGv. The physiological implications of this have been tested only by Rauschecker et al. (1997), who recorded tone responses from R before and after removal of AI. Vigorous tone responses in R persisted in the absence of AI, whereas responses in belt area CM were abolished, implying that AI and R receive independent thalamic input in parallel (though not ruling out that R receives input in series as well).

Proof of parallel input confirms that AI and R share an equivalent level in the anatomical hierarchy, but it does not address the issue of whether they operate in synchrony, or whether the physiological properties of neurons in AI and R are comparable. This chapter will explore the basic physiological properties of the primary and rostral fields, and evaluate the extent to which they should be considered as a unified core.

3.2 Anatomical methods and verification of recording sites

Detailed methods are described in Chapter 2. Briefly, subjects were two male rhesus macaques (designated X and Z), trained to discriminate the laterality of IPD around midline, and accustomed to sitting quietly between behavioral sessions to allow the collection of additional physiological data in the passive state. Stimuli

were generated digitally by MALab software, synthesized by dedicated hardware processors and presented in the closed field via electrostatic speakers coupled to ear inserts (phase and level were calibrated across frequency at the start of each session). An electrode was positioned vertically and advanced with a stepping microdrive, through the parietal cortex to auditory regions of the superior temporal plane (e.g. Pfingst and O'Connor 1980). Entry into auditory cortex was typically preceded by a gap in physiological activity believed to correspond to passage through the lateral sulcus. In animal Z, initial placement of the chamber was too far rostral to capture most of AI; after mapping, the chamber was moved caudally to allow overlapping coverage of the full extent of auditory cortex. After mapping the extent of auditory cortex available within the recording chamber over the left hemisphere in each animal, the chamber was transplanted to the right, allowing a total of four hemispheres to be sampled. Monkey Z has yet to be sacrificed, but anatomical verification of the recording sites has been carried out in monkey X, through histology and magnetic resonance imaging (MRI).

After completion of the physiological survey in the right hemisphere of monkey X, electrolytic lesions were placed at four points in the chamber using an electrode coated in fluorescent dye (di-I, DiCarlo et al. 1996, Snodderly and Gur 1995). Later that day, the animal was deeply anaesthetized by an i.v. injection of sodium pentobarbital (Nembutal), then transcardially perfused, first with heparinized

lactated ringer, followed by four liters of chilled, freshly prepared 4% paraformaldehyde with 0.5 or 1% glutaraldehyde. The brain was post-fixed in 4% paraformaldehyde/ 0.25% glutaraldehyde without being removed from the skull. This allowed the acquisition of MRI images post-mortem, which could not have been obtained with the stainless steel recording chamber in place.

The following day, the tissue was blocked, vibratome-sectioned at a thickness of 50 μm and reacted for 30 minutes in 1% sodium borohydride in 0.1 M phosphate buffer (PB), to stop glutaraldehyde fixation. After 0.1 M PB rinses, alternate sections were stained for Nissl substance, to localize cell bodies, or set aside to be immunoreacted for cytochrome oxidase (CO), a metabolic enzyme previously shown to mark core auditory cortex (Morel et al. 1993, Jones et al. 1995). Cytochrome oxidase processing was commenced no more than 72 hours after perfusion. Sections were pre-incubated at 4°C in a solution of 4% sucrose in 0.1 M PB, on a shaker. They were then transferred into 0.1M PB with 4.5% (w/v) sucrose, .02% catalase (Sigma), .03% cytochrome C (Sigma) and .05% 3´3´-diaminobenzidine (Aldrich) and placed in a shaker oven. Maintained at between 37° and 39°C, the tissue was incubated for up to 48 hours, with refreshed solution after the first 24 hours. After rinses in 0.1 M PB, sections were mounted and coverslipped.

3.2.1 Histology results

Lateral views of both hemispheres (Figure 3.2A) show the scarring evident from chronic electrode penetrations through the somatosensory cortex of the parietal lobe en route to the superior temporal plane (STP), buried within the lateral sulcus (LS). These markings (circled) persisted on the left >2 years after the chamber had been removed, and the more recent penetrations on the right left visible di-I residue on the cortical surface.

Digital images of histological sections were acquired and, when necessary, collaged using Adobe Photoshop. Adjacent sections stained for Nissl and CO are shown on the left and right, respectively, of Figures 3.2B, C, and D. Icons between panels B and C are lateral and coronal views of the brain indicating the plane of section for each panel (vertical lines labeled B, C and D), and the region of magnification (gray boxes). Locations on the interaural axis were estimated by aligning these sections with the rhesus brain atlas of Paxinos et al. (1999) at the widest point of the STP (+4.8 mm interaural). This corresponds to caudal, high frequency AI, as depicted in panel B (from the left hemisphere). Some sections from the right hemisphere, such as those in 3.2C and D, showed clear tracks of electrode penetrations (marked by arrows), indicating the angle at which the electrode passed through cortex; depending on the local convolutions of the STP, this could be nearly orthogonal to the cortical layers (as in 3.2C), or oblique (as in 3.2D). Panel

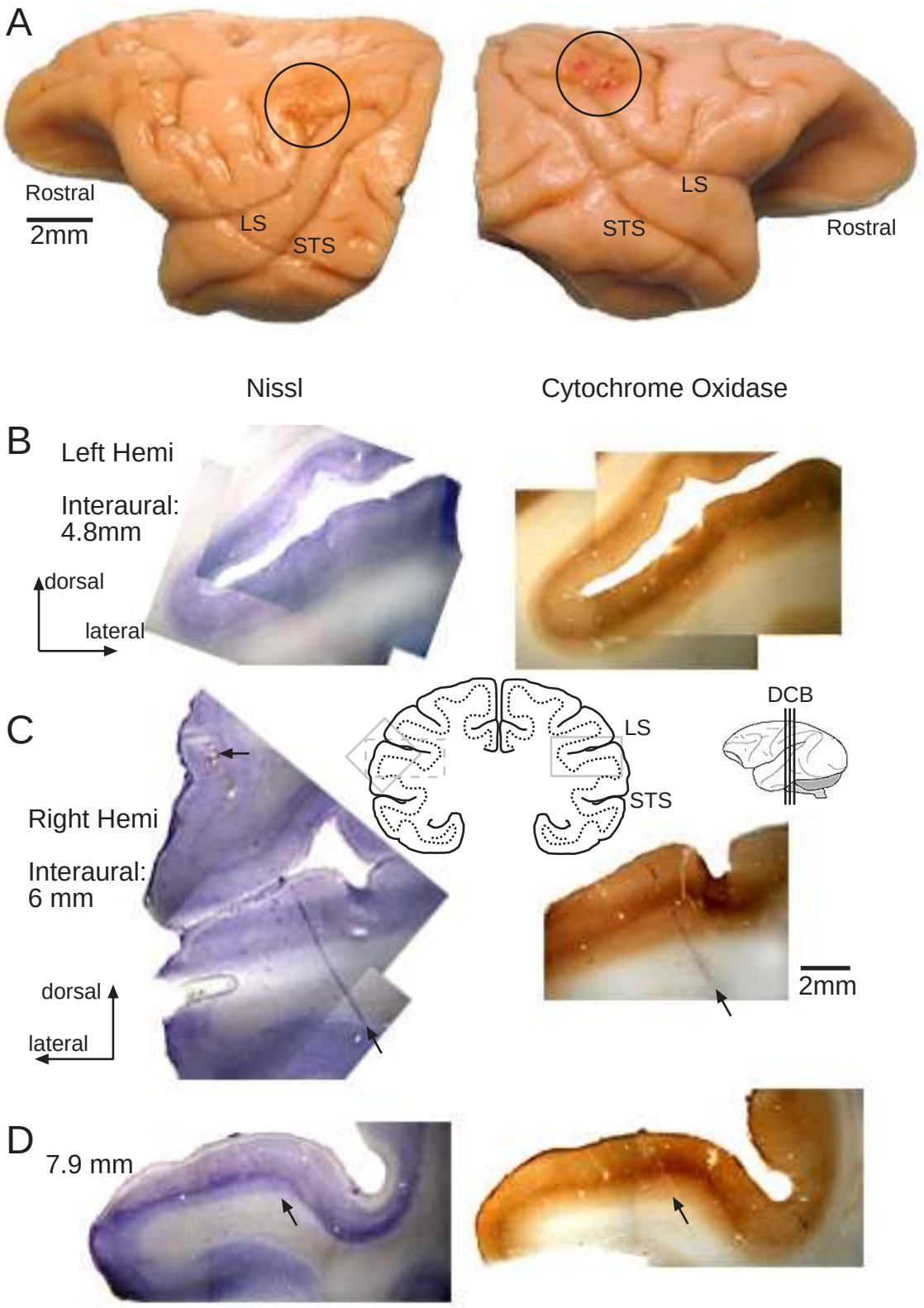

A
Rostral
2mm
LS
STS
LS
STS
Rostral
Nissl
Cytochrome Oxidase
B
Left Hemi
Interaural:
4.8mm
dorsal
lateral
C
Right Hemi
Interaural:
6 mm
dorsal
lateral
LS
STS
DCB
2mm
D
7.9 mm

Figure 3.2 A: Lateral views of the left and right hemispheres of monkey X; circles highlight the scarring from chronic electrode penetrations over the auditory cortex. LS: lateral sulcus, STS: superior temporal sulcus. **B:** Adjacent sections from the left hemisphere stained for Nissl substance (left) or immunoassayed for cytochrome oxidase (right). Gray box on the right side of the coronal icon indicates the area of magnification. Vertical lines on the lateral icon represent the approximate plane of section for panels B, C, and D. These sections represent the widest point of the STP, in caudal AI, and were used to align histological data to interaural coordinates in the rhesus brain atlas of Paxinos et al. (1999). **C:** Adjacent sections from the right hemisphere, conventions as in panel B. Solid gray box on left of coronal icon indicates area of magnification. Arrows mark an electrode track passing through parietal cortex and into medial AI, near the border with R. The track extends well into the white matter. Dark CO staining marks the middle layers of cortex and suggests that this electrode passed nearly orthogonal to the lamina. **D:** Conventions as in C, dashed gray box on coronal icon indicates area of magnification. Two mm rostral to the sections in C, here the curvature of the cortical surface is such that the electrode passes at a more oblique angle to the lamina. A scar at the right of the CO panel suggests that medial penetrations entered insular cortex.

C shows the track left by a di-I coated electrode lowered into the center of the recording chamber, which corresponds to the medial edge of the border between AI and R (the establishment of this border is described below). The track visible in panel D, approximately 2 mm farther rostral, would have entered R. Although it would be impossible to reconstruct the laminar position of each cell recorded, those tracks that were recovered indicate that the auditory cortex was sampled as far medially as the insula in this hemisphere, but not to the lateral belt on the STG, and that recording sites were distributed throughout the cortical layers. Too few tracks were recovered to provide a reliable estimate of cortical depth for each neuron, but the population of neurons presented below samples a relatively even cross-section of the cortical lamina. In one hemisphere, 80 neurons were assigned to the general category of upper, middle, or deep cortical layers on the basis of recording depth and electrophysiological characteristics of the recording (e.g. background activity). By this qualitative assignment, cells were distributed: 43% upper, 21 % middle, and 36% deep. Because of the tight packing density of the granular layers, reliable isolation may have been more difficult to achieve.

Both histological series confirm what was proposed in earlier anatomy studies: the auditory core is a broad, flat plane at the caudal end, tapering into a narrow, folded rostral extension. A series of CO sections from the right hemisphere is shown in Figure 3.3A, arranged from caudal (top left) to rostral (bottom right). The middle

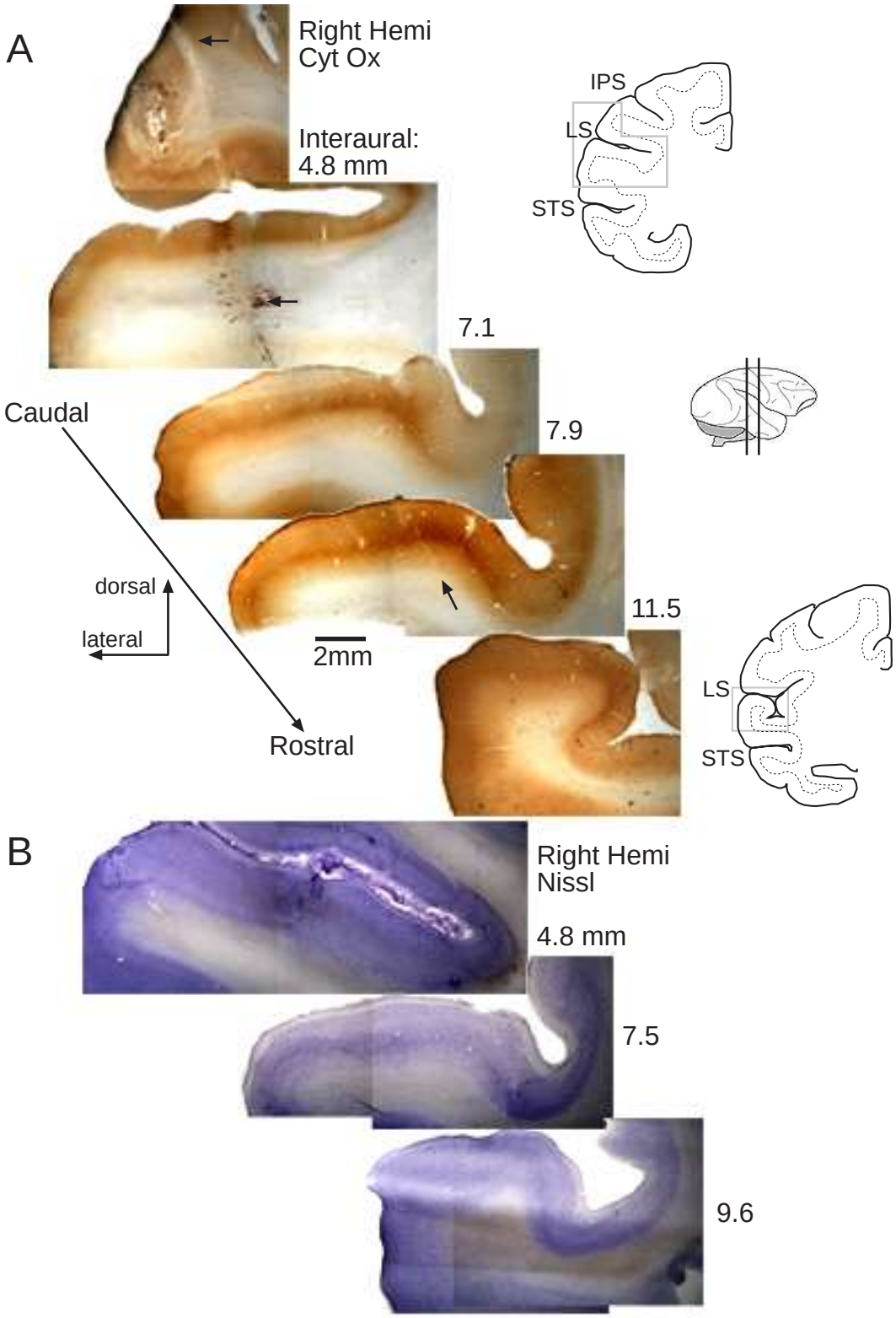

A
Right Hemi
Cyt Ox
Interaural:
4.8 mm
IPS
LS
STS
7.1
7.9
Caudal
dorsal
lateral
2mm
11.5
Rostral
LS
STS
B
Right Hemi
Nissl
4.8 mm
7.5
9.6

Figure 3.3 Histological series from the right hemisphere of monkey X. **A:** CO staining marks the middle layers of the core from caudal AI to R. A prominent electrode track in the most caudal panel marks the center of high-frequency AI (at 0 mm medial/lateral, -5 mm caudal/rostral in the map of Figure 3.10). Staining marks the medial edge of the gyrus in the most rostral panel. Interaural coordinates and icons as in Figure 3.2. **B:** Nissl stain reveals the middle layers of the core to be dense and granular.

cortical layers displayed dense (though sometimes patchy) staining for CO throughout the core, consistent with the anatomy of Jones et al. (1995). The most caudal section shows a clear track through the overlying parietal cortex, passing through the STP and extending well into the white matter. An electrolytic lesion (arrow) marks the deepest point at which an auditory response could be heard in the electophysiological signal, suggesting that some "hash" response was audible from the fibers. (Prior testing at this site showed the best frequency to be around 12 kHz, and this was identified as mid- to high-frequency AI.) Although the staining is uneven in many Nissl sections, granular koniocortex appears to mark the middle layers of the core fields (Figure 3.3B). The trend seen in these sections is confirmed by the CO series from the left hemisphere in Figure 3.4: moving rostrally, the STP narrows as the insula deepens, and the darkest CO staining moves from a horizontal strip of reactivity across the STP to a vertical patch along the medial edge of the gyrus. This is best seen in the penultimate panel, 7.95 mm rostral to the interaural axis, where the extent of the darkest CO immunoreactivity is demarcated by arrows.

Cytochrome oxidase is a marker of metabolic activity, and would be expected to be coextensive with the most active auditory regions. This may explain why, in a two dimensional map of vertical electrode penetrations, tone responsive tissue narrows to a strip only a few mm wide at the rostral end (as seen in the left hemisphere of

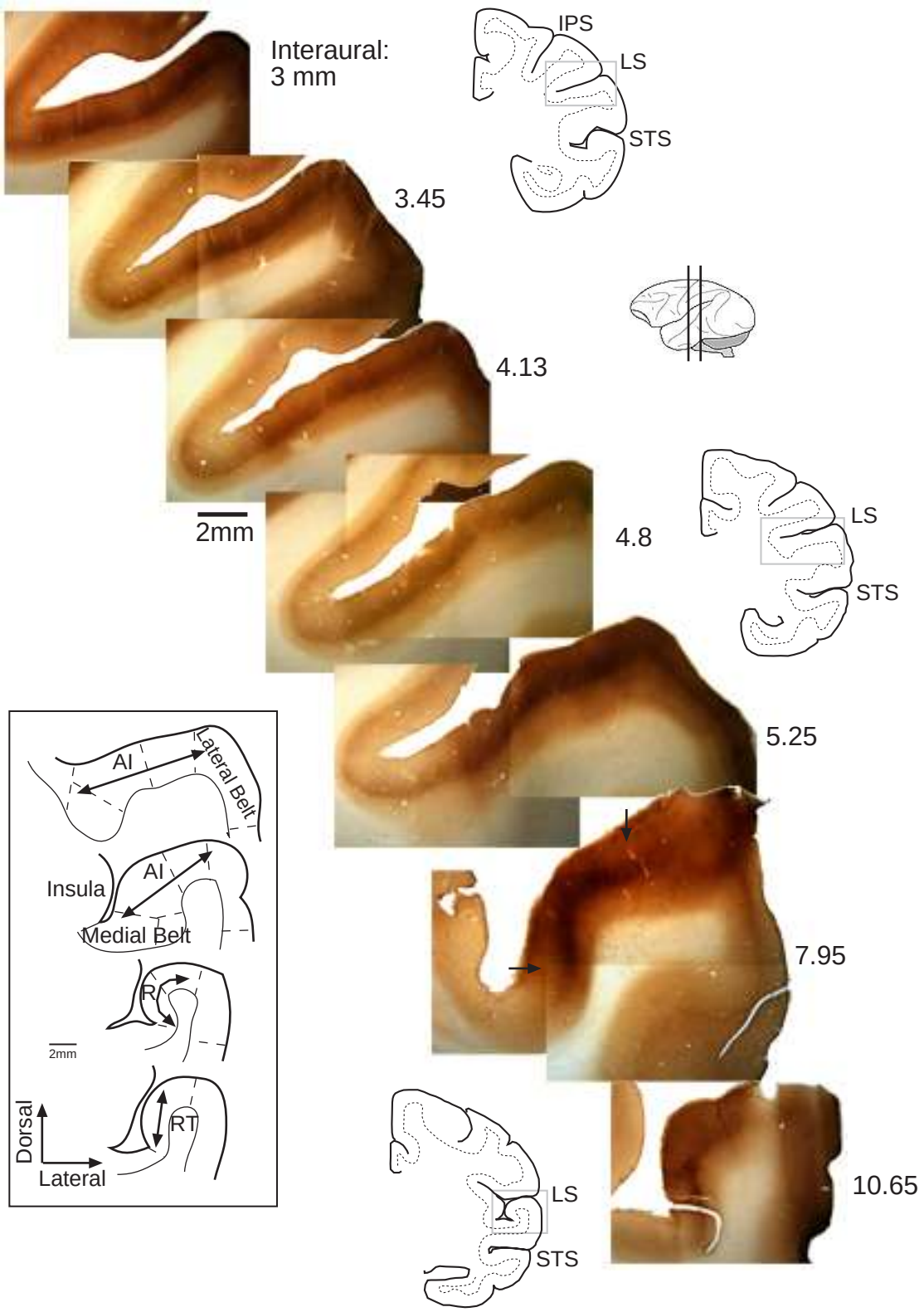

Interaural:
3 mm
IPS
LS
STS
3.45
4.13
2mm
4.8
LS
STS
5.25
7.95
10.65
LS
STS
AI
Lateral Belt
Insula
AI
Medial Belt
R
2mm
RT
Dorsal
Lateral

Figure 3.4 Staining for CO in the left hemisphere of monkey X, covering the full caudal-to-rostral extent of auditory cortex. Interaural coordinates and icons as in Figure 3.2. Upper panels probably show the most caudal extent of AI, where dense CO staining marks a broad extent of the superior temporal plane. Moving rostrally, the staining narrows and migrates toward the medial wall of the superior temporal gyrus (see arrows at 7.95 mm). Since CO is a marker of metabolic activity, densely stained regions would be expected to produce the most vigorous responses to sound.

monkey Z, in which the farthest reaches of the rostral core were sampled – see Figure 3.10). Morel et al. (1993) describe dense CO staining in layer IV as common to both AI and R, whereas Jones et al. (1995) later describe dense CO staining restricted to AI, and less dense in R. Our results seem to confirm consistent CO staining into R, and like all previous studies, show no clear cytoarchitectural boundary between the core fields.

3.2.3 MRI

During the last recording session in which dye-coated electrodes were used to provide histological landmarks, an attempt was made to provide a landmark for MRI images as well. A capillary tube filled with vitamin E (which is highly visible in MRI) was inserted into the brain at the rostral limit of the recording chamber, using the same microdrive with which the electrode was advanced. Although the capillary was removed, the scar and/or trace of vitamin E left behind appears very clearly in the first panel of Figure 3.5. These T2-weighted images were acquired at 0.5 mm isotropic resolution while the tissue was post-fixing. These images span 14 mm along the rostral-caudal axis in 2 mm steps, with the coordinate at the upper right of each panel corresponding to the rostral-caudal coordinates on the x-y stage of the microdrive (and to the corresponding panel of Figure 3.10, below). Moving caudally, the LS (marked by an arrow in each panel) moves dorsally as the STP broadens. The angle of the capillary mark confirms the angle of electrode approach

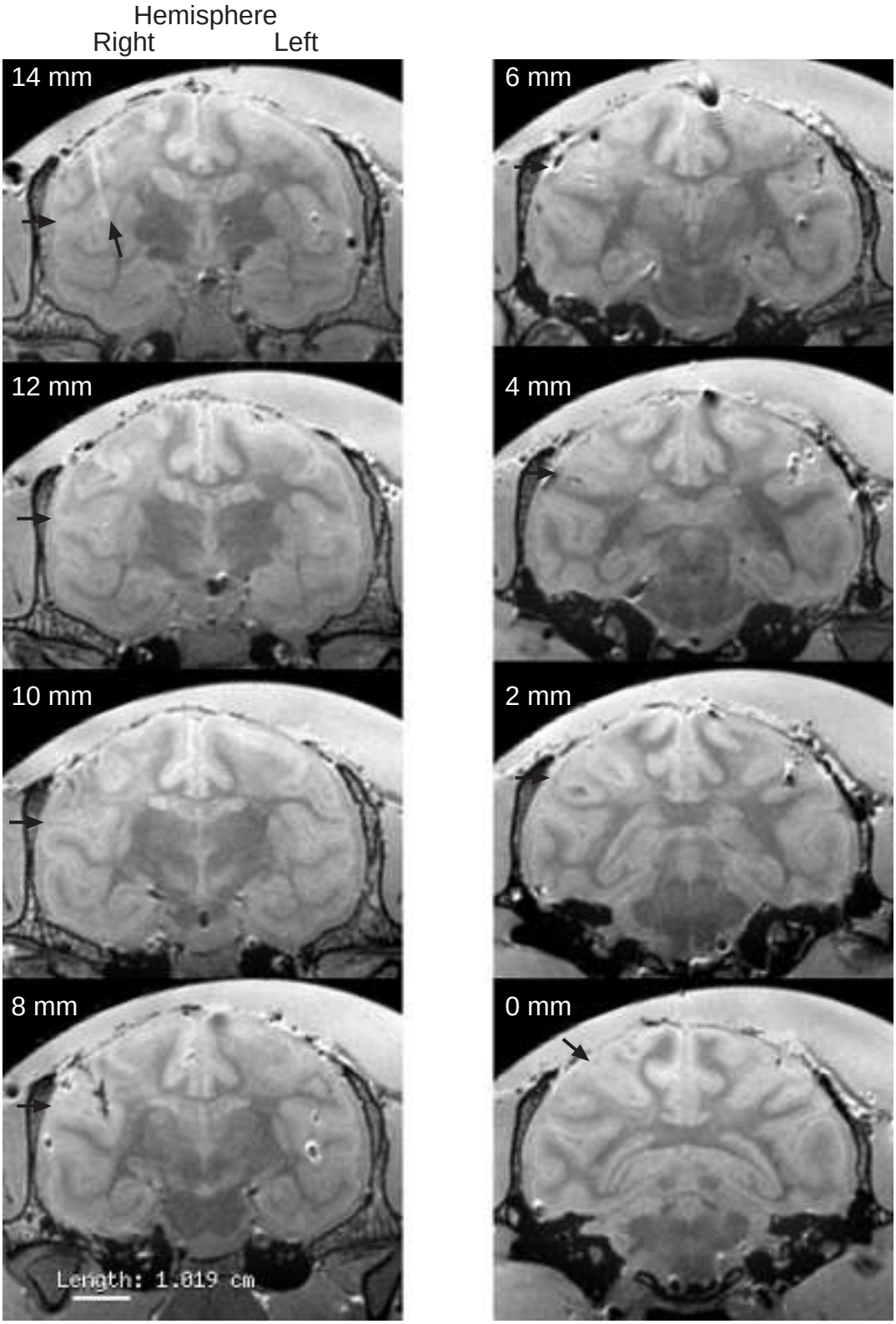

Hemisphere
Right
Left
14 mm
12 mm
10 mm
8 mm
Length: 1.819 cm
6 mm
4 mm
2 mm
0 mm

Figure 3.5 A series of MRI images from monkey X, spanning the full rostral-to-caudal extent of auditory cortex sampled within the recording chambers over the right and left hemispheres. Right and left are “reversed”, per radiological convention; images are T2-weighted and were acquired at 0.5 mm isotropic resolution during post-fixation of the tissue. The upper left panel clearly shows the mark left by an attempt to insert a capillary tube of vitamin E into the brain using the microdrive with which the electrode was advanced; thus the scar accurately reflects the angle of approach for electrode penetrations into the auditory cortex. This landmark identifies the rostral limit of the recording chamber over the right hemisphere; the remainder of the images are spaced at 2 mm intervals, spanning a total of 14 mm in the rostral-caudal dimension. Coordinates given at the upper left of each panel correspond to the rostral-caudal coordinates for the right hemisphere of animal X in Figure 10.3.

seen in histological sections and establishes the limits of the tissue accessible within the recording chamber for the right hemisphere of this animal.

In summary, histology and MRI confirm that recordings in this animal are from the anatomically defined auditory core and surrounding belt fields, extending into the insula in both hemispheres but apparently not the lateral field of the right hemisphere. The relationship between physiological and anatomical properties established in these hemispheres provides a strong basis for assigning recordings in animal Z to their respective cortical fields.

3.3 Physiological methods and measurement of basic response properties

General methods of animal training, stimulus delivery, and physiological recording are detailed in Chapter 2 and section 3.2, above. The electrode was advanced through parietal cortex, usually while the animal was performing a psychoacoustic task (see Chapter 2). Parietal activity was often related to movement of the hand during the task (Merzenich and Brugge 1973). If the animal was not performing, search stimuli were employed to confirm that the electrode was not yet in auditory cortex (tones, noise, or manually generated sounds passed through a speaker into the soundproof room). Entry into the STP was often preceded by a "gap" in neural activity (the absence of action potentials and a quieting of background noise). As the electrode was advanced, and activity began to return, a wide range of search

stimuli was employed to determine if the tissue was auditory, and in what frequency range it was tuned. An optimized search stimulus was then used to identify isolated single units.

3.3.1 Basic response properties of auditory cortex

In contrast to cortex under barbiturate anesthesia, neurons in auditory cortex of the awake monkey frequently exhibit sustained spontaneous firing rates, and complex temporal response profiles. Isolated, triggered spikes were displayed in real time as raster plots and binned into peri-stimulus time histograms (PSTHs). Temporal windows could be defined within the trial, in which spike rate was computed and plotted as a function of the stimulus parameters under study. Because spontaneous firing rates were often high, suppression as well as excitation could be discerned from extracellular recordings. The temporal profile of a neuron's PSTH could include transient excitation or suppression at the onset or offset of the stimulus, as well as a sustained component.

3.3.1.2 Frequency tuning

Before quantitative data were collected, audio-visual testing was used to determine the approximate receptive field of the cell in frequency. Tones were used if they were effective; otherwise, a band-passed noise was presented, and the center frequency of that band roved to find the appropriate range. Sound level (in dB SPL)

was also roved to ensure that a nonmonotonic or high-threshold cell was not dismissed as unresponsive. These tests were performed binaurally, but if the binaural response was poor, the contralateral and ipsilateral ears were tried individually. As will be discussed below, the vast majority of cells responded well to diotic stimulation (with either summative or facilitative interactions), and almost all frequency tuning functions (FTFs) were constructed this way. If a response appeared to be monotonic, the FTF was collected at 60 dB. In cells with strongly nonmonotonic responses, the audio-visually determined best SPL was used.

Rather than using a standardized response area, a set of frequency points was chosen individually for each unit to capture the full extent of its receptive field, often with extra resolution near the apparent BF. Frequency tuning in AI can be very sharp, and to present a standard stimulus set that would capture the BF of each neuron would take a prohibitive amount of time. Stimuli were 100 ms long, with a 5 ms cosine-squared ramp (to avoid spectral artifacts of sudden amplitude transitions), presented every 1000 ms, for ten trials per frequency point. Because stimuli were presented in blocks, the rasters collected may reflect adaptation to the repeated stimulus, which would not be apparent from pseudo-randomized stimuli.

Assigning a BF on the basis of the FTF was not always trivial, as Figure 3.6A illustrates. The PSTHs at right show the spikes accumulated in 10 ms bins over 10

trials at three representative frequencies (the time axis has been truncated at 500 ms, but trials were 1000 ms). The icon below the histograms indicates three time windows over which spike rate could be computed to build a FTF: 100 ms from the onset (black), 100 ms from the offset (gray), or 200 ms capturing both response components (dashed). The FTFs on the left show the diverse functions that can be generated depending on which component of the response is emphasized. The solid black line is computed from a 100 ms window beginning at tone onset, and represents the standard window over which spike rate was computed when assigning BF. This function shows suppression below the spontaneous firing rate at 0.5 kHz, and a BF determined by a strong onset response at 0.9 kHz. The gray line, based on the offset response, is complementary – where the onset response is weakest and suppression during the tone is strongest, the offset is strongest. If a 200 ms window is used to capture both components of the response, the transient excitation, suppression, and after-excitation average to yield no clear BF (dashed line).

3.3.1.3 Rate-level functions

Although the strong onset response in Figure 3.6 allows the identification of a clear BF, spike rate over a fixed window was not always a clear indicator of where a response was strongest. This point is elaborated in Figure 3.6B, which depicts the rate-level function (RLF) of the cell from panel A at its BF of 0.9 kHz (black line,

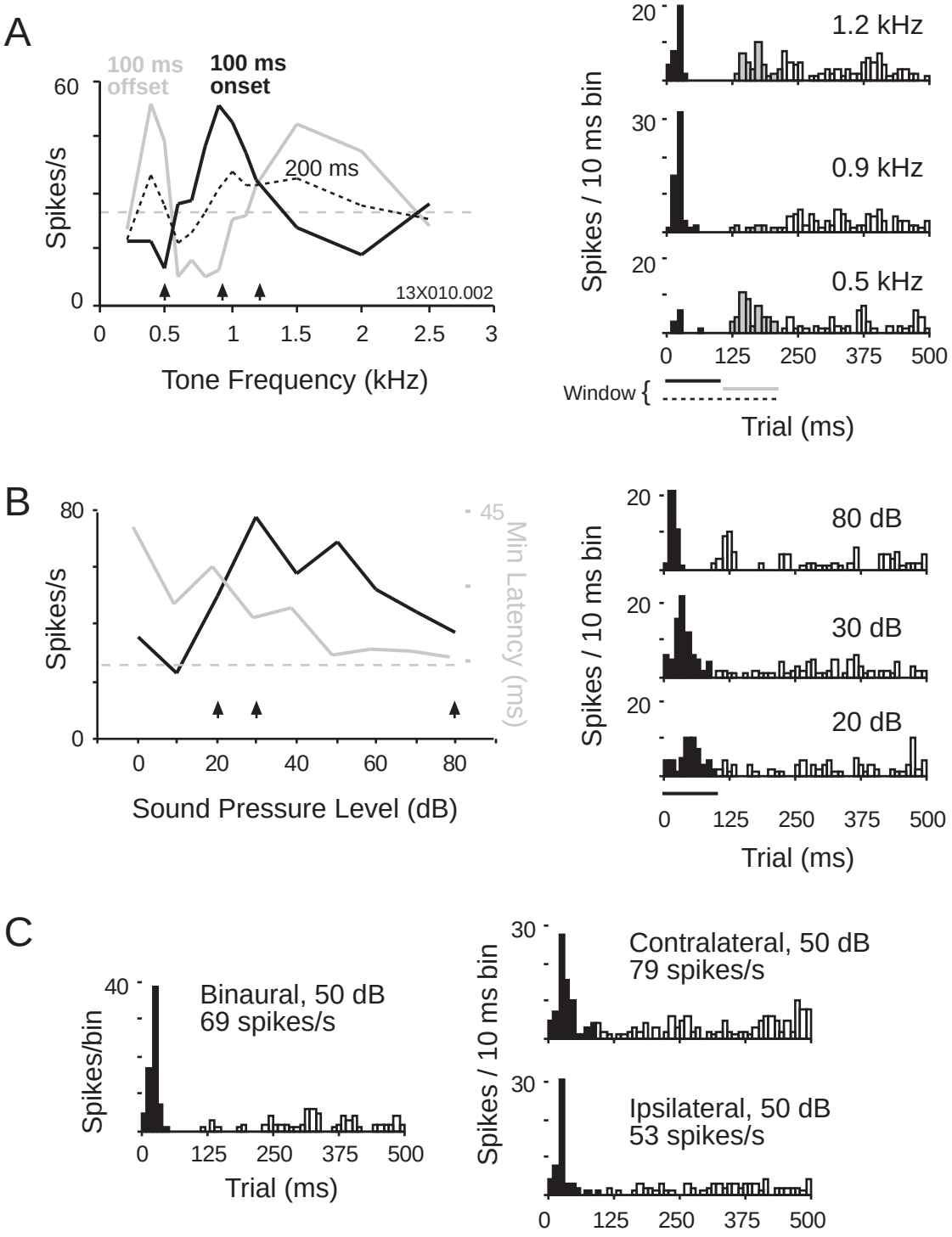

A
100 ms offset
100 ms onset
200 ms
Spikes/s
60
0
0
0.5
1
1.5
2
2.5
3
13X010.002
Tone Frequency (kHz)
1.2 kHz
0.9 kHz
0.5 kHz
Spikes / 10 ms bin
20
30
20
0
125
250
375
500
Window
Trial (ms)
B
80
0
Spikes/s
45
Min Latency (ms)
0
20
40
60
80
Sound Pressure Level (dB)
80 dB
30 dB
20 dB
Spikes / 10 ms bin
20
20
20
0
125
250
375
500
Trial (ms)
C
Binaural, 50 dB
69 spikes/s
Spikes/bin
40
0
125
250
375
500
Trial (ms)
Contralateral, 50 dB
79 spikes/s
Ipsilateral, 50 dB
53 spikes/s
Spikes / 10 ms bin
30
30
0
125
250
375
500

Figure 3.6 An example of general response properties measured from single neurons. **A:** FTFs can vary considerably depending on the temporal window over which spike rates are measured. The three FTFs at left are computed from time windows indicated below the PTSHs on the right (arrows below FTFs indicate responses shown as PSTHs for three example tone frequencies, including the BF of 0.9 kHz). The tone is 100 ms, beginning at time 0, with a 5 ms cosine-squared ramp. Bars of the PSTH are filled to match the window in the icon: black bars and the black FTF are based on the standard 100 ms window beginning at tone onset; gray bars and FTF show the 100 ms window beginning at tone offset; the dashed line is a 200 ms window including both response components. Trials were 1000 ms, PSTH time axis is truncated. Whenever possible, BF was assigned based on the 100 ms window capturing the onset and sustained components of the response (solid black line at left). In this case, the offset shows complementary tuning, peaking where the tone is suppressive. The 200 ms window yields a messy function with no clear BF, because the onset excitation, sustained suppression, and offset excitation average out to a spike rate not much different from the spontaneous firing rate of ~25 spikes/s (dashed gray line). **B:** The RLF for the same cell, taken binaurally at the BF of 0.9 kHz. Responses at three SPLs are shown in the PSTHs at right, black bar indicates 100 ms tone duration and the window over which spike rate was measured. Three arrows mark the responses shown on the right: the threshold of 20 dB, best SPL of 30 dB, and shortest-latency response at 80 dB (typically the highest SPL tested). This cell was classified as “weakly” nonmonotonic; though some cells were completely inhibited at very high SPLs, this one fired fewer overall spikes but with sharper temporal structure. Minimum response latency, overlaid in gray (right-hand axis) continues to decrease with level even as total spike rate drops off, approaching an asymptote of 16 ms at SPLs ≥ 50 dB. **C:** Binaural class was determined by comparison of binaural, contralateral, and ipsilateral presentation of BF tones at best SPL. This cell was clearly excited by either ear alone, and was designated EE (see text). The contra ear evokes a more sustained response, the ipsi ear evokes a sharply time-locked transient response; together, they produce a stronger onset with sustained suppression.

left-hand axis). Threshold for this cell was defined as 20 dB, the first point showing an elevation of activity beyond the spontaneous discharge rate (dashed line; response shown as lower PSTH at right). Threshold could not be defined by a single criterion for all cells – as panel A suggests, a response may not always manifest as more spikes in a given window, but in the temporal rearrangement of those spikes.

Every RLF was classified into one of four categories (or disqualified as "messy") and assigned a best SPL. These categories were: monotonic (continually increasing spike rate with SPL over the full tested range; highest SPL tested is best), plateau (increasing to a point, then leveling off; "knee" of the function is best), weakly nonmonotonic (highest SPL tested generates ≥ 50% of maximal response) and strongly nonmonotonic (highest SPL generates < 50% of maximal response). The function in 3.6B is weakly nonmonotonic with a peak at 30 dB by the spike rate criterion used, but this belies the temporal structure obvious in the uppermost PSTH on the right, showing a sharp onset/suppression/offset response to 80 dB. The timing and magnitude of onset responses, and the shape of the RLF, are dependent on the shape and rise time of the tone onset (Phillips et al. 1995, Heil 1997a, b). Since these parmeters were held constant (5 ms, cosine-squared ramp) when testing each neuron, measures derived from the RLF are valid for comparing

responses across fields, but should not be regarded as fixed, intrinisic characteristics of the neuron.

3.3.1.4 Minimum response latency

Response latency was measured as the average time to first spike across 10 trials, measured from tone onset. Every raster was examined and the temporal window set to exclude spontaneous discharges preceding the response. Response latency generally decreases with increasing SPL, reaching a plateau value for the cell in Figure 3.5B of 16 ms at ≥ 50 dB (solid gray line, right-hand axis). The minimum of this function was reported as the latency for each unit (in the absence of a RLF, or in cells only responsive to noise, the minimum reliable average first spike latency in those functions was used).

3.3.1.5 Binaural class

Many schemes of varying complexity have been proposed to describe how inputs from the two ears interact at the level of the auditory cortex (Imig and Adrian 1977, Middlebrooks et al. 1980, Semple and Kitzes 1993ab, Zhang et al. 2003). Often these schemes reflect an effort to infer binaural interaction when firing rates and spontaneous activity are depressed by anesthesia. The system used here is closest to that applied in the brainstem by Goldberg and Brown (1969), who described binaural interactions in terms of monaural responses. Figure 3.6C shows PSTHs in response to 50 dB BF tones presented binaurally (left), or to the contralateral or

ipsilateral ear alone (right, upper and lower respectively). A two letter classification was assigned where the first letter indicated the response to the contra ear, and the second letter the ipsi ear; E, I, or O was used to denote excitation, inhibition, or no effect. As for FTFs and RLFs, spike rate was measured over a 100 ms window (the same duration as the tone) beginning at tone onset. This cell, driven by both ears, was EE. For population analysis, each cell was classified as binaural (both ears drive, or both inhibit, such as EE or much less common II cells), monaural (EO, OE, IO, or OI), or suppressive (EI or IE, the two ears in opposition). This scheme is simplistic, and ignores such designations as summation or occlusion, but if the binaural response was obviously weaker than a monaural response, the binaural interaction was explored with further manipulations of ILD. An original goal of these experiments, in fact, was to measure neural and behavioral sensitivity to ILD, as we did for IPD (Chapter 2, Chapter 5, Malone et al. 2002), but clear tuning to ILD at the cortical level was uncommon.

The general characterizations described above allowed further testing to be optimized for the frequency, level, and binaural tuning of the neuron under study. Further manipulations included sinusoidal modulations of tone amplitude and frequency (discussed in Chapter 4), and modulation of interaural phase (discussed in Chapter 5).

3.4 Tonotopic organization of the core and delineation of cortical fields

Examination of Figure 3.1 reveals that evidence for tonotopic order rostral to the AI/R border is equivocal. Merzenich and Brugge (1973; my panel A) found a clear progression, and Recanzone (2000a, my panel C) found a reversal with greater overall variability in the awake animal, but the rostral data in panels B and D would not seem like a reversed tonotopy at all, were they not juxtaposed to the steady downward progression in AI. Figure 3.7 presents my data in the same format for two hemispheres, both of which show the relatively clear descending sequence in AI giving way to a less structured representation. The lower panel, from monkey X, shows better evidence of a reversal, whereas BFs in monkey Z seem to bottom out, with no clear trend toward higher BFs over a full 10 mm of tissue rostral to AI. Rostral sampling was not dense enough to conclude whether the most rostral low-frequency points (those at 8-10 mm) might correspond to the hypothesized field RT (Morel et al. 1993, Hackett et al. 1998).

3.4.1 AI vs. R – tonotopic reversal at the low-frequency border

As a caveat for both figures, considerable noise is introduced by collapsing these points onto a single rostral-caudal axis. Averaging the BFs of multiple neurons in a track to create a 2-dimensional map adds little noise when electrode penetrations are normal to the surface, because BF seems to remain relatively constant within

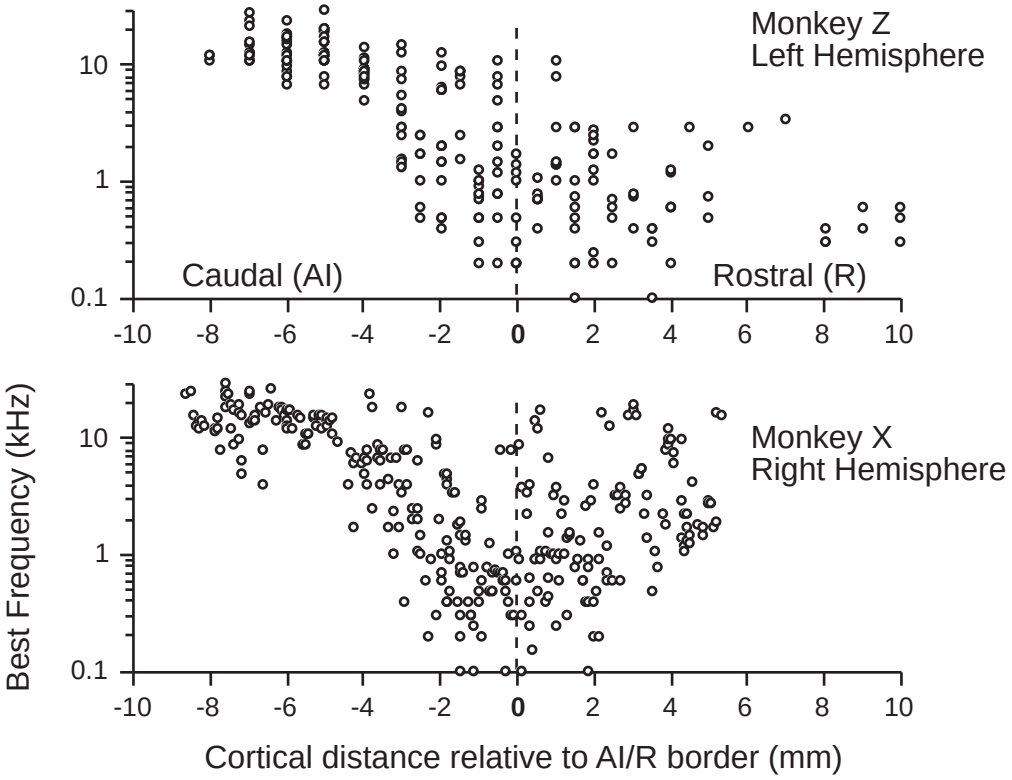


Figure 3.7 The progression of BFs from high (caudal) to low (rostral) in AI, and the degradation of that order in R, are illustrated for two hemispheres in this study. Conventions as in Figure 3.1.

cortical columns (Merzenich and Brugge 1973). However, given the physical constraints of the awake preparation, and the topology of the STP, the angle of the electrode could not be adjusted to keep penetrations normal to the cortical surface (see Figures 3.2, 3.5). For this reason, in one hemisphere, the angles of the electrode and the x-y coordinate system of the microdrive were measured relative to the sagittal, horizontal, and coronal plane, and used to correct the 2-D frequency map for the cortical depth of each recorded cell. Figure 3.8A shows this corrected map for the low-frequency border of AI and R. The angle of approach was such that advancing the electrode in depth moved the tip slightly rostral and medial – up and to the left in Figure 3.8A. Although the global tonotopic order seems messy, BFs within penetrations were generally consistent, or followed an ordered sequence in depth. Upward and downward arrows next to selected penetrations (circled for clarity) indicate the direction of BF change with increasing depth: these sequences were consistently downward in AI, and upward in R, the range of BFs encountered presumably a function of the local curvature of the cortical surface and the angle at which the electrode traversed the cortical lamina. The bold gray line separates the two fields, on the basis of global tonotopic order as well as local BF progressions. Figure 3.8B is an interpolated map of the full data set from which panel A was drawn (lines indicate the area around the AI/R border enlarged in panel A). Figure 3.8C plots the BF without depth correction (averaging the BFs within each penetration), which has the overall effect of "smoothing" that data but reveals the

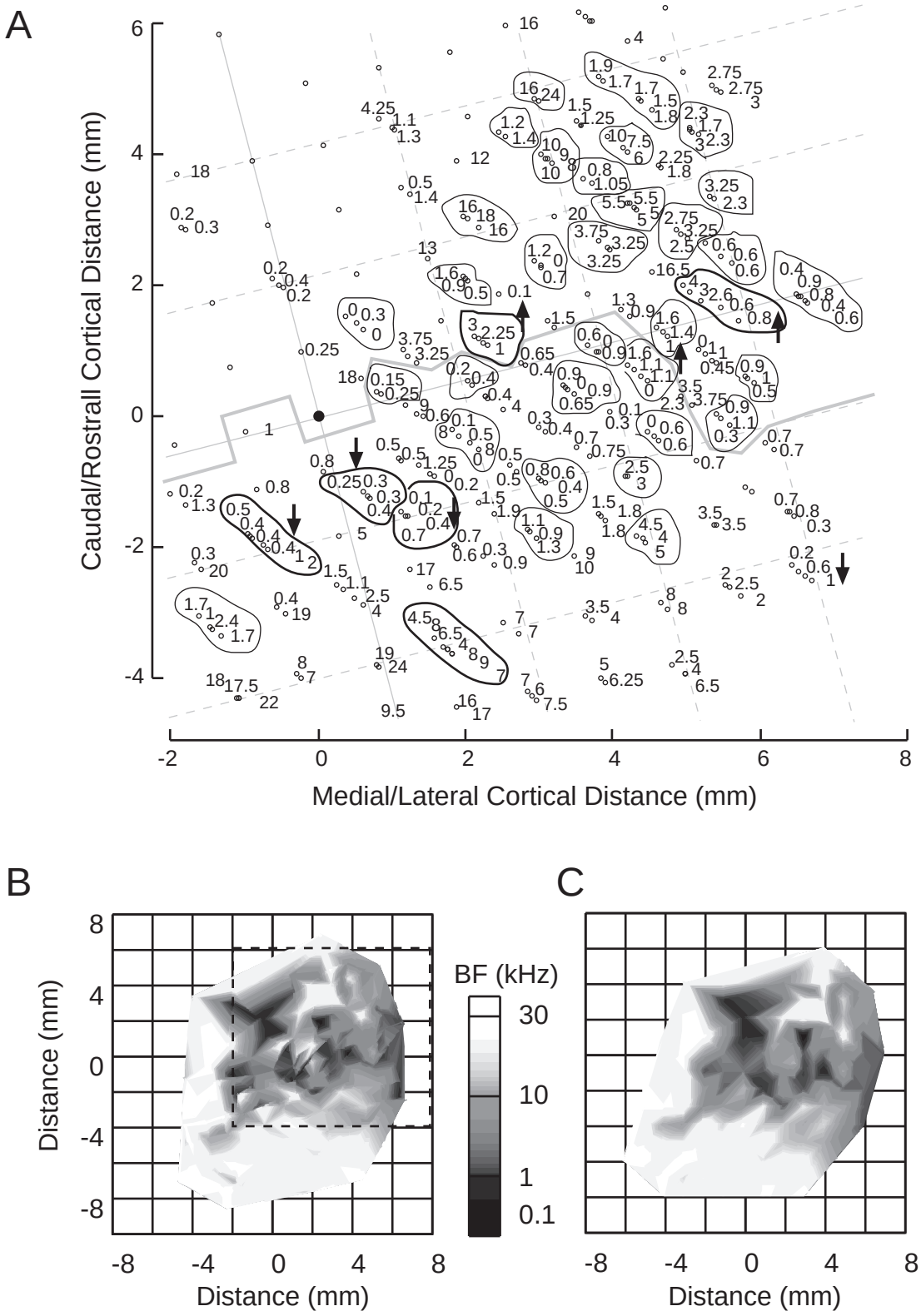

A
Caudal/Rostrall Cortical Distance (mm)
Medial/Lateral Cortical Distance (mm)
B
C
Distance (mm)
Distance (mm)
BF (kHz)
30
10
1
0.1

Figure 3.8 BF maps for monkey X, right hemisphere, illustrating correction for electrode angle and cortical depth. **A:** The axes of the graph are true rostral/caudal and medial/lateral coordinates, the slightly rotated coordinates of the microdrive stage are overlaid as fine gray lines. Each small circle denotes a recording site, the numbers indicate BF (in kHz), and circles have been drawn to clarify certain penetrations showing a reliable sequence of BFs in depth. Within a penetration, deeper cells are slightly more rostral and medial (up and to the left) of shallower cells. Arrows next to circled penetrations denote upward or downward BF sequences with increasing depth; downward sequences were characteristic of AI, reversing to upward sequences in R. They bold gray line separates the two fields.

same pattern: high frequencies caudally (high frequency AI) and medially (probably the medial belt), and low frequencies rostro-laterally. The dark band of lowest frequencies marks the border between AI and R, and some return of higher frequencies at the rostral margin is evident. For those hemispheres in which electrode angle correction is not possible, the cluster of lowest BFs was sufficient to delineate the border between AI and R.

3.4.2 Core vs. Belt – tuning to pure-tone frequency

Drawing the border between core and belt requires a different approach, using not just best frequency but the quality of frequency tuning (Rauschecker et al. 1997, Recanzone et al. 2000a). Figure 3.9 presents several arrays of FTFs from neighboring recording sites at the border of AI. The mosaic represents a grid of sampled points from the left hemisphere of animal X, at a resolution of 1 mm; open squares have been designated AI, black squares at the top are R; going clockwise are the medial belt (M, dark gray), caudomedial field (CM, black), and lateral belt (L, dark gray). The uppermost box of example FTFs compares rostral AI to R; both show sharp tuning to tones, but the sampled points in R are tuned to moderate frequencies of 1-8 kHz, whereas rostral AI is tuned ≤ 1 kHz (this result is almost inevitable, since the reversal or end of the tonotopic gradient in AI is used to define the border). In other words, although a global tonotopic order is less clear in R

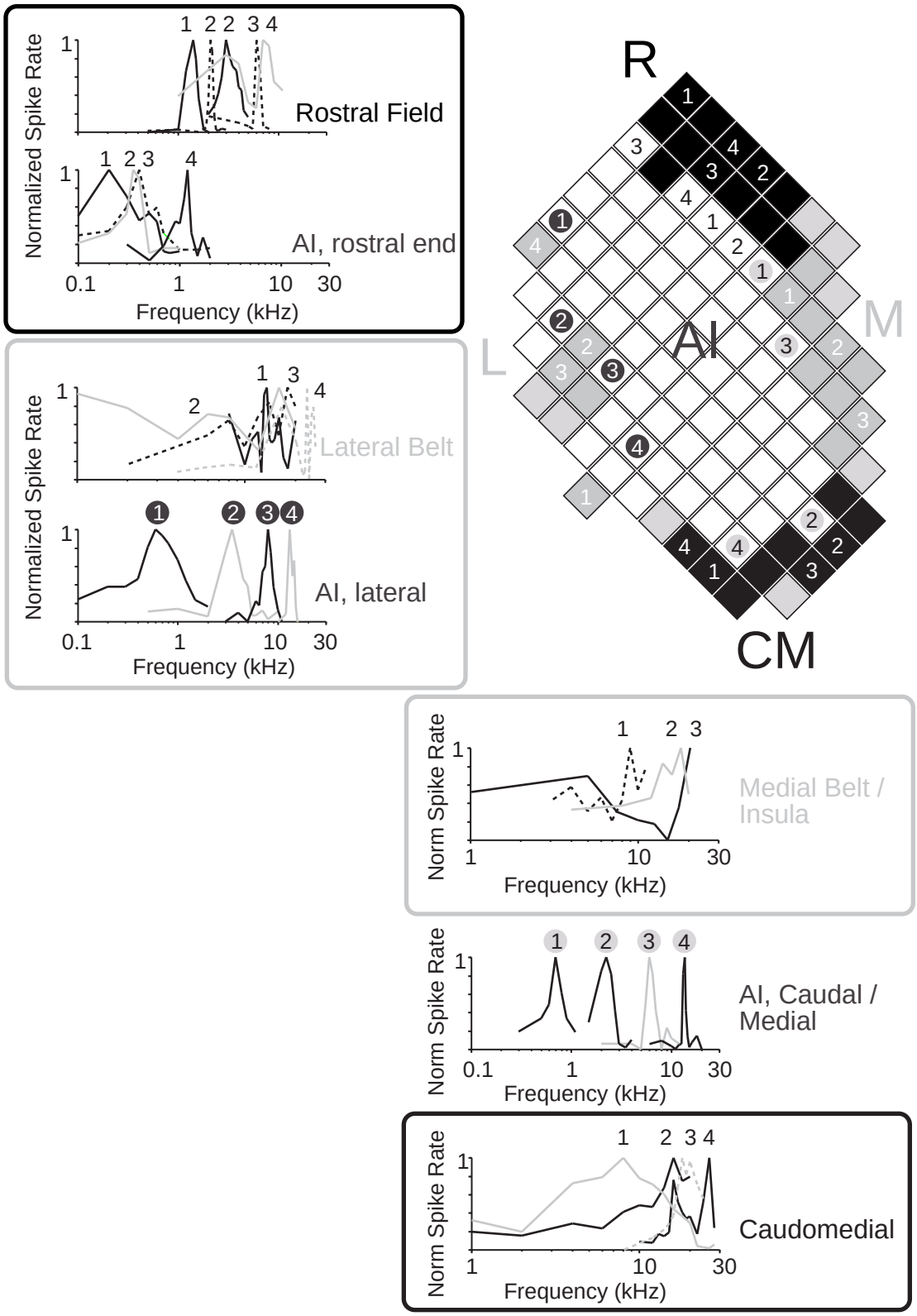

Normalized Spike Rate
Rostral Field
AI, rostral end
Frequency (kHz)
Lateral Belt
AI, lateral
R
L
M
CM
AI
Norm Spike Rate
Medial Belt / Insula
AI, Caudal / Medial
Caudomedial

Figure 3.9 Comparison of tone FTFs at neighboring sites in AI, R, and the belt. The mosaic of squares at the upper right is a partial map of sampled locations in the left hemisphere of monkey X. Open squares are in AI, black and dark grey squares have been assigned to the indicated fields (M: medial belt, CM: caudomedial belt; L: lateral belt). Light gray squares mark unresponsive sites. Numbers identify recording sites from which representative FTFs have been drawn to compare tuning in AI with that of adjacent fields; each FTF is labeled with the number of the site where it was recorded. Tuning in the rostral field (black box at upper left, top panel) is sharp, similar to that in rostral AI but in a higher frequency range (since the reversal or end of the tonotopic sequence in AI was used to draw the border, this is almost certain to be true). The lateral and medial edges of AI are sharply tuned to a range of BFs, but FTFs in the lateral and medial belt are generally messy, with a preference for high frequencies (gray boxes). Caudomedial neurons could show strong responses to tones, but were often broadly tuned and, like other belt areas, biased toward high frequencies (black box at lower right).

than in AI, sharp tuning to tone frequency persists in R, validating its inclusion as part of the auditory core.

Comparison of tone FTFs in core and belt, however, reveals a stark difference in tuning quality. Lateral AI shows sharp tuning at a range of BFs, whereas neighboring sites in the lateral belt show broad, messy tuning clustering at high frequencies. The FTFs at the lower right, from M and CM, show a similar pattern. These differences were the basis for distinguishing core from belt; the separation of belt recordings into M, CM, and L was less rigorous. No clear tonotopy could be defined here, so these borders were drawn such that CM abutted high-frequency AI, with M and L medial and lateral (respectively) to mid-to-low-frequency AI and R. Because these designations are somewhat arbitrary, the belt subregions will not be specifically compared in this chapter.

3.4.3 Defining the fields

Figure 3.10 presents frequency maps for all four hemispheres, produced by averaging the BFs encountered within each penetration and assigning every site to an octave bin (coded by grayscale). Borders of AI and R, determined as described above, are overlaid on each panel. The global progression from high frequencies caudomedially to low frequencies rotrolaterally is well illustrated by the mirror-symmetric maps from monkey X (top panels). The lower left map, from monkey Z, merges maps obtained from two placements of the recording chamber (which was

Average Best Frequency by Recording Site

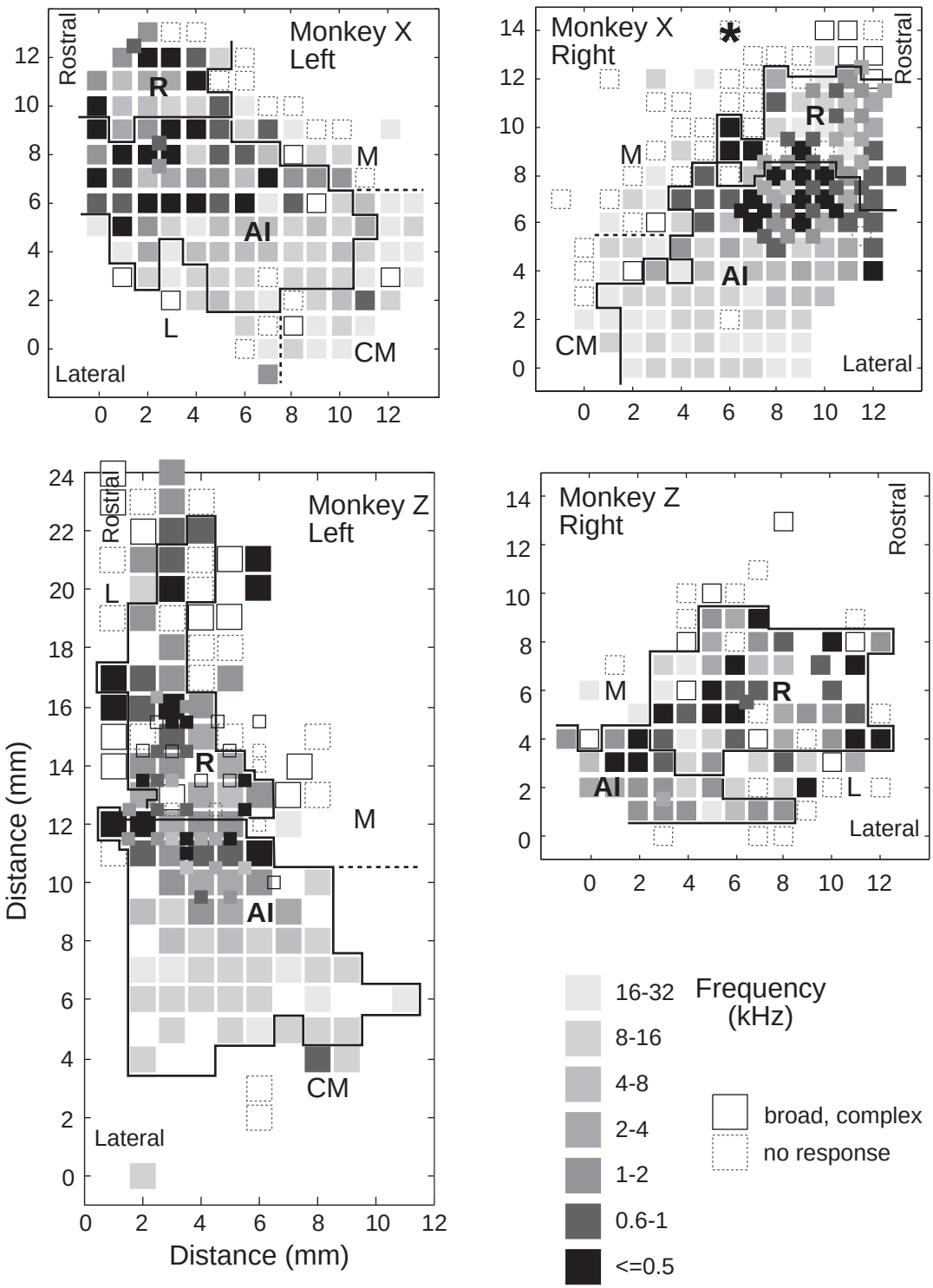


Figure 3.10 Maps of BF from the left and right hemispheres of money X (upper panels) and Z (lower panels). Rostral is up in all panels; medial is toward the midline of the page. Axes are in mm, actual values are arbitrary. The BFs encountered within each penetration are averaged to assign each site to an octave bin, color coded from black (low frequency) to light gray (higher frequency). Open squares denote sites that were auditory but showed no clear BF; dashed squares denote sites from which no auditory response could be evoked. Sampling was generally in a 1 mm grid, but in the right hemisphere of animal X and the left hemisphere of animal Z (upper right and lower left panels), the low-frequency area was sampled at a finer 0.5 mm grain – for clarity, smaller squares mark these sites. The criteria established in Figures 3.8 and 3.9 were used to overlay borders and assign each site to AI, R, or the belt (see text). Dashed lines delineate subregions of the belt, but were not drawn according to any rigid criterion. (Upper right panel replots the data in Figure 3.8 C, without interpolation between points). The asterisk (*) at the top of the upper right panel marks the site of the electrode track evident in the MRI series (Figure 3.5, upper left panel).

only 16 mm in diameter), and provides the best evidence that the rostral field tapers to an elongated strip, as suggested by histochemistry and incorporated into the Hackett et al. (1997) model (see Figure 1.2). As mentioned in regard to Figure 3.4, the folding of the gyrus and the twisting of the histologically defined core from horizontal to vertical may explain the narrow strip of responsiveness in this two-dimensional representation. In this most extensive survey of R, including at least 10 mm of tissue rostral to AI, scant evidence was found for a second reversal of tonotopy marking the proposed field RT.

The total number of single units recorded in these experiments is tabulated in Table 3.1 (next page), subdivided by animal and hemisphere. Not every recorded cell was fully characterized, but at the very least each was quantitatively assessed for frequency tuning; responsiveness and recording stability determined how extensively a given unit was studied. In addition to the 980 single neurons in the core, 34 instances of multi-unit activity were recorded. These recordings were included when constructing BF maps if single units were not available, but are not reported in any population data.

3.5 Comparing basic response properties in AI and R

Before pooling data across hemispheres or animals, a four-way comparison was made to check that no data set was significantly different from the other three. To

Monkey	Hemisphere						
X	Left		Right		Animal Totals		
AI	194	core total	217	core total	AI	411	core total
R	42	236	105	322	R	147	558
M	22	belt total	40	belt total	M	62	belt total
L	16	63	0	49	L	16	112
CM	25		9		CM	34	
Z							
AI	170	core total	63	core total	AI	233	core total
R	74	244	115	178	R	189	422
M	15	belt total	5	belt total	M	20	belt total
L	27	46	13	18	L	40	64
CM	4		0		CM	4	
Hemisphere Totals					**Grand Totals**		
AI	364	core total	280	core total	**AI**	**644**	core total
R	116	480	220	500	**R**	**336**	**980**
M	37	belt total	45	belt total	M	82	belt total
L	43	109	13	67	L	56	176
CM	29		9		CM	38	

Table 3.1: Total number of single units recorded

control for multiple tests, the Tukey-Kramer "honestly significant difference" test was used (JMP statistical sofware, SAS). The only manner in which they differed significantly was the frequency range encountered, but this can be attributed to sampling differences stemming from the placement of the recording chamber in each animal. Many comparisons were made between data sets of different size (e.g., there are more cells from AI than from R), so all statistical tests were re-run

with the larger data set randomly sampled to match the size of the smaller; effects of this correction were slight, suggesting that the tests employed adequately compensated for sample size.

3.5.1 Discharge rates

In general, core auditory cortex in the awake macaque is characterized by an audible population response throughout the depth of the cortex, with a particularly strong and homogeneous "hash" response in the middle layers, where the cytoarchitecture is granular and packing density is high (Pandya and Yeterian 1984, Cipolloni and Pandya 1991, Jones et al. 1995, Hackett et al. 1998). Deep to this, where the koniocortex gives way to a sparser array of large pyramidal cells, background activity is lower and spikes are particularly large. From each recorded cell, spontaneous discharge rate was measured using a 500 ms window over the second half of the 1000 ms trial in the FTF; this generated one estimate at each frequency point, and these estimates were averaged. Spontaneous rates for the population are plotted in the left panels of Figure 3.11 for AI (upper) and R (lower). The median of both populations is 8 spikes/s, and the two are indistinguishable by a non-parametric Wilcoxin test ($p=.67$; the Wilcoxin test will be used for all comparisons of continuous data sets, and does not presume that those data are normally distributed). The peak discharge rate for each cell was measured over 100 ms, from the strongest response in the FTF or RLF, and these distributions are

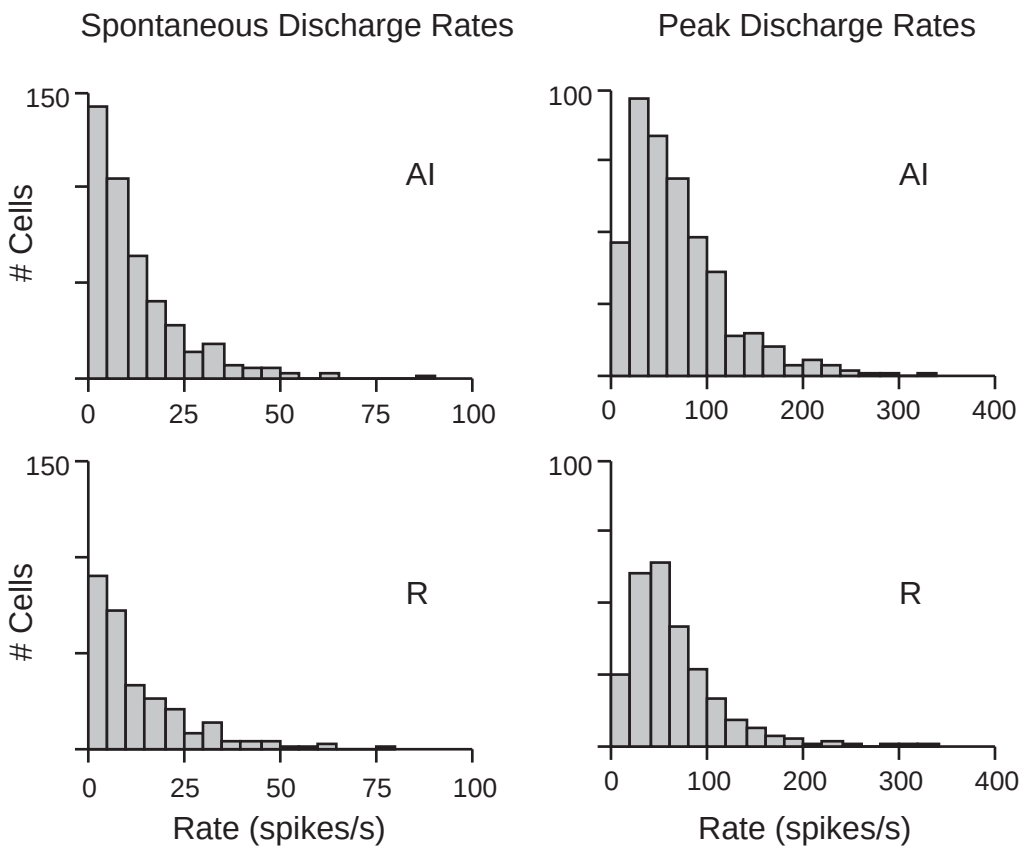


Figure 3.11 Distributions of spontaneous and peak discharge rates in AI and R. Pure tones can evoke strong responses in single units of both fields, often on top of a considerable background firing rate. Spontaneous rate is measured over a 500 ms window during the second half of a 1 s trial; peak rate is measured over a 100 ms window centered on the strongest component of the response. There is no difference between the fields in either measure.

shown on the right of Figure 3.11. Every rate is from a single cell, averaged over 10 trials. These populations are also identical between the two fields (AI: median 58 spikes/s; R: 54 spikes/s; p=.55), consistent with the dark metabolic staining described throughout the length of the core (Figure 3.4). Recanzone et al. (2000a) report lower overall peak rates in AI and R, and find R to be less active than other fields. That study used a standard set of frequencies and levels presented over two trials, then employed local smoothing, whereas stimuli in this study were optimized in frequency and level, and presented over 10 trials. That, plus a greater sample size in R, may explain this disparity.

3.5.2 Distributions of best frequencies

Best frequencies of neurons encountered in AI span nearly the full range of the macaque audiogram (28 Hz to 37 kHz, with peak sensitivity at 4 kHz, as measured by Jackson et al. 1999). In contrast, the distribution in R favors low and midrange frequencies (Figure 3.12A, ANOVA, $p<.0001$ corrected for sample size). The skew toward high frequencies in the AI distribution may reflect a greater proportion of cortical area devoted to this frequency range, but the apparent peak at 1-2 kHz likely results from sampling low-frequency AI more densely in an attempt to maximize the yield of IPD-sensitive cells (analyzed in Chapters 2 and 5). Although the full extent of the rostral core was sampled in only one hemisphere (Z, left), this tissue was tuned almost exclusively to tones ≤ 3 kHz. Other hemispheres showed

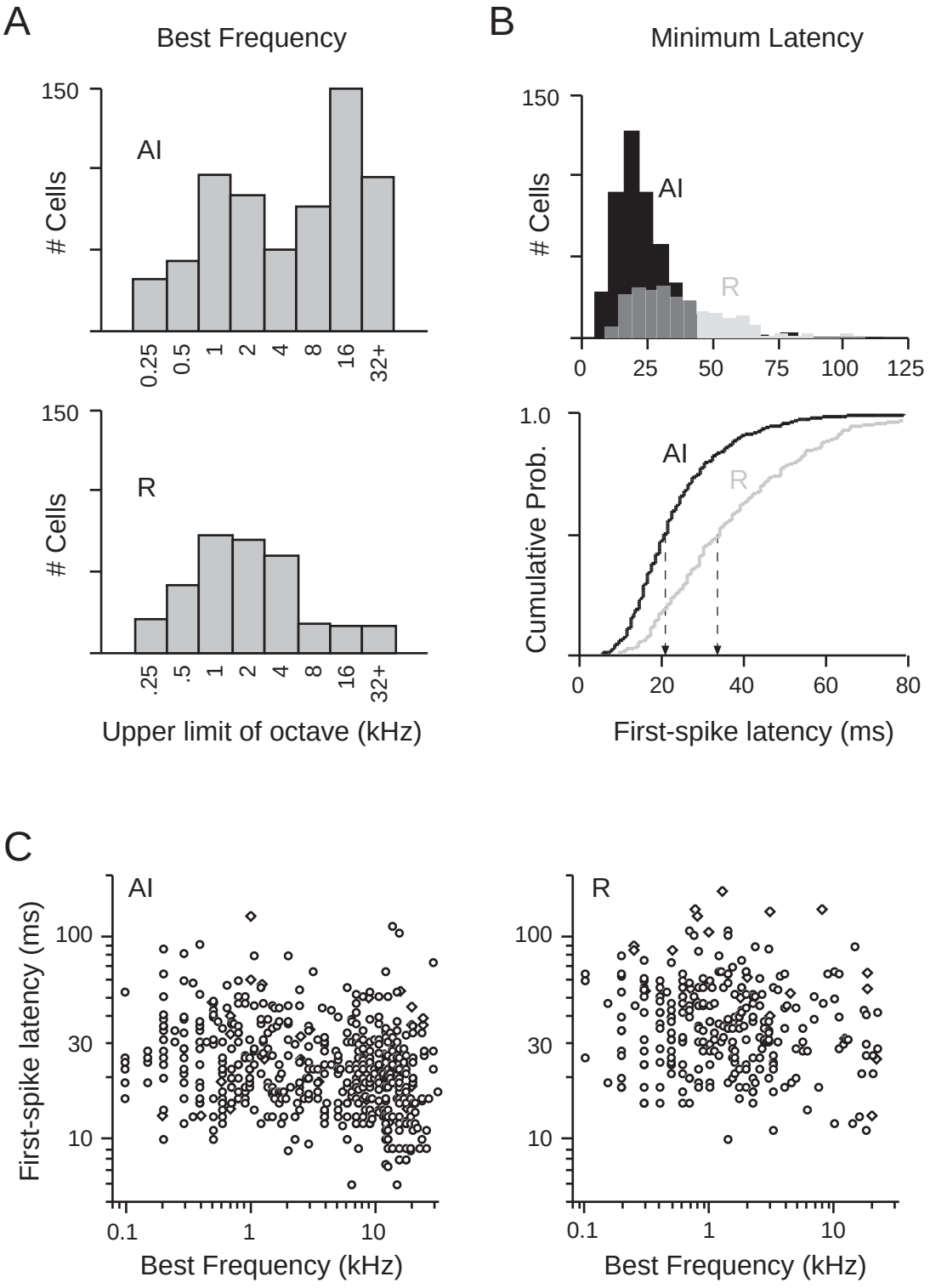

A
Best Frequency
150
AI
Cells
0.25
0.5
1
2
4
8
16
32+
150
R
Cells
.25
.5
1
2
4
8
16
32+
Upper limit of octave (kHz)
B
Minimum Latency
150
AI
Cells
R
0
25
50
75
100
125
1.0
AI
R
Cumulative Prob.
0
20
40
60
80
First-spike latency (ms)
C
AI
First-spike latency (ms)
100
30
10
0.1
1
10
Best Frequency (kHz)
R
100
30
10
0.1
1
10
Best Frequency (kHz)

Figure 3.12 Best frequency and latency distributions in AI and R. **A:** Distribution of BFs encountered in AI (upper panel, N=613) and R (lower panel, N=305). Bins are octaves including BFs ≤ the bin label. Median values are the 8 kHz octave in AI, and the 2 kHz octave in R. **B:** Minimum response latencies in AI and R as overlaid histograms (upper panel) and cumulative probability functions (lower panel). Median latencies are 22 ms in AI (N=480) and 35 ms in R (N=249). **C:** Scatterplots of latency vs. BF in AI and R; open circles indicate latencies measured using pure tones, diamonds measured using band-passed noise.

more variation, but the spectral representation across R is compressed relative to that in AI, and biased toward frequencies between 0.5-8 kHz. This confirms the inference from Figures 3.1 and 3.7 that the reversed tonotopy in R, when present at all, is incomplete (Merzenich and Brugge 1973).

3.5.3 Minimum first-spike latency

The relative timing of responses in AI and R is the aspect in which the two fields most markedly differ. The upper panel of Figure 3.12B overlays distributions of minimum latencies for cells in AI, showing a skew toward short latencies in AI and a broader distribution with a long-latency tail in R. The median latency in R is 13 ms longer than that in AI (AI: 22 ms, N=480; R: 35 ms, N=249; $p<.0001$). The lower panel of 3.12B highlights this difference as a shift in the cumulative distributions of latency. A similar disparity in latency between AI and R is reported by Recanzone et al. (2000b), though they report longer average latencies in both fields; as with peak discharge rates, stimulus optimization and measurement technique may explain the difference.

A modest contributor to this latency shift might be the difference in BFs between the two fields; in the inner ear, activation at the low-frequency apex of the cochlea lags activation of the high-frequency base by a few (~2) ms, this time being required for the traveling wave to reach the apex. Figure 3.12C plots the relation

between latency and best frequency in AI and R; a weak trend exists toward shorter latencies at the highest carrier frequencies in AI, but the relative dearth of high-frequency cells in R makes the trend hard to discern. When latencies are regressed against BF in octave bins (those used in panel A), there is a weak but significant effect in AI of –1.6 ms/octave (ANOVA, $p<.0001$). The effect in R does not quite reach significance, but shows a similar slope (-1.3 ms/octave, $p<.06$). The two BF distributions shown in panel A differ in their median frequency by two octaves, which would predict a latency difference of about 3 ms. The 13 ms latency difference between AI and R appears to reflect a real difference in temporal processing, which will be explored further in the next chapter.

3.5.4 Rate-level functions

Tones often evoked a complex mixture of excitation and suppression in neurons of the auditory core, with the relative magnitude and timing of these response components varying as a function of stimulus level. At least half of the neurons in both fields had nonmonotonic RLFs. Cells exhibiting strong nonmonotonicity (38% in AI, 46% in R) or a weak "rollover" of <50% (15% in AI, 18% in R) outnumbered those with saturating plateau (7% in AI, 5% in R) or strictly monotonic RLFs (31% in AI, 21% in R). Messy, unclassified RLFs accounted for the remaining 8% in AI and 11% in R. When the four categories are condensed into

a dichotomous classification, R has significanlty more nonmonotonic units (53% nonmonotonic in AI, 63% in R, p=.01 by chi square).

Threshold distributions in the two fields were identical (Figure 13.3, left panels, AI: mean 22 dB, R: mean 23.5 dB), confirming that both fields of the core share the same sensitivity to pure-tone stimulation. Strongly nonmonotonic cells tended to cluster in a penetration, consistent with prior reports in awake macaque (Brugge and Merzenich 1973) and other species (e.g. Phillips et al. 1994). This implies that tuning to level may be a columnar property in the auditory cortex.

Best levels differed, with the monotonic functions in AI skewing the distribution toward higher SPLs, whereas best levels in R were distributed nearly uniformly (Figure 13.3, right panels). The median best level in AI was 50 dB (mean 51 dB), the median in R was 40 dB (mean 45 dB; p=.018 corrected for sample size). Related to this, the 100% dynamic range (computed as [best level – threshold]) of neurons in AI was slightly greater than that in R (AI: median 25, R: median 20; p=.0004).

The patterns of distribution of rate-level functions in both fields, and the distributions of best levels, are in accord with the results of Recanzone et al. (2000b), but thresholds in that study were found to be lower in AI, likely a result of

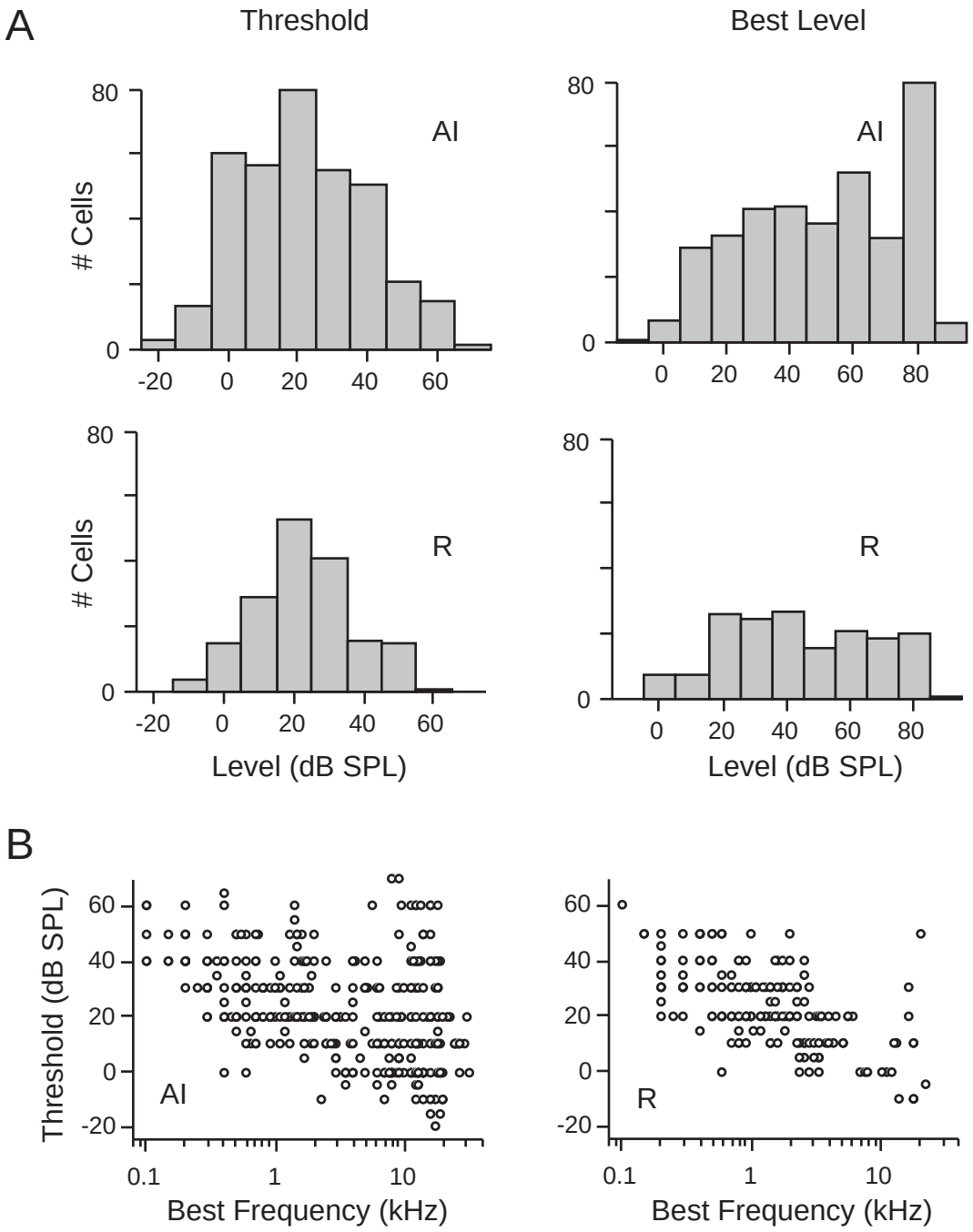

A
Threshold
Best Level
AI
AI
R
R
Cells
Cells
80
0
-20
0
20
40
60
80
Level (dB SPL)
Level (dB SPL)
B
Threshold (dB SPL)
60
40
20
0
-20
0.1
1
10
Best Frequency (kHz)
Best Frequency (kHz)

Figure 3.13 A : Distributions of threshold SPL (left) and best level (right) in AI and R. Thresholds did not differ between fields (AI mean 22 dB, N=360; R mean 23.5 dB, N=174), but best levels were significantly higher in AI (see text). This is consistent with the prevalence of nonmonotonic RLFs in R. The peak in the AI best level histogram is an artifact of often using 80 dB as highest SPL tested, making this the best level for many monotonic neurons. B: Threshold was inversely correlated with BF in both AI (left) and R (right); linear regression indicates an effect of approximately –3.4 dB/octave in AI, and –4.9 dB/octave in R (ANOVA, $p<.0001$). Note the relative paucity of BFs > 4 kHz in R.

disparate sample sizes between the two fields. Threshold in that study was found to be minimal for AI cells with BFs near 1 kHz (the audiogram of the macaque has a minimum at 4 kHz), but minimal thresholds in the current study dropped steadily with SPL in both fields (Figure 13.3B). Since that study used contralateral free-field stimuli, and this used closed-field diotic stimuli, it is possible that thresholds for low-frequency cells appear artificially high because these units are tuned to IPDs other than 0° (see Chapter 5).

3.5.5 Binaural properties

By the criteria established earlier in the chapter, the bulk of the cells studied in both fields were either binaurally driven (59% in AI, 64% in R) or influenced by one ear only (24% in AI, 23% in R); this left only a slim minority of binaurally opponent EI cells (12% in AI, 8% in R). There was a strong effect of BF on binaural class, such that 67% of EI cells in AI had BFs >16 kHz; only 10% fell into the lowest three octaves (≤ 1 kHz). When EI cells were encountered, there was a tendency to find several in a single penetration, suggesting that they may form clusters within the cortex. Brugge and Merzenich (1973) report tuning to ITD that remains consistent within an electrode track in awake macaque AI; there were not enough EI cells present to form bands, but some alternative form of binaural organization may exist (discussed below).

3.6 An alternative approach: mapping responses to standard stimuli across the cortex

The preceding analyses follow a common procedure in sensory neuroscience, whereby the parameters of a stimulus are adjusted to find the "ideal" stimulus to which each cell is tuned. The standard tool for mapping auditory cortex is the isofrequency line, constructed by finding the frequency at which each cell has its lowest threshold (Merzenich et al. 1975, Reale and Imig 1980). This reveals order in the cortex, especially under anesthesia, when the responses recorded in cortex can be attributed largely to thalamic input patterns. In this chapter, frequency tuning was not consistently measured at threshold, but at a moderately loud level (typically 60 dB SPL) or at the best level of strongly nonmonotonic units. Like the isofrequency approach, the stimulus was tailored to each cell. This has been a fruitful approach in many sensory systems, but implies a neural organization in which a given cell functions as a "detector" of a given frequency, at a given level. Both techniques produce topographic mapping of abstracted response properties.

In fact, every stimulus must activate a wide population of cortical neurons: some will be tuned very near the parameters of the stimulus and respond strongly, many others will respond in a graded fashion according to their own tuning. Drawing on the example of Phillips et al. (1994), a standard set of stimuli was applied to every cell across one hemisphere (X, right) in an attempt to map the spread of activation

across the cortex as stimulus parameters are varied. The standard set included contralateral pure tones at eight frequencies (100 ms tones, 800 ms trials, 10 trials per stimulus) at four levels, in addition to a battery of band-passed noise and a modulation of interaural phase (the binaural beat, to be reprised in Chapter 5). This technique was first applied in auditory cortex by Goldstein et al. (1970), where the resulting patterns of activation were taken as evidence for the weakness of tonotopic organization; in fact this apparent lack of order was most likely due to the pooling of data from many hemispheres aligned to sulcal landmarks. Chronic recording introduces uncertainty in cortical mapping (discussed below), but offers an unique opportunity to explore fully a single cortical hemisphere.

3.6.1 Measuring the spread of activation

The extreme variability in response type and overall firing rate complicated the development of a reliable metric for mapping activation. Two metrics were developed, the first based on traditional spike rate measurements. Rates were measured over 50 and 100 ms windows starting at the onset of the 100 ms tone, and a 100 ms window starting at the offset. Spontaneous rate was measured in four 100 ms bins in the second half of the 800 ms trial, to yield an average and a standard deviation from 40 measurements over 10 trials. The maximum absolute rate difference (excitation or suppression, onset or offset) between a response window and spontaneous rate was considered significant at 2 standard deviations away from

the spontaneous rate. Each response was normalized to the maximum significant response of that cell, to yield a metric between zero and one for all units.

A second metric, originally developed in our lab to examine responses to modulated stimuli (see Chapter 4), was used to capture the temporal structure imparted on PSTHs by the stimulus, without regard to overall firing rate. The 10 trials were sorted into two separate PSTHs (the five odd-numbered trials, and the five even-numbered trials); the bin counts in those PSTHs were then correlated, and the correlation coefficient between them taken as a metric of response strength. (Bin number was empirically optimized to 128, and bootstrap simulations of randomly generated PSTHs indicated that a value of ≥ 0.4 was significant at the $p<.001$ level). If correlation in the first 300 ms of the trial (including tone onset and offset) was > 0.4, and greater than that during the last 300 ms (spontaneous activity), it was counted as an activation. Both the rate and correlation measures produced occasional false alarms (confirmed by examining rasters for selected points on the activation maps), but rarely from the same data. To reduce overall noise in the maps, these two metrics (both of which scale between zero and one) were averaged to yield an "activation" axis.

3.6.2 A fractured, level-dependent representation of frequency

Figure 3.14 presents cortical activation maps for two representative frequencies (1 and 8 kHz) across increasing stimulus levels. The upper maps, at 20 dB, show patchy activation at the fringes of AI in response to 1 kHz, and a few islands of activation in more caudal AI in response to 8 kHz. As the SPL of the tone is increased, both representations expand as a greater swath of cortical tissue is engaged by the stimulus. At the maximal level of 80 dB, the activated area is extensive but still complementary between the two frequencies (the 1 kHz map retains "white space" at its caudal extent, and the 8 kHz map shows an absence of activation near the low-frequency border of AI and R.)

Phillips et al. (1994) report no suggestion of isofrequency bands in their standard activation maps, and moreover found that the expansion of activity with increasing SPL can run *orthogonal* to isofrequency contours. In general, the patterns seen in Figure 3.14 echo those of Phillips et al. (1994) in suggesting a patchy distribution of activity, though we have no equivalent maps of isofrequency contours for comparison. The fact that these maps do not suggest isofrequency contours may be attributed to the coarseness of tonotopy (discussed below), or independent tuning to stimulus level and binaurality that prevents any contralateral tone of fixed level from activating the full extent of a contour.

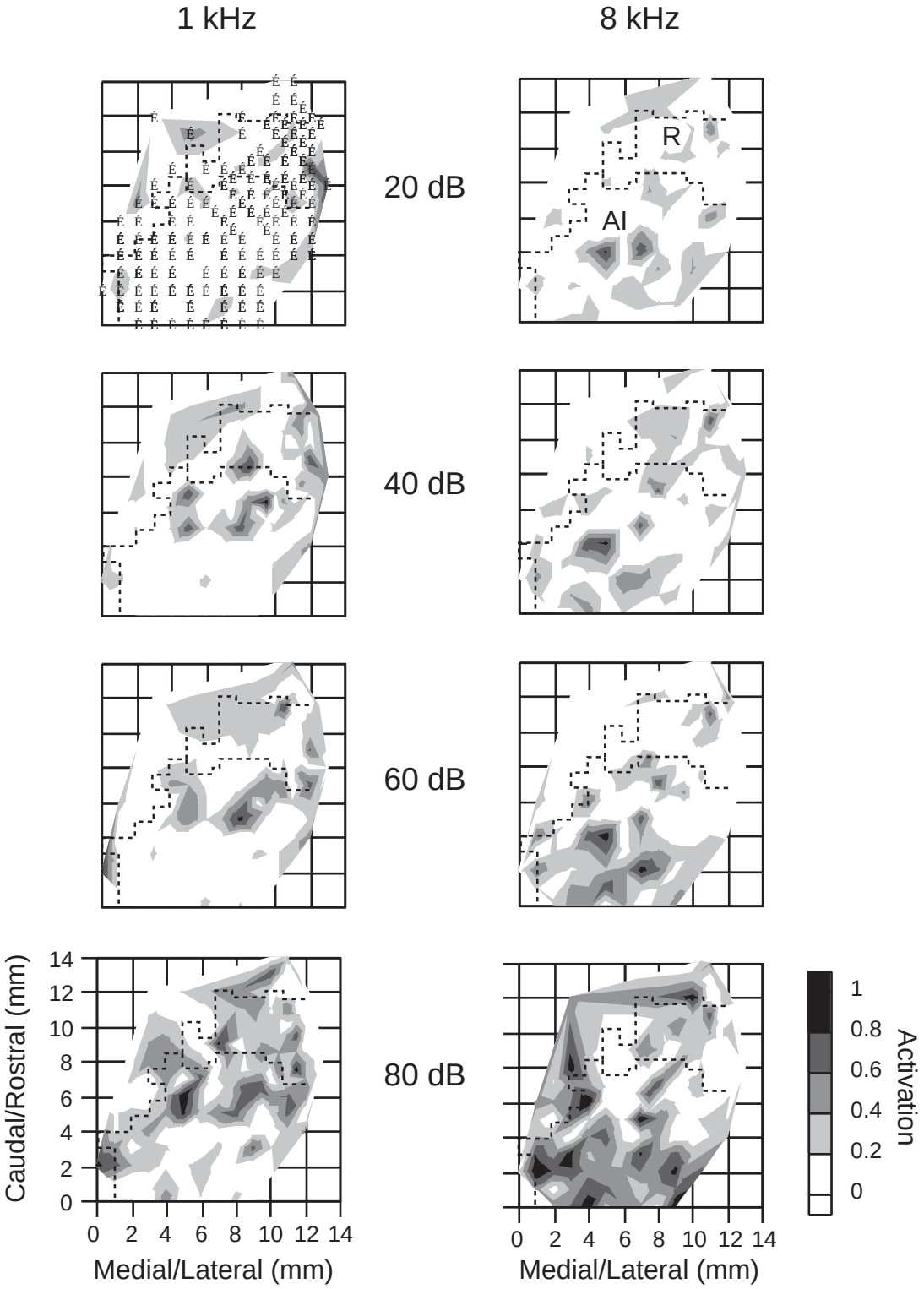

1 kHz
8 kHz
20 dB
40 dB
60 dB
80 dB
R
AI
Caudal/Rostral (mm)
Medial/Lateral (mm)
Medial/Lateral (mm)
0
2
4
6
8
10
12
14
Activation
1
0.8
0.6
0.4
0.2
0

Figure 3.14 Cortical activation maps in response to 1 kHz (left) and 8 kHz (right) contralateral tones presented at a range of SPLs (increasing top to bottom). Grayscale axis represents an average of spike rate contrast and temporal PSTH structure (see text). Open circles at top right mark the recording sites from which maps are interpolated (same for each panel, omitted for clarity), dashed lines mark the boundaries of AI and R; monkey X, right hemisphere, same as in Figure 3.8 and 3.10 (upper right panel). Each frequency elicits a patchy, distributed pattern of activity at low intensity, expanding to a broader activation with increasing SPL. Even at 80 dB, these two tones (3 octaves apart) activate complimentary regions of tissue: 1 kHz fails to drive the most caudal high-frequency tissue, and 8 kHz fails to drive near the low-frequency border of AI and R.

3.6.3 Restricted sensitvity to IPD and broad activation by band-passed noise

Interaural phase sensitivity in low-frequency cells of the core is explored fully in Chapter 5, but a standard set of binaural beat stimuli at four carrier frequencies reliably marked the rostral extent of AI and the border with R (Figure 3.15, left panels). For this stimulus, the grayscale axis indicates correlation alone; a response was significant only if the modulation of phase induced temporal synchrony. The lowest frequency of 250 Hz activated the border region selectively, 500 Hz drove cells in AI and R, and 1 kHz activated select patches in AI – but not the far rostral tissue that gave a rate response to 1 kHz tones at a similar SPL (compare the 1000 Hz map to the 60 and 80 dB maps in Figure 3.14). Phase sensitivity appears especially strong in this border region, and appears not to extend to all low-frequency cells of the core.

Finally, band-passed noise evoked a response across nearly the full extent of the core (Figure 3.15, right panels). Like the tone-derived maps of the previous figure, low and high center frequencies seem to elicit nearly complementary spatial distributions of activation when bandwidth is restricted to 1/3 octave (compare the top panel, 1 kHz, to the third panel, 10 kHz). When bandwidth is increased to a full octave, 10 kHz noise drives the full expanse of the auditory core and surrounding belt, though the response is especially strong in the high-frequency caudal tissue (lower panel).

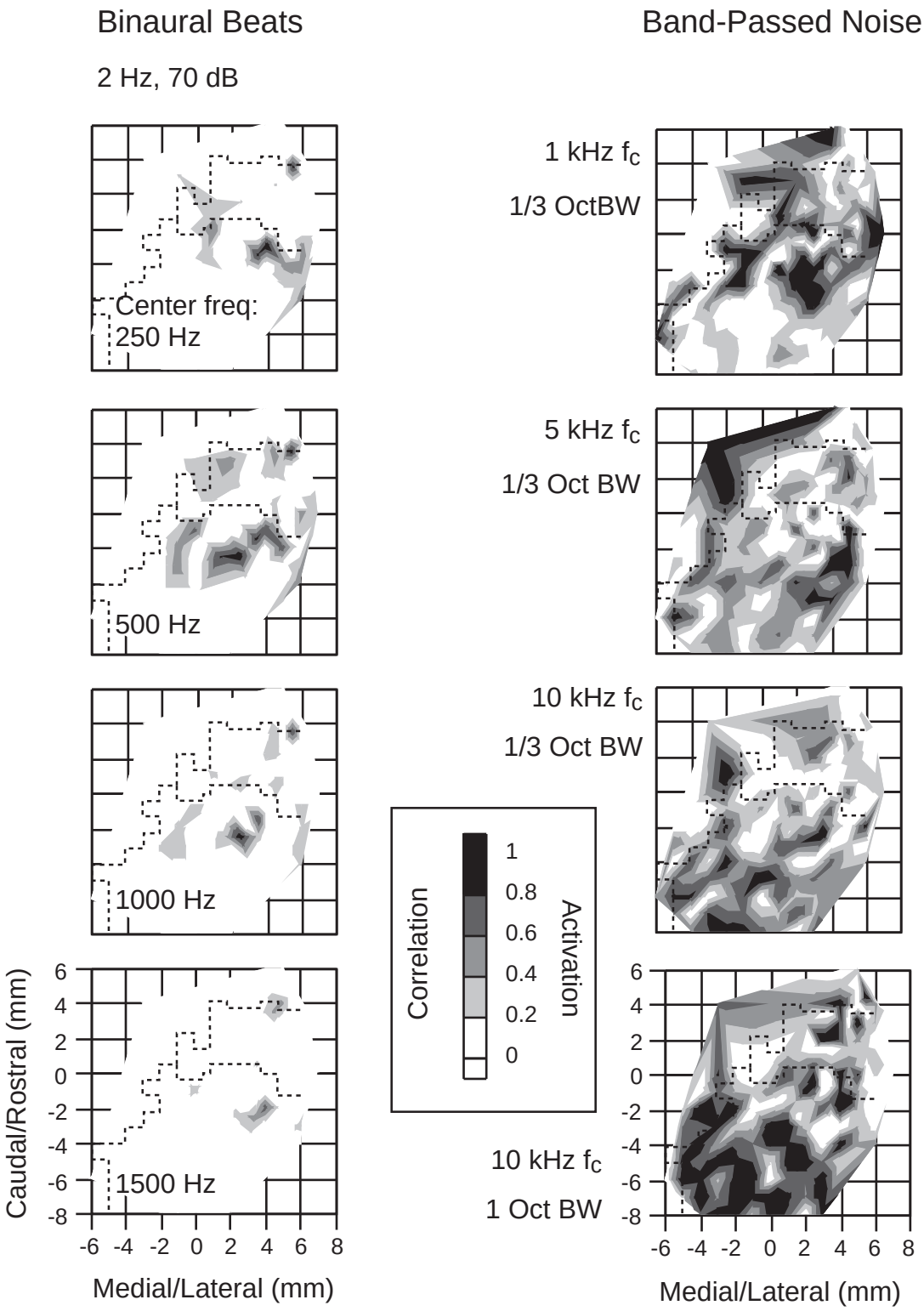

Binaural Beats
2 Hz, 70 dB
Band-Passed Noise
Center freq:
250 Hz
500 Hz
1000 Hz
1500 Hz
1 kHz fc
1/3 OctBW
5 kHz fc
1/3 Oct BW
10 kHz fc
1/3 Oct BW
10 kHz fc
1 Oct BW
Correlation
Activation
1
0.8
0.6
0.4
0.2
0
Caudal/Rostral (mm)
Medial/Lateral (mm)
Medial/Lateral (mm)
6
4
2
0
-2
-4
-6
-8
-6 -4 -2 0 2 4 6 8

Figure 3.15 Cortical activation maps in response to modulations of IPD (binaural beats, left) and band-passed noise (right). Grayscale in noise panels uses same metric as previous figure; in beat panels, it represents temporal synchrony as measured by correlation (see text). At the lowest IPD carrier frequency tested (top left), activation is restricted to the AI/R border, with higher frequencies driving locations in both fields. Noise evoked a wide pattern of activation. Top three panels use approximately 1/3 octave bandpass noise at 1, 5 and 10 kHz center frequency; the lower panel repeats the 10 kHz center frequency with a wider, 1 octave band, an especially effective stimulus in high-frequency caudal AI and medial belt.

3.7 Discussion

3.7.1 Response properties and organization of the core fields

The vast body of work in barbiturate-anesthetized cat describes responses in AI as weak, often restricted to 1-3 onset spikes per stimulus over a very low level of spontaneous activity. Core auditory cortex in the awake macaque is characterized by robust responses to pure tones throughout the depth of the cortex (present study, Brugge and Merzenich 1973, Miller et al. 1974). A high level of spontaneous activity provides the contrast necessary to identify suppression in extracellular recordings. This makes it apparent that sharp tuning to frequency manifests not only as a phasic elevation of discharge rate, but as a complex interplay of excitation and suppression, during and immediately after the stimulus. Tuning to stimulus level, evident in the prevalence of nonmonotonic RLFs, also included temporal restructuring of the response in addition to overall changes in spike rate. Responsiveness to tones, and high spontaneous and driven spike rates, characterized both the primary and rostral fields, in accordance with the consistent staining for the metabolic marker cytochrome oxidase throughout the rostrocaudal extent of the core.

In brainstem auditory structures, neurons have been classified according to the shape of their PSTH in response to a pure tone (Joris 1998). Categories such as

onset, onset-sustained, and offset have been applied to neurons of the awake rhesus cortex (Recanzone 2000), but I made no effort to do so with this data set. Classification of a given *response* may be valid, but classification of the *neuron* is not; PSTH shapes are too variegated and parameter-dependent at the cortical level to hypothesize a structure-function relationship as has been proposed in the cochlear nucleus (Joris 1998). The neuron in Figure 3.6, for example, could be classified in any of several ways depending on the particular combination of frequency and intensity being applied. Such variability was not common to all cells tested (e.g., certain units fired onset responses only), but was common enough to preclude a system of fixed response categories.

3.7.2 Tonotopic organization in awake cortex

Tonotopic order appears coarser in the awake preparation (present study, Pfingst and O'Connor 1981, Recanzone et al. 2000a) than is commonly seen in AI and R of anesthetized macaque (Merzenich and Brugge 1973, Morel et al. 1993) or in AI and the anterior auditory field of cat (Reale and Imig 1980, Lee et al. 2004). The relative degree of order in tonotopic maps depends on the scale of organization being examined. The global order, from high caudally to low rostrally in AI, is very clear in all three hemispheres that allowed a complete mapping of AI (Z, right excepted). Also, a consistent BF or an orderly progression of BFs within an electrode penetration was common in both fields (see Figure 3.8A), implying that

local organization is quite precise. Why, then, is there disorder at the "intermediate" scale, within a few mm along the cortical surface, such as that seen in Figure 3.10?

Any apparent divergence from a smooth cochleotopic gradient may be attributed to at least three experimental factors. First, there is error in electrode placement using the vertical approach employed here, unlike in acute experiments where the overlying cortex may be aspirated to expose the STP. Each penetration is assigned to a point on a map, but its true location could vary by 1 mm in any direction (Pfingst and O'Connor 1980). An additional complication arises because the angle of electrode entry is not necessarily normal to the cortical surface, so a sequence of BFs encountered may be averaged to assign a single BF to each penetration. The depth correction employed in one hemisphere appeared to only add noise to interpolated BF maps (Figure 3.8B and C), suggesting that this may not be the problem. A third possibility is that the absence of anesthesia complicates tonotopic order because the responses recorded reflect the activity of local cortical circuits; components of these circuits lying outside layer IV may be suppressed by anesthesia, producing orderly maps that more accurately reflect patterns of thalamic input than the cortical representation of frequency. The true pattern of frequency response in auditory cortex may not be aligned along isofrequency contours, but

encompass a patchy, distributed activation closer to that seen in the standard-stimulus maps of Figure 3.14.

Frequency tuning functions were collected at or near each cell's best level, so full frequency response areas (in frequency and SPL) are not available for most neurons. This precludes a comparison of tuning width using standard measures such as "Q values" (tuning width at some fixed interval above threshold), as has been done in previous studies of primate cortex (Recanzone et al. 1999, Cheung et al. 2001). The width of neural tuning generally increases with SPL (seen in our own data and many prior studies) and is also dependent on the rise time (Phillips et al. 1995) and spectral bandwidth (Barbour and Wang 2003) of the stimulus, whereas BF is generally invariant with these parameters. This suggests that the BFs mapped in Figure 3.10, and used to delineate the fields of the core, are reliable estimates of each neuron's frequency tuning.

3.7.3 Binaural organization

Binaural interaction has been proposed as a secondary organizational principle for auditory cortex, based on reports that binaurally opponent (EI) cells clustered in bands running orthogonal to isofrequency contours (Imig and Adrian 1977, Middlebrooks et al. 1980). These continuous bands have not been confirmed (perhaps due to methodological considerations, e.g. Phillips and Irvine 1979), but

cells of particular binaural classes do segregate into clusters (Kelly and Judge 1994, Zhang et al. 2004). This is consistent with the observation above that EI cells, when encountered, are found in the same electrode penetration. The tendency for EI cells to be almost exclusively high-frequency is in accordance with the duplex theory of hearing, to the extent that ILD cues are weak at low frequencies, where the brain is sensitive to IPD cues for azimuth. Equating EI cells with ILD sensitivity may be inappropriate given the simple monaural tests applied here, but to subserve ILD processing, EI cells should be distributed across all octaves > 2 kHz, the limit of IPD sensitivity in these animals (see Chapter 2). Two-thirds of EI cells had BFs in the highest two octaves encountered (>16 kHz), but ILD thresholds have never been determined at carrier frequencies > 10 kHz in this species. Whether these cells confer enhanced sensitivity to ILD at very high frequencies remains unknown.

Given the ability of the animal to detect ILD cues (Chapter 2, Wegener 1972), and the prevalence of inhibition in awake cortex, it is perhaps surprising that so few ILD-sensitive cells were described. A clinical study found that some ability to use ILD remained after bilateral lesions of auditory cortex in humans (Yamada et al. 1996), suggesting that the computation of ILD accomplished in the brainstem is accessible in the absence of AI; if the signal is present in cortex, it may be re-coded in such a way that tuning to ILD was not apparent in these experiments. Binaural level differences are created by spectral filtering at the pinna as well as shadowing

by the head, creating location-dependent spectral profiles at each ear (Spezio et al. 2000). Computation of ILD at the cortical level is likely to integrate across frequency bands, and may not be apparent when probed with pure tones.

3.7.4 Hemispheric asymmetries

In the human brain the left hemisphere dominates our processing of speech (Price et al. 1996), perhaps our most intricate acoustic task. Some behavioral evidence from rhesus macaque suggests a possible bias toward the left hemisphere in processing species-specific vocalizations (Petersen et al. 1978, Heffner and Heffner 1984, Hauser and Anderson 1994). Tasks of this complexity are likely to engage many fields beyond the core level of the hierarchy, but the possibility of lateralization of function in rhesus macaque warranted an investigation of hemispheric asymmetries in the primary and rostral fields. Data from the left and right hemispheres were pooled for the above comparisons of AI and R, but comparing each field with its contralateral counterpart revealed some differences at the level of the core. Thresholds in both fields were higher in the left hemisphere (AI: 4 dB difference, $p=.016$, R: 5.7 dB difference, $p=.01$). Spontaneous discharge rate was also lower in left R (13 spikes/s right, 9.7 left, $p=.03$). These differences are slight, but consistent with a positron emission tomography study in rhesus that reported consistently higher activation in the region of the core fields in the right hemisphere relative to the left, for a wide range of sound classes (Poremba et al.

2004). At the temporal pole, species-specific vocalizations (but not other sounds) revealed a bias toward the left hemisphere. If the core fields of the left hemisphere were somehow specialized for vocalization processing, they might be expected to be less responsive to general environmental sounds than the corresponding fields of the right hemisphere. This might explain why the left hemisphere shows elevated thresholds to tones, and lower spontaneous activity.

3.7.5 Conclusions

The basic response properties of single neurons in AI and R confirm their inclusion as core fields, defined by reliable pure-tone responsiveness and separated by the partial reversal of the tonotopic gradient in AI. This reversal is incomplete in that frequency organization breaks down with distance from the low-frequency border between the fields, and the frequency representation rostral to that border is compressed, with a bias toward low- to middle frequencies. Though these fields do operate in parallel, in the sense that each receives direct thalamic input, they do not operate in synchrony. Response latencies in R lag those in AI by 13 ms on average, far longer than would be predicted from synaptic delay even if R received its input in serial, via AI. This implies that AI and R operate at different temporal resolutions, an hypothesis to be explored in the following chapter.

Chapter 4: Temporal envelope processing: amplitude and frequency modulation

4.1 Introduction

The preceding chapter illustrated common response properties of the core auditory fields using simple, static stimuli. On the basis of responses to pure tones, it was shown that the primary auditory cortex (AI) and the adjacent rostral field (R) share similar tuning to frequency and stimulus level, but differ in the timing of their responses. To explore the effects of stimulus dynamics on neural responses in the auditory core, this chapter employs modulations of tone amplitude and frequency to reveal the coding mechanisms and temporal limits of cortical neurons, and potential disparities in the spectral and temporal response properties of AI and R.

4.1.1 Relevance to natural stimuli

Pure tones are a product of the laboratory; real-world sounds consist of complex temporal modulations of spectral content and amplitude, which can be roughly divided into components of "fine structure" and "envelope". Fine structure refers to the spectral content of the stimulus, the amplitude of which fluctuates over time to define the envelope. Environmental noise and communication sounds, including speech, have complex envelopes with many frequency components, but low-frequency modulations may be most important to the comprehension of speech.

Experiments using "chimeric" sounds, which combine the envelope of one sound with the fine structure of another, demonstrate forcefully that envelope cues dominate speech recognition (Smith et al. 2002), particularly for slow modulations in the range of 1-20 Hz (Drullman et al. 1994, Shannon 2002). Ahissar et al. (2001) demonstrated that single-trial success for speech comprehension was best predicted by the correlation between the temporal envelopes (≤ 20 Hz) of the stimulus and the envelope of cortical activity measured by magnetoencephalography, implying that cortical responses to these modulations may carry the representation necessary for perception.

As a model for the processing of complex sounds, modulations of pure-tone stimuli can mimic temporal modulations of amplitude and frequency characteristic of speech and other communication sounds. Sinusoidal amplitude modulation (SAM) and sinusoidal frequency modulation (SFM) offer a simple but flexible tool for investigating the dynamic coding abilities of auditory neurons. For clarity, the introductory discussion will focus on SAM, but the principles established will be applied to both stimulus sets.

4.1.2 Representation of dynamic stimulus features

There is physiological evidence that auditory cortical neurons, and thus the cortical population as a whole, can be sensitive to the shape of modulation waveforms

(Swarbrick and Whitfield, 1972). Nevertheless, the nature of the neural representation of the temporal information within complex sounds remains unresolved. In particular, it is not clear whether the responses of cortical neurons to a commonly studied form of modulation, SAM, are best described as encoding modulation frequency, or alternatively, as encoding the amplitude envelope. Although four parameters are necessary to specify a SAM (or SFM) stimulus, most prior studies have focused on neural sensitivity to the parameter of modulation frequency (to the exclusion of level, depth, and carrier frequency, discussed below), and have emphasized the upper limits of temporal encoding in auditory cortex (e.g., Schreiner and Langner 1988). Support for such focus comes from evidence that summary measures of responses to SAM or SFM, such as the best modulation frequency (BMF) or the overall shape of the modulation transfer function (MTF), are relatively invariant across changes in sound level or modulation depth. These measures can reveal the temporal limits of auditory neurons, and they will be used to do so here, but their invariance has given the impression that temporal modulation, and not the amplitude or frequency modulation itself, is what most auditory neurons appear to extract (Liang et al. 2002).

On the other hand, Krishna and Semple (2000) report that modulation tuning in the inferior colliculus depends substantially upon sound level, supporting the view of Rees and Moller (1987, p. 141) that midbrain neurons "do not function as an array

of stimulus-invariant modulation frequency detectors", but rather "carry a selectively emphasized version of the input signal's amplitude envelope which is modified by the prevailing stimulus conditions." The representation of amplitude modulation at ascending stages of the auditory system is characterized as a progression from isomorphism, in the auditory nerve and cochlear nucleus, to abstraction at the midbrain, thalamus and cortex (Joris et al. 2004). If it is true that there is an increasing tolerance of modulation tuning for other stimulus parameters such as level, carrier, and modulation depth, at higher levels along the neuraxis (Joris et al. 2004), must this mean that these other parameters are not represented in the discharge of cortical neurons?

The tolerance of modulation tuning is related to the more fundamental question of what aspect of SAM and SFM signals cortical neurons actually encode – temporal modulation, or amplitude and frequency envelopes. The two are related insofar as the periodic changes in the stimulus envelope define the modulation frequency. Nevertheless, a code for modulation frequency would suggest that cortical neurons should be relatively insensitive to other parameters of the SAM stimulus, whereas a code for amplitude demands that cortical neurons be robustly sensitive to the features of the amplitude waveform. A neural code for amplitude envelopes also suggests that the nature of the encoding in an individual neuron should also depend on the neuron's tuning to stimulus level. For example, the increasing diversity and

nonmonotonicity of rate-level functions (RLFs) in higher auditory centers (Chapter 3; Semple and Kitzes, 1987) suggests that cortical neurons should respond to SAM stimuli in ways that reflect differences in their tuning for sound amplitude, in addition to differences in their temporal sensitivity. Pure temporal modulation coding, however, would be compromised if a neuron's tuning for stimulus level introduced response features not present in the stimulus envelope (i.e., nonlinearities). If cortical neurons do represent features of the amplitude envelope rather than simply synchronize their discharge to the period of the modulation, conventional measures of SAM coding (such as vector strength to measure synchrony [see Methods]) may fail to capture important aspects of the cortical response (Eggermont, 1991). In other words, we must consider the *fidelity* of the cortical representation of the amplitude envelope, as distinct from the *synchrony* of the response to the modulation frequency.

Detecting this fidelity may not be possible in anesthetized cortex, where the absence of sustained activity may cause neurons to "default" to a sparse synchrony code (Goldstein et al. 1959, Gaese and Ostwald 2001). Relatively few spikes are necessary to provide a synchrony code for modulation frequency; they must occur at a consistent phase within the period, so the primary limit on the extraction of the modulation frequency is the temporal resolution of the auditory pathway up to and including the responding neuron. In order to capture multiple features of the

amplitude envelope, or to track its shape, multiple spikes per modulation cycle are required. Given the limitations on cortical discharge rates, the coding of amplitude or frequency envelopes via changes in instantaneous discharge rates is probably restricted to the low modulation frequency range utilized by animal communication sounds and human speech.

4.2 Methods

4.2.1 Stimulus protocols

Detailed methods are described in Chapter 2. Briefly, subjects were two male rhesus macaques, trained to discriminate the laterality of binaural cues around midline, and accustomed to sitting quietly between behavioral sessions to allow the collection of additional physiological data in the passive state. Stimuli were generated digitally by MALab software, synthesized by dedicated hardware processors and presented in the closed field via electrostatic speakers coupled to ear inserts. Phase and sound pressure level (SPL, dB re: 20 μPa) were calibrated across frequency at the start of each session. An electrode was positioned vertically and advanced with a stepping microdrive, through the parietal cortex to auditory regions of the superior temporal plane. Entry into auditory cortex was typically preceded by a gap in physiological activity believed to correspond to passage through the lateral sulcus; confirmation of recording sites is presented in Chapter 3.

When a responsive single unit was isolated, frequency tuning and rate-level functions (FTFs and RLFs, see Chapter 3) were used to determine the neuron's best frequency (BF) and best level. All stimuli were gated on and off with a cosine-squared ramp of rise time 5 ms (for tones) or 10 ms (SAM and SFM). Modulated stimuli were first presented at best frequency and level; for cells with monotonic RLFs, or those with indeterminate best levels, 60 dB was used. Modulated tones, and an unmodulated pure-tone control (PTC) of equivalent center frequency and SPL, were presented in two long consecutive trials of 10 s, separated by a 2 s interstimulus interval. Unless one ear obviously inhibited the response, stimuli were delivered binaurally (92% of cells). Long stimulus durations were chosen to minimize the effects of onset responses while maximizing the number of modulation periods.

SAM stimuli are defined by a sinusoidal tone of a given carrier frequency (f_c) modulated sinusoidally by a second tone of a given modulation frequency (f_m), such that:

$$s(t) = A[1+m\sin(2\pi f_m t)]\sin(2\pi f_c t)$$

If the carrier frequency is much greater than the modulation frequency ($f_c >> f_m$) then the bracketed term defines the time-varying amplitude of the stimulus. The base amplitude of the stimulus is set by A, and the depth of the modulation is determined by *m*, which varies from 0-1 (described as 0-100% modulation). The

sound level of a SAM stimulus (e.g., 60 dB) shall be referred to in terms of carrier SPL (A), although the actual sound level of the stimulus will vary over the course of the modulation period when m>0%. Because of limitations in the recording time available, not all parameters could be varied in all cells. Most neurons were first presented with 100% modulated SAM at a range of modulation frequencies (0.7, 1, 2, 5, 10, 20, 50, 100 and 200 Hz); higher frequencies were added if necessary to find the limit of discharge synchrony (described below), and intermediate values could be added for greater resolution in the resulting modulation transfer function (MTF). In many cases, effects of modulation depth were explored in an effort to identify the threshold of modulation at which synchrony was evoked. Variations in carrier frequency and level were performed less commonly.

SFM is also defined by a sinusoidal carrier (f_c), the frequency of which is modulated by f_m:

$$S(t) = A*\sin(f_c t + m\sin f_m t)$$

As with SAM, A defines the amplitude (which remains constant within the bound of calibration error) and m defines the depth (maximum deviation of the modulated frequency from the carrier; the total peak-to-trough depth will be twice this number). Manipulations of SFM parameters will be described below.

4.2.2 Data analysis

Calculations of spike rate were based on the entire stimulus duration (10 s). To evaluate the significance of differences in average firing rate, responses were broken into 1 s epochs and the average firing rate (across repeated trials) was calculated for each. Significance was assessed for all comparisons by a heteroschedastic *t*-test. Calculation of spontaneous rates was based on firing rates for 1 second epochs drawn from all interstimulus intervals.

Response synchronization was quantified in terms of vector strength (VS; Goldberg and Brown 1969). Each spike is treated as a unit vector whose angle corresponds to the phase at which it occurs in the modulation cycle. These unit vectors are summed to produce a resultant vector whose length corresponds to the magnitude of the Fourier component of the response at the modulation frequency. Normalizing the resultant vector by the total number of spikes (n) yields the VS, which is bounded from 0 (spike counts are equal at all phases) to 1 (all spikes occur at the same phase). The direction of the resultant vector indicates the mean phase of the response. A weakness of the VS metric is that it is dependent on spike rate – at the extreme, a single spike must produce a VS of 1. To assess the statistical significance of VS, the Rayleigh statistic ($2*VS^2*n$) was computed, and values > 13.816 (Mardia and Jupp, 2000) were considered to be significant at $p<.001$.

Additional details concerning the applicability of the VS metric, especially to SFM responses, will be discussed in the results.

4.3 Results I: SAM

4.3.1 General response properties

The data described in this chapter were obtained as part of an extensive physiological survey of auditory cortex. Table 4.1 (next page) tabulates the number of cells studied with SAM in AI and R, plus some from the surrounding belt regions (not discussed here). Each unit was studied across modulation frequency, but parameters other than modulation frequency may not have been manipulated. Chapter 3 describes the criteria by which cells were apportioned to cortical fields. No compelling differences were found between animals or hemispheres, so these data were pooled.

Carrier levels ranged from -10 to 90 dB, with 48% of cells tested at 60 dB, and carrier frequencies varied from 0.1-32 kHz. At least 40 cells were tested in each octave relative to 0.5 kHz (i.e. <0.5 kHz, 0.5-1 kHz, 1-2 kHz, etc.). A neuron was considered to be responsive to SAM if it exhibited a significantly synchronized response to at least one modulation frequency, or exhibited a significantly different firing rate from the response to the unmodulated control for at least one modulation frequency. By this criterion, 99% of tested cells responded to SAM presented at

their best carrier frequency and level (or 60 dB). Cells generally were not tested with a full range of SAM frequencies if they were deemed unresponsive during initial testing, so this may over-estimate the prevalence of SAM responsiveness in the full cortical population.

	Left Hemi	Right Hemi	Animal Subtotal:
Monkey X			X:
AI	73	145	218
R	16	69	85
Belt	19	21	40
Monkey Z			Z:
AI	67	29	96
R	19	54	73
Belt	13	7	20
Hemi Subtotal			**Grand Totals:**
AI	140	174	**314 in AI**
R	35	123	**158 in R**
Belt	32	28	60
		Core Total:	**472**

Table 4.1: Number of SAM MTFs from single units in four hemispheres

Figure 4.1 illustrates responses to SAM and SFM, and introduces the conventions for display of these responses as modulation period histograms (MPHs). For each cell, the response to an unmodulated tone at best carrier and level was obtained as a control for the responses to modulated tones (4.1A). A striking aspect of cortical responses in awake animals is that roughly one-third (139/468, 30%) had significantly elevated firing rates (relative to the spontaneous rate) when calculated over the full 10 s duration of the control tone. An additional 11% (51/468) were

suppressed over this duration, in opposition to the notion that cortical neurons do not give sustained responses to pure tone stimuli of long duration. Panel 4.1D, at the right, is a schematic illustration of how the SAM stimulus interacts with a cell's receptive field. The arrows indicate the rising and falling phases of sinusoidal modulation along the SPL axis, in relation to monotonic (black line) and nonmonotonic (gray shading) response areas. The length of these arrows would indicate modulation depth, the SPL on which they are centered is the carrier level, and their location along the ordinate indicates carrier frequency (in this case, the BF of the neuron). Note that for nonmonotonic cells, a wide depth or high carrier SPL like that shown here may cause the stimulus to enter and exit the receptive field twice per modulation cycle, a phenomenon that will be examined later in this chapter.

The periodic modulation of spike rate for a 1 Hz SAM stimulus (100% depth) is evident in the trial-based histogram of Figure 4.1B. Folding the responses on the modulation period of 1 s allows the distribution of discharge rates within the modulation period to be more easily seen in the MPH representation of Figure 4.1C, left panel. The stimulus envelope was presented in sine phase, but in the right panel of 4.1C (and all subsequent MPHs), we have rotated the response by 90°, into cosine phase. In this representation, the instantaneous amplitude minimum (at 270°) occurs in the middle of the MPH, and the maximum occurs at the lateral

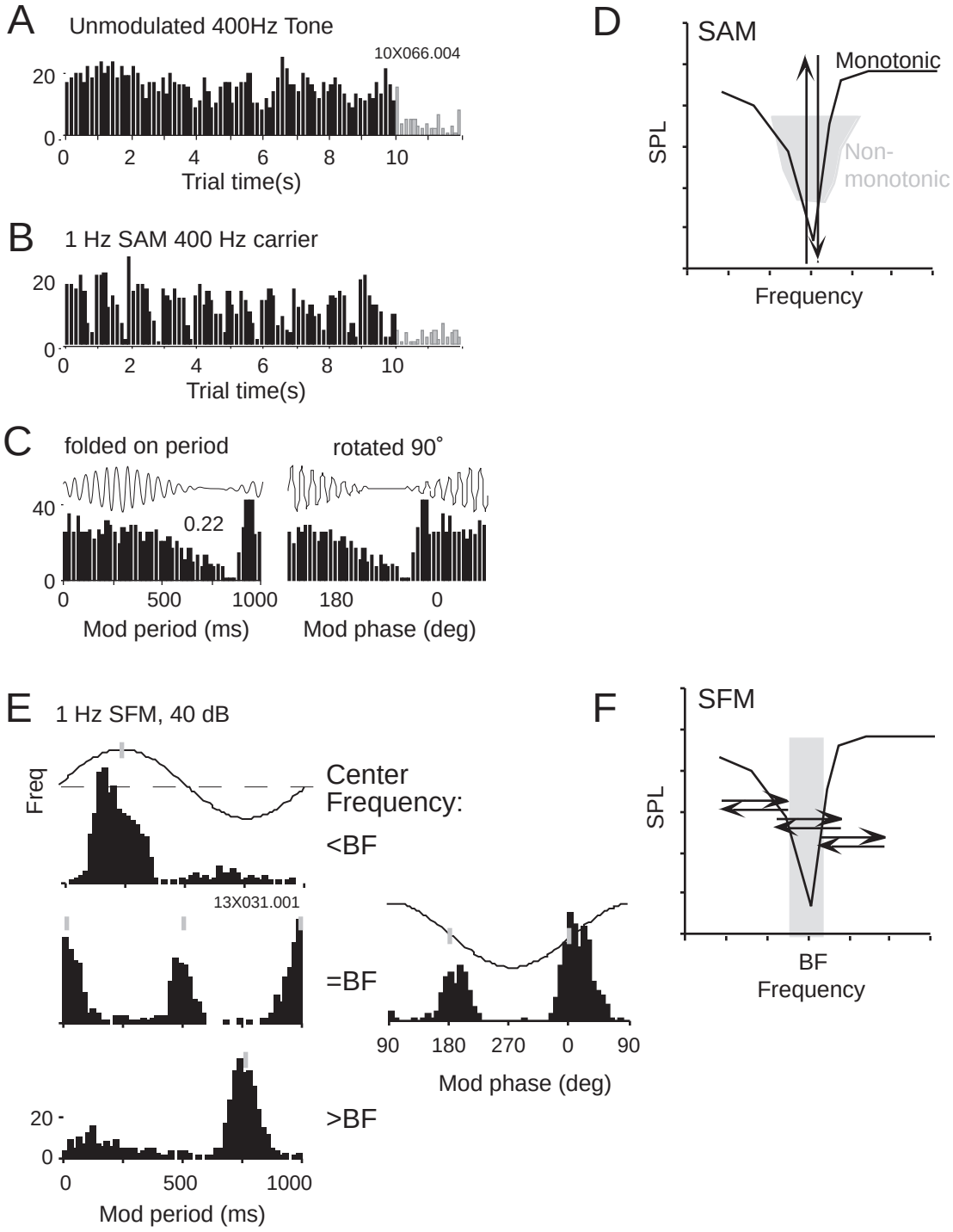

A
Unmodulated 400Hz Tone
10X066.004
20
0
0
2
4
6
8
10
Trial time(s)
B
1 Hz SAM 400 Hz carrier
20
0
0
2
4
6
8
10
Trial time(s)
C
folded on period
rotated 90˚
40
0.22
0
0
500
1000
180
0
Mod period (ms)
Mod phase (deg)
D
SAM
Monotonic
SPL
Non-
monotonic
Frequency
E
1 Hz SFM, 40 dB
Freq
Center
Frequency:
<BF
13X031.001
=BF
>BF
90
180
270
0
90
Mod phase (deg)
20
0
0
500
1000
Mod period (ms)
F
SFM
SPL
BF
Frequency

Figure 4.1 Construction of modulation period histograms (MPHs) **A:** Peri-stimulus time histogram (PSTH) in response to 2 repetitions of a 10 s unmodulated tone, which serves as a control for responses to SAM and SFM. Gray bars beyond 10 s are spikes fired during the 2 s interstimulus interval; the last second of the every trial was used to estimate spontaneous firing rate. **B:** PSTH in response to 1 Hz SAM of the 400 Hz carrier in panel A, at 100% depth. **C:** Response in panel B folded on the modulation period, to create the MPH in sine phase (left panel). Responses to 20 cycles of modulation (2 trials of 1 Hz moduation, 10 s each) are superimposed. Stimulus icon shows the relationship between the SAM envelope and the instantaneous firing rate apparent in the profile of the MPH. For clarity, MPHs were "rotated" by 90° into cosine phase (right panel) to put the envelope minimum at 270° in the center of the axis. The most salient feature of the MPH is the sharp rise in spike rate during the rising phase of the envelope. The vector strength (VS) of this response is 0.22. **D:** Icon depicting the possible relationship between a SAM stimulus and the receptive field of a monotonic (black line) or nonmonotonic (gray area) cell. Arrows indicate the rising and falling phases of the modulation; their length corresponds to the modulation depth, and their center point on the SPL axis corresponds to the carrier level. In this example, the carrier frequency (position of the arrows on the ordinate) is set to the BF of the neuron. **E:** MPHs in response to SFM with a carrier frequency below, at, or above the BF of the cell (upper, middle, and lower panels respectively). Stimulus icon above the MPHs shows the sinusoidal frequency envelope modulating about the center frequency (dashed line). Gray ticks mark when the modulating frequency equals the BF of the cell. As with SAM, MPHs were rotated into cosine phase to place the envelope minimum at the center of the axis; this clarifies that the MPH in response to modulation centered at BF is bimodal, and avoids splitting peaks to either side of the circular axis. **F:** Iconic representation of how SFM may interact with a cell's frequency response area (black line). Arrows indicate the two phases of modulation along the frequency axis; as in panel D, their length corresponds to modulation depth; their center point on the ordinate is the center frequency of the modulation. Leftmost arrows correspond to the upper MPH in panel E, in which the stimulus enters the receptive field only during the rising phase of modulation; middle and right arrows correspond to the center and lower MPHs. (Arrows are offset on the SPL axis for clarity only; all MPHs were gathered at the same level and modulation rate).

extremes of the circular axis (90°). In the "rotated" MPH of Figure 4.1C, the instantaneous spike rate declines as the instantaneous level declines, then rises abruptly when the amplitude increases from its minimum. This feature of the response prevents the MPH from exhibiting the half-rectified sinusoidal profile expected for a linear response to a sinusoidal modulation envelope. Consistent with the sustained response to the pure-tone control, the cell is responsive throughout most of the modulation period, resulting in a relatively low VS of 0.22. Although the nominal synchrony (as measured by VS) of the response is poor, the high spike rate generates a strongly significant Rayleigh statistic. This fact, and inspection of the MPH profile, suggests that despite a low VS, the response represents the stimulus envelope quite well at this low modulation frequency.

Panels E and F of Figure 4.1 present a similar schematic representation of the SFM stimulus, to be discussed below.

As mentioned above, a pure synchrony code for modulation frequency based on phase-locking to the stimulus envelope is achievable with one or fewer spikes per modulation cycle. By contrast, the fidelity of an envelope code based on instantaneous discharge rate is crucially dependent on the number of spikes that a neuron can fire within each modulation period. Whereas a synchrony code of frequency marks a point in the modulation period, a rate code for the envelope

sketches a function through the modulation period, evident in the profile of the MPH. The MPH in Figure 4.1C is defined by an average of >50 spikes per modulation cycle.

Figure 4.2 plots distributions of average spikes/cycle at all tested modulation frequencies, which follow a power function (linear on logarithmic axes) intersecting an average of 1 spike/cycle just above 10 Hz. This limits the range of modulation frequencies over which a neuron may mark more than a single mode within the modulation period to approximately ≤20 Hz – the range of envelopes crucial to comprehension of communication sounds (see above). Typically, multiple response modes in the MPH at modulation frequencies ≤ 5 Hz merge into a single response peak at or above 10 Hz (see Figure 4.4 below), a behavior that is also seen in SFM responses.

4.3.2 Coding in single neurons: dependence of MPH shape on SAM parameters

4.3.2.1 Effects of modulation depth

Although the ability to fire multiple spikes per cycle allows a cortical neuron to track features of the stimulus envelope directly, it does not require that the neuron do so. As has been noted in the auditory nerve and elsewhere (e.g. Joris and Yin, 1992), multiple response modes were typically confined to large depth stimuli (m>80%) presented at moderate to high stimulus level, as shown in Figure 4.3.

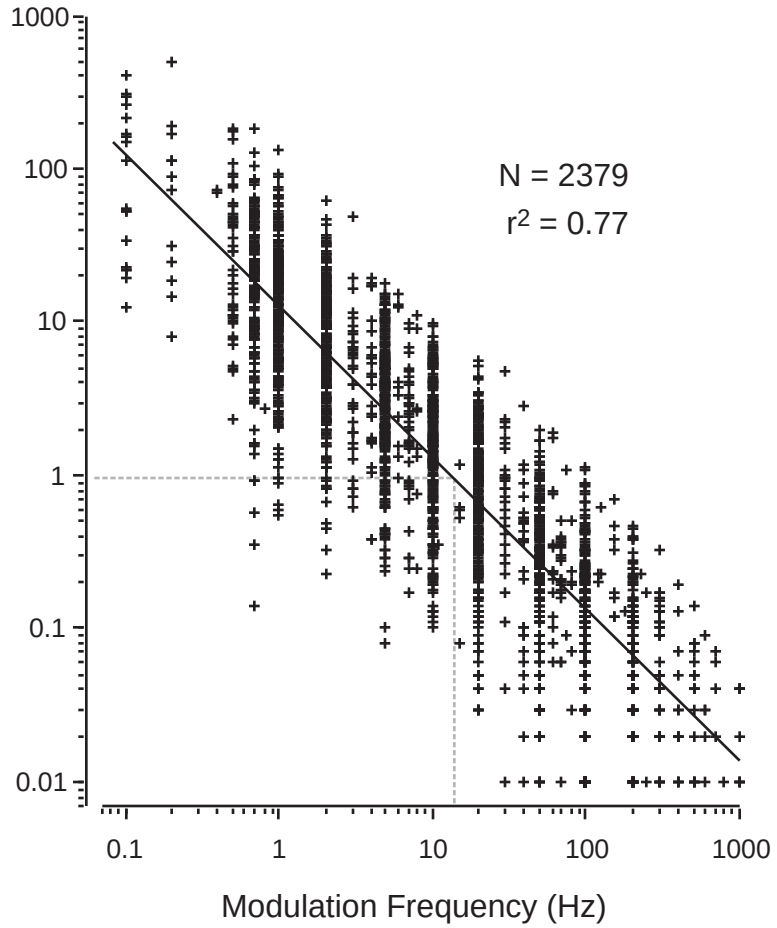


Figure 4.2 In awake macaque cortex, SAM elicits robust responses unlike those in anesthetized preparations. Average number of spikes fired per modulation cycle is dependent on modulation frequency: on logarithmic axes, these spike counts follow a power function, averaging tens of spikes per cycle at low frequencies but firing only one spike per cycle at modulation frequencies > 10 Hz (dashed gray line). In the low-frequency regime, sufficient spikes exist to track the modulation envelope in the shape of the MPH; beyond this regime, the representation defaults to phase-locking at a single point in the envelope.

Sound amplitude is typically represented on a compressed decibel (dB) scale that is representative of our psychophysical sensitivity to sound intensity. Thus, the sinusoidal amplitude envelope of a SAM stimulus is not sinusoidal on a dB axis. In order to reference the SAM envelope to the dB scale commonly used to characterize amplitude sensitivity, the "instantaneous" relative amplitude of the SAM signal was computed as the logarithm of the ratio of the modulated signal to the carrier signal (essentially transforming amplitude into SPL). A family of curves describing the instantaneous SPL of the SAM signal relative to an unmodulated carrier at various depths is shown in Figure 4.3A (curves are shown in cosine phase). With increasing modulation depth, these curves become less sinusoidal, and the falling and rising phases of the envelope become more prominent. There is also an asymmetry in the increases and decreases of SPL within each modulation cycle. When m=10%, the SAM signal increases 0.83 dB (at 0°) and decreases -0.92 dB (at 180°). For m=90%, these differences are 5.5 dB and -20 dB, respectively, and at 100% – the depth used to measure modulation transfer functions – the maximal SPL is double the amplitude of the carrier at 0° (an increase of only 6 dB), and indefinably low at 180° (essentially silent). For the low modulation frequencies emphasized in this chapter, where the envelope is well-defined, instantaneous SPL can be estimated by adding the relative amplitude to the carrier SPL.

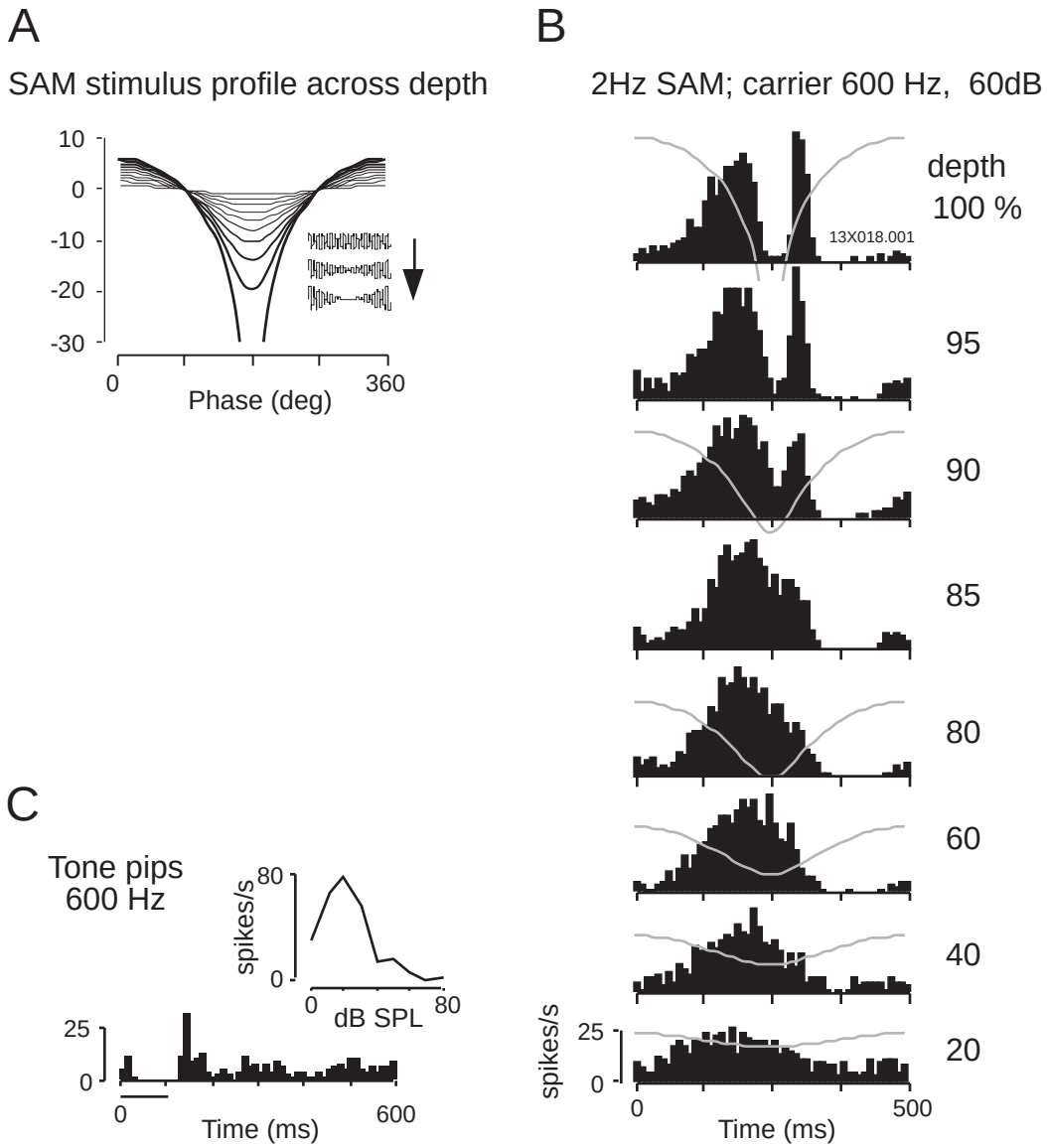


Figure 4.3 Effects of SAM depth on MPH shape. **A:** Sinusoidal amplitude modulation is not sinusoidal on the dB axis used to measure sound intensity (see text). With increasing depth (indicated by line weight), the envelope becomes less symmetric about the carrier SPL (0 on the relative amplitude axis). Inset icons show envelope shapes in amplitude terms, at 0%, 50 % and 100% depth from top to bottom. **B:** MPHs in response to 2 Hz SAM at 60 dB with increasing modulation depth. Stimulus icons in gray show the amplitude envelope transformed onto a dB axis. The best SPL of this cell is 20 dB, well below the 60 dB carrier level; for this reason, the response peak corresponds to the dip in instantaneous amplitude. At high modulation depths (>80%), the falling and rising phases of the modulation elicit distinct peaks in the MPH. C: Rate-level function (RLF) showing the nonmonotoinc tuning of this cell, and a peri-stimulus time histogram in response to a 60 dB tone (100 ms, bar under the time axis). At this SPL, the tone elicits a weak onset response followed by a strong offset response; this is likely a reflection of the same mechanisms that produce the asymmetric response evident in the MPH at high depth.

As a result of this transformation, it can be seen that small modulation depths sample a narrow range of actual stimulus levels. Nevertheless, the cell depicted in Figure 4.3B clearly follows the SAM envelope at 20% depth (+1.58 dB, -1.93 dB), with increasing depths resulting in increasingly robust modulation. Note that in all cases, this neuron responds to the decrease in amplitude (see stimulus icons, in gray) with an increase in firing rate, as would be predicted from the cell's nonmonotonic RLF (Figure 4.3C). The neuron is tuned to tone pips at 20 dB SPL, well below the 60 dB carrier level of the SAM stimulus, and in response to a 60 dB tone pip (peri-stimulus time histogram below the RLF), shows suppression followed by a strong offset response. At 90% depth, a second response mode corresponding to the rising phase of the stimulus envelope begins to emerge. This type of behavior – sudden changes in MPH shape with increases in modulation depth near the top of the range (80-100%) – was common in the sample. It appears to reflect the fact that when the stimulus is of sufficient depth, it moves toward the neuron's preferred range of sound levels twice, once during the falling phase of the envelope, then again during the rising phase (as predicted from the schematic response area in Figure 4.1D).

4.3.2.2 Effects of modulation frequency

Although cortical neurons are capable of encoding stimulus features beyond modulation frequency, this parameter may still reveal the temporal characteristics

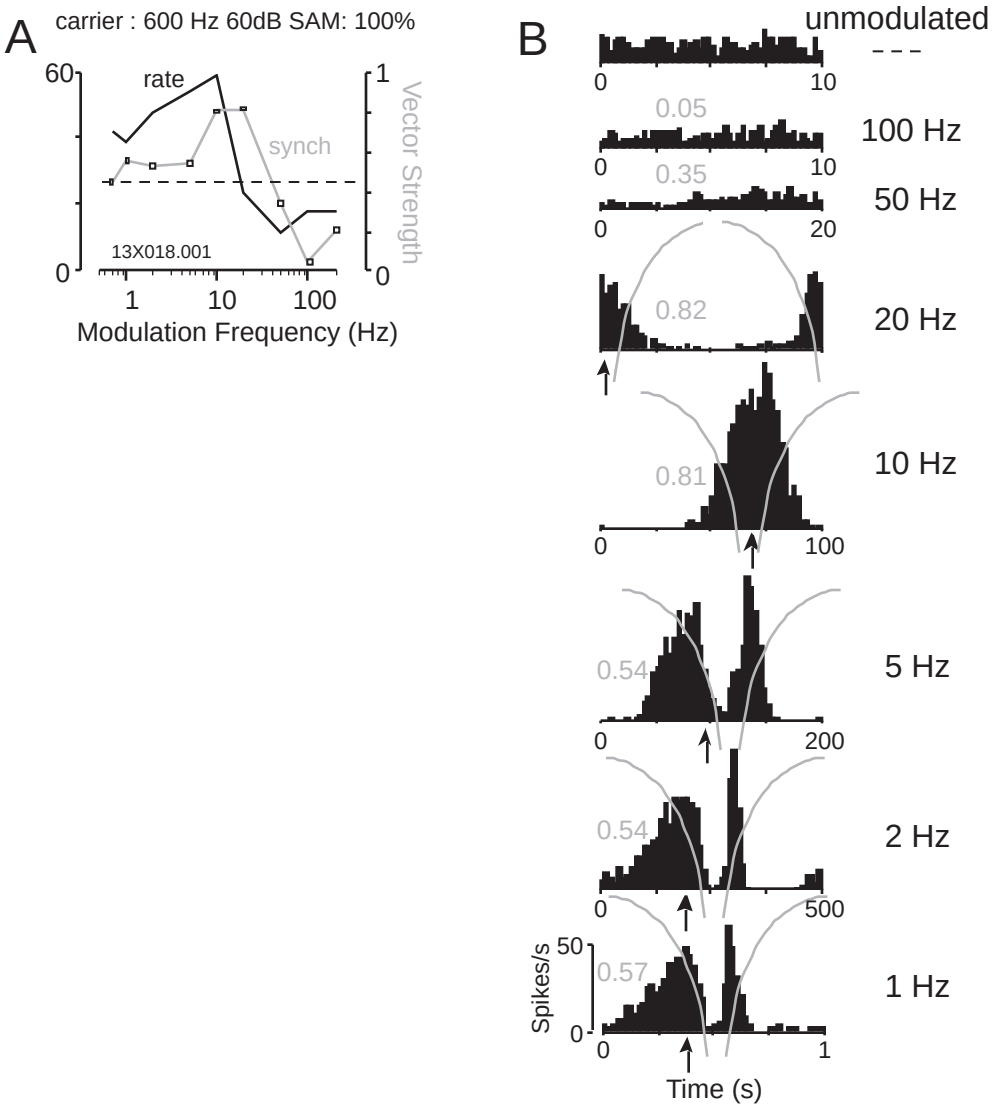


Figure 4.4 Effects of SAM rate on MPH shape and modulation transfer functions (MTFs). **A:** Rate MTF (black line, left axis) plots total spike rate over the trial vs modulation frequency, reaching a peak at 10 Hz and plunging below the spontaneous rate (dashed line) at modulation frequencies > 20 Hz. Synchrony MTF (gray line, right axis) plots vector strengh (VS) against modulation frequency, peaking at 10-20 Hz. **B:** MPH representations of the responses from which the MTFs in panel A are derived. At slow modulation frequencies (1-5 Hz), the rising and falling phases of the envelope are easily resolved in the MPH profile, though these responses merge to a single peak at 10 Hz. Because VS (denoted in gray on each MPH) presumes a unimodal distribution of spikes, the failure of the cell to resolve multiple features of the envelope causes the synchrony MTF to attain its peak. Beyond 20 Hz, synchrony degrades rapidly, and the MPH at 100 Hz is nearly as flat as that for the unmodulated control.

that govern the interaction of a dynamic stimulus with a neuron's receptive field. Responses to changes in modulation frequency (at 100% depth) in Figure 4.4B were obtained from the same cell whose responses are depicted in Figure 4.3. To facilitate interpretation of MPH profiles with regard to the SAM envelopes, stimulus icons have been shifted to reflect the group delay (Eggermont 1999), based on a linear fit of the mean phase (black arrows) from 1 to 5 Hz, where the shapes of the MPHs are relatively constant. This is meant to compensate for the latency of the neuron shifting the response mode within the period as that period becomes shorter. Figure 4.4A shows modulation transfer functions for both rate (rate MTF, black line and left axis) and synchrony (synchrony MTF, gray line and right axis). From these, it is apparent that the best modulation frequencies (BMFs) for rate and synchrony would be 10 and 20 Hz, respectively.

Comparison of the synchrony MTF to the MPHs in Figure 4.4B illustrates how reliance on vector strength (VS) can lead to mistaken assumptions about cortical envelope coding. Although the function is maximal from 10-20 Hz, this apparent increase in response synchrony corresponds to the cell's *inability* to resolve the rising and falling phases of the amplitude envelope. The bandpass shape of the synchrony MTF belies the fact that the fidelity of the representation of amplitude is best below 5 Hz. The presence of narrow response peaks in this range demonstrates that this neuron encodes dynamic changes in amplitude up to the limit of its

temporal tuning properties. These multiple response peaks confound the VS metric, so it appears that the cell is "tuned" for modulation frequencies between 10-20 Hz. Whereas the MPH profile is a representation of what the neuron encodes, the synchrony MTF shape is a representation of what VS encodes: precision of phase-locking for a presumed unimodal distribution of spikes. The MTF is thus an incomplete representation of a cell's coding capacity at low modulation frequencies, though it still may identify the upper limit of temporal synchrony.

4.3.2.3 Effects of carrier frequency

In many studies of SAM processing (including, for the most part, this one) the carrier for SAM signals is chosen in order to elicit the most robust response from the cell, with the assumption that MTFs at neighboring frequencies would be scaled versions of that collected at BF. Although carrier frequency was the parameter varied least often in this study, it was apparent that in cells where the character of the PSTH varied with frequency, those changes were generally reflected in changes of MPH shape. Figure 4.5 provides evidence of this behavior for a 2 Hz modulation presented at carrier frequencies from 13 to 25 kHz. The relative prominence of the response peaks associated with the rising and falling phases shifts in favor of the rising phase near the BF of 19 kHz. This strong response to the rising phase, and weak response to the falling phase, is explicable in relation to the PSTH for a 60 dB tone in panel C, with its strong onset and

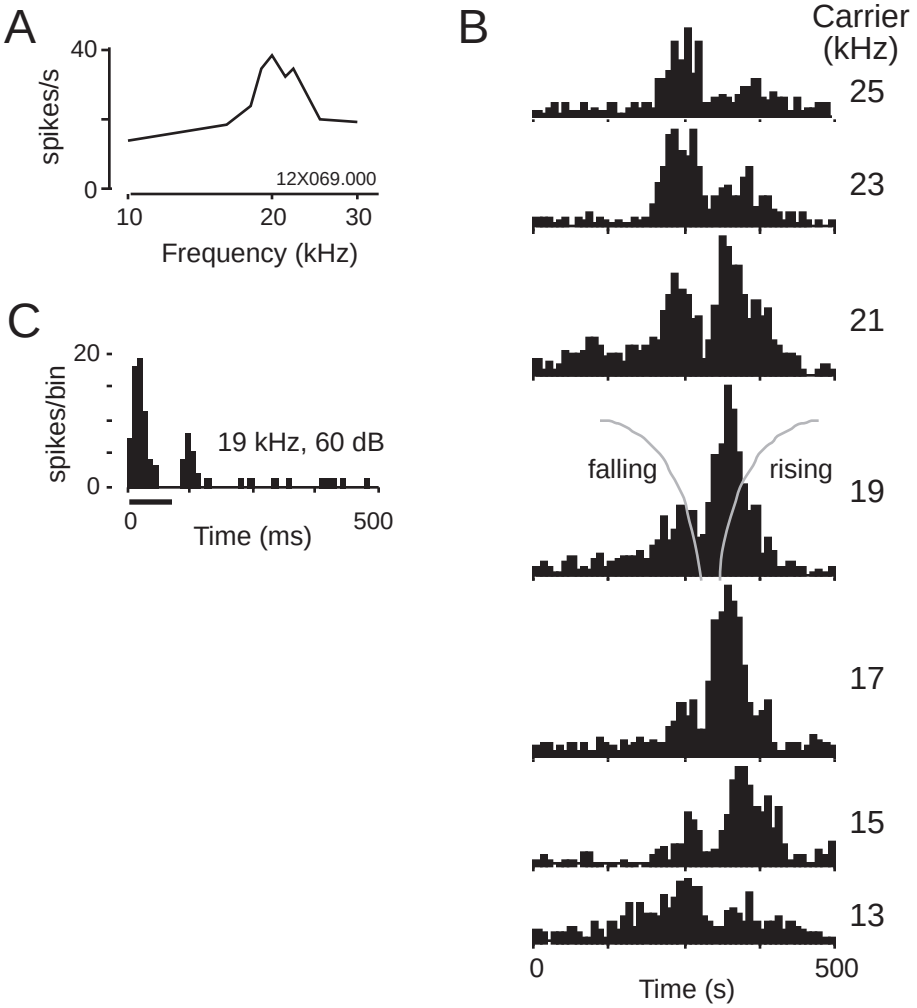


Figure 4.5 Effects of SAM carrier on MPH shape. **A:** Frequency tuning function showing this cell's BF of 19 kHz (determined with tones). **B:** MPHs in response to 2 Hz SAM at 60 dB. The relative prominence of the response peaks corresponding to the rising and falling phases of the modulation is dependent on carrier frequency. **C:** Peri-stimulus time histogram in response to a 60 dB tone at the BF of 19 kHz. The strong onset and weak offset response most likely derive from the same mechanisms that produce the strong response to the rising phase of the amplitude envelope, and the weaker response to the falling phase.

relatively weak offset component. The stronger responses during the falling phase at carriers distant from the BF may reflect sideband inhibition, which leads this cell to represent spectral change via the timing of its response, in addition to its overall discharge rate.

4.3.2.4 Effects of carrier level

The most fundamental description of this unit's tuning for changes in amplitude – i.e., whether it responds to increases or decreases of stimulus level within the modulation period – is jointly determined by depth and carrier level. At 100% depth, the carrier level effectively determines the instantaneous amplitude maximum, because all signals are briefly "off" within each modulation cycle. As shown in Figure 4.6A, changes in the carrier level result in vertical displacements of the signal on the amplitude axis. Unlike modulation depth, changes in the carrier level do not alter the rate of SPL change within the modulation period (the functions appear as scaled, parallel versions of one another).

Figure 4.6C indicates that this neuron was strongly nonmonotonic when tested with pure tones, with a threshold of 20 dB and a best SPL of 30 dB (tone RLF, gray line). As the carrier level for a 1 Hz modulation is increased from 10 to 70 dB, the profiles of the MPHs change profoundly. From 20 to 40 dB, the MPH "trough" corresponding to the lowest instantaneous SPLs (at the center of the axis)

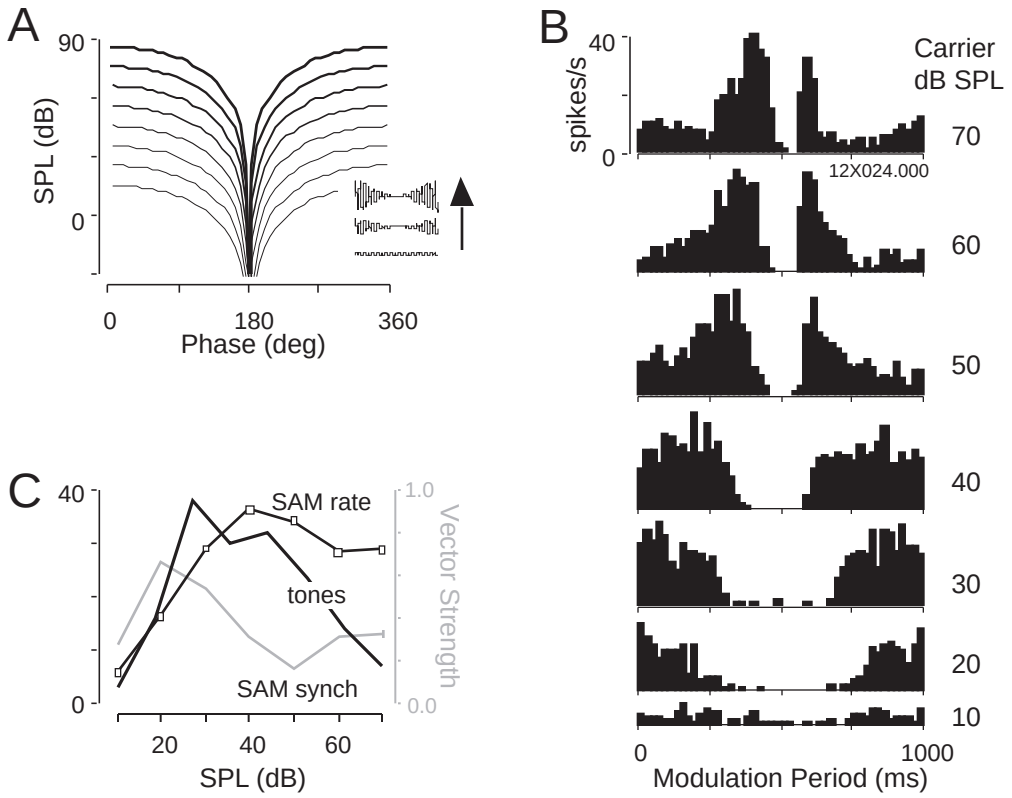


Figure 4.6 Effects of SAM carrier SPL on MPH shape. **A:** Sinusoidal amplitude envelopes transformed onto a dB axis (see text). Inset icons show the original sinusoidal profiles. **B:** MPHs in response to 1 Hz SAM at increasing carrier SPL. The best SPL of the neuron is 30 dB; as SPL rises, the most prominent feature of the response becomes the narrowing notch corresponding to the envelope minimum, and the refinement of the response peaks corresponding to the rising and falling phases of the SPL envelope. **C:** Rate-level functions measured with tones (black line) and SAM (black line, open squares) show that discharge rate in response to SAM does not share the strongly nonmonotonic profile of the tone RLF. Synchrony-level function derived from VS (gray line, right axis) degrades at high SPLs, where peaks and troughs in the MPH track features of the envelope most explicitly.

progressively narrows, as the stimulus spends a greater portion of the modulation cycle above the neuron's dynamic response threshold (which may not be identical to the threshold from the tone RLF). As the carrier level increases through the negatively-sloping range of the RLF (> 50 dB), the cell begins to respond to decreases in SPL with increases in firing rate (left side of the MPH). Conversely, responses during the rising phase of the envelope also become more peaked (right side of the MPH).

The foregoing description would suggest that the cell fires robustly whenever the instantaneous SPL falls within the range of its preferred SPLs. There is some truth to this, but it fails to predict the form of the "SAM RLF" in panel 4.6C (black line, open squares). If the instantaneous firing rate of the cell for SAM stimuli perfectly reflected its static RLF, then the SAM stimuli that spend the most time in this SPL range (20-40 dB) should elicit the largest responses. Relative to the static RLF, the SAM RLF is shifted to higher SPLs, and is also substantially less nonmonotonic. Note that at high SPLs, the cell did not show an onset or offset response, so the static RLF is not distorted by the choice of window over which spike rate was measured.

As described previously, the increasingly phasic responses observed at high carrier SPLs are not captured by VS (gray line and right axis in panel 4.6C). Although this

neuron provides a high-fidelity representation of the SPL envelope, the progressive narrowing of the trough and emergence of dual peaks in the MPH result in a progressive reduction of VS.

4.3.2.5 Parameter interactions and MPH stability

This chapter has argued that cortical neurons in AI and R encode stimulus amplitude via changes in instantaneous firing rate, rather than encode modulation frequency via phase locking to a particular point in the modulation period. Under this view, the MPH is a transformed representation of the SPL envelope, rather than a running tally of stimulus periods. To illustrate how graded changes in the parameters defining SAM produce graded changes in the associated MPHs, Figure 4.7 shows data from 4 slices through the SAM space in a single cell. Variations in modulation frequency at 20 dB are aligned on the z-axis in black; variations in carrier level at 2 Hz are aligned on the x-axis in gray; variations in modulation depth at 20 and 60 dB are aligned on the y-axis as unfilled histograms. The robustness of the MPH profile over repeated trials can be observed in the overlying histograms at the conjunction of these axes (3 repeats at 2 Hz, 30 dB, 100%; 2 repeats at 2 Hz, 60 dB, 100%; each collected several minutes apart).

Along the modulation frequency axis, there is a progressive rounding of the peak corresponding to the SPL maximum within the modulation period (the changing

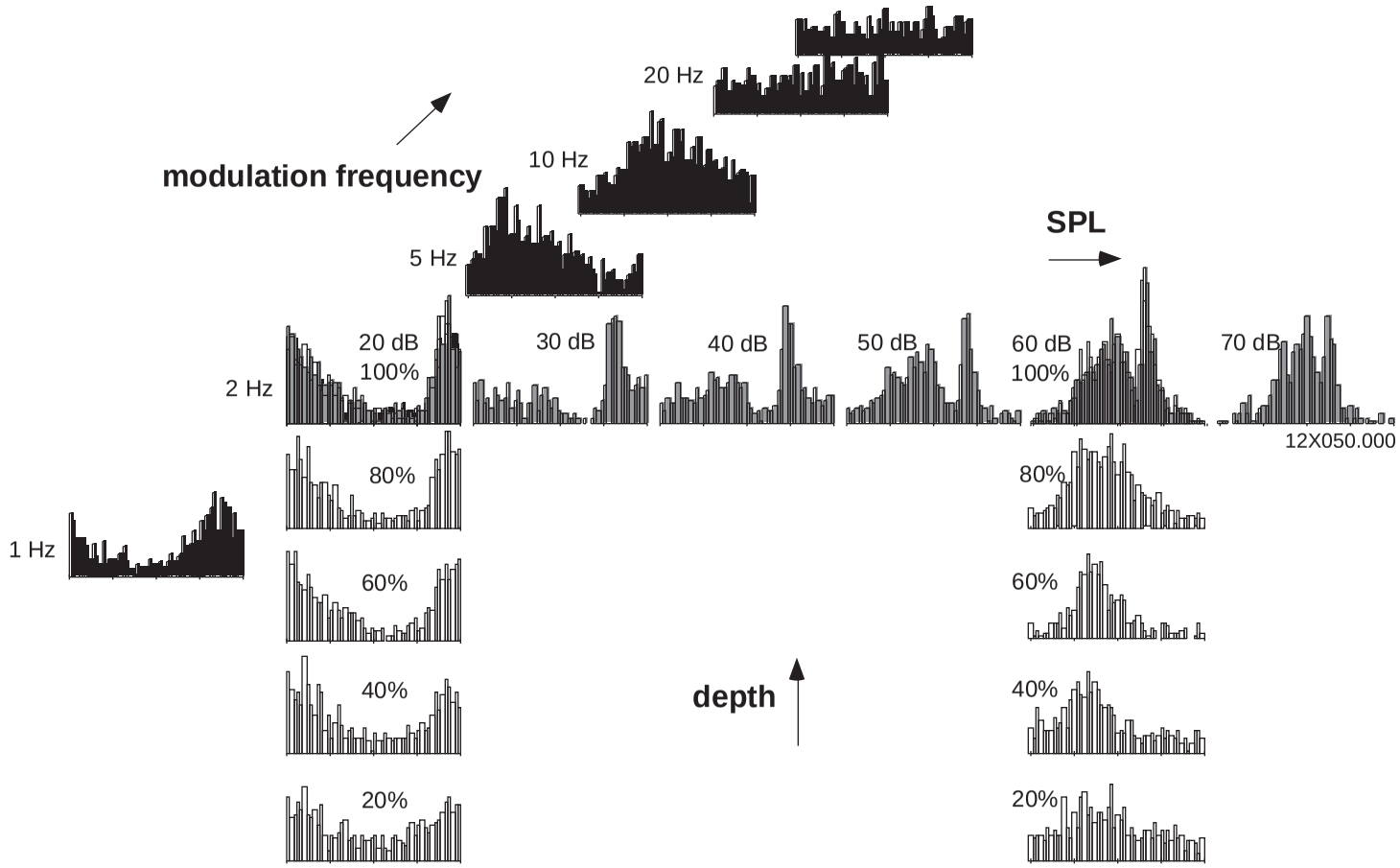

modulation frequency
20 Hz
10 Hz
5 Hz
2 Hz
1 Hz
SPL
20 dB
30 dB
40 dB
50 dB
60 dB
70 dB
100%
80%
60%
40%
20%
depth
12X050.000

Figure 4.7 Interaction of SAM frequency, depth, and level in shaping MPH profiles. Modulation frequency at 20 dB varies along the z-axis (black bars), SPL at 2 Hz varies along the x-axis (gray bars), and depth at 20dB and 60 dB varies along the y-axis (open bars). At the conjunction of the three series, the overlap of the MPHs shows the reproducability of MPH shape. Graded, parameter-dependent changes in MPH profile are discussed in the text.

phase reflects the group delay). Increases in carrier level result in the appearance of a bimodal MPH at 40 dB, as the cell begins to respond to decreases in SPL near the falling phase of the stimulus envelope, suggesting that it is tuned to SPLs below 40 dB. The second response peak, coincident with the rising phase of the envelope, becomes narrow as the SAM signal spends a smaller portion of the cycle within the neuron's preferred range of SPLs (see Figure 4.6), and the response peaks nearly converge at 70 dB. Comparison of the MPHs at 20 dB and 70 dB reveals that the response phases are completely inverted. Nevertheless, these dramatic changes are explicable if one assumes only that the cell is responding when the stimulus approaches its preferred sound levels within the modulation period. If, on the other hand, the cell is encoding only the 2 Hz modulation, the changes in the shape of the MPH are epiphenomena of no further computational consequence.

With the exception of the sudden appearance of the second response peak for 100% modulation at 60 dB – explained by the fact that the stimulus level approaches the best SPL from both above and below when the depth is sufficiently large (Figure 4.3) – the changes with modulation depth are graded and consistent. The reversal in the phase of the responses is apparent even at the lowest modulation depth, where the change in actual stimulus levels is quite small (less than +/- 2 dB at 20% depth). Under these conditions, the cell responds when the stimulus *approaches* the preferred SPL, since it does not actually reach that value at small depths.

4.3.3 Population characteristics

Given the problems with summary measures such as rate and synchrony BMFs, and the unwieldy nature of representations like Figure 4.7, how can one best characterize the population of cortical responses to SAM signals? As a step toward this goal, composite MPHs were constructed by summing all responses (n=61) from the left hemisphere of monkey X to a fully modulated SAM signal at 60 dB. Because MTFs (and the BMFs extracted from them) have been found to be independent of characteristic frequency in auditory cortex (Eggermont 1998, Liang et al. 2002, Schreiner and Urbas 1986, and the current study, described below) it seems reasonable to sum across cells studied at their respective BFs in this analysis. To prevent the most responsive cells from dominating the composite MPH, while preserving differences in discharge rate across modulation frequency within individual cells, MPHs were normalized by the maximum binned spike count across all modulation frequencies. These normalized MPHs were then averaged bin-by-bin across cells to produce the composite MPHs in Figure 4.8, left column.

Examination of these MPHs reveals that for modulations up to at least 10 Hz, there is a clear but broad peak in the composite MPH corresponding to the highest stimulus levels in the modulation period. The composite representation clearly obscures many of the finer details evident in the MPHs of individual neurons (see

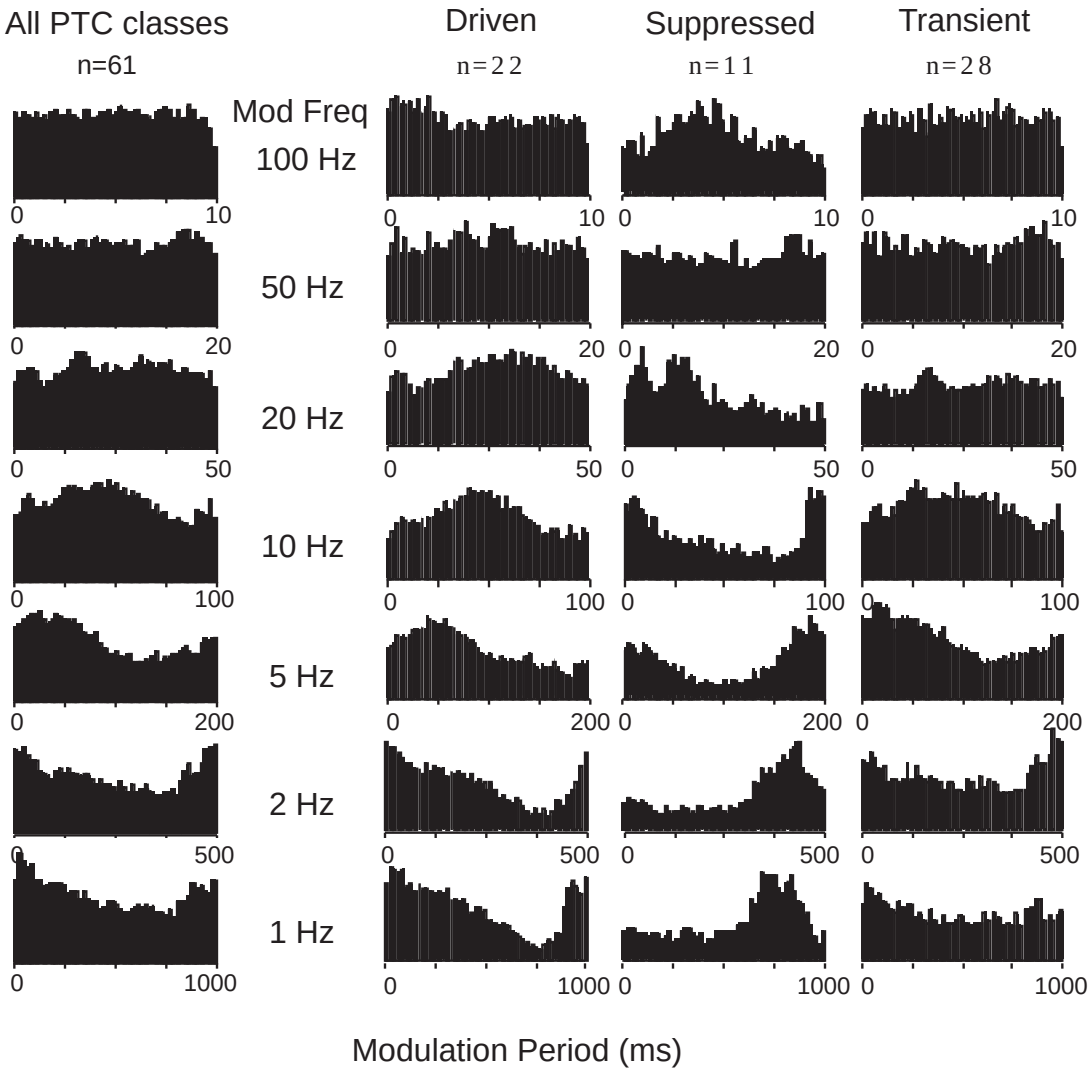


Figure 4.8 Composite MPHs showing phase coherence of cortical response to SAM across the population. MPHs in the left column are the sum of normalized MPHs at 7 modulation frequencies, from 61 neurons in a single hemisphere of one monkey. All data were gathered at 60 dB, at the respective BF of each cell. These MPHs are presented in sine phase. A broad response peak is evident up to 10 Hz, a surprising result given the heterogeneity of RLF shapes and latencies in cortex. In the three columns at the right, MPHs have been segregated on the basis of each cell's response to the pure-tone control (PTC, see text). Cells with a sustained, driven response to the PTC (n=22) show a clear response peak at the maximum of the phase cycle, and a clear trough at the mimimum (270°). Cells exhibiting sustained suppression (n=11) show a perfectly complimentary profile, with a peak at the envelope minimum. The remaining cells did not show a significant sustained response to the PTC, and are deemed "transient" because they most often responded at the onset of the tone only. In the subpopulations defined by PTC response, synchrony persists to at least 20 Hz.

previous figures), but it serves to illustrate an important point regarding the population representation of modulated signals: for a neuron receiving convergent inputs, the relative phase coherence of those inputs will determine the character of its response. Given the diversity of cortical responses to SAM stimuli, and differences in group delay among neurons, it is surprising that a crude summation of responses would leave any residual synchrony.

The population code for sound envelopes is undoubtedly more sophisticated than a normalized average across the population. The composite representation, as discussed so far, is equivalent to an extremely labor-intensive local field potential. A possible refinement of this mechanism is representation of envelopes by multiple, parallel population codes among different cell types – types that can be distinguished in single-unit data. This possibility was explored by categorizing neurons in terms of their response to the unmodulated pure-tone control (PTC). Those neurons whose firing rates differed significantly from the spontaneous rate were considered sustained responders, and were further subdivided into those driven (n=139, 30%) or suppressed (n=51, 11%) by the PTC. The remaining neurons were classified as "transient" (n=278, 59%), because the dominant characteristic was to respond at the onset of the tone; this response was indistinguishable from the spontaneous rate when averaged over 10 s. These classes

were not correlated with carrier frequency or level, nor did they differ in terms of the summary measures derived from MTFs.

The 3 columns at the right of Figure 4.8 show the composite MPHs subdivided by response class. For modulation frequencies <5 Hz, the differences in MPH profiles are particularly clear. Driven responses (n=22) follow the amplitude envelope quite well, showing a clear response minimum at the envelope minimum (these MPHs are *not* rotated as shown in Figure 4.1, thus the amplitude minimum falls at 270°, corresponding to the trough in the 1 Hz driven MPH). Conversely, suppressed responses (n=11) respond maximally at the envelope minimum. The transient class proved the most variable in terms of response phase at the lowest modulation frequencies, but becomes increasingly similar to the driven class above 5 Hz, at which point most MPHs tend to converge on a unimodal response clustered about the rising phase of the envelope. The population representation of SAM signals, while variable by response class, is relatively stereotyped within each class, and the composite MPHs are readily explicable in terms of instantaneous stimulus SPL. Because response class for a given carrier level (e.g. 60 dB) depends on each neuron's tuning for sound amplitude, the phase coherence of responses within a cortical population will depend on each neurons' tuning for sound amplitude, and not merely the distribution of their synchrony cutoffs.

4.4 Results II: SFM

4.4.1 Frequency modulation and the receptive field

In analogy to the introduction of SAM, Figure 4.1F presents an iconic diagram of how SFM stimuli may interact with a cell's receptive field. The middle set of arrows indicate modulation centered about the BF of the cell, traversing the most sensitive region of the response area (shaded in gray). The flanking arrows represent modulations centered around frequencies adjacent to the BF (they are offset on the SPL axis for clarity). As is the case for carrier level of SAM, center frequency of SFM interacts with the depth parameter (analogous to the length of the arrows) to determine whether a given modulation cycle includes one or two passes through the receptive field. The MPHs in panel 4.1E were collected at three carrier frequencies: the upper panel is centered below BF, the center at BF, and the bottom panel above BF. The stimulus icon shows the frequency modulation envelope, analogous to the amplitude modulation envelope in panel 4.1C, with gray tick marks indicating crossings of the carrier frequency. As would be predicted from the icon, modulation centered below the BF elicits a response only during the rising phase of the envelope, whereas modulation centered above BF elicits a response only during the falling phase. When centered at BF, each modulation cycle traverses the receptive field twice, producing a bimodal response that (if centered exactly on a symmetrical receptive field) has its peaks 180° apart (center panel). As with SAM data, the MPH is rotated to cosine phase, placing the

minimum of the modulation envelope at the center, and eliminating the false impression of three peaks caused by the circular axis in sine phase.

In general, response characteristics to SFM were much like those to SAM. The same carrier frequency and SPL that was used for SAM MTFs was used to measure SFM MTFs, to ensure the compatibility of measures derived from these functions within each cell. Depth was manipulated on-line and chosen to maximize both discharge rate and synchrony. By the same criteria established for SAM above, >99% of tested cells showed some synchrony or rate response to SFM, 97% showed synchrony at some modulation frequency, and only 11% of cells showed rate tuning for modulation frequency beyond the synchronized range. Table 4.2 tabulates the total number of single units from which SFM MTFs were obtained.

	Left Hemi	Right Hemi	Animal Subotal
Monkey X			X:
AI	31	108	139
R	10	51	61
Belt	9	20	29
Monkey Z			Z:
AI	39	19	58
R	10	37	47
Belt	4	6	10
Hemi Subtotals			**Grand Totals**
AI	70	127	**197 in AI**
R	20	88	**108 in R**
Belt	13	26	39
		Core Total	**305**

Table 4.2 Number of SFM MTFs from single units in four hemispheres

As with SAM, each of the parameters defining the SFM signal can influence the resulting MPH. The following sections will explore effects of carrier frequency, modulation depth, and modulation frequency, revealing coding mechanisms for instantaneous frequency comparable to those encoding the amplitude envelope.

4.4.2 Coding in single neurons: dependence of MPH shape on SFM parameters

4.4.2.1 Carrier frequency and synchrony metrics

The rasters of Figure 4.9A depict the timing of spikes in response to two 10 s trials of SFM at 2 Hz, spanning a range of carrier frequencies about this cell's BF of 2.9 kHz. Above and below BF (e.g. 2.5 and 3.3 kHz), the stimulus enters the receptive field only once per modulation period, causing 20 discrete bursts of spikes. As the carrier approaches BF, these bursts bifurcate into 40 responses representing two passes through the receptive field per modulation cycle. Tracing one response in the raster vertically highlights the shift in response phase with carrier, which produces the overall "triangle wave" pattern in the raster array. Panel 4.9B shows these same data binned into MPH form.

As would be predicted in any frequency-tuned neuron, responses on the edge of the receptive field are weak, but the effect of carrier frequency from 2.4 to 3.3 kHz seems to manifest as a shift in response phase, and the transformation from

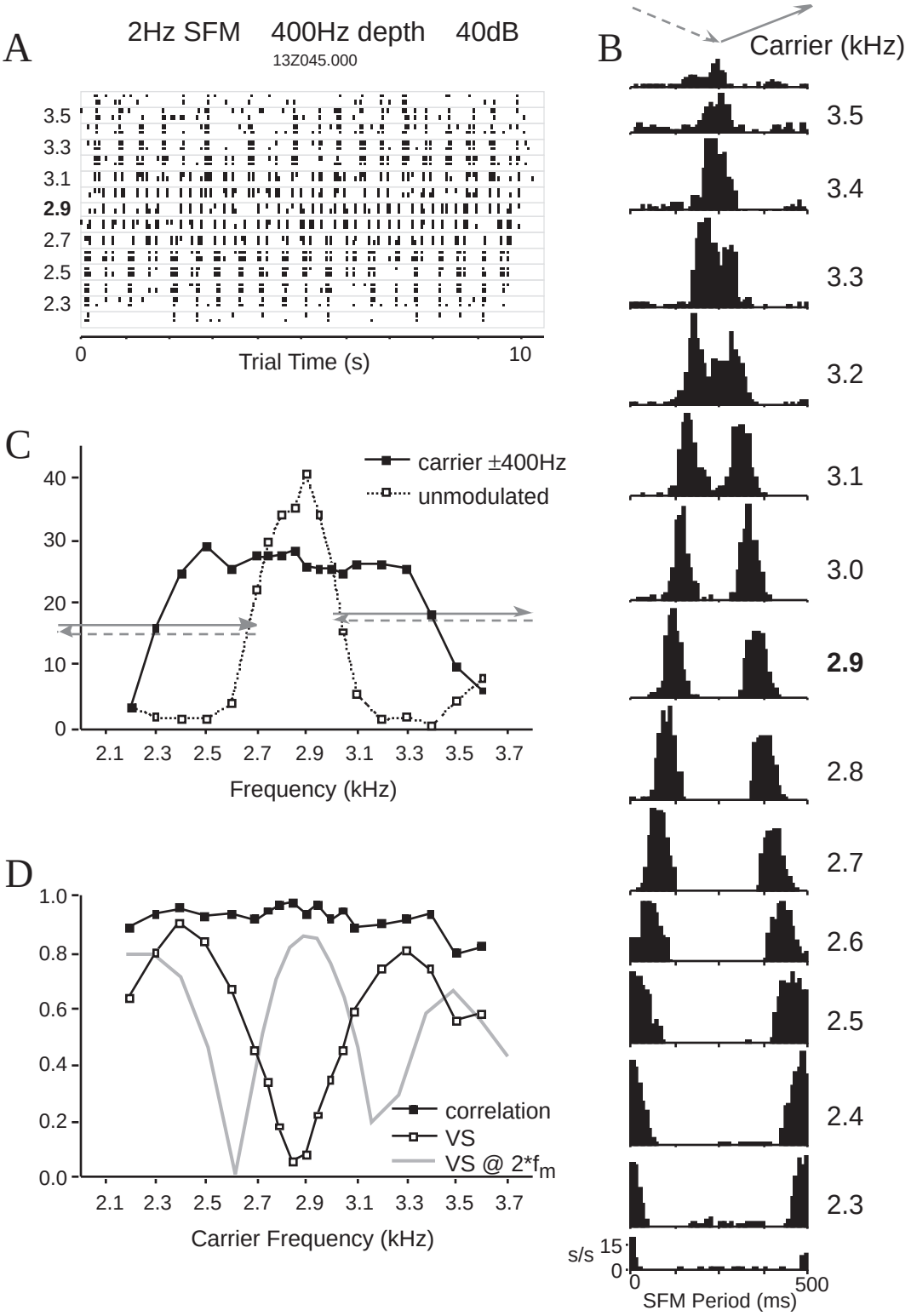

A
2Hz SFM 400Hz depth 40dB
13Z045.000
3.5
3.3
3.1
2.9
2.7
2.5
2.3
0
10
Trial Time (s)
B
Carrier (kHz)
3.5
3.4
3.3
3.2
3.1
3.0
2.9
2.8
2.7
2.6
2.5
2.4
2.3
s/s
15
0
0
500
SFM Period (ms)
C
carrier ±400Hz
unmodulated
40
30
20
10
0
2.1 2.3 2.5 2.7 2.9 3.1 3.3 3.5 3.7
Frequency (kHz)
D
1.0
0.8
0.6
0.4
0.2
0.0
correlation
VS
VS @ 2*fm
2.1 2.3 2.5 2.7 2.9 3.1 3.3 3.5 3.7
Carrier Frequency (kHz)

Figure 4.9 Effects of SFM carrier frequency. **A:** Spike time rasters in response to two 10 s trials of 2 Hz SFM (at 400 Hz depth and 40 dB SPL) across a range of carrier frequencies. On the fringes of the response area, responses are unimodal; as carrier frequency approaches BF, the response bifurcates and shifts in phase. **B:** Spikes from panel A binned into MPHs, showing the shift from a unimodal response at the rising phase of modulation below BF (2.3-2.5 kHz; response appears bimodal as an artifact of the circular axis) to a bimodal response separated by 180° at BF (2.9 kHz). Stimulus icon over the uppermost MPH indicates the risng and falling phases of the modulation. When carrier frequency is above BF, the response shifts back to a unimodal peak, now centered on the envelope minimum. **C:** Frequency tuning functions measured from the unmodulated control (open squares) and the carrier frequency of SFM with 400 Hz depth (filled squares). Predicatably, SFM spike rate rises when the carrier moves within 400 Hz of the edge of the receptive field (as defined by the control response), but remains surprisingly flat across the 800 Hz range of center frequencies spanning BF. Even when the carrier is exactly equal to the BF, the overall spike rate is below that elicited by the control (which drives this cell in a sustained manner). **D:** MPH shape can confound standard metrics of response synchrony, such as VS (open squares), which is maximal when the response is unimodal and falls almost to zero as the response peaks shift to 180° out of phase. Folding the modulation period again, and measuring VS at $2*f_m$ (gray line), is effective when the peaks are 180° out of phase, but fails (for the same reason) when the peaks are 90° apart. The correlation metric makes no assumptions about the form of temporal structure in the histogram, and reliably detects strong fidelity in the MPH across the range of tested carriers.

unimodality (for $f_c \leq 2.5$ kHz and $f_c \geq 3.3$ kHz) to bimodality near BF. This impression is confirmed by the frequency tuning functions in panel 4.9 C, derived from the unmodulated pure-tone control (open squares) and the SFM signal modulated with 400 Hz depth (filled squares). At the fringes, neither stimulus enters the receptive field; when the SFM center frequency is 400 Hz below BF, the extreme of the modulation cycle (depicted as the gray arrows) enters the receptive field as defined by the PTC, and the modulated signal elicits a constant overall spike rate from 2.5 to 3.3 kHz – a range of frequencies spanning 400 Hz on either side of the BF, which (not coincidentally) is the depth of the SFM modulation. Even when the f_c=BF, the total spike rate in response to SFM is lower than that elicited by the PTC, presumably because the 400 Hz modulation depth places the stimulus outside of the receptive field for some portion of the period. This behavior is probably unique to the sustained, driven class of PTC responders; cells that adapt to the PTC tend to show an elevated spike rate in response to temporal modulation, despite the time the stimulus spends outside the receptive field.

Synchrony is detected by the VS metric (black line, open squares) at the edges of the response area, where MPHs are unimodal, but decays to zero as the response peaks shift to 180° out of phase – at this point, because VS is based on vector addition, they cancel one another. A common technique to circumvent this problem is to fold the MPH and measure VS at twice the modulation frequency (gray line),

which is effective so long as the carrier is exactly equal to the BF; just as VS fails when the response peaks are 180° apart, this metric falls to zero when the peaks are 90° apart.

An alternative metric was developed specifically for these SFM responses, to measure temporal structure with no assumptions about the number or position of response peaks. Spikes from the first and second half of the 10 s trial are binned into separate MPHs (52 bins, as shown in all figures), and these vectors of bin counts are correlated with one another. The resulting correlation coefficient reflects the similarity in bin values between these two MPHs, presumed to result from temporal structure imparted by the modulating stimulus. The only prior assumption is knowledge of the modulation frequency, used to create the MPH, although the metric can be applied to PSTHs without folding on modulation frequency (in Chapter 3, it was used to evaluate responses to pure tones and noise for the construction of the standard stimulus activation maps). As seen in panel 3.9D, the correlation metric (black line, filled squares) reliably detects the temporal structure in the MPH across all center frequencies. Bootstrap simulations on randomly distributed spikes indicate that correlations >0.4 generally can be considered significant at the $p<.001$ level, but the significance is dependent on spike count – for this reason, significance of correlation is not as neatly defined as VS via the Rayleigh statistic.

4.4.2.2 Effects of modulation depth

Increasing modulation depth sharpens the temporal coherence of the cortical response to SFM, as can be seen in the rasters of Figure 4.10A (data are from the same cell depicted in the previous figure). The MPHs in panel B clearly register the modulation of the signal at a depth of 100 Hz, and these peaks sharpen with increasing depth up to 1000 Hz (at which point the depth is about 1/3 of the carrier frequency). Beyond the threshold of modulation, total spike rate (panel C) drops and the peaks sharpen as the total time spent in the receptive field decreases. This sharpening is clearly registered by the correlation metric (panel D), but not detected by VS unless the period is doubled.

Because modulation period is held constant, the *rate* at which instantaneous frequency changes is confounded with the total *time* spent in the receptive field. An interesting experiment for this cell would have been to define diagonals in modulation rate/depth space that hold one of these parameters constant while changing the other. As with instantaneous level coding of SAM, the rate of change of instantaneous frequency may trigger temporal mechanisms that emphasize these envelope features in a nonlinear manner.

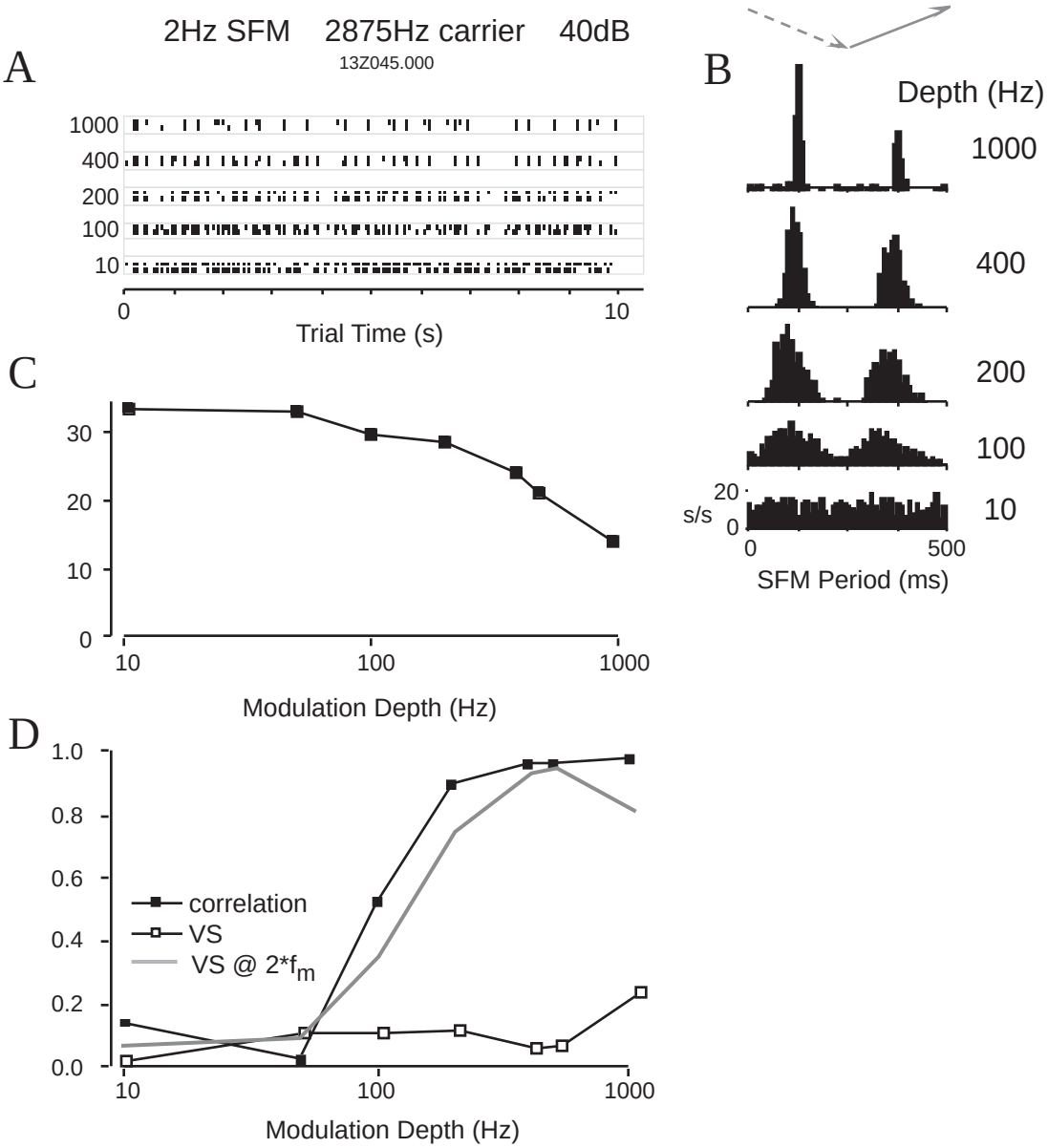


Figure 4.10 Effects of SFM depth, in the same cell depicted in previous figure. **A:** Rasters in response to 2 Hz SFM, showing temporal precision increasing with depth. **B:** MPHs clearly show the emergence of synchrony at 100 Hz depth (approx. 3% of the carrier frequency). The response peaks sharpen with increasing depth, accompanied by a drop in the overall discharge rate (panel **C**). **D:** Synchrony-depth functions indicate increasing temporal precision with depth, as measured by correlation (filled squares) or VS@2*fm (gray line).

4.4.2.3 Modulation frequency and MTF consistency

The efficacy of SFM at distributing spikes within the modulation period is apparent in the rasters of Figure 4.11A, in which synchrony in the response is visible over a wide range of modulation frequencies. From 0.1-5 Hz, MPHs are indistinguishable and symmetrical, demonstrating an equivalent response during the rising or falling phase of the modulation, which is centered on BF with a depth of 400 Hz (Figure 4.11B, same cell as previous two figures). Between 5 and 10 Hz, an asymmetry begins to emerge as the falling phase comes to dominate the response, which becomes unimodal by 40 Hz. (Between 10 and 30 Hz, the latency of the response becomes large enough relative to the modulation period to "wrap" the response around, so the dominant peak appears on the right side of the axis.) A small vestige of synchrony persists at 100 Hz, but by this point the time spent in the receptive field is so brief (or, by the same token, the rate of frequency change is so fast) that the cell's overall discharge rate drops dramatically. This is evident in the rate MTF of panel 4.11C, which shows a rate BMF of 30 Hz and a steep cutoff beyond that point. Because depth is held constant, the stimulus is within the receptive field for the same *proportion* of the cycle at each modulation frequency; furthermore, because the total trial length is held constant, the *total* time spent in the receptive field is also constant. The drop in discharge rate must therefore be attributed to the speed of instantaneous frequency change "outrunning" the temporal integration window of the neuron. Thus the rate MTF should not be taken

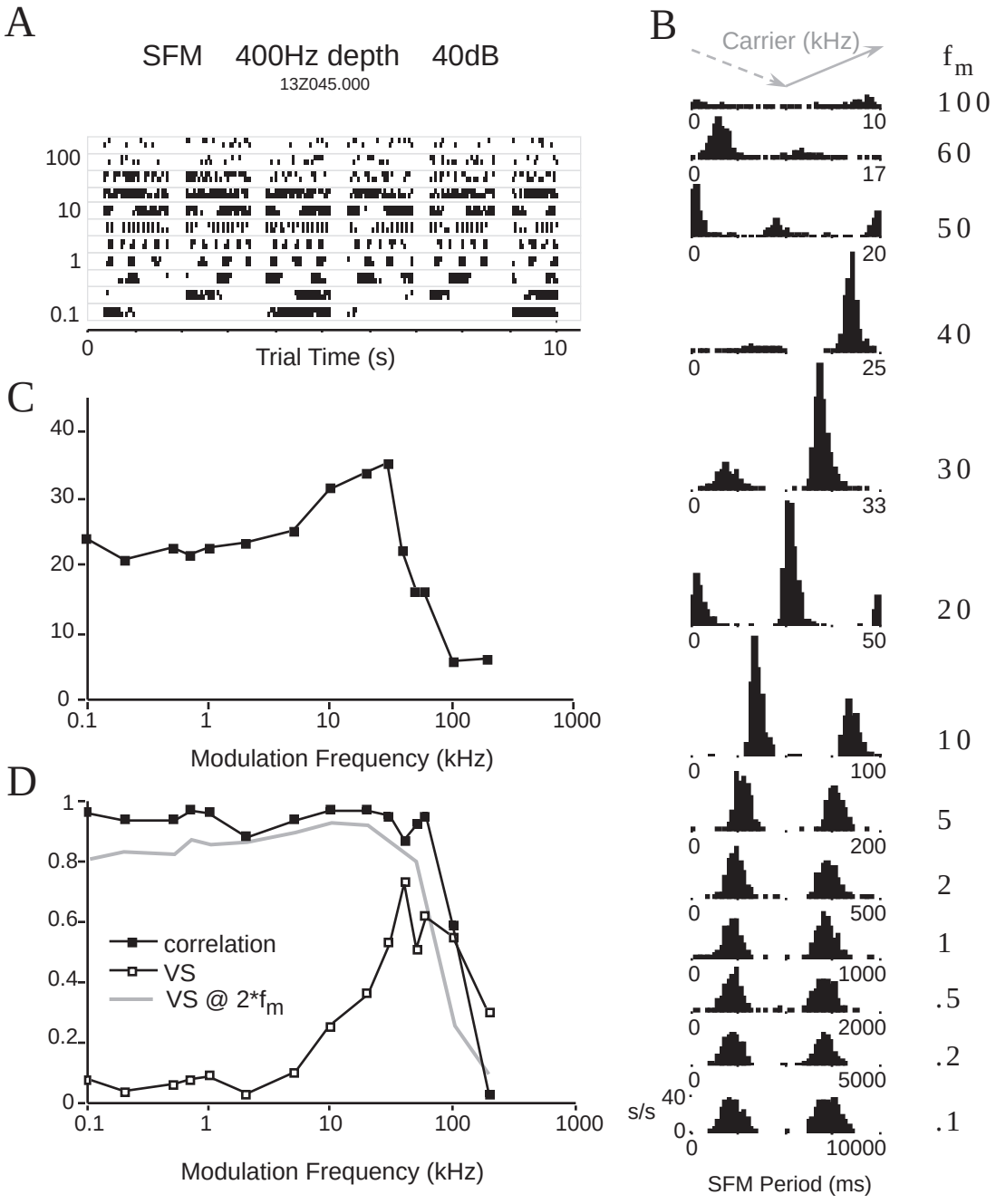

A
SFM 400Hz depth 40dB
13Z045.000
100
10
1
0.1
0
10
Trial Time (s)
B
Carrier (kHz)
f_m
100
60
50
40
30
20
10
5
2
1
.5
.2
.1
0
10
17
20
25
33
50
100
200
500
1000
2000
5000
10000
s/s
40
0
SFM Period (ms)
C
40
30
20
10
0
0.1
1
10
100
1000
Modulation Frequency (kHz)
D
1
0.8
0.6
0.4
0.2
0
correlation
VS
VS @ 2*f_m
0.1
1
10
100
1000
Modulation Frequency (kHz)

Figure 4.11 Effects of SFM frequency, in the same cell depicted in previous two figures. **A:** Rasters in response to SFM at 400 Hz depth from 0.1 to 200 Hz modulation frequency. **B:** MPHs below 5 Hz are symmetrical and unchanging across frequency. Above 5 Hz, the response peak corresponding to the falling phase comes to dominate the response. Between 10 and 30 Hz, the latency of the response becomes large enough relative to the period length to "wrap" the response around the axis; for this reason, the dominant peak shifts to the right, but still corresponds to the falling phase of the frequency envelope. At 40-60 Hz the response is nearly unimodal, and the last vesitge of synchrony appears at 100 Hz. **C:** The rate MTF reaches its peak at 30 Hz, with a sharp cutoff above the rate BMF. **D:** Synchrony MTFs measured with correlation or VS@2*fm are perfectly flat lowpass functions with cutoff frequencies of 100 Hz; VS reaches its peak at 40 Hz, the nominal synchrony BMF, because this is where a single mode comes to dominate the MPH.

to indicate that this cell is "tuned" for 30 or 40 Hz – it clearly tracks the envelope over a much wider range of modulation frequencies – but it does reflect the interaction of this neuron with the dynamics of its input.

The correlation metric captures the temporal structure in the MPH up to and including the synchrony cutoff at 100 Hz, and because the modulation is centered around BF, the double-frequency VS metric is effective as well. Standard VS produces a sharply bandpass synchrony MTF with a BMF of 40 Hz, the frequency that produced the most unimodal response. As was illustrated for SAM in Figure 4.4, the peak in the VS function actually coincides with the breakdown of the cell's ability to encode the instantaneous value of the stimulus envelope. (When applied to the SAM data in Figure 4.4, the correlation metric is lowpass, with a consistent value >0.9 up to the synchrony cutoff of 20 Hz.)

The interaction of modulation frequency and depth can be seen in the array of MPHs in Figure 4.12C. These parameters define a response area in this 2-dimensional SFM space, such that the upper limit of effective modulation depth falls with increasing modulation frequency (hence the flat MPHs at the upper right of the array). The MTFs and modulation depth functions in the upper panels show the interaction of these parameters in terms of spike rate (panel A) and spikes/cycle (panel B). The family of MTFs in 4.12A shows a consistent rate cutoff beyond the

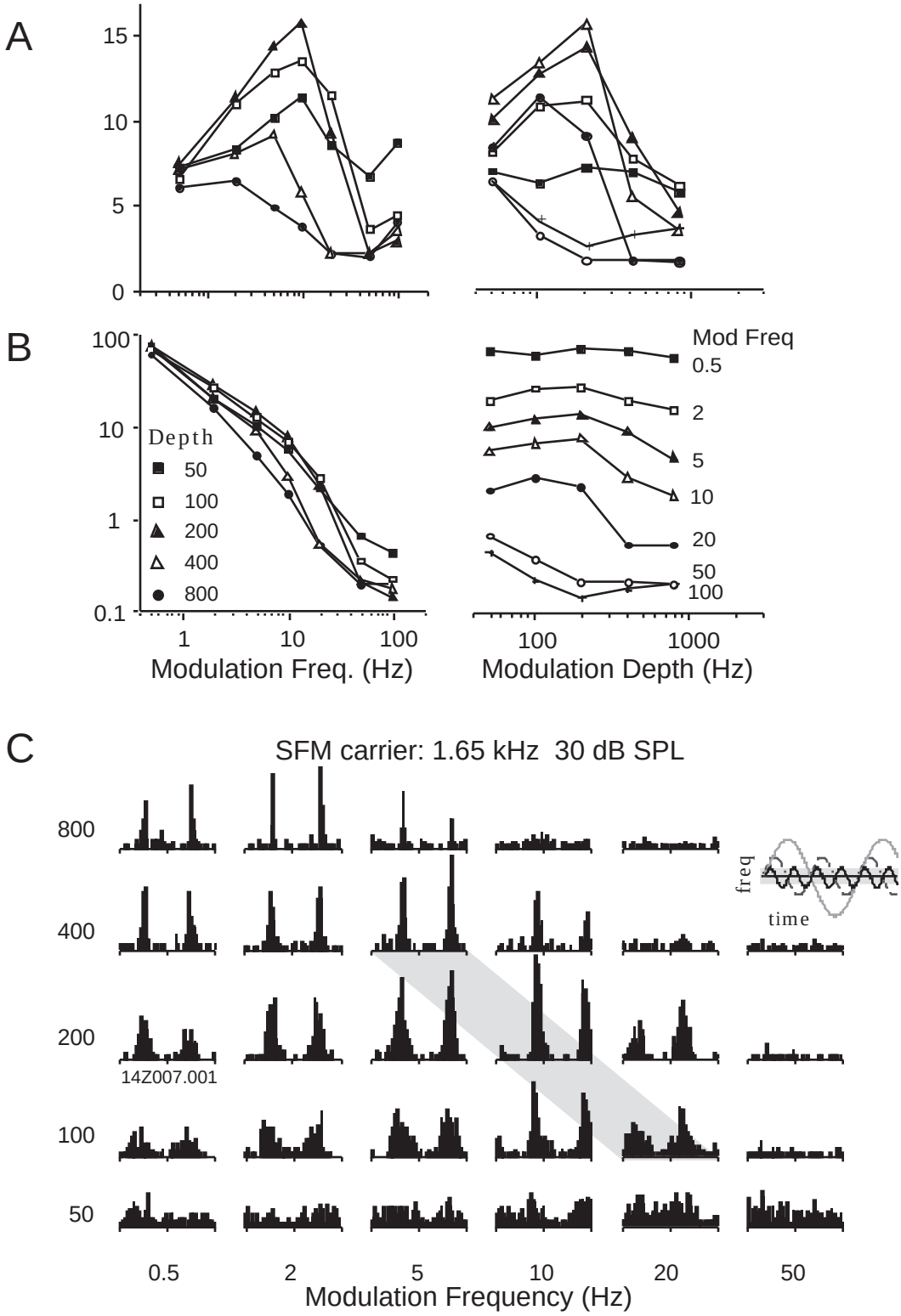

A
15
10
5
0
B
100
10
1
0.1
Depth
50
100
200
400
800
1
10
100
Modulation Freq. (Hz)
Mod Freq
0.5
2
5
10
20
50
100
100
1000
Modulation Depth (Hz)
C
SFM carrier: 1.65 kHz 30 dB SPL
800
400
200
100
50
14Z007.001
freq
time
0.5
2
5
10
20
50
Modulation Frequency (Hz)

Figure 4.12 Interaction of SFM frequency and depth. **A:** Rate MTFs (left panel) are robust across a range of modulation depths, as are rate vs. depth functions (right panel) across a range of modulation frequencies. At the extremes of modulation rate or depth, the functions collapse (see text). **B:** Average spikes/cycle decreases as a power function with modulation frequency, regardless of depth (left panel). At low modulation frequencies, spikes/cycle is independent of modulation depth (right panel), and these functions become increasingly lowpass up to 20 Hz modulation frequency. Beyond this frequency, responses are very weak regardless of depth. **C:** An array of MPHs at increasing modulation frequency (from left to right) and modulation depth (from bottom to top). The gray shading connects MPHs along a diagonal in stimulus space such that when modulation frequency halves, modulation depth doubles. This maintains a more or less constant rate of frequency change through the receptive field, as the icon inset at the upper right of the panel is intended to show (see text).

10 Hz BMF at low to moderate depths (50-200 Hz), though the peak and cutoff shift lower at the highest tested depths (400 and 800 Hz). These data are re-plotted in the right panel to show that the most effective depth is consistently 200 Hz at moderate modulation frequencies (2-10 Hz), but rates drop off markedly at higher modulation frequencies of 50-100 Hz. Spikes/cycle follows a power function with modulation frequency, as demonstrated for SAM in Figure 4.2, regardless of modulation depth. When plotted in terms of modulation depth, spikes/cycle functions follow a graded transition from flat at low rates (0.5-2 Hz), to bandpass at moderate rates (10-20 Hz, where high depths can be seen to take effect in the MPHs of panel C), to lowpass at the highest rates were synchrony and firing rate are very low overall.

The gray region in panel C connects three MPHs along a diagonal in this depth/frequency space, such that each move up the diagonal corresponds to a doubling of modulation depth, and a halving of modulation rate. The icon inset at the upper right of the panel compares the frequency envelopes of these stimuli, relative to a hypothetical frequency response area shaded in gray. When these curves align, their paths are parallel, suggesting that each traverses the receptive field at the same rate. In theory, a linear model would predict that these stimuli elicit identical responses, though in fact the response can be seen to degrade at the upper left and lower right of the MPH array.

Why does the coding of instantaneous frequency break down at these extremes? If one imagines increasing SFM depth but holding the period constant (i.e. stretching a sine wave from the icon vertically), the slope of the function through the receptive field will increase, and the total time in the receptive field will decrease, possibly becoming too fast for the cell to integrate the spectral content of the signal. Conversely, if the frequency is increased while holding depth constant (i.e., squeezing a sine wave from the icon horizontally) the rate at which the stimulus traverses the receptive field increases, and the time between each traversal decreases, leaving the cell less time to recover between cycles. In other words, looking along the diagonal, the decay in the representation moving up and to the left is due to *integration*; the decay moving down and to the right is due to *adaptation*. Between these extremes, the temporal integration properties of the cell are best matched to the stimulus envelope, and coding by instantaneous discharge rate attains its highest fidelity.

4.5 Comparing SAM and SFM

Although the MTF is not an all-encompassing descriptor of what a cell may encode, it can provide a useful index of a cell's temporal characteristics – the modulation frequency at which it can best integrate the signal passing through its

receptive field, or the temporal limit beyond which it can no longer synchronize its discharges to features of the modulation envelope.

For every cell from which both SAM and SFM transfer functions were collected (n=288), rate and synchrony BMFs were determined. A rate BMF was initially assigned as the modulation frequency eliciting the highest spike rate; a *t*-test was then applied to identify other frequencies at which the rate response was statistically equivalent. The geometric mean of these values was taken as the rate BMF (Liang et al. 2002), and these values are plotted in Figure 4.13A. Rate BMFs for SAM and SFM are weakly but significantly correlated (ANOVA, $p<.001$, $r^2=0.15$), and despite the appearance that these points are biased to lie above the line of unity, they are indistinguishable by a paired *t*-test ($p=0.47$).

To determine the upper limit of temporal synchrony, a point was interpolated from the synchrony MTF by fitting a line between the Rayleigh statistic at the highest modulation frequency showing significant synchrony, and the value at the next tested frequency. The point where this line crossed the Rayleigh significance threshold was taken as the synchrony cutoff (Liang et al. 2002). For SAM data, the VS metric was used, whereas for SFM, the procedure was run on VS at f_m and $2*f_m$, and the higher value used. Because MTFs were always taken with the carrier frequency matched to the cell's BF, and MPHs usually merged into unimodal

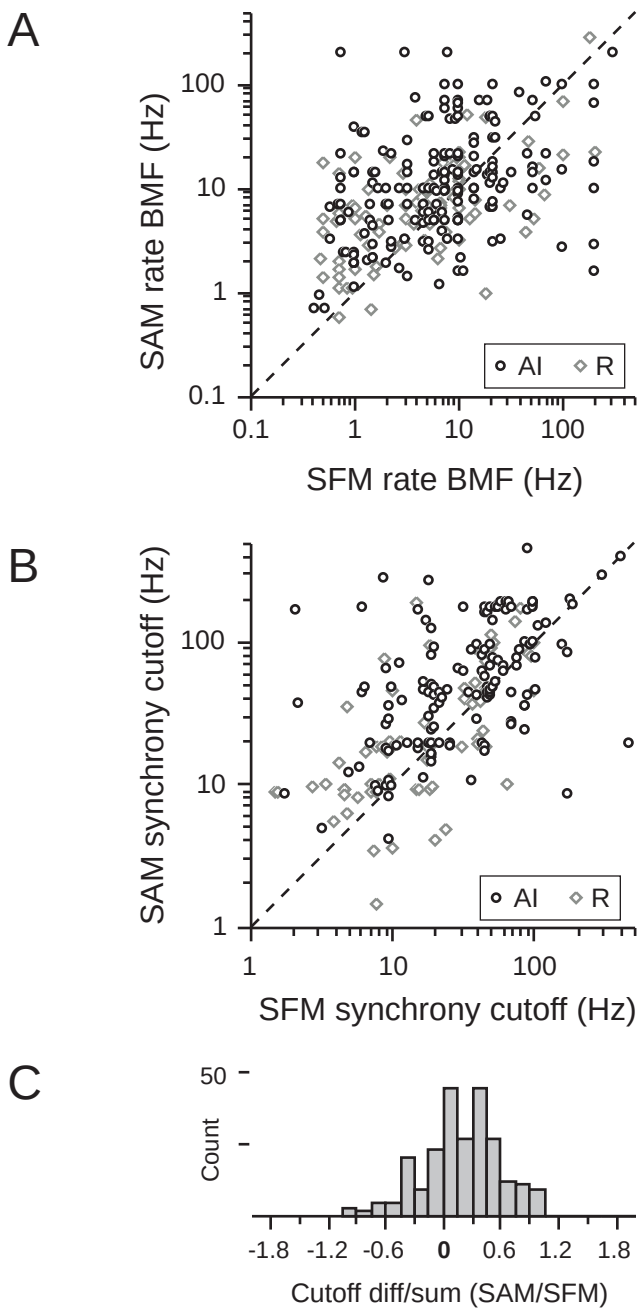

A
SAM rate BMF (Hz)
SFM rate BMF (Hz)
0.1
1
10
100
AI
R
B
SAM synchrony cutoff (Hz)
SFM synchrony cutoff (Hz)
1
10
100
AI
R
C
Count
50
-1.8
-1.2
-0.6
0
0.6
1.2
1.8
Cutoff diff/sum (SAM/SFM)

Figure 4.13 Comparison of response characteristics derived from MTFs for SAM and SFM. **A:** Rate BMFs for SAM and SFM (n=288) are correlated and not significantly different by a paired t-test. Black circles mark cells in AI (n=186) and gray diamonds mark cells in R (n=102). Dashed line is unity. **B:** Interpolated cutoffs from synchrony MTFs (n=216) are significantly higher for SAM than for SFM (see text; AI n=142, R n=74). Conventions as in panel A. **C:** Distribution of difference/sum values for cutoff frequencies for SAM and SFM (AI and R combined). The shift of the distribution to the right of zero confirms that cutoff frequencies measured with SAM are higher than those measured with SFM.

responses above 5-10 Hz, one of these two metrics should have provided a reliable estimate of the limits of synchrony. These cutoffs (n=216) are plotted in Figure 4.13B, and like the rate BMFs they are correlated between SAM and SFM (ANOVA, $p<.001$, $r^2=0.26$), but SAM cutoffs are significantly higher (paired *t*-test, $p<.001$). To illustrate the extent of this disparity, a cutoff difference/sum metric was computed for each cell as

$$(\text{SAM cutoff} - \text{SFM cutoff}) / (\text{SAM cutoff} + \text{SFM cutoff})$$

and the distribution of these values plotted in panel C. The distribution is significantly shifted to the right of zero ($p<.001$ by *t*-test or Wilcoxin), indicating a higher cutoff value for the SAM stimulus. The most obvious explanation for this is that the SFM stimulus, as typically applied to measure MTFs, passes through the receptive field twice per cycle; in this sense, the cell is being asked to respond at twice the duty cycle of the SAM stimulus, which typically was positioned to produce a unimodal response. (This disparity can be attributed to the compressive nature of the dB scale as demonstrated in Figure 4.3 – if the best SPL of the cell is chosen for carrier level, a SAM stimulus at 100% modulation is only 6 dB louder at its maximum, which will only produce a bimodal MPH in the most nonmonotonic of neurons). A SAM cutoff twice as high as its corresponding SFM cutoff will produce a difference/sum of 0.33; the actual mean of the distribution in panel C is 0.21, suggesting that SAM cutoffs, across the population, are not quite twice as high as SFM cutoffs. Factoring in the intrinsic differences between the stimuli, and

the similarity of the rate BMFs, it appears that both stimuli reveal underlying neural processing on equivalent time scales.

4.6 Comparing AI and R

The marked differences in response latency between the two fields of core auditory cortex demonstrated in Chapter 3 would predict a similar disparity in temporal processing of dynamic stimuli like SAM and SFM. The scatter plots in the previous figures included neurons from both fields, using different symbols for AI (black circles) and R (gray diamonds). The rate BMF values for R (n=102) appear to be interspersed among those for AI (n=186), but with a tendency to cluster at lower values. This is confirmed by examining the medians of the two populations. The median rate BMFs for SAM are 10 in AI, 6 in R ($p<.0001$, Wilcoxin signed-rank), and for SFM are 7 in AI, 4.7 in R ($p=.003$).

A similar trend is evident in Figure 4.13B, and is also borne out by statistics: the median synchrony cutoffs for SAM are 46 in AI, 10.3 in R ($p<.0001$, Wilcoxin); for SFM they are 31 in AI, 17.8 in R ($p=.0003$). When the difference/sum distribution in 4.13C is re-evaluated for each field, the mean values for AI and R are 0.26 and 0.14, respectively – the mean for R is significantly greater than 0 ($p=.002$), but significantly less than that of AI ($p=.015$, Wilcoxin). In other words, cutoffs for SAM are higher than those for SFM in R, but the disparity is not as

great as that in AI. Thus it appears R may have a "selective impairment" for processing SAM relative to AI, perhaps because it filters amplitude envelopes on a longer time scale than frequency envelopes.

To compare the temporal differences between AI and R across the population, cumulative distributions of cutoff frequency in each field are plotted in Figure 4.14 for SAM (panel A) and SFM (panel B). For both stimuli, values for R (gray line) rise more steeply and saturate at a lower frequency, indicating the relative shift of this distribution to lower values. The temporal acuity of R is lower than that of AI for both stimuli, but as suggested in the above analysis, R is more like AI for SFM stimuli than for SAM.

The difference in temporal acuity between AI and R demonstrated with SAM and SFM is consistent with the longer minimum response latencies revealed in Chapter 3. Cutoff frequencies for SAM and SFM were inversely correlated with minimum latency to tones (ANOVA, $p<.0001$) with a slope of –1.8 Hz/ms for SAM cutoffs (for the full population, regardless of cortical field). The 13 ms difference in median latency between AI and R would predict a difference in cutoff frequencies of 23 Hz; the actual difference in the medians is 36 Hz, implying that temporal filtering mechanisms in R may be more sensitive to SAM than to the sharper attack of pure tones. The slope of the effect for SFM is –1 Hz/ms, predicting a difference

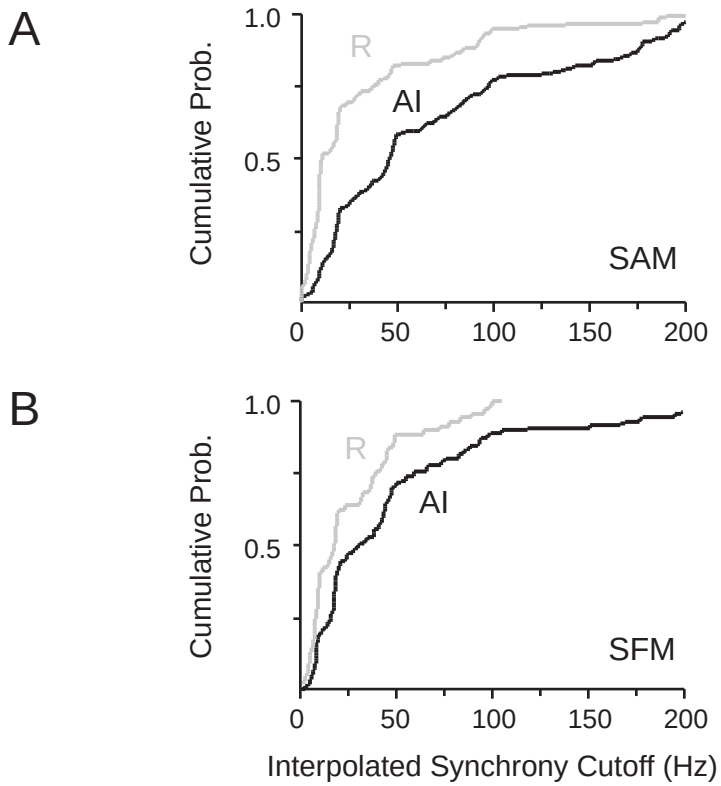


Figure 4.14 Cumulative distributions of interpolated synchrony cutoffs in AI (black line) and R (gray line) measured with SAM (**A**) and SFM (**B**). Distributions in R rise more steeply and saturate at lower frequencies, confirming that neurons in R lose synchrony at lower modulation frequencies than neurons in AI.

of 13 Hz – exactly the difference measured between AI and R. Thus is seems that the latency and SFM cutoff difference in R are manifestations of the same temporal disparity between the core fields. These differences are summarized in Table 4.3.

	SAM	SFM
Median cutoff, AI	46 Hz	31 Hz
Median cutoff, R	10.3 Hz	17.8 Hz
p-value b/w AI and R	<.0001	.0003
% of difference predicted by latency	64%	100%

Table 4.3: Summary of synchrony cutoff frequencies between AI and R

The preceding population analyses comparing the two fields were based on single points drawn from the MTF of each cell: one modulation frequency eliciting the highest spike rate, and one frequency marking the upper limit of synchrony to the modulation envelope. These distributions are not illustrative of how the cortical population *as a whole* encodes the envelope across frequency. From all MTFs collected using either stimulus, a "population MTF" was constructed by evaluating the proportion of cells that were responding at each of the 9 most commonly tested modulation frequencies. Rate responses were evaluated as a significantly elevated discharge rate relative to either the spontaneous rate of the cell, or the response to the pure-tone control. Synchrony was evaluated using the Rayleigh statistic on VS and/or VS at $2*f_m$.

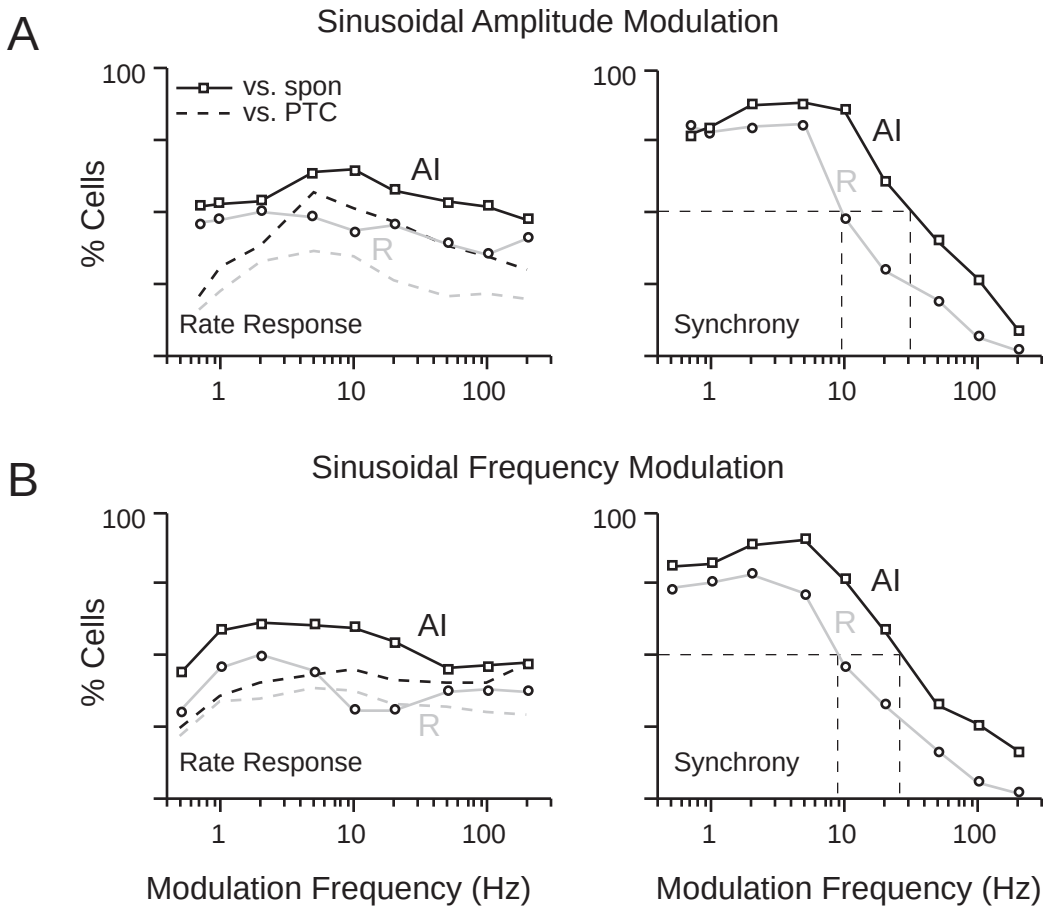


Figure 4.15 Population MTFs. **A:** Percentage of cells in AI (black lines) and R (gray lines) showing significant changes in discharge rate relative to spontaneous rate (solid lines) and the pure tone control (dashed lines). Relative to spontaneous discharge, about 50% of cells exhibit a rate response at any given modulation frequency; the curve measured relative to the pure tone control is more tuned, with a broad peak centerd at 5 Hz. Curves in R follow the same general shape as those in AI, but the percentage of cells responding is consistently lower across modulatin frequency. Analogous population synchrony MTFs in the right panel plot the percentage of cells showing significant synchrony (by VS) at each modulation frequency. In both AI and R, >75% of cells fire in synchrony to the modulation envelope up to 5 Hz; beyond this point, populaton synchrony degrades in R, and synchrony in AI degrades with a similar slope beyond 10 Hz. Dashed lines mark the median point at which 50% of the population responds in synchrony (30 Hz in AI, 10 Hz in R). **B:** The same analysis for SFM responses. The rate response relative to spontaneous discharge is very weakly bandpass with a peak around 2 Hz, but as for SAM, the functions are very flat overall (even more so relative to the pure tone control). The population synchrony MTFs in the right panel are nearly identical to those measured with SAM.

Population MTFs are plotted in Figure 4.15 for SAM and SFM, and confirm the similarity in the population responses to the two types of modulation. Perhaps surprisingly, no modulation frequency elicits a rate response in more than about 60% of cells (left panels), and the rate response curves for SAM are almost entirely flat when measured relative to the spontaneous rate (panel A, solid lines). Rates relative to the unmodulated control (dashed lines) exhibit a peak at 5 Hz in AI, but both curves in R are flatter and lower overall than their AI counterparts. The SFM population MTFs (panel B) look much the same: they are generally flat and weakly tuned, with a slight peak at 2 Hz in R relative to the spontaneous rate but no tuning whatsoever relative to the pure tone control. By the criterion established by Liang et al. (2002), approximately 75% of neurons are bandpass tuned in rate (true for SAM and SFM, with no significant difference between fields), so the lack of structure in the population MTF must derive from heterogeneity in the peaks of these functions.

Population synchrony MTFs (right panels of Figure 4.15) were consistently lowpass in shape, and reflected the diminished temporal acuity of the rostral field. At least 75% of neurons synchronized their discharges to some component of the envelope for both SAM and SFM in AI, up to a modulation frequency of 10 Hz; beyond this point, synchrony declines steadily among the population. The drop occurs at lower frequencies among the population of cells in R, as would be

predicted from the cutoffs in the previous figure. The median point, at which only half the population retains synchrony to the envelope, is approximately 30 Hz in AI, and 10 Hz in R, for both SAM and SFM.

4.7 Discussion

4.7.1 Envelope coding across modulation frequency

Studies of the neural processing of SAM and SFM signals generally have focused on the representation of modulation frequency, embodied in the MTF. This chapter takes a complimentary approach to the question by focusing instead on the representation of the stimulus envelope, manifest in the profile of modulation period histograms. At progressively higher modulation frequencies (>20 Hz), where cortical neurons typically fire less than once per modulation cycle, the representation of modulated signals converges on a single MPH peak, corresponding to the rising or falling phase of the envelope. For SAM, this is typically the rising phase of the level cycle, where SPL changes most rapidly (Figure 4.4). Below this range, however, MPH profiles contained a range of response features consistent with an interaction of the instantaneous amplitude or frequency of the modulated signal interacting with the amplitude or frequency tuning of the neuron. Some of these response features, such as distinct MPH peaks corresponding to the rising and falling phase of the modulation cycle, have been

described in the midbrain (Rees and Moller 1983) and are likely to be partially inherited from thalamic input.

There is no reason to believe that what cortical neurons encode is fundamentally different at higher modulation frequencies, although the relationship between the envelope and the MPH is strained not only by biophysical limits on temporal precision (Blackburn and Sachs 1989), but also by limits on instantaneous discharge rates. As the modulation period shortens, the instantaneous rate code for stimulus frequency or level that is evident at low modulation frequencies degrades to an apparent synchrony code as multiple features in the period can no longer be resolved. A single response feature may come to dominate the MPH by "stealing" spikes from the others due to refractory periods, and/or adaptation within the modulation period. Terminologically, the distinction to be drawn is not between synchrony and rate coding (Lu et al. 2001a), but rather between within-cycle rate coding, and across-cycle rate coding. These data indicate that in auditory cortex of awake macaques, the distribution of spike rates within the modulation period encodes the amplitude or frequency envelope of the stimulus, rather than extracting the modulation frequency. This may even be the dominant code for low-frequency sound envelopes in the macaque, because little evidence was found for across-cycle rate coding in the *absence* of synchrony. Not only were significant differences in firing rate for modulation frequencies beyond the synchrony cutoff uncommon,

most of the variation in average firing rate across modulation frequency occurred within the synchronized regime.

Changes in carrier frequency, level, and modulation depth all produce clear and consistent changes in the MPH profile, which are captured poorly by summary measures derived from MTFs. Changes in MPH shape can result in minimal or misleading changes in VS (see Figures 4.4 and 4.11), and the distribution of firing rates within the modulation period is generally more informative than the distribution of firing rates across periods (such as rate BMF). Taken together, these findings indicate that the coding of low modulation frequencies is more directly concerned with providing high-fidelity representations of multiple envelope features rather than a high-synchrony representation of the dominant modulation frequency.

4.7.2 Coding of real-world envelopes

From the perspective that auditory cortex acts as a modulation filterbank (Green and Kay 1974, Kay and Matthews 1972), perfect synchrony would represent an effective neural code for modulation frequency, but an extremely poor code for features of the envelope such as instantaneous amplitude. Envelopes of natural sounds contain power at a range of modulation frequencies with a distribution following a power law, emphasizing the low end (Attias and Schreiner 1998). The

data in this chapter indicate that low-frequency amplitude modulations are those least likely to produce high-synchrony, unimodal MPHs that are invariant with changes in carrier level or modulation depth. Instead, cortical MPHs are most diverse in this range, posing a problem for any cortical network attempting to detect synchrony (bearing in mind that the brain, unlike our analysis software, does not have a reference for "time zero"). In addition, the envelopes of natural sounds can be broken into sinusoidal components via the Fourier transform, but these envelopes themselves are seldom periodic. The MPHs in this chapter are constructed using knowledge of the modulation period, but the ergodic assumption (and the composite MPHs of Figure 4.8) suggest that within a subset of cortical cells, instantaneous discharge rate will vary coherently across the population.

4.7.3 Mechanisms of envelope sensitivity

In many neurons, particularly those with sustained responses to pure tones, tuning to the level or frequency of pure tones provides a useful model for mapping the modulation envelope to the MPH. But, to take amplitude as an example, responses to SAM cannot be predicted simply by treating the rate-level function as a "lookup table" for instantaneous firing rate. It is most likely inappropriate to expect the response to SAM to be fully explicable by the RLF, or the response to SFM to be a direct product of the frequency tuning function measured with pure tones. One is not *determined by* the other; rather, both are reflections of the *same underlying*

process. As Kay (1984) has argued, a tone is not a static stimulus, but rather a trapezoidal modulation of amplitude, which invokes sensitivity to envelope parameters. Additional mechanisms sensitive to the time scale of stimulus dynamics may further shape the MPH, and determine how their shapes change as modulation frequency is varied (see discussion of adaptation in Figure 4.12).

A recent review of SAM processing (Joris et al. 2004) poses the question of whether specialized neural mechanisms exist to extract AM information. There may be synaptic specializations contributing to the processing of temporally modulated signals in particular cortical regions (Varela et al. 1997, Eggermont 1999), or classes of neurons within a single cortical area (Atzori et al. 2001), but the adaptive mechanisms themselves may be general. Such mechanisms would operate at each successive stage of the ascending auditory pathway, so their effects will accumulate as the output of one stage serves as the input to the next.

4.7.4 Comparison with previous studies

Only one prior study has examined both SAM and SFM in awake primate auditory cortex (Liang et al. 2002). The similarity between the findings of the two studies (this one in macaque, the other in marmoset) outweigh the differences, and for that reason this chapter has made minimal use of MTF-based analysis except where it may illuminate differences between cortical fields. Both studies found rate BMFs

to be indistinguishable between SAM and SFM, though that study reached the same conclusion with regard to synchrony cutoffs (this chapter reports higher cutoffs for SAM). Also, when population synchrony MTFs were constructed from marmoset data, no modulation frequency elicited synchrony in >60% of cells, and these functions were bandpass because of an absence of synchrony at 2 Hz; Figure 4.15 shows low-pass functions with ~75% synchrony at 1 and 2 Hz. Figure 4.15 examines only those neurons studied with the full range of modulation frequencies, not all neurons encountered, so the abcissa is better interpreted as the percent of the SAM-responsive population. Differences in the shapes of the functions between this figure and Liang et al. (2002) may be attributed to the relative length of the modulation period and the trial (1 s for 10-20 trials in that study, 10 s for 2 trials in this one), but interspecies differences may play a role as well.

Another study in awake marmoset cortex identified two neural populations, one of which encoded the temporal modulation in click trains by synchronizing to the clicks, the other of which did not synchronize but was tuned to click frequency in overall spike rate (Lu et al. 2001a). No comparable unsynchronized population of rate-coding neurons was identified using SAM or SFM in awake macaque cortex. Although cells could show an elevated rate response at high modulation frequencies (usually ≥100 Hz) as demonstrated in the inferior colliculus (Krishna and Semple 2000), these cells invariably showed significant synchrony to the

envelope at some lower modulation frequency. If this rate-coding population is unresponsive to tones, it may have been under-sampled in the SAM and SFM population reported here.

There are several reasons why the direct coding of modulation envelopes may not have been emphasized in previous studies of cortical SAM and SFM processing. Few prior studies used awake animals (Evans and Whitfield 1964, Ribaupierre et al. 1972, Goldstein and Abeles 1975, Creutzfeldt 1980, Beiser and Muller-Preuss 1996, Malone et al. 2000, Liang et al. 2002). Our results are inconsistent with Schreiner and Langner's (1998) view that there is "no evidence that anesthesia exerts a strong detrimental influence on the cortical representation of amplitude modulated signals". Because responses in anesthetized cortex rarely exceed a single spike per modulation cycle, it would appear that the cortical rate representation of SAM signals is profoundly distorted, and much of the diversity in MPH shapes reported here could not be observed. This may explain why Schreiner and Langner (1988) concluded that the "repetition rate of transients plays a more important role in the cortical representation of signals than do details in the envelope shape". A second factor is the focus on the limits of temporal coding at relatively high modulating frequencies, where the underlying rate code for amplitude or frequency becomes indistinguishable from a synchrony-based code for modulation frequency. In this regime, MPH shapes are likely to be dominated by static temporal filtering

rather than the adaptive gain control mechanisms that shape responses to more slowly modulated signals (Malone et al. 2002). One study has explicitly examined the sensitivity of neurons in awake primate cortex to envelope shape (Lu et al. 2001b), revealing rate-based selectivity for temporally asymmetric waveforms, but these stimuli were periodic at 40 Hz, beyond the range where MPH shape would be expected to track the envelope (though figure 12 of that paper presents data from a 10 Hz repetition rate, where the phase of the response is clearly envelope-dependent).

As mentioned above, a third reason previous cortical studies of SAM and SFM have not emphasized envelope coding at low modulation frequencies is the reliance on measures derived from modulation transfer functions. In the auditory nerve and cochlear nucleus, where tuning properties for frequency and level (and MPH and MTF shapes) are more homogeneous, the MTF approach has been relatively successful (Joris and Yin 1992, Rhode and Greenberg 1994). This chapter shows that cortical neurons also can be exquisitely sensitive to changes in the depth, carrier level, and carrier frequency of SAM and SFM signals when MPHs are considered in detail. Thus the increasing tolerance of modulation tuning for changes in these stimulus parameters at higher levels of the auditory pathway appears to reflect the diminished utility of summary measures at more central

auditory structures, in awake animals, rather than a diminished sensitivity to these parameters.

4.7.5 Implications for perception

The phase coherence of responses to modulated signals affects how the populations of auditory cortical neurons could represent the envelopes of complex stimuli. In the composite MPHs of Figure 4.8, no attempt was made to correct for the group delay (the latency of each individual cell), in part because it is not clear how the brain could "correct" for this difference across neurons. In a synchrony code for modulation frequency, the phase of the response could be arbitrary. In order to generate a coherent population representation of the signal, however, some common facet of the modulation must be selected as the basis for phase locking. If this feature is expressed in terms of amplitude (e.g., the rising phase of the envelope), it is difficult to see why the resulting representation should not be described as coding amplitude. In fact, considerable coherence was evident in AI responses to SAM signals at low modulation frequencies, particularly within classes as defined by responses to an unmodulated tone. Psychophysical studies of have shown envelope peaks to be the predominant cue to speech intelligibility (Drullman 1995), perhaps reflecting the enhanced salience of their neural representation.

Physiological studies in human cortex still endorse a modulation filterbank model of auditory processing, in which successive stages of the auditory pathway exhibit bandpass MTFs centered at lower and lower SAM frequencies. The findings of these studies are in fact compatible with the range of sensitivities exhibited in awake macaque, but the interpretation of the results is clouded by the techniques available in human subjects, such as auditory evoked potentials from intracerebral electrodes (Giraud et al. 2000), and functional magnetic resonance imaging (fMRI; Liegeois-Chauvel et al. 2004). The temporal grain of fMRI restricts analysis to time-averaged hemodynamic responses, presumed to underlie synaptic activity, which would obscure synchrony even in the low range of modulation frequencies critical to speech comprehension. Nevertheless, activation in the primary auditory area was maximal in the range of 8 Hz for a SAM white noise, intermediate between the best modulation frequencies in auditory thalamus (16 Hz) and putative belt areas surrounding Heschl's gyrus (4-8 Hz) (Giraud et al. 2000). These values are consistent with the broad peak of the population MTF (5 Hz) measured relative to spontaneous rate in Figure 4.15A, an analogous method of pooling responsiveness across a broad cortical area. Evoked potential recordings are capable of detecting rapid modulations of population activity, and indicate that "only the lowest AM frequencies (<30 Hz) are preserved" at the level of human cortex (Liegeois-Chauvel et al. 2004). Because single units could not be discriminated, these data are more analogous to the composite MPHs of Figure 4.8,

in which synchrony is entirely absent above 10 Hz when responses of all units are summed (left panel); when the responding population is subdivided, synchrony is apparent up to 20 Hz in two of three response categories, but this synchrony was masked by pooling cells responding to different phases of the modulation envelope. When methodological constraints are taken into account, the core fields of auditory cortex in alert macaque appear to offer a tenable model of human cortical processing.

The phase coherence of responses across the cortical population potentially could inform the perceptual boundaries for SAM signals. In human listeners, SAM tones are perceived as "fluctuations" in amplitude at very low modulation frequencies (< ~5Hz), remain resolved as individual periods up to 20 Hz, and are heard as "roughness" between 20 and 300 Hz. These perceptual boundaries correspond well to the changes in MPH shapes observed with increasing modulation frequency, and the global degradation in synchrony apparent in the population MTFs of Figure 4.15. In auditory cortex of awake squirrel monkeys, modulation of cortical activity by SAM has been described with a dependence on modulation frequency, depth, and carrier level comparable to that of perceived "fluctuation strength" in human listeners (Fastl et al. 1986). Other psychophysical studies bolster the idea that the physiological transformation from tracking the envelope to marking the period corresponds to a perceptual blurring of envelope shape, and a confusion between

AM and FM (Kay 1982). For example, square-wave AM waveforms were shown to adapt the detectability of sinusoidal AM above – but not below – 10 Hz. More generally, Kay (1982) concludes that "coincident modulation rate is not of itself a sufficient stimulus for adaptation". These psychophysical studies support the argument, based on cortical physiology, that the processing of modulated signals is better described as envelope shape discrimination than modulation frequency extraction.

4.7.6 Implications for cortical organization

Single neurons in AI and R share the mechanisms of low-frequency envelope coding described above, but differ in the upper limits of their ability to synchronize (Figure 4.14) and, as a population, cells in R are less likely to be firing in synchrony to the envelope at most modulation frequencies (Figure 4.15). The degradation of temporal synchrony in R, perhaps due to temporal integration of the acoustic signal, makes this field a candidate for rate-based coding of modulation frequency beyond the (restricted) synchronized regime, as has been described for click trains in auditory cortex of awake marmoset (Lu et al. 2001a). Yet the prevalence of rate coding beyond the temporal cutoff was equal in AI and R for SAM (16%) and SFM (13% in AI, 11% in R, p=0.12). Bandpass rate tuning in general, as defined by Liang et al. (2002), also did not differ significantly between cortical fields for either stimulus (SAM: 78% in AI, 74% in R, p=.42; SFM: 78% in

AI, 69% in R, p=.07). The population MTFs of Figure 4.15 indicate that at any given modulation frequency, fewer cells in R show a significant rate response than in AI, further suggesting that R is not the cortical manifestation of the modulation filterbank model.

The properties of field R in the New World monkey seem comparable to those described in macaque, including a diminished synchrony to SAM stimuli, leading Beiser and Muller-Preuss (1996) to conclude that AI is among "the areas where the envelope fluctuations of the monkey's calls may be encoded… other fields have insufficient time resolution to manage such information." This statement may draw too close an analogy between the simplistic, repetitive SAM stimulus and the complex envelopes of monkey vocalizations: the diminished responsiveness in R may instead reflect adaptation to a stimulus whose envelope is not flat, but is periodic and (beyond the time scale of that period) unchanging. No acoustic stimulus is truly static, since even a pure tone must be gated on and off; in this sense, static vs. dynamic is a matter of time scale. Just as many auditory neurons will quickly adapt to a pure tone, higher-level neurons in the rostral core may adapt to simple envelopes because of a temporal filtering mechanism – informed by adaptation – that disregards unchanging envelopes.

Synchrony cutoffs derived from modulation transfer functions suggest that in R, relative to AI, synchrony in response to SAM is degraded further than that in response to SFM. If true, this could lead to testable hypotheses about the time scales over which neurons in R integrate spectral energy, or amplitude envelopes. For example, reversed sounds, or chimeric sounds that meld the amplitude envelopes of one sound with the spectral content of another, might be indistinguishable to neurons in AI but be selectively represented by neurons in R. Thus, the degradation of temporal synchrony beyond the range of frequencies crucial to speech and communication signals need not render the rostral field an inferior site for auditory object identification.

In summary, close examination of responses to slow amplitude and frequency modulations reveals mechanisms of envelope coding by instantaneous rate. The next chapter will employ periodic modulations of a binaural cue to sound location, further elucidating the potential functional roles of the core cortical fields in auditory scene analysis.

Chapter 5: Binaural Beats: Temporal and Spatial Processing

5.1 Introduction

The exploration of cortical responses to amplitude and frequency modulation in the preceding chapter illustrates a potentially general coding principle in AI: temporal or spectral dynamics of the stimulus may be reflected in the instantaneous discharge rate of the neuron, up to some limiting rate of modulation. Beyond this cutoff frequency, overall discharge rate (without regard to the modulation period) may continue to carry information about the modulation frequency of the stimulus. This mechanism, enacted across the cortical population, would constitute the representation of spectrotemporal stimulus features within real-world sounds such as animal vocalizations. This same model could extend to the representation of any dynamic acoustic parameter to which auditory neurons are tuned. This chapter will take a similar approach toward cortical responses to modulation of binaural differences.

Spectral and temporal sensitivity, as exemplified by responses to SAM and SFM in Chapter 4, is suspected to underlie one of the two putative processing pathways emanating from the auditory core, the rostral "what" projection (Rauschecker and Tian 2000). The other, "where" subsystem is thought to emerge from caudal auditory cortex, and project to frontal and parietal regions involved in spatial processing (Romanski et al. 1999a,b). Physiological evidence for this spatial

pathway has been sought using free-field stimuli in lateral belt regions of the anesthetized monkey, and in core and belt of the behaving monkey. Neurons in awake AI are sensitive to sound source location and movement (Ahissar et al. 1992), with caudo-medial belt showing possible enhancement of tuning for sound source location, relative to core or rostral belt areas (Tian et al. 2001, Recanzone et al. 2000). The question of what binaural cues might underlie this spatial selectivity at the level of single cells remains unanswered. Anatomical data (Hackett et al. 1998) suggest that cells in the belt receive their input from primary regions in the core (AI and the rostral field, R), and thus may construct their spatial tuning from sensitivity to binaural cues in neurons of the primary field. In Chapter 2, closed-field stimuli that mimic acoustic motion were presented to awake macaques trained to lateralize sounds based on binaural cues. This chapter presents additional data collected during passive listening to characterize more fully the responses of neurons in core auditory cortex to these binaural disparities, allowing the comparison of single-unit responses to psychophysical discrimination in the same animals.

The binaural beat is a cyclical modulation of interaural phase disparity (IPD, see below), the rate of which can be manipulated to explore the temporal coding of binaural cues in analogy to the spectral and temporal properties revealed by SAM and SFM. We report an extensive, low-frequency, contiguous field of auditory

cortex that responds to binaural beats with discharges synchronized to the beat period at low frequencies of modulation. In contrast to earlier studies that found transient responses to the beat cycle in anesthetized cortex (Reale and Brugge 1990), neurons could respond robustly throughout the period of the beat. The fidelity of this single-neuron representation is compared to the psychophysical sensitivity of macaques to dynamic IPD demonstrated in Chapter 2.

5.2 Background

A prominent cue underlying sound localization along the azimuth is the difference in arrival time of a sound at the two ears. Physiological sensitivity to this interaural time difference (ITD) arises from a coincidence-detection mechanism in the medial superior olive (MSO) of the brainstem, where cells are tuned for a particular temporal delay – on the order of microseconds (μs) – between sound arriving form the contralateral and ipsilateral ear (Boudreau and Tsuchitani 1968, Goldberg and Brown 1969, Yin and Chan 1990). For pure tones in the low-frequency range where the auditory nerve can phase-lock to the period of the stimulus, the same mechanism is sensitive to an ongoing difference in phase between the ears, the interaural phase disparity (IPD) (Rayleigh 1907, Spitzer and Semple 1995). Neural tuning to ITD, once established in the brainstem, is preserved through the inferior colliculus (IC) in the midbrain (Rose et al. 1966, Kuwada and Yin 1983, Spitzer

and Semple 1993), and relayed via the medial geniculate nucleus (MGN) of the thalamus to AI (Brugge and Merzenich 1973, Fitzpatrick and Kuwada 2001).

The preferred phase and tuning sharpness of a neuron may be determined from repeated presentations of dichotic tones spanning the full range of IPD, an effective but time-consuming method. A more efficient stimulus for the assessment of IPD tuning is the binaural beat, an ongoing modulation of IPD produced by presenting tones of slightly different frequencies to the two ears (see Methods). This produces a percept of auditory motion in human listeners, a tone that appears to circle the head at a rate determined by the frequency difference between the tones. As the difference in the two tones is increased, the perceived beat loses its spatial quality and becomes a fluctuation or roughness (Licklider et al. 1950). Physiological responses to this stimulus therefore can reveal spatial, spectral, and temporal aspects of the binaural system. The advantage of the binaural beat stimulus over static IPD stimuli was demonstrated by Yin and Kuwada (1983ab) in cat IC, where they generated characterizations of neural tuning equivalent to those obtained with static IPD stimuli in far less time.

Anesthetic state can have drastic effects on discharge properties at several levels of the auditory system (cochlear nucleus: Evans and Nelson 1973, IC: Kuwada et al. 1989), but the depressive effects of anesthesia may be most disruptive in the cortex,

where not only the level of responsiveness (Erulkar et al. 1956, Pfingst et al. 1977, Pelleg-Toiba and Wollberg 1989) but tuning properties of individual neurons can be affected (Gaese and Ostwald 2001). Cortical responses to binaural beats have been studied only in anesthetized cats, in which responses tend to be transient (Reale and Brugge 1990), allowing only a limited characterization of IPD representation in a brain region known to be critical in sound source localization (Heffner and Heffner 1990). Even less is known in primates, where only one previous study has examined sensitivity to ITD in the awake monkey (Brugge and Merzenich 1973). Macaque AI has a large low-frequency representation (Merzenich and Brugge 1973), and these animals have been trained to discriminate IPD in low-frequency stimuli (Houben and Gourevitch 1979, Wegener 1974, see Chapter 2), but the IPD sensitivity of this low-frequency neural population has never been described in the alert state.

5.3 Methods

5.3.1 Subjects, recording procedure, and stimuli

Detailed methods are described in Chapter 2. Briefly, subjects were two male Rhesus macaques, trained to discriminate the laterality of IPD around midline, and accustomed to sitting quietly between behavioral sessions to allow the collection of additional physiological data in the passive state. Stimuli were generated digitally by MALab software, synthesized by dedicated hardware processors and presented

in the closed field via electrostatic speakers coupled to ear inserts (phase and level were calibrated across frequency at the start of each day's session). An electrode was positioned vertically and advanced with a stepping microdrive, through the parietal cortex to auditory regions of the superior temporal plane. Entry into auditory cortex was typically preceded by a gap in physiological activity believed to correspond to passage through the lateral sulcus.

When a responsive cell was isolated, threshold and best frequency (typically at 60 dB) were determined using pure tones (See Chapter 3). Phase sensitivity was initially assessed by binaural beats, an ongoing modulation of interaural phase produced by presenting two tones of slightly different frequency to the two ears. In our paradigm, a positive phase value corresponds to a phase lead at the left ear. Thus, a 2 Hz binaural beat at a carrier frequency of 1000 Hz is produced by presenting 999 Hz to the right ear and 1001 Hz to the left, evenly spacing the actual tone frequencies about the nominal "carrier", or center, frequency. Because cells were recorded from the auditory cortices of the left and right hemispheres, these data have been transformed into compatible "contralateral" and "ipsilateral" reference frames where appropriate. Beats were typically presented as two 10 s pips, presented during 12 s trials (allowing for a 2 s "off" time between presentations). Cells that synchronized their discharge to the resultant beat could be further characterized using trapezoidal modulations of phase (Malone et al. 2002),

and/or recorded while the animal discriminated IPD (Chapter 2). Center frequency and SPL were optimized for a robust beat response before other parameters (e.g. beat frequency) were manipulated; in some cases the best beat response was not at the tone-derived BF of the cell.

5.3.2 Statistical Analysis of synchrony

The binaural beat is a cyclical modulation of IPD: over one cycle of a 1 Hz beat, the signals begin in phase (IPD=0°), modulate to a maximal left ear lead at 0.5 s (IPD=180°) and then re-align after 1 s to begin the cycle again (IPD=360°=0°). This circular axis simplifies analysis by confining the stimulus space, but requires the use of circular statistics. Each spike fired is treated as a unit vector in a polar coordinate system with the angle representing the instantaneous phase of the beat, and the direction of the sum of those vectors defines the mean phase of the response. Normalized to the number of spikes, the length of that mean vector is dependent on how tightly the spikes cluster about that one point in the stimulus cycle. This vector strength (VS) is a measure that varies between zero (a flat period histogram) and one (all spikes in one histogram bin) for unimodal circular distributions. (This is the inverse of the measurement typically used in vision, circular variance.) Because VS depends on spike rate, VS alone cannot be used to determine if the cell is firing in synchrony with the modulation of the stimulus (in an extreme example, a single spike would give the maximum VS of 1 at whatever

phase it happened to occur; or, two spikes falling 180° apart would cancel one another to yield a VS of 0). Rayleigh (1880, cited in Fisher 1993) developed a statistical criterion for the nonuniformity of a circular distribution:

$$R = 2n*v^2$$

where the Rayleigh statistic R is computed from the number of spikes in the period histogram [n] and the vector strength [v]. A response was considered significantly synchronized if the Rayleigh statistic was >13.816, the $p<.001$ level. In Chapter 4, an alternative measurement of synchrony within the modulation period (correlation) was introduced to accommodate responses at multiple modes within the stimulus cycle; such responses were not observed in binaural beat data set, so VS is a sufficient index. Also, "synchrony" in this chapter is always in reference to the stimulus modulation cycle, not the concerted firing of multiple neurons.

5.3.3 Recording Sites

A systematic mapping of response properties in the horizontal plane, and the depths of those recordings, reveal a coarse tonotopic trend within a circumscribed region of the superior temporal plane (Scott et al. 2000; discussed in detail in Chapter 3) that corresponds to previous studies on the functional and anatomical organization of primate auditory cortex (Hackett et al. 1998, Merzenich and Brugge 1973). This map has provided a framework in which to analyze responses to binaural beats in the primary auditory field (AI) and adjacent rostral field (R) in both animals. For

the analyses presented below, the population has been defined physiologically and has not been segregated by animal, hemisphere, or cortical field unless otherwise stated. (Such comparisons were made wherever possible, to ensure that pooling of the data sets across animals and hemispheres was appropriate.)

5.4 General Results

Vigorous responses to binaural beats were obtained over a large swath of the superior temporal plane in both animals. Electrode penetrations yielding beat responses formed a largely contiguous cluster around rostral, low-frequency AI and appeared to extend into the adjacent rostral field R. Entry into low-frequency AI was often marked by an audible "hash" in response to a binaural beat stimulus, suggesting homogeneous responsiveness among the local population of neurons. This was especially true in the middle layers of cortex, although responses could be obtained throughout the depth of the tissue (typically 2500 μm, estimated from auditory responsiveness, though tracks could be longer – presumably when the electrode was not normal to the cortical surface).

Outside of the AI/R border, occasional units in the belt could be isolated that responded to low-frequency tones or band-passed noise (see Chapter 3). These units were not found to be phase-sensitive when tested with binaural beats, suggesting that the border between AI and R is the ultimate representation of IPD

in the auditory system. Beyond the cortical core, spatial information may no longer be carried in frequency-specific channels sensitive to binaural disparities in simple stimuli (see Discussion, below, and Chapter 6).

	Monkey X	Monkey Z	Total
Total cells w/ BF<2.5 kHz	290	256	546
Qualitatively studied	226	171	397
Quantitatively studied	186	97	**283**
Showing synchrony to beat	178	90	**268**
			(168 AI, 91 R, 9 belt)

Table 5.1: Number of cells studied with binaural beats

Table 5.1 tabulates the number of cells that had best frequencies in the range where sensitivity to IPD might be expected, and the number of those cells tested with the binaural beat stimulus. Quantitative data were collected from 283 cells (186 in monkey X, 97 from Z) from AI and R. (Recording chamber positions in monkey X were optimal for recording from the low-frequency border of AI and R, resulting in a disproportionate sample of phase-sensitive neurons from this animal.) Of cells studied quantitatively, 268 (95%) discharged with significant synchrony (Rayleigh test, $p<.001$) to a binaural beat, though these cells represent only 68% of the total number of cells tested (cells that showed no sensitivity to the qualitative test were less likely to be studied in further detail). Recording conditions were often quite stable, without interference from pulsation or mechanical artifacts of animal movement, allowing a cell to be held in isolation for a broad characterization lasting up to an hour or more.

An example of a single-unit response to a 2 Hz binaural beat is shown in Figure 5.1A as a spike raster, binned into a post-stimulus time histogram (PSTH). Cycling through IPD induced a synchronized firing pattern in the cell, evident in the PSTH and emphasized in the period histogram (5.1B), in which responses to forty cycles of the beat (two 10 s presentations at 2 cycles/s) are superimposed on a common time axis by "folding" on the beat duration of 500 ms. As in Chapter 4, this representation can be called the Modulation Period Histogram (MPH). This profile was typical of responses to binaural beats in the awake cortex, and differs markedly from previous observations in anesthetized cat AI, where cells responded transiently with a single spike per beat cycle at comparable rates of modulation (Reale and Brugge 1990).

The binaural beat proved to be an efficient stimulus for measuring the phase tuning of neurons in the primate cortex, as was demonstrated in cat IC by Yin and Kuwada (1983b). Overlaid in gray on the period histogram (Figure 5.1B) is a tuning curve obtained from presentations of static phase offsets at 8 values of IPD. The mean phases computed from the beat and static IPD responses are nearly identical, indicating that the same underlying phase selectivity is revealed by either method. The degree of selectivity for phase (and, in the dynamic case, synchrony to the beat cycle) is measured by the vector strength of the response (see Methods), which is

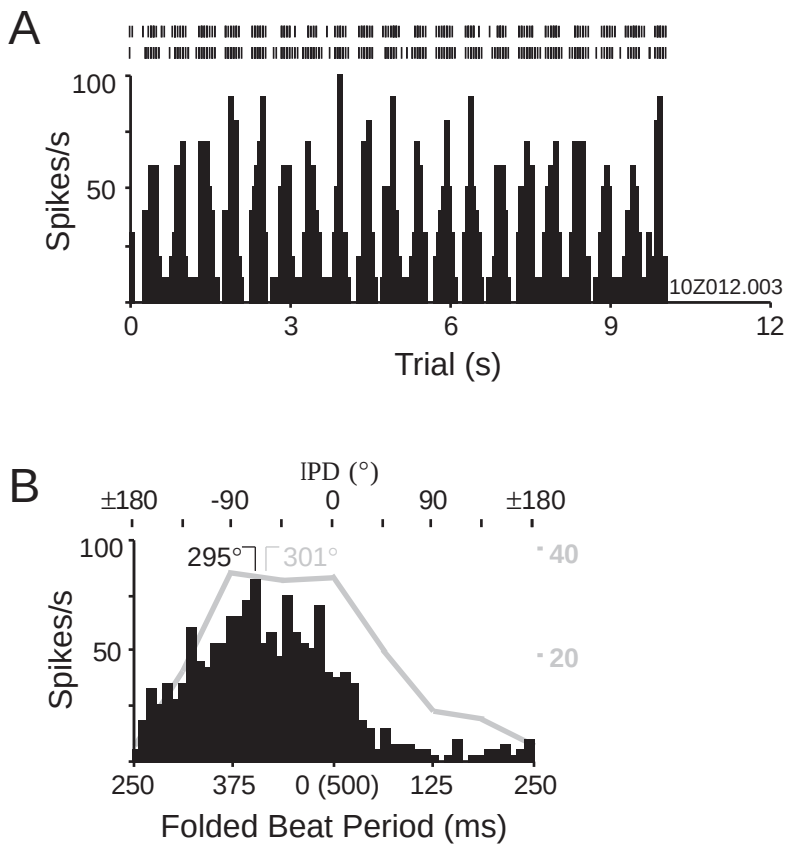


Figure 5.1 **A:** Single-unit response to a 2 Hz binaural beat presented for 2, 10 s pips (2 s intertrial interval). Spike raster at top is binned into a post-stimulus time histogram (PSTH) showing 20 peaks of spike discharge each corresponding to one cycle of the beat. Spike rate on the abscissa is computed within each 50 ms bin (overall spike rate for the 10 s trial was 29 spikes/s). Spike rate does not adapt over the course of the trial. Stimulus was presented at 80 dB SPL at the cell's BF of 1600 Hz (1599 Hz to the right/contra ear, 1601 Hz to the left/ipsi). **B:** Histogram folded on the period of modulation, showing spikes from all 40 cycles of modulation (2 trials x 10 s x 2 Hz) in 10 ms bins. Overlaid in gray is the tuning curve derived from presentations of static IPD at 8 values (spike rates on right ordinate, in gray). The phase of both responses has been wrapped to illustrate more clearly the relative phase and vector strengths of the beat and static responses. The lower abscissa indicates ms from the beginning of the beat period, the upper abscissa indicates the corresponding instantaneous phase of the beat and the phase of the static IPD stimulus (because the stimulus is circular, ±180 are the same point, and -90° is equivalent to 270°). As is typical for the population, the mean phase of these two responses are very close (295° and 301° for the beat and static IPD, respectively). As a reference, the vector strengths for the beat and static responses are 0.56 and 0.42 respectively.

sharper for the binaural beat (beat: 0.56; static: 0.42). In 46 units, phase sensitivity was studied with both static IPD and binaural beats, allowing a comparison of the mean phase and vector strength estimates afforded by the two stimuli across the population. Figure 5.2A confirms that a 2 Hz beat gives a reliably similar estimate of mean phase when compared to estimates from static IPD stimuli presented at the same carrier frequency (r^2=.92, slope=.97; arrow indicates cell from Figure 5.1). Although a given cell's mean phase estimate could differ by 45° or more between the two measures, there was no consistent tendency for the binaural beat mean phase to lead or lag the static (visually, the points are equally distributed above and below the diagonal; insignificant by *t*-test). When vector strengths of the two measures are compared (Figure 5.2B), they are correlated but consistently higher for beat responses (r^2=.45, slope=.64, paired *t*-test $p<.0001$) as suggested by the example unit in Figure 5.1B. This may be attributed to nonselective transient responses that could be present at the onset of every tone used in a static IPD procedure, but in only 2 of the 40 beat cycles folded into the period histogram. Alternatively, there could be a more interesting effect of gliding smoothly through phase rather than presenting static IPDs in isolation that enhances a neuron's selectivity for phase, in a manner similar to the dynamic conditioning effects described by Malone et al. (2002). The generality of this sharpening at higher rates of IPD change is addressed in the next section.

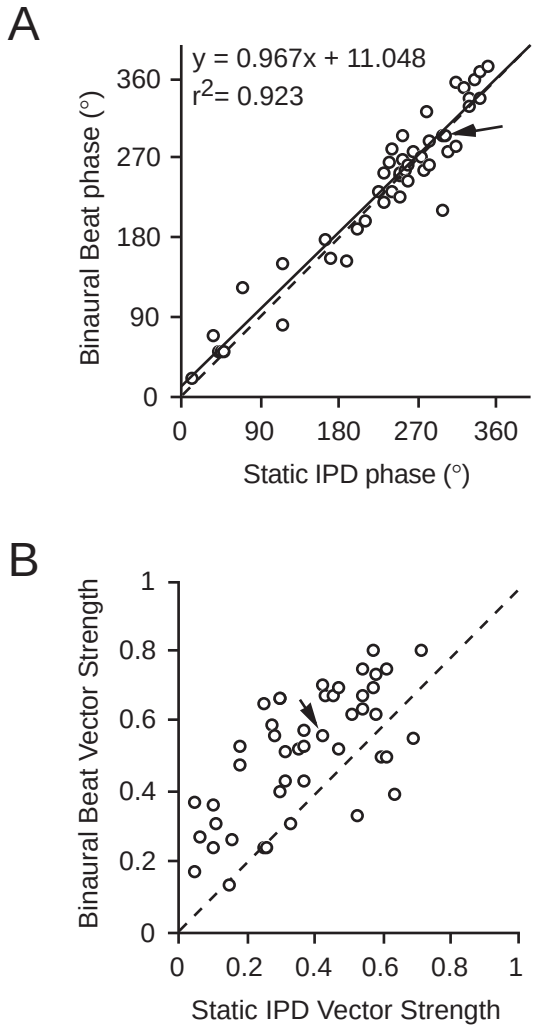


Figure 5.2 46 units were tested with binaural beat and static IPD stimuli, generating similar measures of mean phase but consistently higher vector strengths in response to beats. Arrows indicate the cell from Figure 5.1, dashed lines indicate identity (slope=1). **A:** Mean phase of 2 Hz binaural beat responses vs. mean phase of static IPD tuning functions at the same carrier frequency and SPL. Solid line is a linear least-squares fit (slope=0.97, r^2=0.92). A paired *t*-test revealed no consistent shift in mean phase as measured by the two stimuli. **B:** Vector strength of binaural beat responses vs. static IPD tuning functions. The two are roughly correlated (fit not shown; slope=0.64, r^2=0.45), but beat vector strengths are consistently higher (paired *t*-test p<.0001).

5.5 Tuning to IPD I: Effects of binaural beat frequency

From a constant center frequency, the disparity between the ears can be widened to produce a range of binaural beat frequencies. Responses from two cells tested across beat frequency are shown in Figure 5.3, illustrating a number of features to be addressed in the following section. The effect of beat frequency on the vector strength of each cell's response is evident from these MPHs. At low binaural beat frequencies (lower panels), synchrony to the beat is seen in the deeply modulated profile of the spike counts. At high beat frequencies (50-100 Hz, upper panels), the MPH of the cell on the left has flattened, whereas the cell on the right continues to synchronize its discharges. Also apparent, particularly in the left panel, is the shift in the mean phase of the response: the MPH profile seems to slide leftward (later in the period) and at 20-50 Hz "wraps around" to center. In addition, the total firing rate can be affected by the beat frequency, with both of these units showing a drop in overall activity at high rates of modulation.

The effect of beat frequency on each of these response characteristics can be summarized as a modulation transfer function (MTF); a family of such functions is shown as an inset below each cell's MPHs in Figure 5.3. Vector strength as a function of modulation frequency is termed the temporal MTF (tMTF, Krishna and Semple 2000), shown in the upper panel of each inset. The tMTF of the left cell is a lowpass function rolling off above 20 Hz (typical of the population), whereas the

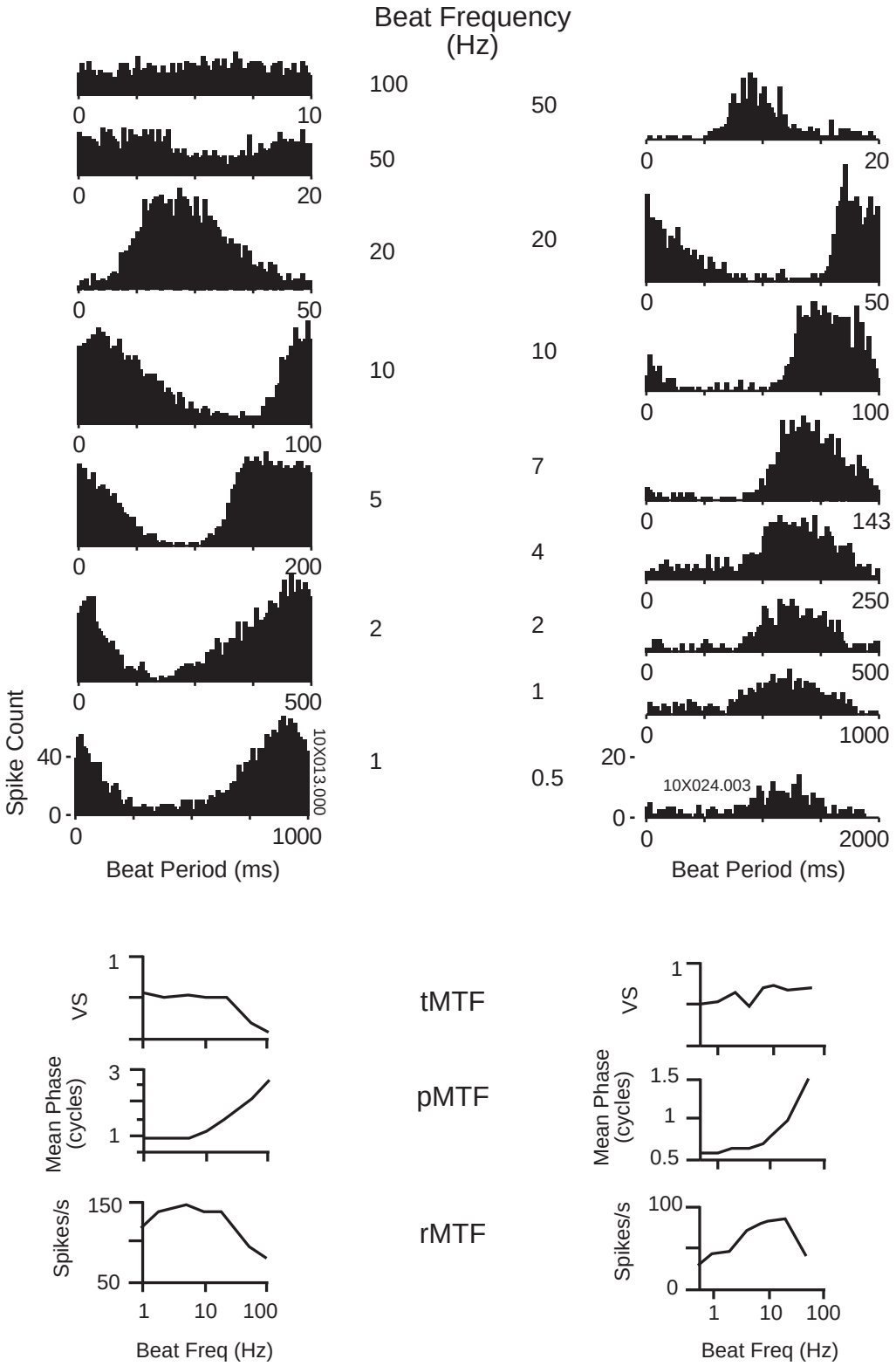

Beat Frequency
(Hz)
100
50
20
10
5
2
1
0.5
Spike Count
40
0
10X013.000
10X024.003
Beat Period (ms)
Beat Period (ms)
VS
Mean Phase
(cycles)
Spikes/s
tMTF
pMTF
rMTF
Beat Freq (Hz)
Beat Freq (Hz)

Figure 5.3 Period histograms from two single units tested across binaural beat frequency. Each period is divided into 100 bins, with the ordinate indicating spike count in each bin. All stimuli were of equal duration (10 s, presented twice in left panel and once in right panel), so the 10 Hz histogram represents ten times as many IPD cycles as the 1 Hz histogram, etc. Functions inset in center illustrate the effects of beat frequency on vector strength, mean phase, and spike rate (tMTF, pMTF, and rMTF, respectively), each of which is visible by eye in the arrayed histograms.

cell on the right has a messy tMTF but maintains synchrony up to at least 50 Hz. Beyond 5 Hz, beat rate causes a shift in the mean phase of the response; phase MTFs (pMTFs) are shown in the center panel of each inset (see also Figure 5.5 and 5.6). Overall spike rate, regardless of vector strength, also can also be affected by the rate at which the binaural beats cycle through IPD, as demonstrated in the rate modulation transfer functions (rMTFs) in the lower panel of each inset (and Figure 5.7).

5.5.1 Effects of beat frequency on vector strength

The 2 Hz beat used to generate the data in Figures 5.1 and 5.2 was effective in all cells that showed synchrony, but a range of higher frequencies (up to at least 20 Hz) was tested to determine how synchrony between the beat and the neural discharge degraded at higher rates of IPD modulation. Predictably, the advantage in tuning sharpness relative to the static IPD stimulus exists over only a limited range of beat rates: when the differences between VS measured from beats and static IPD pips were compared, VS was higher for beats from 2-10 Hz (data not shown; Wilcoxin signed-rank, $p<.05$). To assess the effect of beat frequency on VS across the population, all recorded measurements were pooled from those cells in which beats were presented across modulation frequency (N=183, beat frequencies tested were 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100 and 200 Hz). Figure 5.4A plots the median VS, along with the 10th and 90th percentiles, across beat frequency. The

median degrades from 0.5 to 0.2 with increasing beat frequency, and the 90th percentile traces a weakly bandpass curve peaking at 5-10 Hz but falling toward the median at higher rates, where synchrony becomes uniformly poor.

The beat rate at which cells can no longer synchronize their firing to the beat may indicate the maximal velocity of IPD change which can be detected. Cutoff frequency was determined in 97 cells where beat frequency was tested beyond the range of statistically significant synchrony. (A linear interpolation was performed between the Rayleigh statistic at the highest synchronized frequency, and the adjacent lowest unsynchronized frequency; the frequency at which this line crossed the Rayleigh threshold of 13.816 was taken as the cutoff [Liang et al. 2002].) Figure 5.4B is a survival plot of these interpolated cutoffs, essentially plotting the percentage of cells that could fire in synchrony at or above the beat frequency on the ordinate. The median cutoff, at which the survival plot crosses 50%, is 20 Hz. Cutoff frequency was not dependent on the best frequency of the cell.

Each cell could be assigned a best modulation frequency (BMF) on the basis of firing rate, synchrony, or Rayleigh strength (a balance of the two). The distributions of BMFs generated from each measure were essentially identical, with a mode at 5 Hz. As with synchrony cutoff, there was no dependence of BMF on best frequency (ANOVA, $p>.89$). As suggested by the MPHs in Figure 5.3 and the distribution of

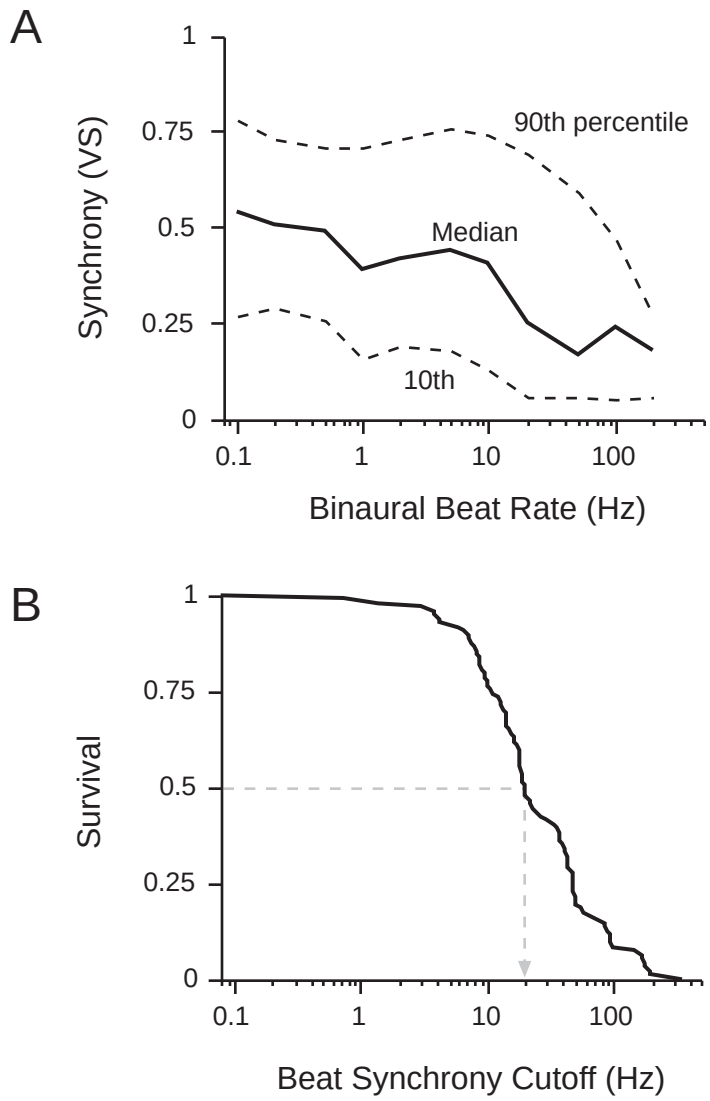


Figure 5.4 Effects of binaural beat rate on vector strength across the population. **A:** All recorded VS measurements were pooled from those cells in which beats were presented across modulation frequency (N=183, beat frequencies tested were 0.1, 0.2, 0.5, 1, 2, 5, 10, 20, 50, 100 and 200 Hz). Median VS (heavy line) is plotted along with the 10th and 90th percentiles (dashed lines), across beat frequency. The median degrades from 0.5 to 0.2 with increasing beat frequency, and the 90th percentile traces a weakly bandpass curve peaking at 5-10 Hz but falling toward the median at higher rates, where synchrony becomes uniformly poor. **B:** Survival plot of synchrony cutoff. For 97 cells tested beyond the limit of significant synchrony, a line was fit between the Rayleigh statistic at the highest significant point, and that at the next highest frequency; the frequency at which this point intercepted the $p<.001$ significance criterion was used as the interpolated cutoff. Survival indicates the proportion of cells that could synchronize at or below the frequency on the ordinate. The median cutoff (gray dashed line) was 20 Hz, beyond which only half the tested population could synchronize to the beat envelope.

cutoffs, cells could code IPD in their instantaneous spike rates over a wide range of modulation frequencies, so the assignment of a BMF is not especially informative.

5.5.2 Effects of beat frequency on mean phase

Figure 5.2 asserts that the binaural beat is an accurate estimator of a neuron's preferred IPD, but as with VS, mean phase is similarly dependent on the binaural beat frequency employed. Whereas a slow, 2 Hz beat generates a mean phase equivalent to that estimated from static IPD stimuli, increasing beat rates elicit a phase lag deriving (presumably) from the conduction time from ear to cortex. For example, a latency of 20 ms is only 4% of the 500 ms cycle of a 2 Hz beat (i.e., by the time the cell fires, the instantaneous IPD has shifted 14°), but the lag will be tenfold larger relative to the shorter period of a 20 Hz beat. Figure 5.5A plots the difference in mean phase between the beat response and static tuning curve for the same 64 cells from Figure 5.2 (arrow indicates the 2 Hz data from Figure 5.2), a difference which is significantly >0 at all beat frequencies ≥5 Hz for the population overall ($p<.001$, Wilcoxin signed-rank). Data from three individual units are plotted in 5.5B to emphasize the constancy of mean phase estimates in the low-frequency regime (0.1 – 5 Hz), and the smooth, linear progression as beat frequency increases and latency effects become more prominent. For the cells illustrated, and the population in general, binaural beats in the "negative" (right ear lead) direction produced mean phase plots that were the mirror image of those produced by

“positive” frequencies. Figure 5.6 illustrates this phase lag for eight units tested across beat frequency in the positive and negative directions, plotted on logarithmic (A and C) and linear axes (B and D). The four units in panel A showed a diversity of mean phases at low beat frequencies, and the curves remain roughly parallel as the phase lag manifests. The cells in panel C had similar mean phase values and were tested to greater extremes, but showed consistent mean phase shifts to modulations as fast as 200 Hz in either direction.

The slope at which mean phase changes with beat frequency should give an estimate of the fixed delay in the system (the neural latency) causing the phase lag. Least-squares linear fits, like the examples in 5.6 B and D, were weighted by the Rayleigh statistic at each point. Latency estimates from these slopes were compared against average first-spike latencies in response to diotic pure tones for 124 single units. The two measures are roughly correlated (r^2=.22, ANOVA: $p<.0001$), with first-spike latency estimates showing a slight but insignificant tendency to be shorter (paired t-test, $p=0.08$). The poor fit between the two measurements suggest that onset responses may be attributed to a different, possibly monaural mechanism distinct from that which induces the phase lag. Adaptation to repeated stimulus cycles could explain a longer estimate of delay, as proposed by Eggermont (1999).

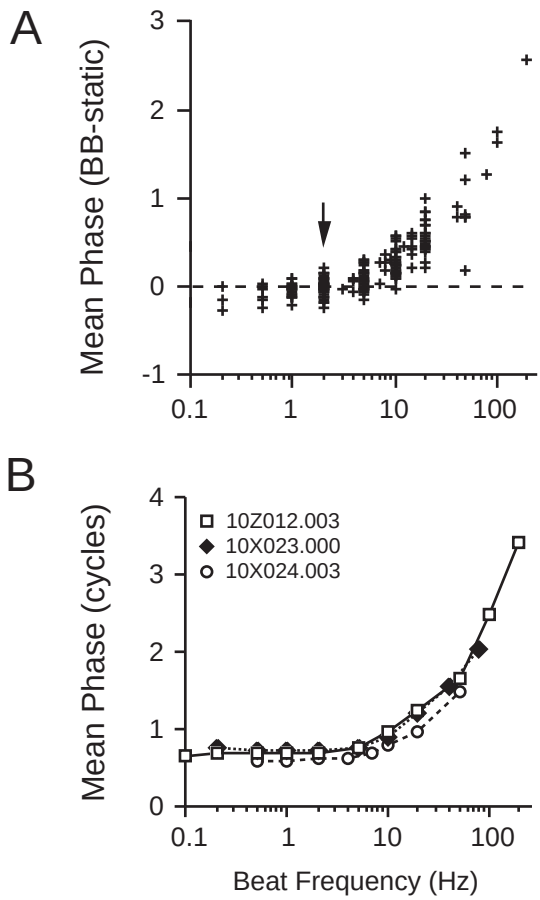


Figure 5.5 **A:** Binaural beats accurately predict the static preferred IPD at low beat rates, but higher rates of IPD modulation elicit a lag in the phase of the response across the population. The difference in mean phase estimates (beat – static) is plotted vs. beat frequency for the population. Dashed line is zero (no difference), arrow indicates 2 Hz data from Figure 2A. At all beat frequencies ≥5 Hz, the difference is significant (Wilcoxin signed-rank, $p<.001$). **B:** Three units from panel A are plotted to emphasize the constancy of mean phase estimates below 5 Hz before the transition to a linear shift in phase with beat frequency.

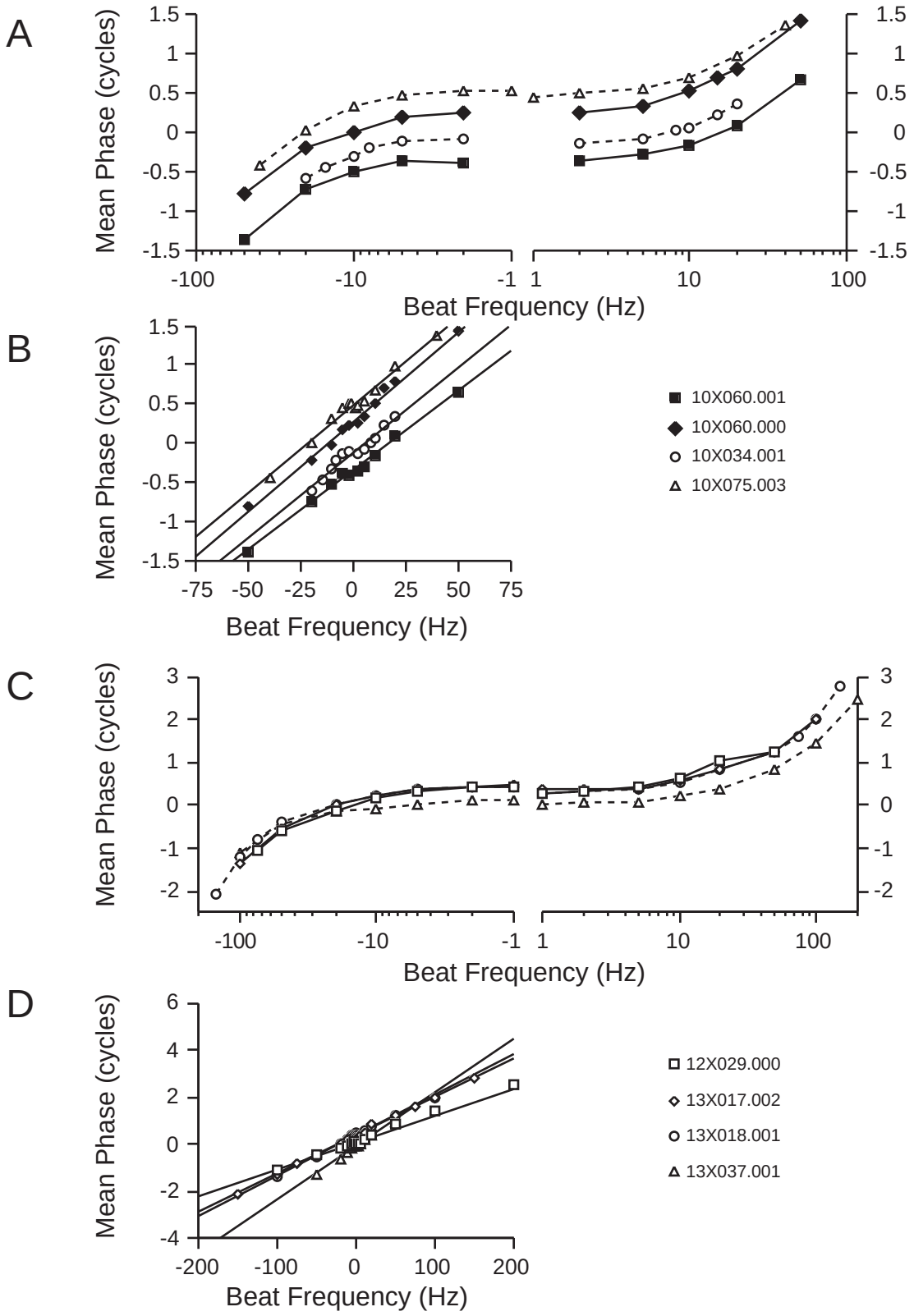

A
Mean Phase (cycles)
Beat Frequency (Hz)
B
Mean Phase (cycles)
Beat Frequency (Hz)
10X060.001
10X060.000
10X034.001
10X075.003
C
Mean Phase (cycles)
Beat Frequency (Hz)
D
Mean Phase (cycles)
Beat Frequency (Hz)
12X029.000
13X017.002
13X018.001
13X037.001

Figure 5.6 Phase shifts are linear with beat frequency in the positive or negative direction. **A:** These four units show a variety of mean phase values at low beat frequencies, and their phase/frequency functions remain parallel out to 50 Hz in either direction. **B:** Data from A on linear axes, showing the linear fits used to measure group delay. **C, D:** Additional units tested to greater extremes of beat frequency, showing more diversity in the slopes of theiir phase shifts.

5.5.3 Effects of beat frequency on discharge rate

The overall spike rate at which a neuron fires may carry information about the binaural beat stimulus in addition to, or in the absence of, synchronized firing at a preferred phase. This would be unlikely in anesthetized cortex, where previous studies showed that cells typically fire only one spike per cycle, making firing rate a monotonic function of beat frequency (Reale and Brugge 1990). Cortex in the awake animal can respond throughout the IPD cycle (see Figure 1, 3), so spike rate need not increase linearly with beat rate. Figure 5.7A shows normalized rate-based modulation transfer functions (rMTFs, Krishna and Semple 2000) from six cells tested with beat frequencies from 0.1 to 100 Hz. All are nonmonotonic functions, indicating an optimal range of beat frequencies between 5 and 20 Hz where the cell discharges maximally over the 20 seconds of stimulus presentation. Figure 5.7B shows the data from the same cells in terms of average spikes fired per beat period. In almost every case, the number of spikes per cycle falls off as a power function. At a low beat rate of 1 Hz, the cells fire on the order of 10-100 spikes/s (or spikes/beat), but at higher rates (20-100 Hz in these cells) they begin to behave as described in anesthetized cortex, firing only one or two spikes per cycle, then skipping cycles as the curves dive below 1 spike/beat at very high rates of modulation (dashed gray line).

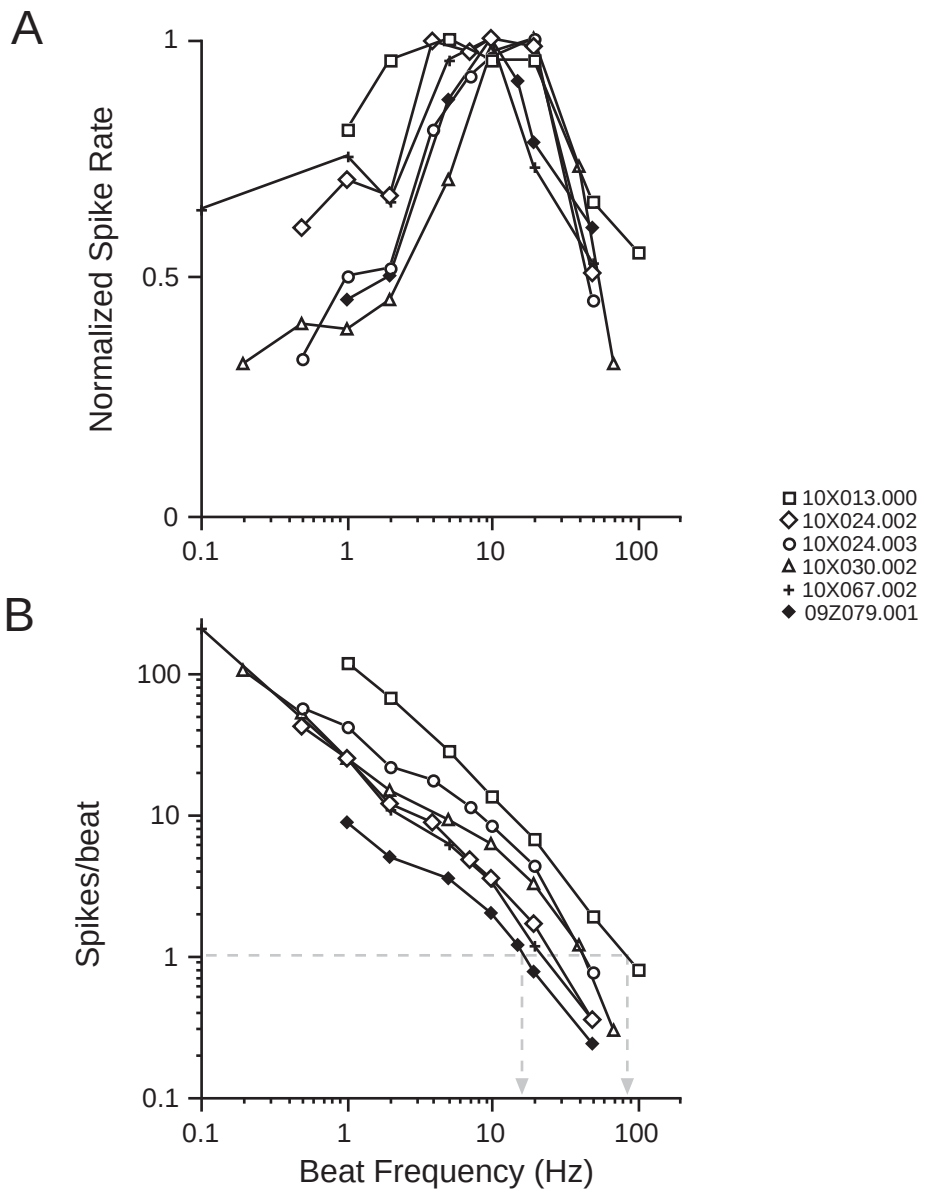


Figure 5.7 A: Spike rate modulation transfer functions (rMTFs) for six units, normalized to maximum spike rate for each run (actual rates: 142, 85, 63, 35, 33, 20 spikes/s). All show some degree of band-pass tuning of overall spike rate (measured over the full 20 seconds of stimulation) to binaural beat frequency, a phenomenon rarely observed in midbrain. All peaks occur at beat rates between 5-20 Hz. **B:** The same data plotted in terms of average spikes per beat cycle, which falls off as a power function (linear on log axes). Values <1 spike/beat (gray dashed lines) indicate the cell is skipping cycles. This contrasts sharply with responses to binaural beats in anesthetized cortex (see text).

Spitzer and Semple (1998) examined the effect of beat frequency on overall spike rate in the IC, where responses generally resemble those in awake cortex: spikes are fired throughout the phase cycle. In all but a handful of IC neurons (35/44), beat rate (and direction) did not affect spike rate. To obtain a similar estimate in this sample, an index of firing rate contrast was computed for each unit tested across beat frequency:

$$(FR_{max} - FR_{min}) / (FR_{max} + FR_{min})$$

where FR_{max} and FR_{min} denote the maximum and minimum evoked firing rate; a contrast of 0.33 would indicate a doubling of spike rate at one beat frequency relative to another. Of 183 cells, 73% had a rate contrast of 0.3 or higher, which corresponded well to band-pass rate tuning by visual inspection of rMTFs. Band-pass rate tuning to beat frequency in awake cortex, like that seen in Figure 5.7A, is the rule rather than the exception, a seemingly emergent property observed rarely in the IC and never in the superior olivary complex (Spitzer and Semple 1998).

5.6 Tuning to IPD II: Effects of center frequency

Tuning to IPD, as to any other parameter in the auditory system, is multiplexed with sensitivity to several other features of the acoustic stimulus. The fundamental tuning property of auditory neurons is frequency selectivity, established in the cochlea and refined in central auditory structures up to and including the cortex. All other stimulus parameters are thus analyzed in frequency-specific channels of a

particular bandwidth. The low-frequency, IPD-tuned neurons of core auditory cortex respond over a finite range of carrier frequencies, and carrier frequency can exert a predictable influence on the measured properties of the binaural beat response discussed in the previous section: firing rate, synchrony, and particularly mean phase.

5.6.1 The phase/frequency response area

The dual tuning of low-frequency cells in auditory cortex to frequency and IPD is exemplified in Figure 5.8 (left panels, A and C), where spike rate is represented as a contour map in phase-frequency space for two cells. A vertical slice through the contours at any frequency would be the beat period histogram at that center frequency. Significant mean phase points and a linear fit are overlaid. The cell in the upper panel shows little effect of frequency on mean phase, though the response area shown by the phase/frequency contours is well circumscribed. The range of IPDs over which the cell responds is broadest when carrier is at the (rate-defined) best frequency of the cell, 1.6 kHz (thus VS might show an inverse relationship, being higher in the carrier frequency range above and below BF where the IPD response area is restricted). The cell in the lower example shows a clear phase advance with increasing frequency, and a broader range of effective carrier frequencies. On the right, the same data are plotted in raw form as polar spike rasters to emphasize the circular nature of the stimulus. Mean phase points are

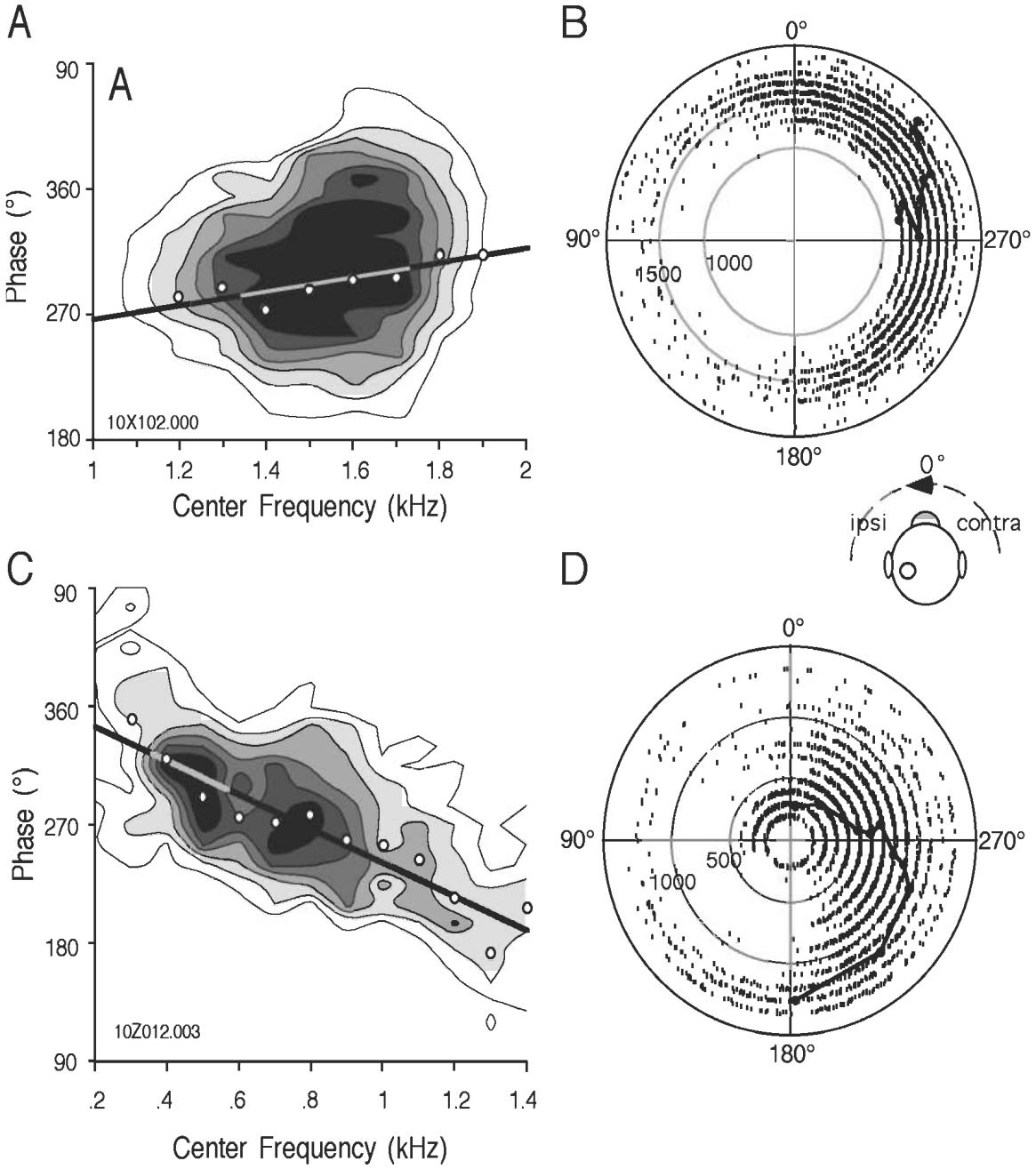

A
A
90
360
270
180
Phase (°)
10X102.000
1
1.2
1.4
1.6
1.8
2
Center Frequency (kHz)
B
0°
90°
270°
180°
1500
1000
0°
ipsi
contra
C
90
360
270
180
90
Phase (°)
10Z012.003
.2
.4
.6
.8
1
1.2
1.4
Center Frequency (kHz)
D
0°
90°
270°
180°
500
1000

Figure 5.8 A, C: The dual tuning of cortical cells in the carrier frequency and IPD dimensions represented as a contour plot in phase/frequency space, where black indicates the maximal response. The maximal spike rate in both examples was approximately 68 spikes/s (55 [A] or 56 [C] spikes in a 20 ms bin, averaged over 40 total beat periods). Significant mean phase points and a linear fit (weighted by Rayleigh strength) are overlaid. **B, D:** Corresponding raw data presented as spike rasters in polar coordinates, where angle indicates instantaneous IPD and carrier frequencies are arrayed along the radius. The 0° point at the top of each plot corresponds to midline for the monkey, as suggested by the inset schematic: an overhead view of the monkey showing the direction of phase modulation relative to the recording chamber over the left auditory cortex. Although IPD does not map directly to azimuthal space (frequency and head size are factors), the spikes cluster to the right side of both plots, corresponding to the auditory hemifield contralateral to the recording site. As in A and C, significant mean phase points are overlaid.

overlaid. As can be seen from the contour plots, spikes in the upper raster cluster around 270° regardless of frequency, whereas the spikes of the lower cell are initially nonselective for phase at 300 Hz, but form a clockwise spiral moving outward as frequency increases. The inset shows the equivalent overhead view of the monkey. Presuming it hears the stimulus as a human would, the beat creates an illusion of auditory motion in the azimuth beginning at the midline, rotating counterclockwise around (or within) the head 90°, and reappearing at the right ear after some period of ambiguity before returning to midline. In both examples, the spikes cluster at an IPD corresponding to azimuthal space contralateral to the recorded (left) hemisphere, though true apparent azimuthal location is a function of both IPD and carrier frequency.

5.6.2 Linear phase/frequency functions: Characteristic Delay and Characteristic Phase

If the IPD-tuned cells of the auditory cortex ultimately derive their sensitivity from coincidence detection of a fixed interaural delay in the MSO of the brainstem, a change in the center frequency of the input should affect the phase of the output in a linear manner. As carrier frequency is increased, the period of the signal becomes shorter and a given delay will account for a greater proportion of the cycle. The slope of the function tracking mean phase across frequency expresses the change in mean phase per cycle of frequency, the units of which (cycles/[cycles/s]) reduce to

s. This time is called the characteristic delay (CD) of the cell. The intercept of that linear plot – the phase of the response at "zero frequency" – represents the cell's preferred phase regardless of frequency, called the characteristic phase (CP) (Yin and Kuwada 1983b). All cells showing significant synchrony to the binaural beat at three or more carrier frequency points were subjected to linear regression. To take the strength of frequency tuning into account when determining CD and CP from these plots, regression lines were weighted by the Rayleigh statistic at each point. A simple bootstrap technique was used to determine statistical significance of each fit (Yin and Kuwada 1983b, appendix).

Linear phase/frequency functions were collected from 150 cells (83% of 180 tested, 30 cells failed to yield satisfactory linear fits despite showing significant synchrony at 3 or more carrier frequencies). Of these, 100 cells (67%) had negative slopes indicating a progressive phase lead with increasing frequency. Figure 5.9 shows a collection of phase-frequency plots with both negative (A) and positive (B) slopes. Negative slopes were not only more common, but steeper on average, suggesting that cells failing to show a phase advance showed only a weak phase lag (mean negative slope=-0.00021, positive=0.00015, significantly different by Wilcoxin signed-rank, $p<.005$).

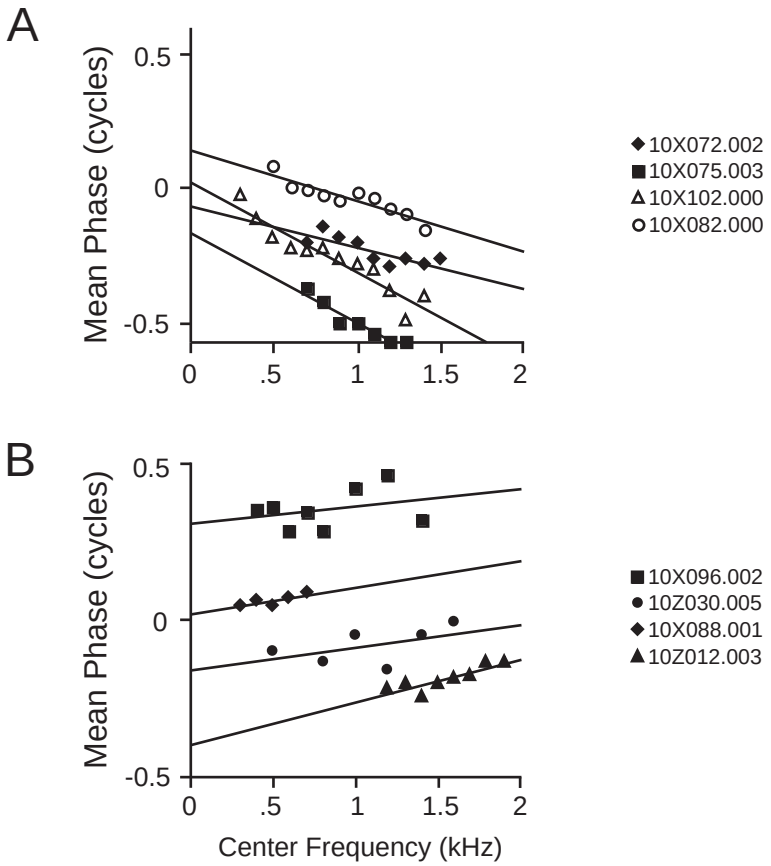


Figure 5.9 **A:** Linear phase/frequency plots with negative slopes, which were common among the population. The slope of each function indicates the cell's characteristic delay (CD), and the "zero-frequency" intercept its characteristic phase (CP). **B:** Cells with positive slopes to their phase/frequency functions were less common, and fits were often poorer.

The populations of CD (slope, converted to μs of the beat period) and CP (intercept, in cycles) values are plotted against one another in Figure 5.10A, with marginal distributions for both metrics; an inset circular histogram replots the CP values to emphasize the clustering of their distribution at a mean phase of 333° on the circular axis (the significance of this clustering can be assessed using the same metrics used for spikes: VS=0.34, significant by the Rayleigh test; the mean CP and VS of the left and right hemisphere distributions were almost exactly identical, data have been pooled). This corresponds to an azimuthal location contralateral to the recorded hemisphere. The distribution of CDs (upper histogram) shows the bias toward negative values, again indicating a preference for a lead at the contralateral ear. The two metrics are not independent, as CPs from cells with negative (contralateral) CDs are much more tightly clustered than those from cells with positive slopes: points on the left side of the scatterplot are clearly centered around 0, while points on the right (ipsilateral) side of the plot are scattered in phase. This reinforces the suggestion apparent from the example functions in Figure 5.9: the intercepts of negative functions cluster at the origin, whereas the intercepts of positive functions are varied. No effect of best frequency on CD was apparent in these data (but c.f. McAlpine et al. [2001] using delayed noise in the guinea pig IC).

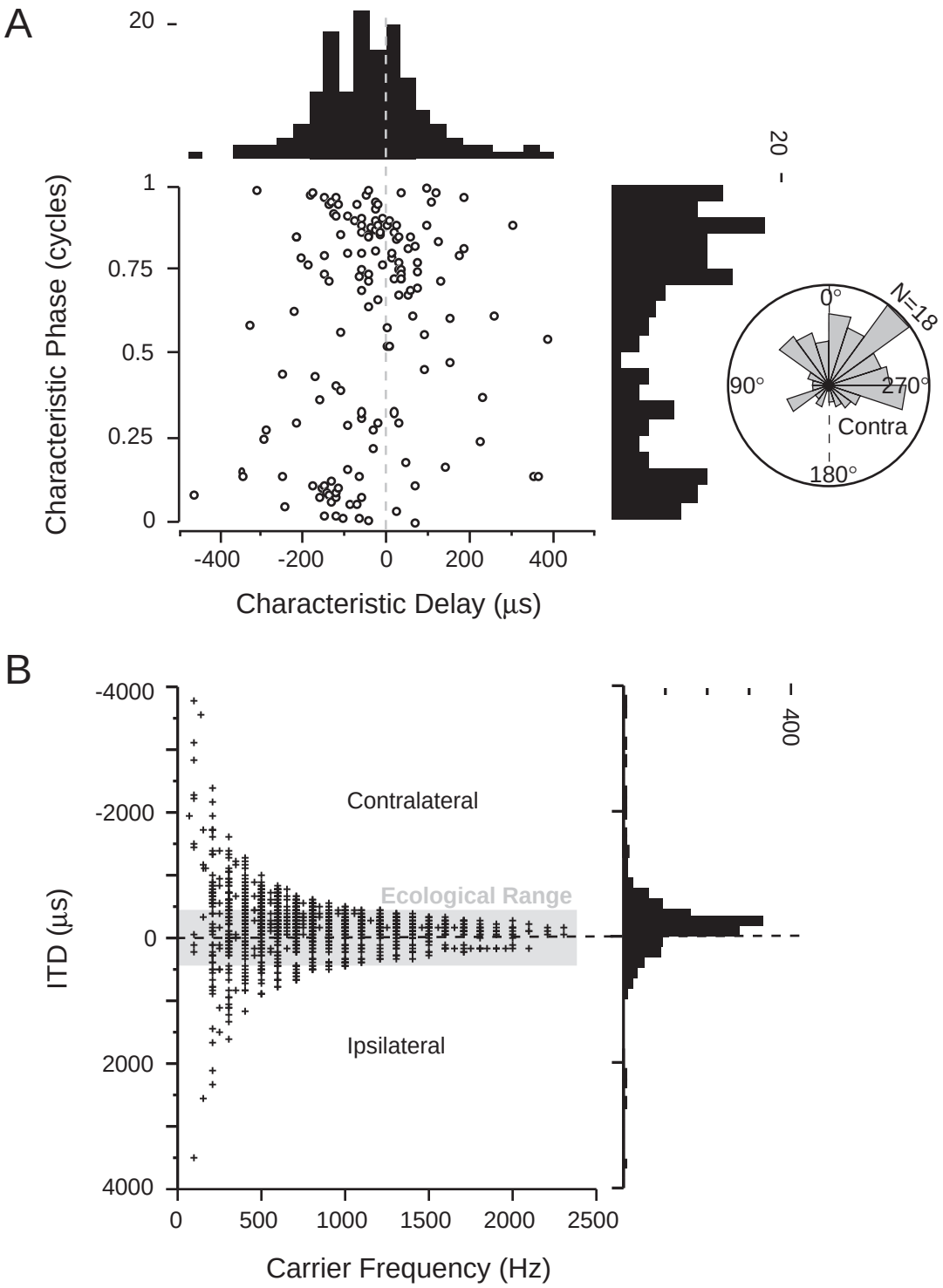

A
20
20
1
0.75
0.5
0.25
0
Characteristic Phase (cycles)
-400
-200
0
200
400
Characteristic Delay (µs)
0°
N=18
90°
270°
Contra
180°
B
-4000
-2000
0
2000
4000
ITD (µs)
Contralateral
Ecological Range
Ipsilateral
400
0
500
1000
1500
2000
2500
Carrier Frequency (Hz)

Figure 5.10 A: Characteristic phase and characteristic delay (n=150) plotted against one another, with marginal distribtuions for both. Dashed gray line separates cells with negative (contralateral) and positive (ipsilateral) CDs, emphasizing that cells with negative CDs have CPs clustered around 0°. The overall distribution of CDs is biased toward negative, contralateral values (upper histogram), with 98% of CD values falling within the plausible ecological range of the macaque. CPs have a mean value of 333° (right histogram, VS=0.34, significant by Rayleigh test). The rose plot inset at right shows this distribution in polar coordinates, making the clustering of CP values more apparent. **B:** Distribution of ITDs computed from the mean phase of every significant binaural beat response (n=1280 responses from 248 single units). The maximum ITD at each carrier frequency is equal to half the modulation period, which constrains the distribution to the shape observed. The marginal distribution at right clarifies the bias toward negative (contralateral) ITDs. The shaded area marks the maximum ITD (±400 μs) for a macaque, given its head size. Only 71% of points fall within this range, though the proportion is 94% for carriers ≥ 800 Hz.

The range of interaural delays that might be encountered in the environment is restricted by the head size of the animal, and the speed of sound (Kuhn 1977). Of 150 cells that showed a CD, 147 (98%) fell into the range of ecologically plausible ITDs for macaques, about ±400 μs (as computed from the measured head sizes of our animals, and consistent with direct measurements by Spezio et al. [2000]).

5.6.3 ITD across carrier frequency: coding by peaks vs. slopes

The range of CDs measured in macaque auditory cortex appears to fall within the ecological range of ITD. But CD is an abstracted measurement of a neuron's response to ITD independent of frequency, whereas the tones (and other stimuli) processed by the auditory system consist of energy at a given frequency or frequencies. Examination of the distribution of preferred ITDs across all carrier frequencies reveals whether the peaks of these tuning functions span the ecological range of ITDs, or whether an alternative coding scheme is more plausible. Figure 5.10B plots all measured ITDs, computed from all significant mean phases of binaural beat responses (1280 responses from 162 cells in monkey X, 86 from Z; 425 from the left hemisphere, 855 right). Effective carrier frequencies spanned 70 – 2300 Hz (median 700), SPL was optimized for each cell (20-85 dB, median 60), and all points were collected at beat frequencies <5 Hz. For values over .5 cycles, 1 cycle was subtracted to yield a negative ITD – there is no way to disambiguate ITDs longer than half a period using a cyclical stimulus. The expected bias toward

negative (contralateral) values is clarified by the marginal distribution at right. The maximum range of ITDs at each frequency is constrained to half the stimulus period, defining the outer edges of the distribution. Ecological range, however, is constant (constrained by head size), such that the proportion of ITD peaks falling within the ecological range is highly dependent on carrier frequency. Overall, 71% of points fell within the approximate ecological range for the macaque (shaded area), 94% if only points above 800 Hz are considered. Reale and Brugge (1990, their Figure 11A) report 85% and 99%, respectively, for a their analysis in anesthetized cat (the discrepancy is most likely attributable to a dearth of points below 400 Hz in their sample). Like McAlpine and colleagues (2001), we take these seemingly "unrealistic" tunings at low frequencies as evidence that IPD-tuned cells are not feature detectors for their respective mean phases, but that the slopes of these functions, falling within the ecological range, provide the dynamic range to code IPD.

5.7 Directional sensitivity and behavioral relevance

Chapter 2 demonstrated the ability of macaques to discern the direction of dynamic IPD, then examined the role of behavioral state in shaping neural responses to those stimuli. This chapter, thus far, has presented a broader survey of the IPD-sensitive neurons at the low-frequency border between AI and R in the auditory cortical core. Figure 5.11A reproduces the psychophysical thresholds of both animals for

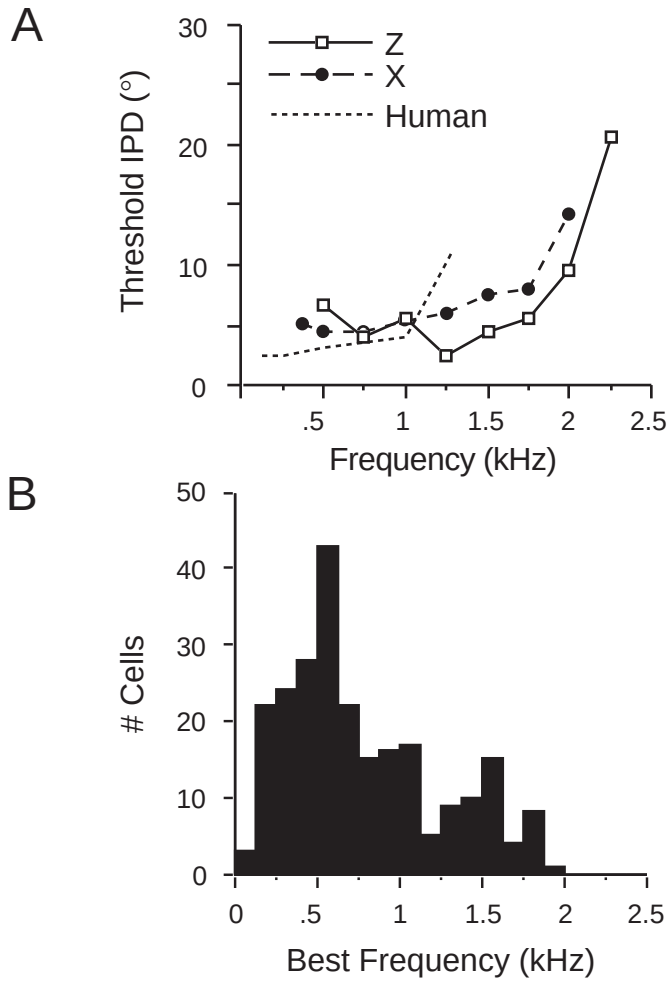


Figure 5.11 **A:** Thresholds for lateralization of trapezoidal shifts in IPD from midline across carrier frequencies in two macaques (see Chapter 2 for threshold determination). Dashed curve shows thresholds for human lateralization of static phase offsets (Klump and Eady 1956). Both animals performed very poorly or refused to work at frequencies outside the range shown, and both show elevated thresholds at frequencies > 1.5-1.75 kHz. **B:** Distribution of BFs for all binaural beat-sensitive cells (n=268, median 0.65 kHz). Sensitivity to IPD among the neural population seems to decline with the rise in thresholds.

lateralizing the direction of dynamic IPD sweeps across carrier frequency; below, 5.11B presents the distributions of best frequencies for the binaural-beat-sensitive population. Thresholds are lowest in the frequency range where the bulk of beat-sensitive neurons are tuned, and thresholds become less precise as the IPD sensitivity of the population wanes at higher carrier frequencies. This relationship is only circumstantial in the absence of deactivation (ablation or cooling) experiments selective for the low-frequency cortical core, but it merits a closer look at the coding properties of the beat-sensitive population as it regards psychophysical discrimination of IPD.

5.7.1 Effect of beat direction on discharge rate and synchrony

Cells that are selective for binaural beats in a particular direction would provide an obvious substrate for the dynamic IPD task in Chapter 2. But the prevailing characteristic among cells tested at "positive" and "negative" beat frequencies was symmetry, of both rMTFs and tMTFs. (By convention, "negative" means the apparent motion of the stimulus is heading *into* the contralateral hemifield as it passes midline.) In other words, firing rate and synchrony were dependent on absolute beat frequency (sections 5.5.1 and 5.5.3) but largely independent of which ear was leading in phase. Figure 5.12A plots rMTFs for four cells tested over a wide range of beat frequencies in both directions, and regardless of overall firing rate, the functions are symmetric about zero (symbols at center are firing rates to an

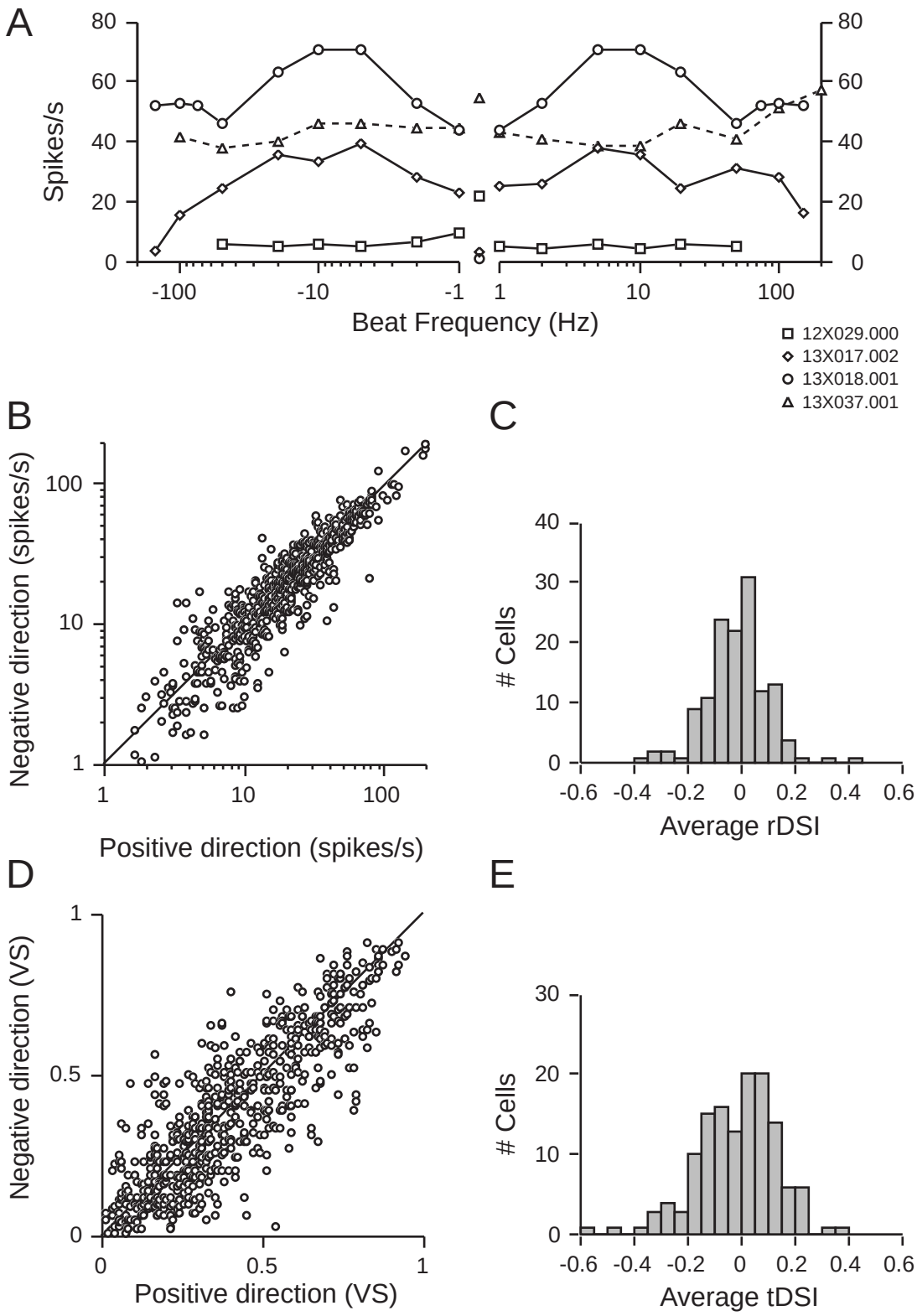

A
Spikes/s
Beat Frequency (Hz)
12X029.000
13X017.002
13X018.001
13X037.001
B
Negative direction (spikes/s)
Positive direction (spikes/s)
C
Cells
Average rDSI
D
Negative direction (VS)
Positive direction (VS)
E
Cells
Average tDSI

Figure 5.12 A: rMTFs for cells tested with binaural beats in the positive and negative directions are generally symmetrical, showing no rate-coding of IPD modulation direction. Points between the positive and negative sides of the graph mark the diotic tone control (0 Hz). **B:** Population summary of spike rates evoked by beats of equal frequency in the positive and negative direction (N=661, from 135 neurons). Dashed line indicates identity; there is no trend across the population for points to fall above or below the line (paired *t*-test, p>.56). **C:** Distribution of the rate Direction Selectivity Index (see text) averaged for each neuron used in panel B. The mean of the distribution is not different from zero. A rDSI value of 0.3 indicates a doubling of spike rate (on average) in one direction, but very few cells show this behavior. **D:** Population of VS values from the same responses shown in panel B, again with no consistent directional bias in synchrony. **E:** Temporal DSI values, averaged for each neuron as in C. The distribution of tDSI is broader than rDSI, but the mean is not different from zero, and only 6 cells exhibit a doubling of vector strength in one direction over the other.

unmodulated control tone at 0° IPD). For all beat frequencies tested in two directions (N=661), the firing rates elicited in the positive and negative directions are plotted against one another in 5.12B, without any apparent bias (insignificant by paired *t*-test). To examine direction selectivity cell-by-cell, a rate Direction Selectivity Index (rDSI) was computed at each beat rate, using a difference/sum metric similar to that used to assess firing rate contrast in section 5.5.3:

$$rDSI=(FR_{pos}-FR_{neg})/(FR_{pos}+FR_{neg})$$

where FR is firing rate, and the subscript indicates the positive or negative beat direction. An average rDSI was computed for each neuron (N=135), the distribution of which is shown in Figure 5.12C. Values of +/- 0.33 would indicate a doubling of firing rate in one direction relative to the other, but the bulk of the values fall below +/- 0.2, and there is no directional bias in the distribution (mean not significantly different from 0). The same analysis was carried out for all vector strength points (5.12D) and the temporal DSI showed a somewhat broader distribution but no significant bias away from 0 (5.12E). Although 6 of 135 cells showed a doubling of firing rate or VS in one direction, there was no consistent trend toward direction selectivity in the population, suggesting that most cortical neurons encode IPD modulation within their receptive fields regardless of the direction of apparent motion.

How, then, may these animals reliably detect phase shifts as small as 5° given the observed cortical representation? And what has been gained by observing responses in the awake cortex, if all these extra spikes don't carry information? The fact that responses in the awake cortex can be continuous throughout the phase cycle may have wide-ranging implications for the encoding of IPD, as the next section will explore.

5.7.2 A d′ analysis of peaks vs. slopes of IPD tuning functions

Cells in the anesthetized cortex (Reale and Brugge 1990) are almost digital in nature, in that their phasic, phase-locked responses provide information about only the preferred phase of the cell (i.e., the instantaneous IPD is exactly the preferred phase, or it is not). Cells in the awake preparation can respond to IPD over a wide range, allowing a large population of cells encoding a given stimulus. The binaural beat responses reported above were used to construct simple neurometric functions of discriminability (Tolhurst et al. 1983, Parker and Hawken 1985, Britten et al. 1992, Skottun et al. 1998, 2001), allowing a comparison between the IPD discrimination of the animal and individual neurons of the auditory cortex.

Example cells were chosen with strong, well-tuned beat responses, and the spike times were sorted into matrices of 16 columns, each corresponding to a phase bin of 22.5°. Each row of the matrix represented spike counts from one period, where a

1 Hz beat presented for 2 10 s trials produced 20 complete IPD cycles. Treating each period as an individual "trial" – as opposed to folding them into a single period histogram – allowed a computation of the mean spikes fired and an estimate of the standard deviation of spike count within each phase bin. Figure 5.13A graphs these mean spikes/period with error bars representing one standard deviation (same cell used in Figure 5.1). D-prime was computed between every phase bin and the following bin as

d′ = [difference in mean spikes] / [average standard deviation]

and is overlaid in gray (right-hand axis). This response peaked in the bin from 270°-292.5°, near the average characteristic phase of the population, but the d′ statistic shows that discriminability does not peak near the peak of the response. To the contrary, d′ dips at the maximum and minimum firing rates, and peaks on the slopes of the response at 0° and 180°, where local dynamic range is maximal. Thus a cell like this one, with a strongly tuned beat response and d' maximal at 0°, should be an ideal neural unit for the monkey's psychophysical task of detecting IPD modulation at the midline.

Figure 5.13B shows a sample of ten d′ functions, with maximal discriminability spanning a range of IPDs. When these functions are aligned to the peak response of each cell (Figure 5.13C), it is apparent that the properties described in the example unit from panel A apply to the population at large: discriminability is greatest on

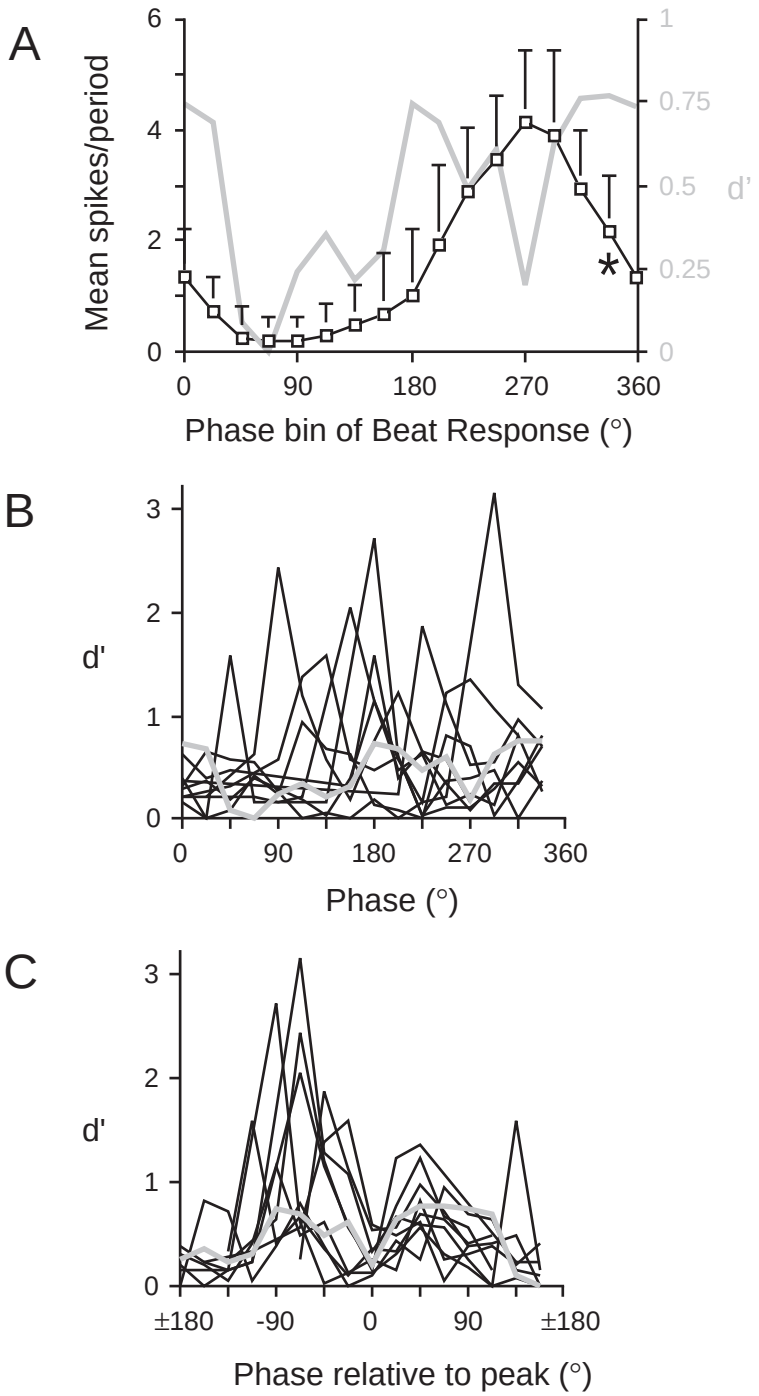

A
Mean spikes/period
6
4
2
0
1
0.75
0.5
0.25
0
d'
*
0
90
180
270
360
Phase bin of Beat Response (°)
B
d'
3
2
1
0
0
90
180
270
360
Phase (°)
C
d'
3
2
1
0
±180
-90
0
90
±180
Phase relative to peak (°)

Figure 5.13 A: Example cell with the flank of its beat response at 0°, ostensibly ideal for the midline discrimination task. Spikes have been binned in 22.5° increments, each point is the mean number of spikes per period, error bars are +1 SD (mean and standard deviation computed from 20 repetitions of the beat cycle, each treated as one "trial" – see text). Gray line plots d' discriminability between each point and the one following (right-hand ordinate). As predicted, d' is maximal at 0° (midline) and 180°, and has local minima at the peak (mean phase) and trough of the response. Asterisk designates the point taken as the "noise" distribution for comparison with the three "signal" distributions before and after in the ROC analysis (see text). **B:** A collection of d' functions from cells with some of the strongest Rayleigh strengths in the population. The example in A is the bold gray line, and although its maximal d' overall is comparatively weak, it has the highest d' at the 0° midline, where the psychophysical task is centered. **C:** The same d' functions aligned to the peak of each underlying response. As predicted, d' functions are minimal at the peak (0°) and 180° from the peak (±180°), and have their maxima about 90° before the peak, on the flanks of the response. This bias toward higher d' values on the left of the graph signifies that the rising phase of the response (as instantaneous IPD approaches the peak) is steeper than the falling phase (as IPD moves away from the peak).

the slopes of the response, and is minimal at the peak and trough of the response (the peak is 0° on this axis, the trough 180° away). What is also revealed is a tendency for the rising slope of the response to be steeper than the falling slope (e.g., d′ is greater on the left side of the graph, when the instantaneous IPD is *approaching* the peak, and lower on the right, as IPD moves *away* from the peak). The same effect is seen when d′ is computed from beats in the negative direction (data not shown), suggesting that this is not a result of inherent directional preference.

5.7.3 Neurometric analysis

To compare the phase discrimination of the cell to that of the animal, signal detection theory (Green and Swets 1974) was applied to produce the Receiver Operator Characteristic (Bradley et al. 1987, Britten et al. 1992, Skottun 2001; Chapter 2 above) for two example neurons with maximal d′ near the midline, where they would be most valuable in the animals' psychophysical task. (The advantage of ROC analysis over d′ is that no assumptions are made about the distribution of spike rates; because spike rates can not go negative, they are seldom normally distributed around low mean rates.) The bin in Figure 5.13A marked with an asterisk was used as the "noise" distribution of spike counts, and counts from the three bins before and after were the "signal" compared at a range of criterion values spanning the full dynamic range of the response. The probability of a spike count

from the signal being higher than the criterion was defined as P(hit), and the probability of a spike count from the noise being higher than the criterion was P(false alarm). Plotting P(hit) against P(false alarm) for each criterion on double-probability coordinates produces the ROC curves in Figure 5.14A, one for each "signal" value being compared to the "noise" (i.e., response at each adjacent phase being compared against response at 0°). Integrating under each curve yields a point in the neurometric function, producing the curve in Figure 5.14B (open squares). The phase axis was reversed to allow fitting with a cumulative Weibull (Quick) function. The neurometric functions from this cell and another, similarly tuned, cell from the same track are displayed along with the psychometric function for dynamic IPD discrimination at the closest carrier frequency tested.

The animal's behavioral function is far steeper than either of the neurometric functions, even though these cells were among the strongest in the sampled population. The same analysis for two cells of lower BF in the other animal (data not shown) again revealed that the animal's ability to discriminate changes in phase around the midline far outstrips its best cortical cells. This could be taken as evidence against a phase representation by firing rate, or as an indication that a large number of cells with uncorrelated noise must pool their responses to attain the demonstrated behavioral acumen.

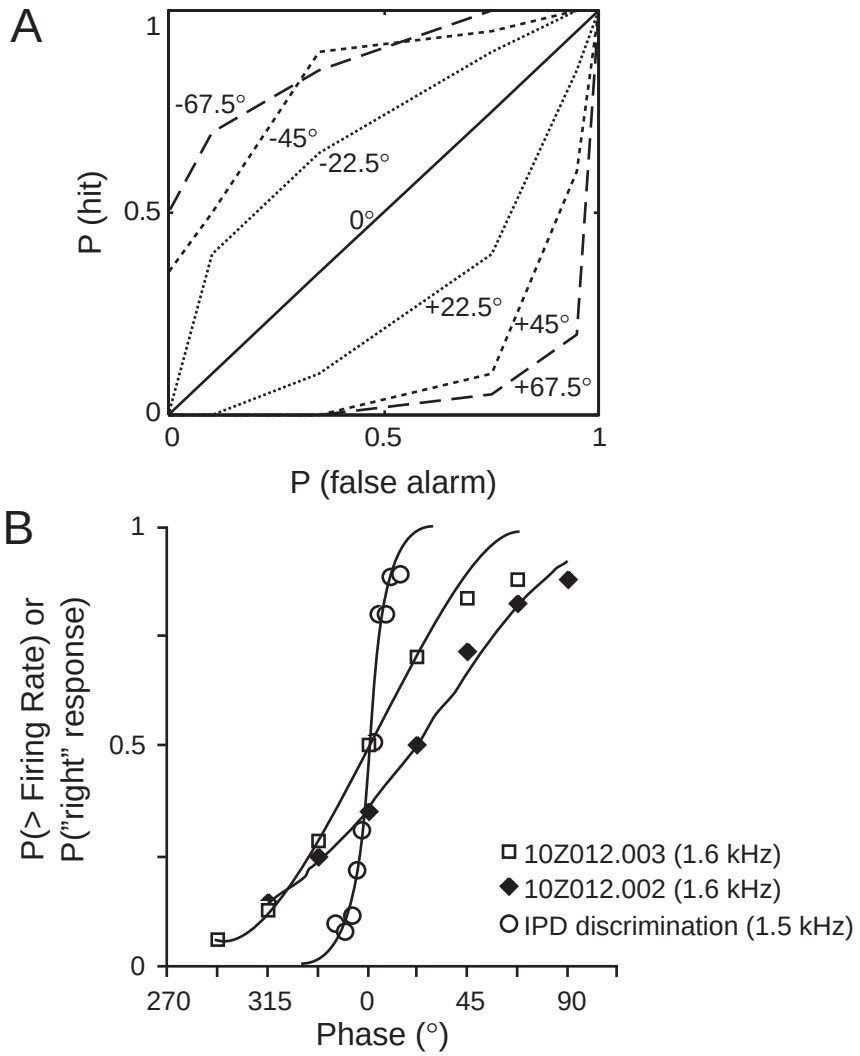


Figure 5.14 A: ROC curves comparing spike count distributions in the "noise" bin and six flanking "signal" bins. Abscissa, P(hit) is the probability of the spike count in the signal distribution being greater than a criterion value; ordinate, P(false alarm), is the probability of the spike count in the noise distribution being above that criterion. Each curve joins probabilities at a number of criterion values spanning the range of spike counts. The diagonal is a curve comparing the noise distribution to itself. The area under these curves depends on the separation of the signal and noise spike rate distributions – if they do not overlap at all, the curves will follow the edges of the plot, and integrate to 1 (if signal is higher than noise) or 0 (if signal is lower than noise); if they overlap entirely, the curve will follow the diagonal and integrate to 0.5. **B:** Neurometric functions obtained by integrating under ROC curves from two cells, and the psychometric function of the monkey in the same frequency range (open squares are from the cell in panel A). For the neurometric functions, the abscissa represents the overall probability that the spike count was greater for that stimulus than for the noise (thus the curve is constrained to go through P=0.5 at the phase value used as the noise bin – 0° for the open squares and 22.5° for the filled diamonds). For the psychometric function, the abscissa is the probability that the monkey responded by pressing the button on the right (indicating that IPD moved clockwise). The ordinate has been reversed to allow fitting with a cumulative Weibull function.

5.8 Discussion

Sensitivity to ITD was first reported in the alert macaque by Brugge and Merzenich (1973), who found cells in AI selective for ITDs of pure tones over a range of frequencies comparable to that reported here. Although the binaural beat stimulus has been applied in the IC (Yin and Kuwada 1983ab) and the cortex (Reale and Brugge 1990) of anesthetized cats, our data reveal several features previously unobserved in the neural representation of continuously-varying IPD. Bandpass tuning of discharge rate as a function of binaural beat frequency is nonexistent in the SOC and rare in the IC of the anesthetized gerbil (Spitzer and Semple 1998), but common in our sample of cortical responses. In the absence of anesthesia, cortical neurons can modulate their rate of spike discharge throughout the IPD cycle. Qualitatively, the audible modulation of the hash response in response to binaural beats in the awake cortex suggested a homogeneity of tuning in the local neural population, especially in the densely packed middle layers, the site of thalamic input. These high discharge rates and lack of transient, strictly phase-locked responses force a reconsideration of the mechanisms by which azimuthal location may be extracted from the response profile of the cortical pool, a population of neurons that – much like the primate visual cortex – appears to operate in a "high-input regime" (Shadlen and Newsome, 1997).

Cortical cells appear not to be tuned so that the distribution of mean phases covers the full range of IPD, but are generally broadly tuned to contralateral space (Figure 5.10), a model of low-frequency sound localization in agreement with that of McAlpine et al. (2001). Coupled with the demonstration of the macaque's psychoacoustic acuity in Chapter 2, this comprises the first comprehensive survey of IPD representation – neural and perceptual – in the alert primate.

5.8.1 Binaural beat frequency

The standard beat rate of 2 Hz used to characterize the tuning sharpness and mean phase of IPD tuning was an effective compromise between speed (number of beat cycles presented in a limited recording time) and accuracy, since beat rate did not affect the vector strength or mean phase at rates < 5 Hz. Although the beat was an unbiased estimator of static mean phase, vector strength of the beat response was higher than that for static IPD responses, consistent with observations in anesthetized cat cortex by Reale and Brugge (1990). Their population entrained to beat frequencies up to 35 Hz, though they show few examples of cells entraining beyond 20 Hz, the median cutoff frequency in this study (Figure 5.4B). As would be predicted for a phasic response, total spike rate under anesthesia was effectively a monotonic function of frequency (a constant number of spikes was fired per cycle), falling off at higher beat rates. This contrasts with our finding of band-pass

rate tuning within the synchronized range, and the power-law decrease in the average number of spikes fired per cycle of IPD (Figure 5.7).

Langner and Schreiner (1988) predict a transformation in the representation of amplitude modulated (AM) sounds from the auditory nerve to central structures, in which a temporal code (synchrony to the stimulus envelope) is replaced by a rate code as synchrony breaks down. Although IPD is not encoded prior to binaural convergence in the MSO, it could be argued that synchrony to time-varying IPD is re-coded between brainstem and cortex as the spatial aspects of the stimulus are abstracted. Given the expected effect of successive low-pass filtering of the modulated signal at each synapse between brainstem, midbrain, thalamus and cortex, it is surprising that the cutoffs reported for beats in this study are as high as in any other structure, indicating that the emergence of discharge rate tuning need not occur simultaneously with the degradation of temporal fidelity. An additional factor limiting the response to binaural beats, but not AM signals, is that faster beat rates are created by widening the frequency difference between the tones at each ear, eventually to the extent that no common channel exists to cross-correlate the signal. A given beat rate requires a greater frequency difference, in octaves, around a lower center frequency, but this does not appear to be the limiting factor in our population because synchrony cutoff was not correlated with best frequency (data not shown).

Many of these beat rate effects have analogs in human psychophysics. Listeners report hearing two individual tones, rather than a beat, at frequency separations between 25-30 Hz, though the threshold is frequency-dependent and peaks for carrier frequencies of about 300-600 Hz, near the mode of the distribution of BFs for beat-sensitive cells (Licklider et al. 1950, my Figure 5.11B). The decay of synchrony in the population at beat frequencies >20 Hz corresponds well to Licklider's observed transition between perceived "fluctuations" and "roughness" around the same point in human listeners (Licklider et al. 1950), but is well above the threshold for clear perception of auditory motion, around 2-5 Hz (Blauert 1972, cited and corroborated by Grantham and Wightman 1978). Subjective differences in the quality of this roughness, below the limen for detection of separate tones at the two ears, may be attributed to residual rate tuning in the absence of synchrony among IPD-sensitive neurons of the cortex.

The evidence for phase lag in the binaural system for varying ITDs at rates greater than 5 Hz was suggested by the psychophysical study of Grantham and Wightman (1978), who concluded that "the binaural system is quite sluggish in its response to fluctuating [ITDs]". Listeners detected a static target tone in a noise masker that "moved" in azimuth by varying sinusoidally the ITD of the noise at the two ears, but release from masking lagged behind the position of the noise when the masker

oscillated at rates >5 Hz. Even 5 or 10 Hz beats, however, seem well beyond the expected range of ecological relevance for auditory motion – few, if any, objects in the environment circle the listener at 3600°/s. One product of the binaural system that is not purely "spatial" is the reduced threshold for the detection of a tone in a noise masker on the basis of IPD in the signal, commonly called the binaural masking level difference (BMLD). Grantham and Wightman (1988) show that simulated motion (using a binaural beat instead of a static IPD in the target tone) does *not* enhance the detectability of the tone in a BMLD task. So synchrony to very fast rates of dynamic IPD may have no obvious utility to sound source separation (e.g., "cocktail party" situations) or acoustic motion detection in the real world. A more likely explanation is that synchrony to dynamic IPD up to (and beyond) 20 Hz is an epiphenomenon of the precision of the binaural system in detecting static IPDs and, from those, computing acoustic motion. Although an animal may never have to locate an object moving in azimuth at these speeds, the ability of its cortical neurons to follow such rates could arise from the precision of the cross-correlator in the brainstem, leaving vestigial synchrony at the cortical level.

A model of responses to dynamic phase in the IC (Borisyuk et al. 2002) replicates many characteristics observed *in vivo* (Spitzer and Semple 1998) and predicts several response properties observed in cortex. In accordance with a previous

model (Cai et al. 1998), firing rate adaptation and IPD-tuned excitatory and inhibitory inputs can produce sharper tuning in response to beats than static IPD stimuli, and predicts the phase lag seen at faster beat frequencies. The model also predicts a phase lead at very low beat rates (i.e., the curves in the right-hand panels of Figure 5.6A and C should sag before going upwards), as the latter half of the modulation histogram is eroded by adaptation, effectively advancing the mean phase of the response. This discrepancy could be accounted for in the model by the inclusion of a transmission delay that masked phase advance at low frequencies (Borisyuk et al. 2002). Under no circumstances did the IC model predict band-pass tuning of spike rate to beat frequency.

5.8.2 Center frequency and mean phase

The concepts of characteristic delay and characteristic phase stem from linear systems analyses of the auditory system (Yin and Kuwada 1983c), a system that is profoundly nonlinear from transduction onward. It seems surprising, then, that we find 83% of cells tested to possess an identifiable CD and CP at the level of the auditory cortex. A linear model of binaural coincidence detection would predict that optimal summation for excitatory (EE) interactions would occur for signals perfectly in phase, whereas excitatory-inhibitory (EI) summation would yield maximal responses when signals are perfectly out of phase. Thus CPs would cluster at the peaks and troughs of the input waveform, i.e. 0° and 180°. While there is

some evidence of this in the SOC (Spitzer and Semple 1995 observed excitatory interactions only), we find our CPs did not show a bimodal distribution, but instead formed a messy cluster centered at 333° (Figure 5.10A). This is consistent with the phase/frequency plots shown by Reale and Brugge (1990), which show a variety of intercept values.

If the ultimate stimulus feature derived from IPD is azimuthal location, tuning to a fixed interaural delay regardless of carrier frequency generates ambiguity, since azimuthal angle is dependent on IPD and the carrier of the stimulus. When mean phases of the binaural beat responses are converted, via carrier frequency, to ITDs, their distribution is not constrained by the range of ITDs encountered in nature (Figure 5.10B), suggesting that the peaks of these functions do not form a map of azimuthal angle. Similar evidence also undermines the idea of a mapping of azimuth in the peaks of IPD tuning curves. McAlpine and colleagues (2001) found tuning in low-BF cells of the guinea pig IC to ITDs that were well outside the range of delays a small-headed animal could experience. Although the peaks of their noise-delay functions could fall outside the ecological range, the slopes of these functions were steepest at the midline – corresponding to the "auditory fovea" where azimuthal acuity is greatest (but c.f. Middlebrooks and Green [1991], who find only modest improvement in localization at the midline). Though our experiments differ in stimulus, species, structure, and anesthetic state, the

conclusion drawn may be similar: we should reject the notion of a labeled-line code for ITD (or azimuth) in favor of broadly tuned spatial channels (discussed below).

5.8.3 Neural sensitivity to auditory motion

One previous study has examined sensitivity to auditory motion in AI of the alert macaque, using noise stimuli from a moving free-field speaker (Ahissar et al. 1992). Although the tuning of single units in azimuth was not sharp enough to account for the ability of monkeys to localize similar stimuli under free-field conditions (Brown et al. 1978), the authors did observe that most neurons showed the largest modulation in response magnitude for changes in sound location around the midline regardless of their preferred azimuth, consistent with the notion that the *slopes* of these tuning functions provide the acuity demonstrated around midline. We report no overall sensitivity to the direction of simulated motion (i.e., positive or negative beat frequencies) among the population when assessed by spike rate or synchrony (section 5.7.1, Figure 5.12). On the issue of neural preference for leftward or rightward movement, Ahissar et al. (1992) draw a different conclusion from a qualitatively similar dataset (cf. Figure 5.12B, their Figure 5B). They report a preponderance of units preferring movement toward the contralateral hemifield, but only among the subpopulation of units that show a directional preference in the first place (as computed by a *t*-test on ≥20 trials in each direction). Our failure to reveal an overarching directional preference among the population may stem from

the differences between the stimuli and methods of analysis in the two studies. Their white-noise stimulus, presented in the free field, would engage interaural level difference mechanisms, as well as providing spectral cues from filtering by the pinna.

Although spike rate and synchrony in response to the beat failed to reveal a clear directional preference in either hemisphere, the d′ analysis (section 5.7.2 and Figure 5.13C) did reveal a sharper change in spike rate as the stimulus *approached* the preferred IPD, compared to the change in spike rate as the stimulus left the receptive field. Thus the shape of the MPH as the stimulus passed through the IPD receptive field was not symmetrical. A similar effect has been observed in a model of primary visual cortex neurons (Chance et al. 1998), in which rapid synaptic depression in the absence of inhibition engendered a band-pass frequency response function, analogous to the rate-tuned rMTFs emergent in cortex (e.g. Figure 5.6A). Depression also elicited a phase lead of unusual origin: unlike an advance due solely to spike rate adaptation eroding the falling edge of the response, depression actually *steepened* the rising phase of the response, in addition to causing 15° worth of phase shift at a stimulus drift rate of 2 Hz. Synaptic depression in the afferent pool driving our AI neurons could account for the bias toward greater d′ values (steeper slopes) on the rising edge of the response. This would imply that the lack of directionality discussed in the previous section is a result of simplifying MPH

shapes to scalar values like firing rate and vector strength, a point also evident from analysis of the SAM and SFM responses in Chapter 4.

5.8.4 Comparing neural and psychophysical discriminability

The neurometric analysis employed here (section 5.7.3) rests on the presumption that thresholds for perception of change in IPD may be predictable from the neural response to those changes – a relationship which has been posited previously in the visual domain (Tolhurst et al. 1983, Britten et al. 1992). In the auditory system, Skottun (2001) found the best thresholds in the population of guinea pig IC neurons to be comparable to ITD thresholds in humans, when tested with sufficient resolution and sampling, but this seem atypical in the literature. Our finding generally echoes those of several investigations comparing psychophysical and neuronal performance in the visual system (reviewed, along with data from other modalities, in Parker and Newsome 1998) that found the sensitivity of most cortical neurons to fall short of psychophysical sensitivity for contrast, orientation, spatial frequency, position and acuity. Our study, like some of these, is limited by comparing slightly different stimuli (beats vs. small trapezoidal phase shifts), and data collected at different times (though our data are from not only the same species, but the same animals, in the awake condition).

Partially, the disparity between neural and perceptual threshold can be attributed to recording from primary sensory cortex, not a specialized structure where an ideal stimulus has been identified. This presupposes that neural sensitivity commensurate with lateralization acuity may emerge in areas downstream of AI. For example, discrimination of the direction of random dot motion by neurons in visual are MT has been modeled to fit the simultaneous psychophysical responses of macaques, but MT had been previously identified with the representation of visual motion (Britten et al. 1992). Once the basic response properties and organization of awake auditory cortex (primary and otherwise) are better understood, we can pursue the question of the psycho-neural link with consummate rigor. Convergence of primary cortical IPD-tuned inputs to a secondary cortical area could improve signal-to-noise by integrating across units with uncorrelated noise (Britten et al. 1992, Shadlen et al. 1996). A pooling model has been developed for ITD discrimination in SOC, IC and thalamus by Fitzpatrick et al. (1997), who show a progressive sharpening in ITD tuning at sequentially higher centers. In our cortical population, we have no evidence that noise is uncorrelated, and in fact it seems that neurons within the cortical nexus fire in synchrony as a result of their common (thalamic) inputs and dense interconnections. An alternative to the pooling model is the lower-envelope hypothesis (reviewed in Parker and Newsome 1998), whereby psychophysical threshold represents the tuning of only the most selective neurons in the population. We, like McAlpine et al., looked at the flanks of the spike rate vs. IPD function as

the region of the response most likely to be indicative of threshold-level discrimination, where the relationship between spike rate and the parameter of interest is steepest. Perhaps we have not encountered these neurons in our limited sample, or perhaps those cells we chose to analyze with steep slopes around the midline under-represent the overall tuning of the population.

An alternative scenario is that the comparatively broad tuning we see is not the relevant measure for the comparison at hand. Middlebrooks has postulated a "panoramic" code among broadly tuned neural units as the basis for sound localization, and has implemented network models based on spike timing in the anesthetized cortex (Middlebrooks et al. 1995). This idea is better supported than the notion of coding azimuth in the distribution of sharp IPD tuning peaks as discussed above, and is compatible with other rate-based schemes (McAlpine et al. 2001, Fitzpatrick et al. 1997) in that first-spike latency and spike rate are likely to be correlated in unanesthetized cortex (the optimal stimulus typically produces the strongest response *and* shortest latency). McAlpine et al. (2001) endorse a population code based on an analogy to the fine hue discrimination of which the visual system is capable given just three broadly wavelength-tuned cones.

Tuning of cells in the auditory cortex to complete modulations of the IPD cycle may be expected to misrepresent discrimination of small modulations in IPD, such

as those presented in our psychophysical task (Chapter 2). Malone et al. (2002) emphasized the pitfalls of estimating responses to stimuli in an unique temporal context from traditional tuning curves, particularly for the case of IPD in macaque cortex. It seems quite possible that the difference in discharge rate between two steady-state IPDs presented in the psychophysical task may be greater than the difference computed from the continuous variations in IPD that constitute a binaural beat. Context-induced changes in rate coding may enhance the information conveyed by *single neurons*, independent of later network pooling (Malone et al. 2002).

5.8.5 Organization of auditory cortex

The region of binaural beat sensitivity reported here may serve as a valuable landmark for establishing the location of AI and R. In the awake preparation, where immediate histological verification of recording sites is not available, this physiological criterion could be combined with MRI to establish the locations of the surrounding belt regions as well. Belt regions surrounding the core contain neurons that could be driven by low-frequency stimuli, but these cells were not responsive to binaural beats, suggesting that whatever spatial information is extracted from IPD in pure-tone stimuli must be re-coded beyond the auditory core. The hypothesized spatial area downstream from AI would most likely abstract further from the coincidence-based representation of azimuth in the brainstem to

meld sensory localization into truly spatial coordinates. This is consistent with ablation studies that have shown auditory cortex to be unnecessary for conditioned avoidance based on sound location, but indispensable in tasks requiring an animal move toward the perceived location of a sound in space (Heffner and Heffner 1990).

Cells in low-frequency AI and R possess tuning properties consistent with the origin of a spatial stream of cortical fields, but provide inputs that are crucial to the identification of sound sources as well. Existing tracer studies either did not use physiology to determine the frequency region where injections were made (Hackett et al. 1998) or made injections only into low-frequency areas in the lateral belt (Morel et al. 1993), so the patterns of inputs from the low-frequency core to downstream cortical fields are not yet fully established. Projections from the low-frequency area described here most likely target multiple fields in the auditory belt and parabelt, including the proposed non-spatial "what" streams (Romanski et al. 1999ab). For example, primate vocalizations contain low-frequency components that modulate in frequency and amplitude, and beat-sensitive cells were often responsive to the SAM and SFM stimuli of Chapter 4 as well as to IPD. Also, binaural interactions can contribute to phenomena like auditory streaming and the "cocktail-party effect" in humans, in which separation of sound sources aids speech intelligibility (Cherry 1953, Bregman 1994). Confirmation of the "what vs. where"

model of auditory processing awaits further evidence from fields beyond the core, but the IPD-sensitive population described here should contribute to both.

Chapter 6: Conclusions

6.1 Summary

6.1.1 Behavior

The awake-trained macaque offers a valuable animal model with which to explore the physiology and functional organization of auditory cortex. Without the depressive effect of anesthesia, stable recordings from single neurons are possible throughout the depth of the cortical lamina. Rhesus macaques can be trained on a range of psychophysical tasks, and are shown in these studies to have binaural discrimination thresholds on par with humans (Chapter 2). Although extended recordings would almost certainly have been impossible in untrained animals, the tuning of neurons in core auditory cortex was not reliably enhanced by behavioral context, implying that data collected during passive listening provides a veridical sampling of cortical responses. Passive physiology, carried out in the context of behavioral training, offers unparalleled flexibility to explore the cortical representation of an array of acoustic stimuli, in a species with a brain structure homologous to that of humans (Pandya and Yeterian 1984b).

The firing rates of single neurons, measured during behavior or during passive listening to binaural beats, were insufficient to explain the thresholds of the animals for discriminating the direction of change in interaural phase disparity (Chapter 2,

Chapter 5). A similar result has been reported using sounds presented in the free field (Ahissar et al. 1992), and confirms the very broad tuning for stimulus location described in anesthetized cat AI (Middlebrooks and Pettigrew 1981). Despite the strong, sustained, sharply tuned responses elicited by IPD, the cortex appears to represent stimulus location – or the interaural cues to azimuthal location – through an ensemble code, and quite likely employs spike timing information ignored in the receiver-operating characteristic technique employed here (e.g. Middlebrooks et al. 1994).

Future experiments aimed at revealing the functional role of auditory cortex in behavior should move beyond discrimination tasks, to engage the animal in some form of pattern recognition more likely to require the complex circuitry of the cortex. Early ablation experiments cast doubt on the functional necessity of the auditory cortex, perhaps because the stimuli and behavioral paradigms involved were adequately represented by subcortical structures in the tested species (reviewed in Whitfield, 1984). By contrast, ablation studies in monkey have shown the primary auditory cortex to be necessary for the discrimination of acoustic patterns (reviewed in Heffner and Heffner 2003), suggesting that the cortex excels at grouping components of the acoustic stream. By breaking down sound into its frequency components, the auditory periphery performs sound *analysis*; by the level of cortex, investigations should be aimed at understanding *synthesis*. This

distinction, and the effect of behavioral context on neural responses, are likely to be of greater importance in cortical fields beyond the core.

6.1.2 Physiological organization of the core

In accordance with prior anatomy and mapping studies (Jones et al. 1995, Hackett et al. 1998, Merzenich and Brugge 1973, Morel et al. 1993, Kosaki et al. 1997), a core region could be defined physiologically by reliable, frequency-tuned responses to pure tones (Chapter 3). This frequency tuning defined a rough rostral-caudal gradient within AI, and identified a low-frequency border rostrally, beyond which the gradient partially reverses as tonotopic order degrades. Although early studies were ambivalent on whether the reversal truly defined a separate field (Jones et al. 1995), comparisons in the present study identify the rostral field as physiologically distinct.

Although R shares the tuning for pure tones found in AI, the frequency representation within R is compressed relative to that in AI, with a bias toward lower frequencies. When tested across SPL using pure tones, cells in R are more likely to produce nonmonotonic rate-level functions, and most strikingly, the minimum latencies of cells in R are 1.5 times longer than those in AI. Despite these differences between the fields, organization within fields was difficult to discern. Even tonotopy, the long-standing organizational principle of auditory

cortical organization, appears rough except when comparing adjacent cells within an electrode track. When effects of SPL are considered, as in the standard activation maps of Figure 3.14, the distribution of activity across the cortical surface in response to a pure tone can appear broad, or patchy, or both – with no evidence of isofrequency lines. Tuning to SPL, along with classes of binaural interaction, appears to cluster within an electrode track as well, but without any apparent similarity between neighboring tracks when sampling in a ~1 mm grid.

In unanesthetized monkey auditory cortex, using supra-threshold stimuli, there is scant evidence for a place code representing frequency, level, or binaurality in smooth orthogonal gradients (e.g. Schreiner 1998). But the tuning of each neuron for these three parameters could affect not only its response *magnitude* (i.e. firing rate), but also its response *phase*. The relative timing of cortical activation within the patchy distribution of tuning properties must carry information about multiple parameters of the stimulus. These temporal dynamics may be key to understanding how the cortex represents complex signals.

6.1.3 Temporal and spectral modulation

Multiple features of low-frequency temporal and spectral envelopes may be represented in single neurons by spike rate profiles within the modulation period. The range of modulation frequencies over which this behavior is observed (< 20

Hz) corresponds to the range of envelope modulation frequencies crucial to speech comprehension in human listeners (Drullman et al. 1994, Shannon 2002). The features of the envelope to which a neuron is sensitive can be predicted in part by its receptive field, as measured with static tones. However, the fact that a response to amplitude modulation (as an example) cannot be fully explained by its RLF is not necessarily a result of nonlinearities introduced by the application of dynamic stimuli, but instead may reflect the fact that both SAM and the RLF are imperfect measures of the same underlying sensitivity to SPL. This tuning is likely to produce nonmonotonic RLFs, and modulation period histograms with multiple peaks corresponding to multiple traversals of a circumscribed receptive field. As shown for the composite modulation period histograms in Figure 4.10, separation of single units on the basis of their pure-tone response reveals tighter structure in the firing rate profile than was apparent in the pooled response. A cell's tuning to carrier frequency and SPL will affect the latency of its response, and all three of these factors interact to influence the phase coherence of the population response to modulating envelopes. If basic tuning properties cluster within the cortex, cells within these clusters may be expected to respond in phase with both the stimulus envelope, and one another.

The difference in temporal characteristics between AI and R is also apparent in responses to modulated stimuli. Best modulation frequencies as determined by rate

are higher in AI for both stimuli, as are synchrony cutoff frequencies. For SAM, the difference in cutoffs is greater than would be predicted by the difference in latency to tones, suggesting that cortical circuits in R may be sensitive to amplitude changes of a particular rate, via a mechanism that is overcome by the very short (5 ms) rise times of pure tones. Cutoff frequencies for SFM in R match what would be predicted by the latency difference between the fields, and likely reflect a more general temporal filter.

6.1.4 Interaural phase modulation

Like other forms of temporal modulation, binaural beats evoke strong, tuned discharges over a substantial portion of the modulation cycle. In contrast, responses observed under anesthesia were phasic, generally restricted to a single spike per beat cycle (Reale and Brugge 1990). A distribution of multiple spikes throughout the modulation period opens the possibility of coding mechanisms based not on the peak of the rate vs. IPD function (embodied by the MPH), but on the rising and falling slopes of that function. The importance of these slopes is illustrated by a d′ analysis showing that discriminability of IPD, based on firing rate, is greatest where the rate of change in the MPH profile is steepest (Figure 5.13; Skottun 2001). Moreover, the peaks of the MPH often correspond to interaural delays that cannot be experienced by a monkey, as limited by the animal's head size. Thus the best IPD of each cell, at a given frequency, cannot form a place code for azimuthal

location across the cortical surface. As with temporal and spectral modulation, instantaneous changes in firing rate seem to convey the most information about changes in interaural delay.

The quality, tuning, and temporal cutoff of binaural beat responses did not differ between AI and R, validating their analysis as a unified, physiologically defined population. The binaural substrate for computation of azimuthal location based on fine-structure cues is shared between both fields of the core.

6.2 Integration

6.2.1 Temporal response properties between the core fields

The ultimate limit of temporal synchrony in any cell may be determined by intrinsic properties (such as membrane conductances), or the level of temporal structure in its inputs (weak synchrony to the stimulus could be enhanced by thresholding or averaging out noise, but in the absence of input synchrony, none can be created). Comparison of synchrony cutoffs in response to three different modulations – SAM, SFM, and binaural beats – may suggest the extent to which this property is intrinsic to a cortical neuron, or dependent on the stimulus parameter being modulated.

Synchrony modulation transfer functions in response to all three stimuli are presented for two cells in Figure 6.1A. In the cell at left, overall vector strength and cutoff frequency are highest for binaural beats, which is unusual in the population. Beats also produce the strongest synchrony in the MTFs on the right, though beats and SFM share an overlapping high frequency cutoff, and SAM synchrony is comparatively poor. Despite this, the synchrony cutoff for SAM is higher than that for the other two modulations, illustrating the limitations of MTF-derived measures. The relation of SAM and SFM cutoffs for the population is addressed in Chapter 4, and the relation of each to binaural beat cutoffs is plotted in panel 6.1 B. Cutoffs for beats and SAM are correlated (ANOVA, p=.002) and higher for SAM on average (median for beats=19.5 Hz, for SAM=44.5 Hz, p=.033 by Wilcoxin signed-rank). Cutoffs between beats and SFM did not correlate or differ significantly. This evidence does not clearly support intrinsic properties or stimulus-dependent mechanisms as the limiting factor in temporal synchrony, but implies a combination of the two.

Best modulation frequency, determined by discharge rate, was higher for SAM and SFM than for beats (means of 19 Hz and 32 Hz respectively, vs. 8 Hz for beats), though neither was correlated. The reason for this may be spectral rather than temporal: SAM and SFM retain power at the carrier frequency as modulation frequency increases, but beat rate is increased by separating the frequencies at the

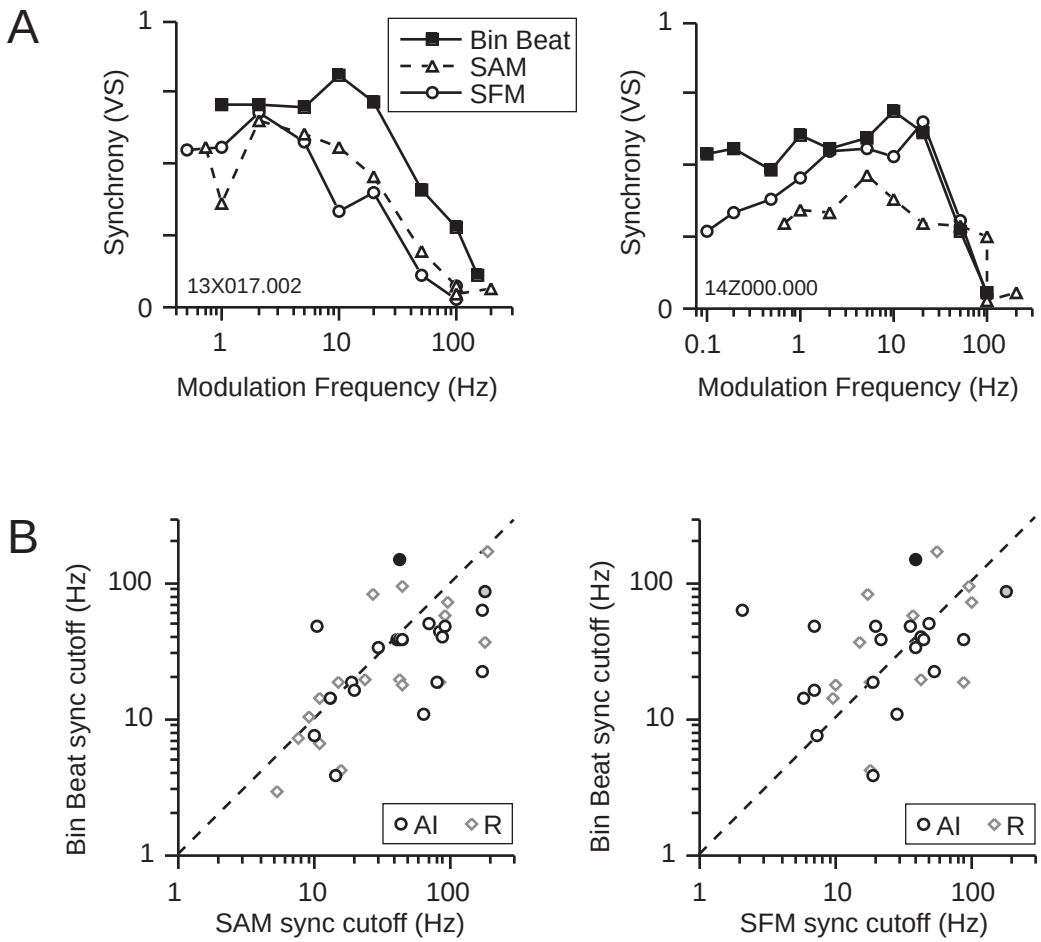


Figure 6.1 Comparison of synchrony cutoff frequencies for binaural beat, SAM, and SFM stimuli. **A:** Synchrony MTFs for all three modulations in two example cells (VS used for beats and SAM, VS at $2*f_m$ used for SFM). The cell at left is unusual in that the cutoff for beats is higher than that for SAM or SFM; this characteristic is reversed in the cell at right, but the cutoff metric does not reflect that fact that beats produce better synchrony than SAM at modulation frequencies <50 Hz. **B:** Across the population, cutoff frequencies for beats and SAM (left panel) were significantly correlated (ANOVA, p=.002, r^2=.25) and higher for SAM (median for beats=19.5 Hz, for SAM=44.5 Hz, p=.033). Cutoffs between beats and SFM were not correlated and did not differ significantly (median for SFM=39 Hz). Circles mark cells in AI, diamonds mark cells in R; the filled black circle indicates the cell in panel A, left; filled gray circle indicates the cell in panel A, right.

two ears. Thus the amount of energy within the receptive field drops with increasing beat frequency.

A future experiment combining binaural and modulation sensitivity could use binaural beats created by offsetting the phase of the *envelope* at the two ears, rather than the fine structure. This would be accomplished by using slightly different SAM frequencies at each ear, just as beats are created by different carrier frequencies. Because sensitivity to this cue would not be limited by the ability of the auditory nerve to phase-lock to fine structure, it could be applied across a wider range of carrier frequencies than the binaural beat (Batra et al. 1989, Joris 2003). Sensitivity to this cue, and cutoff frequencies commensurate with those to binaural beats, could allow a re-evaluation of the idea that the binaural system is comparatively "sluggish" in its response to rapid temporal modulations (Grantham and Wightman 1978).

The temporal differences between AI and R, illustrated with static and dynamic tones, generalize to stimuli of broad spectral content. Multiple repetitions of long duration band-passed noise reveal that neurons in AI discharge in synchrony with particular temporal features of the noise (constructed here with constant equivalent SPL), whereas neurons in R show less precision. The finer grain of temporal structure in AI is illustrated for two representative cells in Figure 6.2: the relative

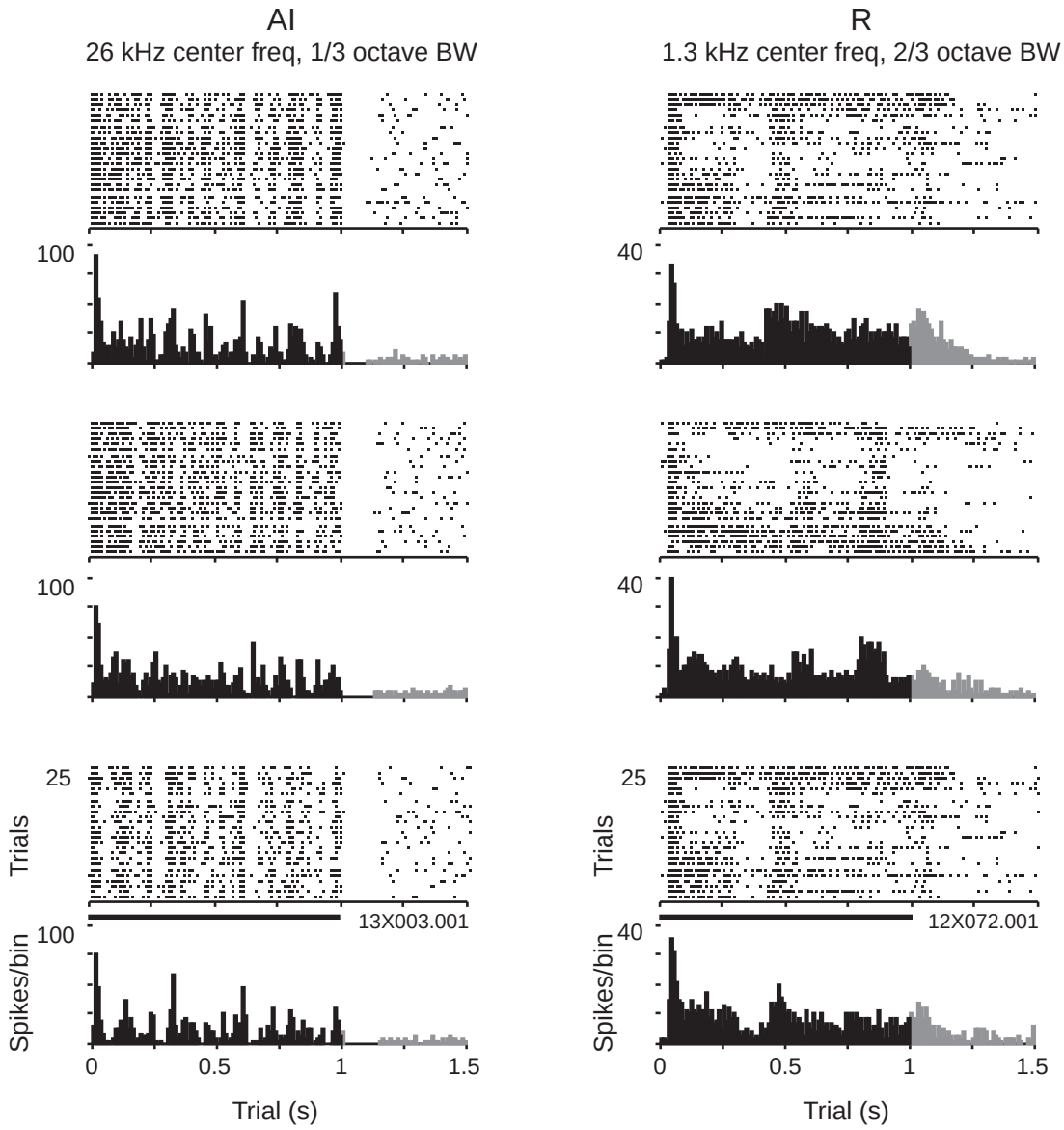


Figure 6.2 Responses to long-duration band-passed noise reveal temporal differences between AI and R. Rasters and underlying peri-stimulus time histograms (10 ms bins) depict the response to 1 s duration band-passed noise at the optimal center frequency and bandwidth of each cell (10 ms rise time, 25 trials of 2 s duration, time axis truncated at 1.5 s). Each column represents the same cell tested with three different noise stimuli of identical center frequency and bandwidth, pseudo-randomly generated from different seed values. The temporal structure in each response is thus stimulus-specific, and not an intrinsic property of the cell. This example is from high-frequency AI, but equivalent fine-grain temporal sensitivity was evident at low frequencies as well.

degree of temporal structure is evident in spike rasters, and peri-stimulus time histograms, collected in response to band-passed noise optimized in center frequency, bandwidth, and level. For each neuron, random noise was generated using three different "seed" values, confirming that the temporal structure in the response is attributable to the stimulus, and not the intrinsic dynamics of the cell.

The response of the cell in R also varies more across trials, habituating after about 10 trials for each stimulus, then returning to full strength for the last few trials. This effect is most evident in the first 250 ms after noise onset, but does not apply to the initial onset response. Whether this is a general property of cells in R is not known, since most stimuli were presented over so many long, consecutive trials. For this cell, habituation begins about 20 seconds into the stimulus block, and must represent adaptation on a much longer time scale than that implicated in shaping MPH responses to dynamic stimuli.

Evidence from the cat auditory system suggests that differences between AI and R may be at least partially inherited from the thalamus. The ventral division of the medial geniculate shows a gradient of responsiveness orthogonal to the frequency progression, such that neurons at the anterior end of the nucleus are in more precise tonotopic order, respond with shorter latency, lock their discharges more precisely to repetitive clicks, and are more likely to have monotonic RLFs (Rodgrigues-

Dagaeff et al. 1989). Retrograde labeling indicates that the core fields of feline auditory cortex receive inputs that are biased in a graded manner, from anterior to posterior (the homology of these fields to macaque cortex is discussed in section 6.3.1, below). Physiology in primate thalamus is less complete, but it is likely that the differences observed between AI and R reflect a combination of disparities in thalamic inputs, and local cortical processing.

6.2.2 Rate and temporal coding by single neurons

Any discussion of rate versus temporal coding comes down to the matter of time scale: these are not so much two competing ideas, as a continuum based on the duration over which spike rates are computed. In the MTFs of chapters 4 and 5, responses to modulation were measured by total spike rate averaged over 20 s of stimulus presentation (firmly at the "rate" end of the spectrum), and the precision with which single spikes phase-locked to the envelope (closer to a pure "temporal" code). As a compromise between these two, responses to low-frequency modulations were analyzed in terms of instantaneous discharge rate, evident in MPH profiles. These profiles offer the least ambiguous representation of dynamic stimulus envelopes over a restricted (though important) frequency range, leaving the traditional mechanisms of across-period spike rate, and within-period synchrony, to encode faster modulations.

Bandpass rate tuning for all three modulations is common in AI and R, but the peaks of rate MTFs are almost always within the synchronized range (rate BMF < synchrony cutoff, $p<.0001$ by paired *t*-test for SAM, SFM and beats). Rate tuning to SAM has been described in the midbrain (e.g. Krishna and Semple 2000), but rate tuning to the modulation frequency of binaural beats seems to be emergent at the level of thalamus or cortex (Chapter 5; Spitzer and Semple 1998). The possibility exists that rate tuning within the synchronized regime arises as an epiphenomenon of the receptive field properties and adaptation that shape MPHs. This does not preclude the rate-based coding of modulation frequency beyond the range emphasized in Chapter 4 (e.g. > 100 Hz). The spectral sidebands introduced by SAM or SFM may play a role in recruiting an elevated rate response at high modulation frequencies, where modulation envelopes induce the perception of pitch (Langner and Schreiner 1988, but c.f. Krishna and Semple 2000).

If temporal coding by instantaneous spike rate is the mechanism by which neurons of the auditory core signal amplitude, frequency, and interaural phase of dynamic stimuli, what is the utility of this representation, and what limits its fidelity? Most neurophysiology rests on the ergodic assumption that sampling responses from one neuron many times (as in the last figure) is equivalent to sampling a homogeneous population of neurons responding to a single presentation. The PSTH or MPH is assumed to reflect the temporal precision of that local population, which responds

similarly because of its common thalamic input, and dense interconnections. Simulations of interconnected cortical neurons challenge this view by demonstrating that even weak correlation between neurons increases the variability in their spike timing, implying that histogram representations from single neurons may over-estimate the temporal coherence of instantaneous firing rate among cells in a cortical ensemble (Mazurek and Shadlen 2002). Why, then, would the cortex employ such an interconnected columnar architecture? The authors argue that a cortical neuron receives "a barrage of synaptic input for each spike it produces, which furnishes it with a nearly continuous estimate of incoming rate signals, even when individual inputs are firing at low rates" (p. 469). Thus, the redundant connectivity that limits ensemble synchrony may allow the rate-profile code apparent in responses to low-frequency modulated stimuli.

Given each cell's sensitivity to multiple stimulus parameters, it is possible that two very different stimuli (e.g., SAM and a binaural beat) could be adjusted to produce nearly identical MPHs. The scenario is even more likely for rate-based coding, in that any bandpass-tuned function must have equivalent stimulus values above and below the peak that produce an equivalent discharge rate. How might these "stimulus metamers" be disambiguated? The answer is likely to lie not in the responses of individual neurons, but in the pattern of activation within or across cortical ensembles. For example, the relative activation in two ensembles of

equivalent frequency and SPL tuning, but different binaural characteristics, may distinguish a SAM response from a beat response. The character of such a representation would depend on the organization of stimulus parameters across the cortex.

6.2.3 Cortical maps: gradients and clusters

The only consistent map identified across the extent of AI in this study was that of best frequency, but even this map became messy across the border into R, or when measured using standard stimuli at suprathreshold SPL. Recanzone et al. (2000b) reported a similar finding in awake rhesus AI, and despite a detailed analysis could not identify a global organization of latency, threshold, or spectral tuning width like that described in anesthetized cat (Read et al. 2002). The present study, and Recanzone's, confirm the clustered distribution of response properties like binaurality and nonmonotonicity alluded to by Brugge and Merzenich (1973).

Mapping studies in anesthetized AI of other primate species confirm a non-random distribution of tuning properties other than BF, but these properties were distributed in patches rather than smooth maps. Cheung et al. (2001) found a patchy distribution of tuning sharpness in low-frequency AI of squirrel monkey, and a relationship between threshold and BF that matches the data in this report (Chapter 3) more closely than the comparable awake rhesus data (Recanzone et al. 2000b).

The data reported here also reveal some similarities to a multi-unit mapping study in owl monkey AI (Recanzone et al. 1999): a comparable threshold-BF relationship, and in at least one animal, a small population of EI cells clustered in high-frequency AI. In owl monkey, frequency tuning sharpness, threshold, best level, and latency were non-uniformly distributed, independent of the BF gradient, and independent of one another. These distributions may only be revealed by anesthesia and/or dense multi-unit mapping, or may reflect a difference between these two New world species and the Old world macaque.

Anatomy and connectivity studies support the idea of a patchy, distributed mosaic of response properties across the cortical surface, partially inherited from thalamic inputs, and refined by intrinsic cortico-cortical connections. Velenovsky et al. (2003) unified prior observations by demonstrating that clusters of homogeneous responses in cortex align with discrete thalamocortical projection patterns. Anterograde labeling identified terminations from the ventral MGN, within which BF and binaural class were consistent; the edges of these patches often coincided with abrupt transitions in physiological characteristics. Within the cortex, retrograde tracing experiments by Matsubara and Phillips (1988) reveal a patchy, anisotropic pattern of cortico-cortical connectivity constrained by BF gradients. Physiology confirmed that patches connected regions of common or higher BF, but

did not follow binaural class, such that "only a fraction of an isofrequency strip projected to a given patch".

Taken together, these data may explain why awake cortex could appear more patchy or disordered than anesthetized cortex. As mentioned above, sampling density in awake cortex is generally coarser and less well controlled than is possible under anesthesia: patches on the order of 0.5-1 mm wide (as described in the above studies) may not be resolved in a 1 mm grid. This allows clustering to be observed within a track, but not between adjacent tracks. The other experimental concern is anesthesia, which often restricts responsiveness to the middle thalamorecipient layers of cortex. Frequency-specific intracortical connections, likely to involve projections from layer III, may serve to refine inherited tuning properties via local cortical circuits. Thus, tuning bandwidth and threshold at CF may appear as a global gradient under anesthesia, but appear patchy (or disorganized) in the awake animal due to the engagement of cortico-cortical connections. Callosal connections are also specific for frequency and binaurality (Imig and Brugge 1978), and offer yet another dimension in which local activation patterns can be compared.

Application of a standard stimulus across the cortex, irrespective of each cell's tuning for frequency and level (Goldstein et al. 1970, Phillips et al. 1994), suggests

that tonotopy breaks down into a broad and discontinuous distribution at high SPLs (Chapter 3). A recent report using optical imaging (Nelken et al. 2004) challenges this notion, producing smooth BF maps at 75 dB in ferret auditory cortex. In addition to the considerable temporal averaging involved in this technique, the optical signal (like the hemodynamic signal in fMRI) is thought to reflect synaptic activity, not the spiking output of cortical neurons. Therefore the patchiness observed in electrophysiological maps, whether in awake or anesthetized animals, is likely to reflect processing intrinsic to the cortex.

To the established patchiness of cortical organization across space, I propose the added dimension of activation in time. Any stimulus will activate a broad region of cortex, but the relative timing of that activation between patches of similar, and disparate, tuning properties may represent a wealth of stimulus features.

6.3 Implications

6.3.1 Homology to auditory cortex of other species

Although the alert primate is gaining popularity as a model for the study of auditory cortex, the broadest survey of auditory cortical physiology has been carried out in the anesthetized cat. Despite many differences between these preparations highlighted in the above chapters, understanding the homology between these species offers a valuable insight into common motifs of auditory

processing. The most logical place to look for a homolog of R in cat is not necessarily rostral of AI – the rostral-caudal direction of tonotopy is reversed between cats and monkeys, such that the anterior auditory field (AAF) in cat shares a high-frequency border with AI (Phillips and Irvine 1982). Like R, AAF shares input from the MGv with AI, but its BFs are biased toward high frequencies, and its latencies are as short as those in AI. The same relation between AAF and AI is described in ferret (Kowalski et al. 1995), where latencies were actually shorter than in AI. With the exception of receiving direct input from MGv, this sounds like belt area CM in the macaque: it abuts high-frequency AI, seems biased toward high frequencies, and has latencies at least as short as those in AI (my unpublished data, Recanzone et al. 2000b).

A better candidate for the feline homolog of R is the posterior field, P, which shares a low-frequency border with AI. It also shares a sharp tuning for frequency, but shows much longer latencies (20-50 ms, vs. <20 ms in AI). Whereas up to 75% of neurons in AI have monotonic RLFs (Brugge and Reale 1984), nonmonotonicity is the rule in P: estimates range from 70% (Phillips et al. 1995) to 86%, with 90% of those showing strong nonmonotonicity (>50% reduction from the peak at high SPLs; Phillips and Orman 1984). This parallels the longer latencies and prevalence of nonmonotonicity in R (74% nonmonotonic), though the differences between AI and R by both criteria are less extreme than those between AI and P in cat.

Sensitivity to interaural time difference is also equivalent between AI and P (Orman and Phillips 1984), as was shown by the distribution of binaural beat responses across the AI/R border. The only inconsistency in this analogy is that thresholds in P are 5-15 dB higher than in AI (Phillips and Orman 1984), whereas they are indistinguishable between AI and R, but if some of the input to P arrives in series from AI, this may be an artifact of anesthesia.

The posterior field in cat receives a major input from the caudal half of the MGv, and AI receives a major input from the rostral half, but AAF receives a lesser input from this nucleus (Rodrigues-Dagaeff et al. 1989). So the proposed homology of R with P, and CM with AAF, is not farfetched – over the course of whatever evolutionary changes reversed the tonotopy of the temporal lobe, a re-weighting of thalamic inputs would be a relatively minor change. This idea is similar to what was proposed by Woolsey and Walzl (1982, cited in Brugge and Reale 1984), who described two fields rostral to AI and theorized that they were homologous to the posterior and ventral fields in cat. They also equated what would now be called the caudolateral belt with AAF, and proposed a homolog of cat AII in the caudal insula, medial to AI. A few of my recovered electrode tracks indicate that this area was recorded, and these cells may have been assigned to area CM.

6.3.2 Functional implications of cortical organization

The primary and rostral fields provide multiple, asynchronous representations of acoustic input, the former with high temporal fidelity, the latter with some degree of abstraction related to temporal filtering. As was demonstrated with acoustic "chimeras", conflicting interaural disparity of a sound's envelope and fine structure causes human listeners to hear the speech encoded by the envelope, coming from the location suggested by the fine structure (Smith et al. 2002). Sensitivity to this fine structure cue is shared between low-frequency AI and R, but the envelope representation differs between the two.

Amplitude and frequency envelopes are also used in the identification of species-specific calls by many primate species, including macaque monkeys (LePrell and Moody 1997, Ghazanfar et al. 2001), and cortex is indispensable to this task (Heffner and Heffner 1984). The fidelity of responses in AI suggests that neurons here would follow the envelope of a vocalization, or any similar sound, to the limits of its temporal precision. This is not the case in marmosets, however, where AI discharges favor amplitude transients at the expense of spectral modulations, and respond to vocalizations more strongly than to similar synthesized sounds (Wang et al. 1995). This reduces the overall temporal complexity of the population response, but enhances synchrony among cells. In fact, this synchrony extends to widely distributed patches of AI. Such a detailed analysis has not been performed

in macaque, but it is reasonable to assume that the temporal precision of AI responses would similarly highlight amplitude transients, as it appears to do for SAM signals. By presenting simultaneous modulation of amplitude and frequency, it could be determined whether neurons in macaque AI give amplitude envelopes preference over frequency envelopes. This is likely to be the case, since these neurons are sharply tuned in frequency, but often respond over a wide range of SPLs.

Neurons in R are also tuned in frequency, but show less temporal precision to the amplitude envelope, suggesting that responses to vocalizations in this field would be abstracted yet further from their acoustic envelopes. An intriguing potential specialization is that the distribution of BFs in R (Figure 3.12A) closely matches the distribution of spectral energy in rhesus vocalizations, which is almost entirely ≤ 4 kHz.[1] Because they integrate instantaneous amplitude on a longer time scale than neurons in AI, cells in R are likely to be yet more selective for natural calls over their reversed or synthesized substitutes. A quantitative approach to this selectivity, in the temporal and spectral domains, may be offered by decomposing the salient features of vocalizations (Averbeck and Romanski 2004).

---

[1] http://www.wjh.harvard.edu/~mnkylab/media/rhesuscalls.html
Spectrographic analysis of 8 calls using Amadeus II software, Hairersoft.com.

Call identification is not the only task associated with vocalizations: identification of the individual caller appears to depend on harmonic structure and high-frequency amplitude modulation (Fitch 1997), and is represented in cortical responses (Ghazanfar 2003). Fine structure cues certainly convey more about vocalizations (monkey or human) than merely location.

Nelken et al. (2003) have argued that AI neurons can not be understood in terms of feature detection, but as processors of previously derived features, working on multiple time scales. These various time constants interact in the formation of auditory objects, to which such qualities as timbre or location can be assigned by higher-order cortical areas. The diversity of time scales available in AI (Nelken et al. 2003) is further elaborated in the rostral field of the core.

6.3.3 Models of cortical function

The fields of the belt are preferentially connected with their adjacent core field (Hackett et al. 1998), implying that the rostral belt should share the characteristics of R, and the caudal belt should share the characteristics of AI. Studies in awake macaque, using free-field stimuli, suggest that the caudomedial field contains neurons with a heightened sensitivity for spatial location, relative to neurons in AI (Recanzone et al. 2000a). A study of lateral belt responses in anesthetized macaque provides weak evidence that the caudal end of the lateral belt is sensitive to the

spatial origin of vocalizations, whereas the rostral belt may be more selective for the type of vocalization (Tian et al. 2001).

These belt regions are the origin for the proposed functional streams emanating from auditory cortex to the frontal and parietal lobes. The caudal belt has strong reciprocal connections with the dorsolateral prefrontal cortex (PFC), an area that has been shown to respond to acoustic stimulation during a localization task in an alert monkey (Vaadia et al. 1986), and is also implicated in visual spatial processing (Wilson et al. 1993). The tracer studies of Romanski et al. (1999a, b) show that rostral belt communicates with the rostral and ventral prefrontal cortex, apparently non-spatial regions (Wilson et al. 1993). These connections, though not entirely exclusive, are interpreted as forming the basis for "what" and "where" pathways, with rostral auditory cortex serving as the beginning of the spectrotemporal recognition pathway, and caudal auditory cortex beginning the spatial processing stream.

Presuming that rostral belt receives its major inputs from R, it would benefit from temporal filtering of the envelope in R, in accord with the hypothesis. But single neurons in R are also sensitive to IPD, a fine structure cue to azimuthal location. As discussed in Chapter 1, spatial cues aid in speech intelligibility and auditory object formation (Cherry 1953, Yost 1991), so a cortical system for sound identification

would benefit from the interaural sensitivity evident in R. In other words, a “what” pathway would make use of “where” information. By the same token, spatial sensitivity is not derived only from interaural cues – judgment of elevation relies on spectral cues imposed by the pinna, the same sort of spectral analysis that would be required for sound identification. At the level of the core, both fields represent both aspects of a stimulus, and both types of information are likely to be passed to the adjacent belt regions.

In the visual system, an alternative hypothesis for the functions of the ventral and dorsal stream has been proposed: instead of what and where (both perceptual concepts), these systems may subserve perception and action (Goodale and Milner 1992). Transposing this hypothesis to audition, one system (possibly involving the rostral core and belt) may be dedicated to the recognition of acoustic objects, using a combination of spectral, temporal, and spatial cues. Another system, with probable projections to parietal cortex and the frontal eye fields, could use spatial information to direct the eyes (Grunewald et al. 1999, Linden et al. 1999) or hands (Cohen and Andersen 2000) to the location of a sound.

6.3.4 Relevance to human hearing

In Chapter 1, the areas surrounding the primary auditory cortex in human were equated to the belt regions in monkey. Several of these regions are activated by

sound movement in the vertical and horizontal plane, anterior and posterior to AI (Pavani et al. 2002). Sensitivity in both dimensions requires inputs with both binaural and spectral sensitivity, as has been demonstrated in both fields of the core. A clinical study found that bilateral temporal lobe damage abolished the ability to use ITD cues, but some ILD sensitivity persisted, suggesting that the cortex is crucial to the perception of time delay. The ubiquitous sensitivity to binaural beats in low-frequency AI and R, and the relative paucity of ILD-sensitive neurons, is consistent with this finding. The growing body of electrophysiological data from awake macaque, and the ever-improving resolution of functional imaging techniques, hold the potential to synthesize our understanding of auditory cortical processing in human and non-human primates.

The auditory system is capable of representing a daunting array of sound qualities, by virtue of exquisite sensitivity to frequency, amplitude, and time – from the level of peripheral analysis, to the complex computation performed by central auditory structures. By virtue of these computations, dry acoustic parameters like frequency, envelope modulation, and interaural disparity are experienced as pitch and timbre; melody and harmony; vibrato and tremolo; depth and space. Ultimately, these qualities of auditory experience are constructed from the spiking activity of cortical neurons. The representation of sound in the cortex bridges the gap between vibration and perception.

References

Abeles M, Goldstein MH, Jr. (1970) Functional architecture in cat primary auditory cortex: columnar organization and organization according to depth. J Neurophysiol 33:172-187.

Ahissar M, Ahissar E, Bergman H, Vaadia E (1992) Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. J Neurophysiol 67:203-215.

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98:13367-13372.

Attias H, Schreiner CE (1998) Blind source separation and deconvolution: the dynamic component analysis algorithm. Neural Comput 10:1373-1424.

Atzori M, Lei S, Evans DI, Kanold PO, Phillips-Tansey E, McIntyre O, McBain CJ (2001) Differential synaptic processing separates stationary from transient inputs to the auditory cortex. Nat Neurosci 4:1230-1237.

Averbeck BB, Romanski LM (2004) Principal and independent components of macaque vocalizations: constructing stimuli to probe high-level sensory processing. J Neurophysiol 91:2897-2909.

Batra R, Kuwada S, Stanford TR (1989) Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit. J Neurophysiol 61:257-268.

Beaton R, Miller JM (1975) Single cell activity in the auditory cortex of the unanesthetized, behaving monkey: correlation with stimulus controlled behavior. Brain Res 100:543-562.

Beitel RE, Schreiner CE, Cheung SW, Wang X, Merzenich MM (2003) Reward-dependent plasticity in the primary auditory cortex of adult monkeys trained to discriminate temporally modulated signals. Proc Natl Acad Sci U S A 100:11070-11075.

Benson DA, Hienz RD (1978) Single-unit activity in the auditory cortex of monkeys selectively attending left vs. right ear stimuli. Brain Res 159:307-320.

Benson DA, Hienz RD, Goldstein MH, Jr. (1981) Single-unit activity in the auditory cortex of monkeys actively localizing sound sources: spatial tuning and behavioral dependency. Brain Res 219:249-267.

Bieser A, Muller-Preuss P (1996) Auditory responsive cortex in the squirrel monkey: neural responses to amplitude-modulated sounds. Exp Brain Res 108:273-284.

Blackburn CC, Sachs MB (1989) Classification of unit types in the anteroventral cochlear nucleus: PST histograms and regularity analysis. J Neurophysiol 62:1303-1329.

Borisyuk A, Semple MN, Rinzel J (2002) Adaptation and inhibition underlie responses to time-varying interaural phase cues in a model of inferior colliculus neurons. J Neurophysiol 88:2134-2146.

Boudreau JC, Tsuchitani C (1968) Binaural interaction in the cat superior olive S segment. J Neurophysiol 31:442-454.

Bradley A, Skottun BC, Ohzawa I, Sclar G, Freeman RD (1987) Visual orientation and spatial frequency discrimination: a comparison of single neurons and behavior. J Neurophysiol 57:755-772.

Bregman AS (1990) Auditory Scene Analysis: the perceptual organization of sound. Cambridge, MA: MIT Press.

Britten KH, Shadlen MN, Newsome WT, Movshon JA (1992) The analysis of visual motion: a comparison of neuronal and psychophysical performance. J Neurosci 12:4745-4765.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci 13:87-100.

Brosch M, Schulz A, Scheich H (1999) Processing of sound sequences in macaque auditory cortex: response enhancement. J Neurophysiol 82:1542-1559.

Brosch M, Schreiner CE (2000) Sequence sensitivity of neurons in cat primary auditory cortex. Cereb Cortex 10:1155-1167.

Brown CH, Beecher, M.D, Moody, D.B., and Stebbins, W.C. (1978) Localization of pure tones by Old World monkeys. J Acoust Soc Am 63:1484-1492.

Brugge JF, Merzenich MM (1973) Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. J Neurophysiol 36:1138-1158.

Brugge JF, Reale RA (1984) Auditory Cortex. In: Cerebral Cortex (Peters A, Jones EG, eds). New York: Plenum Press.

Cai H, Carney LH, Colburn HS (1998) A model for binaural response properties of inferior colliculus neurons. II. A model with interaural time difference-sensitive excitatory and inhibitory inputs and an adaptation mechanism. J Acoust Soc Am 103:494-506.

Chance FS, Nelson SB, Abbott LF (1998) Synaptic depression and the temporal response characteristics of V1 cells. J Neurosci 18:4785-4799.

Cherry EC (1953) Some experiments on the recognition of speech with one and with two ears. Journal of the Acoustical Society of America 25:975-979.

Cheung SW, Nagarajan SS, Bedenbaugh PH, Schreiner CE, Wang X, Wong A (2001) Auditory cortical neuron response differences under isoflurane versus pentobarbital anesthesia. Hear Res 156:115-127.

Cheung SW, Bedenbaugh PH, Nagarajan SS, Schreiner CE (2001) Functional organization of squirrel monkey primary auditory cortex: responses to pure tones. J Neurophysiol 85:1732-1749.

Cipolloni PB, Pandya DN (1991) Golgi, histochemical, and immunocytochemical analyses of the neurons of auditory-related cortices of the rhesus monkey. Exp Neurol 114:104-122.

Clarke S, Bellmann A, Meuli RA, Assal G, Steck AJ (2000) Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. Neuropsychologia 38:797-807.

Cohen YE, Wessinger CM (1999) Who goes there? Neuron 24:769-771.

Cohen YE, Knudsen EI (1999) Maps versus clusters: different representations of auditory space in the midbrain and forebrain. Trends Neurosci 22:128-135.

Cohen YE, Andersen RA (2000) Reaches to sounds encoded in an eye-centered reference frame. Neuron 27:647-652.

Creutzfeldt O, Hellweg FC, Schreiner C (1980) Thalamocortical transformation of responses to complex auditory stimuli. Exp Brain Res 39:87-104.

De Ribaupierre F, Goldstein MH, Jr., Yeni-Komshian G (1972) Cortical coding of repetitive acoustic pulses. Brain Res 48:205-225.

DiCarlo JJ, Lane JW, Hsiao SS, Johnson KO (1996) Marking microelectrode penetrations with fluorescent dyes. J Neurosci Methods 64:75-81.

Drullman R, Festen JM, Plomp R (1994) Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am 95:2670-2680.

Eggermont JJ (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. Hear Res 56:153-167.

Eggermont JJ (1998) Representation of spectral and temporal sound features in three cortical fields of the cat. Similarities outweigh differences. J Neurophysiol 80:2743-2764.

Eggermont JJ (1999) The magnitude and phase of temporal modulation transfer functions in cat auditory cortex. J Neurosci 19:2780-2788.

Erulkar SD, Rose, JE, Davies, PW (1956) Single unit activity in the auditory cortex of the cat. Bull Johns Hopkins Hosp 99:55-86.

Evans EF, Whitfield IC (1964) Classification of Unit Responses in the Auditory Cortex of the Unanaesthetized and Unrestrained Cat. J Physiol 171:476-493.

Evans EF, Nelson PG (1973) The responses of single neurones in the cochlear nucleus of the cat as a function of their location and the anaesthetic state. Exp Brain Res 17:402-427.

Fastl H, Hesse A, Schorer E, Urbas J, Muller-Preuss P (1986) Searching for neural correlates of the hearing sensation fluctuation strength in the auditory cortex of squirrel monkeys. Hear Res 23:199-203.

Fisher NI (1996) Statistical analysis of circular data. Cambridge [England]: Cambridge University Press.

Fitzpatrick DC, Batra R, Stanford TR, Kuwada S (1997) A neuronal population code for sound localization. Nature 388:871-874.

Fitzpatrick DC, Kuwada S (2001) Tuning to interaural time differences across frequency. J Neurosci 21:4844-4851.

Fritz J, Shamma S, Elhilali M, Klein D (2003) Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. Nat Neurosci 6:1216-1223.

Gaese BH, Ostwald J (2001) Anesthesia changes frequency tuning of neurons in the rat primary auditory cortex. J Neurophysiol 86:1062-1066.

Ghazanfar AA, Smith-Rohrberg D, Hauser MD (2001) The role of temporal cues in rhesus monkey vocal recognition: orienting asymmetries to reversed calls. Brain Behav Evol 58:163-172.

Ghazanfar AA, D.A. L, Hauser MD, N.K. L (2003) Acoustic parameters underlying vocalization-sensitive neural responses in rhesus monkey auditory cortex. Program 18114 Society for Neuroscience Abstract Viewer/Itinerary Planner.

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

Goldberg JM, Brown PB (1968) Functional organization of the dog superior olivary complex: an anatomical and electrophysiological study. J Neurophysiol 31:639-656.

Goldberg JM, Brown PB (1969) Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. J Neurophysiol 32:613-636.

Goldstein MH, Jr., Abeles M, Daly RL, McIntosh J (1970) Functional architecture in cat primary auditory cortex: tonotopic organization. J Neurophysiol 33:188-197.

Goldstein MH, Jr., Abeles M (1975) Note on tonotopic organization of primary auditory cortex in the cat. Brain Res 100:188-191.

Goodale MA, Milner AD (1992) Separate visual pathways for perception and action. Trends Neurosci 15:20-25.

Grantham DW, Wightman FL (1978) Detectability of varying interaural temporal differences. J Acoust Soc Am 63:511-523.

Grantham DW (1986) Detection and discrimination of simulated motion of auditory targets in the horizontal plane. J Acoust Soc Am 79:1939-1949.

Grantham DW, Luethke LE (1988) Detectability of tonal signals with changing interaural phase differences in noise. J Acoust Soc Am 83:1117-1123.

Grantham DW (1995) Spatial hearing and related phenomena. In: Hearing (Moore BCJ, ed). San Diego: Academic Press.

Green DM, Swets, J.A. (1966) Signal detection theory and psychophysics. New York: Wiley.

Green GG, Kay RH (1974) Proceedings: Channels in the human auditory system concerned with the wave form of the modulation present in amplitude and frequency-modulated tones. J Physiol 241:29P-30P.

Griffiths TD, Green GG, Rees A, Rees G (2000) Human brain areas involved in the analysis of auditory movement. Hum Brain Mapp 9:72-80.

Grunewald A, Linden JF, Andersen RA (1999) Responses to auditory stimuli in macaque lateral intraparietal area. I. Effects of training. J Neurophysiol 82:330-342.

Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. J Comp Neurol 394:475-495.

Hackett TA, Stepniewska I, Kaas JH (1999) Callosal connections of the parabelt auditory cortex in macaque monkeys. Eur J Neurosci 11:856-866.

Hashikawa T, Molinari M, Rausell E, Jones EG (1995) Patchy and laminar terminations of medial geniculate axons in monkey auditory cortex. J Comp Neurol 362:195-208.

Hauser MD, Andersson K (1994) Left hemisphere dominance for processing vocalizations in adult, but not infant, rhesus monkeys: field experiments. Proc Natl Acad Sci U S A 91:3946-3948.

Heffner H, Masterton B (1975) Contribution of auditory cortex to sound localization in the monkey (*Macaca mulatta*). J Neurophysiol 38:1340-1358.

Heffner HE, Heffner RS (1984) Temporal lobe lesions and perception of species-specific vocalizations by macaques. Science 226:75-76.

Heffner HE, Heffner RS (1990) Effect of bilateral auditory cortex lesions on sound localization in Japanese macaques. J Neurophysiol 64:915-931.

Heffner HE, Heffner RS (2003) Audition. In: Handbook of Research Methods in Experimental Psychology (Davis S, ed), pp 413-440: Blackwell.

Heil P (1997) Auditory cortical onset responses revisited. II. Response strength. J Neurophysiol 77:2642-2660.

Heil P (1997) Auditory cortical onset responses revisited. I. First-spike timing. J Neurophysiol 77:2616-2641.

Hind JE, Anderson DJ, Brugge JF, Rose JE (1967) Coding of information pertaining to paired low-frequency tones in single auditory nerve fibers of the squirrel monkey. J Neurophysiol 30:794-816.

Hocherman S, Benson DA, Goldstein MH, Jr., Heffner HE, Hienz RD (1976) Evoked unit activity in auditory cortex of monkeys performing a selective attention task. Brain Res 117:51-68.

Houben D, Gourevitch G (1979) Auditory lateralization in monkeys: an examination of two cues serving directional hearing. J Acoust Soc Am 66:1057-1063.

Hubel DH, Henson CO, Rupert A, Galambos R (1959) Attention units in the auditory cortex. Science 129:1279-1280.

Imig TJ, Adrian HO (1977) Binaural columns in the primary field (A1) of cat auditory cortex. Brain Res 138:241-257.

Imig TJ, Brugge JF (1978) Sources and terminations of callosal axons related to binaural and frequency maps in primary auditory cortex of the cat. J Comp Neurol 182:637-660.

Jackson LL, Heffner RS, Heffner HE (1999) Free-field audiogram of the Japanese macaque (*Macaca fuscata*). J Acoust Soc Am 106:3017-3023.

Jones EG, Dell'Anna ME, Molinari M, Rausell E, Hashikawa T (1995) Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. J Comp Neurol 362:153-170.

Joris PX, Yin TC (1992) Responses to amplitude-modulated tones in the auditory nerve of the cat. J Acoust Soc Am 91:215-232.

Joris PX (1998) Response classes in the dorsal cochlear nucleus and its output tract in the chloralose-anesthetized cat. J Neurosci 18:3955-3966.

Joris PX, Schreiner CE, Rees A (2004) Neural processing of amplitude-modulated sounds. Physiol Rev 84:541-577.

Kay RH, Matthews DR (1972) On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. J Physiol 225:657-677.

Kay RH (1982) Hearing of modulation in sounds. Physiol Rev 62:894-975.

Kilgard MP, Merzenich MM (1998) Cortical map reorganization enabled by nucleus basalis activity. Science 279:1714-1718.

Klump RG, and Eady, H.R. (1956) Some measurements of interaural time difference thresholds. Journal of the Acoustical Society of America 28:859-860.

Kowalski N, Versnel H, Shamma SA (1995) Comparison of responses in the anterior and primary auditory fields of the ferret cortex. J Neurophysiol 73:1513-1523.

Krishna BS, Semple MN (2000) Auditory temporal processing: responses to sinusoidally amplitude- modulated tones in the inferior colliculus. J Neurophysiol 84:255-273.

Kuhn GF (1977) Model for interaural time differences in the azimuthal plane. J Acoust Soc Am 62:157-167.

Kulkarni A, Colburn HS (1998) Role of spectral detail in sound-source localization. Nature 396:747-749.

Kuwada S, Yin T (1983) Binaural interaction in low-frequency neurons in inferior colliculus in the cat. I. Effects of long interaural delays, intensity, and repetition rate on interaural delay function. J Neurophysiol 50:981-999.

Kuwada S, Batra R, Stanford TR (1989) Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit: effects of sodium pentobarbital. J Neurophysiol 61:269-282.

Langner G, Schreiner CE (1988) Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. J Neurophysiol 60:1799-1822.

Le Prell CG, Moody DB (1997) Perceptual salience of acoustic features of Japanese monkey coo calls. J Comp Psychol 111:261-274.

LeDoux JE, Iwata J, Cicchetti P, Reis DJ (1988) Different projections of the central amygdaloid nucleus mediate autonomic and behavioral correlates of conditioned fear. J Neurosci 8:2517-2529.

Lee CC, Imaizumi K, Schreiner CE, Winer JA (2004) Concurrent tonotopic processing streams in auditory cortex. Cereb Cortex 14:441-451.

Liang L, Lu T, Wang X (2002) Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates. J Neurophysiol 87:2237-2261.

Licklider JCR, Webster, J.C., and Hedlun, J.M. (1950) On the frequency limits of binaural beats. Journal of the Acoustical Society of America 22:468-473.

Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. Cereb Cortex 14:731-740.

Linden JF, Grunewald A, Andersen RA (1999) Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation. J Neurophysiol 82:343-358.

Lu T, Liang L, Wang X (2001a) Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat Neurosci 4:1131-1138.

Lu T, Liang L, Wang X (2001b) Neural representations of temporally asymmetric stimuli in the auditory cortex of awake primates. J Neurophysiol 85:2364-2380.

Malhotra S, Hall AJ, Lomber SG (2004) Cortical control of sound localization in the cat: unilateral cooling deactivation of 19 cerebral areas. J Neurophysiol 92:1625-1643.

Malone BJ, Scott, B.H., and Semple, M.N. (2002) Context-dependent adaptive coding of interaural phase disparity in the auditory cortex of awake macaques. J Neurosci 22:4625-4638.

Mardia KaJ, PE (2000) Directional Statistics. New York: Wiley.

Matsubara JA, Phillips DP (1988) Intracortical connections and their physiological correlates in the primary auditory cortex (AI) of the cat. J Comp Neurol 268:38-48.

Mazurek ME, Shadlen MN (2002) Limits to the temporal fidelity of cortical spike rate signals. Nat Neurosci 5:463-471.

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

Merzenich MM, Brugge JF (1973) Representation of the cochlear partition of the superior temporal plane of the macaque monkey. Brain Res 50:275-296.

Merzenich MM, Knight PL, Roth GL (1975) Representation of cochlea within primary auditory cortex in the cat. J Neurophysiol 38:231-249.

Middlebrooks JC, Dykes RW, Merzenich MM (1980) Binaural response-specific bands in primary auditory cortex (AI) of the cat: topographical organization orthogonal to isofrequency contours. Brain Res 181:31-48.

Middlebrooks JC, Pettigrew JD (1981) Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. J Neurosci 1:107-120.

Middlebrooks JC, Clock AE, Xu L, Green DM (1994) A panoramic code for sound location by cortical neurons [see comments]. Science 264:842-844.

Miller JM, Sutton D, Pfingst B, Ryan A, Beaton R, Gourevitch G (1972) Single cell activity in the auditory cortex of Rhesus monkeys: behavioral dependency. Science 177:449-451.

Miller JM, Beaton RD, O'Connor T, Pfingst BE (1974) Response pattern complexity of auditory cells in the cortex of unanesthetized monkeys. Brain Res 69:101-113.

Molinari M, Dell'Anna ME, Rausell E, Leggio MG, Hashikawa T, Jones EG (1995) Auditory thalamocortical pathways defined in monkeys by calcium-binding protein immunoreactivity. J Comp Neurol 362:171-194.

Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. J Comp Neurol 335:437-459.

Nakamoto KT, Zhang J, Kitzes LM (2004) Response patterns along an isofrequency contour in cat primary auditory cortex (AI) to stimuli varying in average and interaural levels. J Neurophysiol 91:118-135.

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

Nelken I, Bizley JK, Nodal FR, Ahmed B, Schnupp JW, King AJ (2004) Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals. J Neurophysiol.

Orman SS, Phillips DP (1984) Binaural interactions of single neurons in posterior field of cat auditory cortex. J Neurophysiol 51:1028-1039.

Pandya DN, Yeterian EH (1984) Proposed neural circuitry for spatial memory in the primate brain. Neuropsychologia 22:109-122.

Pandya DN, Yeterian EH (1984) Architecture and connections of cortical association areas. In: Cerebral Cortex (Peters A, Jones EG, eds). New York: Plenum Press.

Parker A, Hawken M (1985) Capabilities of monkey cortical cells in spatial-resolution tasks. J Opt Soc Am A 2:1101-1114.

Parker AJ, Newsome WT (1998) Sense and the single neuron: probing the physiology of perception. Annu Rev Neurosci 21:227-277.

Pavani F, Macaluso E, Warren JD, Driver J, Griffiths TD (2002) A common cortical substrate activated by horizontal and vertical sound movement in the human brain. Curr Biol 12:1584-1590.

Paxinos GH, X. Toga, A.W. (1999) The Rhesus Monkey Brain in Stereotaxic Coordinates. New York: Academic Press.

Pelleg-Toiba R, Wollberg Z (1989) Tuning properties of auditory cortex cells in the awake squirrel monkey. Exp Brain Res 74:353-364.

Penfield W, Perot P (1963) The Brain's Record of Auditory and Visual Experience. A Final Summary and Discussion. Brain 86:595-696.

Petersen MR, Beecher MD, Zoloth SR, Moody DB, Stebbins WC (1978) Neural lateralization of species-specific vocalizations by Japanese macaques (Macaca fuscata). Science 202:324-327.

Pfingst BE, O'Connor TA, Miller JM (1977) Response plasticity of neurons in auditory cortex of the rhesus monkey. Exp Brain Res 29:393-404.

Pfingst BE, O'Connor TA (1980) A vertical stereotaxic approach to auditory cortex in the unanesthetized monkey. J Neurosci Methods 2:33-45.

Pfingst BE, O'Connor TA (1981) Characteristics of neurons in auditory cortex of monkeys performing a simple auditory task. J Neurophysiol 45:16-34.

Phillips DP, Irvine DR (1979) Methodological considerations in mapping auditory cortex: binaural columns in AI of cat. Brain Res 161:342-346.

Phillips DP, Irvine DR (1982) Properties of single neurons in the anterior auditory field (AAF) of cat cerebral cortex. Brain Res 248:237-244.

Phillips DP, Orman SS (1984) Responses of single neurons in posterior field of cat auditory cortex to tonal stimulation. J Neurophysiol 51:147-163.

Phillips DP, Semple MN, Kitzes LM (1995) Factors shaping the tone level sensitivity of single neurons in posterior field of cat auditory cortex. J Neurophysiol 73:674-686.

Poirier P, Miljours S, Lassonde M, Lepore F (1993) Sound localization in acallosal human listeners. Brain 116:53-69.

Poirier P, Lepore F, Provencal C, Ptito M, Guillemot JP (1995) Binaural noise stimulation of auditory callosal fibers of the cat: responses to interaural time delays. Exp Brain Res 104:30-40.

Poremba A, Malloy M, Saunders RC, Carson RE, Herscovitch P, Mishkin M (2004) Species-specific calls evoke asymmetric activity in the monkey's temporal poles. Nature 427:448-451.

Price CJ, Wise RJ, Warburton EA, Moore CJ, Howard D, Patterson K, Frackowiak RS, Friston KJ (1996) Hearing and saying. The functional neuro-anatomy of auditory word processing. Brain 119 (Pt 3):919-931.

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

Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of "what" and "where" in auditory cortex. Proc Natl Acad Sci U S A 97:11800-11806.

Rayleigh (1907) On our perception of sound direction. Phil Mag 13:214-232.

Read HL, Winer JA, Schreiner CE (2002) Functional architecture of auditory cortex. Curr Opin Neurobiol 12:433-440.

Reale RA, Imig TJ (1980) Tonotopic organization in auditory cortex of the cat. J Comp Neurol 192:265-291.

Reale RA, Brugge JF (1990) Auditory cortical neurons are sensitive to static and continuously changing interaural phase cues. J Neurophysiol 64:1247-1260.

Recanzone GH, Schreiner CE, Sutter ML, Beitel RE, Merzenich MM (1999) Functional organization of spectral receptive fields in the primary auditory cortex of the owl monkey. J Comp Neurol 415:460-481.

Recanzone GH, Guard DC, Phan ML, Su TK (2000) Correlation between the activity of single auditory cortical neurons and sound-localization behavior in the macaque monkey. J Neurophysiol 83:2723-2739.

Recanzone GH, Guard DC, Phan ML (2000) Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. J Neurophysiol 83:2315-2331.

Recanzone GH (2000) Response profiles of auditory cortical neurons to tones and noise in behaving macaque monkeys. Hear Res 150:104-118.

Rees A, Moller AR (1983) Responses of neurons in the inferior colliculus of the rat to AM and FM tones. Hear Res 10:301-330.

Rees A, Moller AR (1987) Stimulus properties influencing the responses of inferior colliculus neurons to amplitude-modulated sounds. Hear Res 27:129-143.

Rhode WS, Greenberg S (1994) Encoding of amplitude modulation in the cochlear nucleus of the cat. J Neurophysiol 71:1797-1825.

Rodrigues-Dagaeff C, Simm G, De Ribaupierre Y, Villa A, De Ribaupierre F, Rouiller EM (1989) Functional organization of the ventral division of the medial geniculate body of the cat: evidence for a rostro-caudal gradient of response properties and cortical projections. Hear Res 39:103-125.

Romanski LM, Bates JF, Goldman-Rakic PS (1999) Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J Comp Neurol 403:141-157.

Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. Nat Neurosci 2:1131-1136.

Rose JE, Gross NB, Geisler CD, Hind JE (1966) Some neural mechanisms in the inferior colliculus of the cat which may be relevant to localization of a sound source. J Neurophysiol 29:288-314.

Rose JE, Brugge JF, Anderson DJ, Hind JE (1967) Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J Neurophysiol 30:769-793.

Ryan AF, Miller JM, Pfingst BE, Martin GK (1984) Effects of reaction time performance on single-unit activity in the central auditory pathway of the rhesus macaque. J Neurosci 4:298-308.

Schreiner CE, Urbas JV (1986) Representation of amplitude modulation in the auditory cortex of the cat. I. The anterior auditory field (AAF). Hear Res 21:227-241.

Schreiner C, Langner G (1988) Coding of temporal patterns in the central auditory nervous system. In: Auditory Function: Neurobiological bases of hearing (Edelman GM, Gall WE, Cowan WM, eds). New York: Wiley.

Schreiner CE (1998) Spatial distribution of responses to simple and complex sounds in the primary auditory cortex. Audiol Neurootol 3:104-122.

Scott BH, B.J. Malone and M.N. Semple (2000) Physiological delineation of primary auditory cortex in the alert rhesus macaque. Soc for Neurosci Abstracts 358.20.

Semple MN, Kitzes LM (1987) Binaural processing of sound pressure level in the inferior colliculus. J Neurophysiol 57:1130-1147.

Semple MN, Kitzes LM (1993) Focal selectivity for binaural sound pressure level in cat primary auditory cortex: two-way intensity network tuning. J Neurophysiol 69:462-473.

Semple MN, Kitzes LM (1993) Binaural processing of sound pressure level in cat primary auditory cortex: evidence for a representation based on absolute levels rather than interaural level differences. J Neurophysiol 69:449-461.

Semple MN, Scott BH (2003) Cortical mechanisms in hearing. Curr Opin Neurobiol 13:167-173.

Shadlen MN, Britten KH, Newsome WT, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. J Neurosci 16:1486-1510.

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870-3896.

Shannon RV (2002) The relative importance of amplitude, temporal, and spectral cues for cochlear implant processor design. Am J Audiol 11:124-127.

Skottun BC (1998) Sound localization and neurons. Nature 393:531.

Skottun BC, Shackleton TM, Arnott RH, Palmer AR (2001) The ability of inferior colliculus neurons to signal differences in interaural delay. Proc Natl Acad Sci U S A 98:14050-14054.

Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87-90.

Snodderly DM, Gur M (1995) Organization of striate cortex of alert, trained monkeys (Macaca fascicularis): ongoing activity, stimulus selectivity, and widths of receptive field activating regions. J Neurophysiol 74:2100-2125.

Spezio ML, Keller CH, Marrocco RT, Takahashi TT (2000) Head-related transfer functions of the Rhesus monkey. Hear Res 144:73-88.

Spitzer MW, Semple MN (1991) Interaural phase coding in auditory midbrain: influence of dynamic stimulus features. Science 254:721-724.

Spitzer MW, Semple MN (1993) Responses of inferior colliculus neurons to time-varying interaural phase disparity: effects of shifting the locus of virtual motion. J Neurophysiol 69:1245-1263.

Spitzer MW, Semple MN (1995) Neurons sensitive to interaural phase disparity in gerbil superior olive: diverse monaural and temporal response properties. J Neurophysiol 73:1668-1690.

Spitzer MW, Semple MN (1998) Transformation of binaural response properties in the ascending auditory pathway: influence of time-varying interaural phase disparity. J Neurophysiol 80:3062-3076.

Swarbrick L, Whitfield IC (1972) Auditory cortical units selectively responsive to stimulus 'shape'. J Physiol 224:68P-69P.

Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19:109-139.

Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. Science 292:290-293.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23:775-785.

Ungerleider LG, Mishkin M (1982) In: *Analysis of Visual Behavior* (Ingle DJ, Goodale MA, Mansfield RJW, eds), pp 549-586. Cambridge, MA: MIT Press.

Vaadia E, Benson DA, Hienz RD, Goldstein MH, Jr. (1986) Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. J Neurophysiol 56:934-952.

Varela JA, Sen K, Gibson J, Fost J, Abbott LF, Nelson SB (1997) A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. J Neurosci 17:7926-7940.

Velenovsky DS, Cetas JS, Price RO, Sinex DG, McMullen NT (2003) Functional subregions in primary auditory cortex defined by thalamocortical terminal arbors: an electrophysiological and anterograde labeling study. J Neurosci 23:308-316.

Wagner H, Takahashi T (1992) Influence of temporal cues on acoustic motion-direction sensitivity of auditory neurons in the owl. J Neurophysiol 68:2063-2076.

Wang X, Merzenich MM, Beitel R, Schreiner CE (1995) Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. J Neurophysiol 74:2685-2706.

Wegener JG (1974) Interaural intensity and phase angle discrimination by rhesus monkeys. J Speech Hear Res 17:638-655.

Weinberger NM, Bakin JS (1998) Learning-induced physiological memory in adult primary auditory cortex: receptive fields plasticity, model, and mechanisms. Audiol Neurootol 3:145-167.

Whitfield IC (1984) The role of auditory cortex in behavior. In: Cerebral Cortex (Peters A, Jones EG, eds). New York: Plenum Press.

Wilson FA, Scalaidhe SP, Goldman-Rakic PS (1993) Dissociation of object and spatial processing domains in primate prefrontal cortex. Science 260:1955-1958.

Yamada K, Kaga K, Uno A, Shindo M (1996) Sound lateralization in patients with lesions including the auditory cortex: comparison of interaural time difference (ITD) discrimination and interaural intensity difference (IID) discrimination. Hear Res 101:173-180.

Yin TC, Kuwada S (1983) Binaural interaction in low-frequency neurons in inferior colliculus of the cat. III. Effects of changing frequency. J Neurophysiol 50:1020-1042.

Yin TC, Kuwada S (1983) Binaural interaction in low-frequency neurons in inferior colliculus of the cat. II. Effects of changing rate and direction of interaural phase. J Neurophysiol 50:1000-1019.

Yin TC, Chan JC (1990) Interaural time sensitivity in medial superior olive of cat. J Neurophysiol 64:465-488.

Yin PT, H. Moorhead, K. Mishkin, M. (2003) Neuronal responses to unpredicted reward in rostral superior temporal gyrus of the monkey. Program 1817 Society for Neuroscience Abstract Viewer/Itinerary Planner.

Yost WA (1991) Auditory image perception and analysis: the basis for hearing. Hear Res 56:8-18.

Zhang J, Nakamoto KT, Kitzes LM (2004) Binaural interaction revisited in the cat primary auditory cortex. J Neurophysiol 91:101-117.

Zurita P, Villa AE, de Ribaupierre Y, de Ribaupierre F, Rouiller EM (1994) Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. Neurosci Res 19:303-316.

Zwislocki J, Feldman RS (1956) Just noticeable differences in diotic phase. J Acoust Soc Am 28:860-862.