Title
Auditory Processing and Perception in Songbirds

Author
Nagel, Katherine

Publication Date
2007-06-29

Peer reviewed|Thesis/dissertation

Auditory Processing and Perception in Songbirds

by

Katherine Nagel

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of
DOCTOR OF PHILOSOPHY

in
NEUROSCIENCE

in the
GRADUATE DIVISION

Acknowledgements

I would like to thank Mimi Kao, Kamal Sen, and Maneesh Sahani for teaching me chronic recording, Matlab, and reverse correlation, Sarah C. Woolley for companionship in both recording and behavioral experiments, Tatyana Sharpee for shaping the experimental design in chapters 1 and 2, as well as for general computational guidance, Adria Arteseros for histology and for keeping the lab running, Jonathan Kaplan for suggesting the bi-variate Mexican hat model in chapter 2, Ram Ramachadran for his enthusiasm for the auditory system, Lazlo Bocskai for designing and building the operant cages used in chapter 3, Ken McGary for assistance with their electrical wiring, Alex, Yvonne, and Rebecca Nagel for mathematical, technical, and emotional support, and David Schoppik for the questions that launched and maintained this project over many years.

Chapter1 appeared originally in Neuron 51: 845-851

Auditory Processing and Perception in Songbirds

by

Katherine Nagel

Abstract

Songbirds, like humans, learn to produce and to recognize complex, species-specific sounds, providing a biologically tractable model to study the neural mechanisms of speech production and perception. I used chronic recording from single neurons, and operant behavioral techniques to ask how complex sounds are represented in the songbird forebrain, and how this representation may be related to the birds' perception of song. I found that neurons in field L, the avian analog of the human primary auditory cortex, represent three different types of modulations found in natural sounds: spectral modulations, temporal modulations, and spectro-temporal modulations. Neurons specialized for different modulations have different physiological properties and are localized to different parts of field L. The response properties of these neurons depend nonlinearly on the average intensity of the stimulus. At high intensities, they respond only to differences in sound energy between nearby frequency or times, while at low intensities they integrate information from nearby frequencies and times. This nonlinearity is shared with the visual system and may represent a computational principle of sensory encoding. Finally, I used operant techniques to ask whether songbirds could generalize a learned song discrimination task to songs altered in pitch, duration, or volume. I found that birds generalized correctly to songs altered in duration but not to those altered in pitch or volume. These data suggest that birds use the spatial pattern of neurons activated by a song rather than the temporal pattern of neural activation to determine what song they heard.

Table of Contents

List of Figures

General Introduction

Our ability to speak depends on our ability to perceive and categorize complex sounds. A normal adult speaker of English can recognize tens of thousands of words (Kirkpatrick, 1891; Hartman, 1946), despite variation in their volume, pitch, prosody, and intonation. Many conversations take place in a background of other speakers or poor telephone connections. How the human brain executes robust acoustic pattern recognition in these circumstances is poorly understood.

Like humans, songbirds learn to produce and recognize complex, species-specific sounds (Konishi 1985, Doupe and Kuhl, 1999). Songbirds have been used as a biologically tractable model of vocal learning for many years, while their ability to recognize and perceive song has received less attention. While vocal learning in many species is restricted to young male birds (Konishi, 1985, Doupe and Kuhl, 1999), both males and females learn to recognize new songs throughout life (Miller 1979, Stripling et al, 2003). Song has a rich statistical structure that can share many features with human speech (Singh and Theunissen, 2003). The major auditory areas of the avian brain have been identified and mapped anatomically (Fortune and Margoliash, 1992, Fortune and Margoliash, 1995, Vates et al, 1996). Songbirds thus provide an excellent model for asking questions about the neural basis of complex sound recognition.

The avian auditory system is composed of many levels. Auditory information enters the brain through the brainstem cochlear nuclei, angularis (NA) and magnocellularis (NM, Sullivan and Konishi, 1984). These two areas project to other brainstem nuclei, including nucleus laminaris (NL), and nucleus ventralis laterale pars posterior (VLVp), which in turn project to inferior colliculus in the midbrain (Takahashi et al, 2003), often called MLd in songbirds (Woolley et al, 2005). MLd projects to nucleus ovoidalis (Ov), the auditory region of the avian thalamus, and ovoidalis projects to the forebrain area field L (Woolley et al, 2005). Field L is composed of 3 regions, the thalamo-recipient input area L2, and two flanking areas, L1 and L3. L2 is further subdivided into L2a anteriorly, and the more diffuse L2b posteriorly. L1 and L3 are reciprocally connected, and form reciprocal connections with several higher order forebrain auditory areas, including the mediocaudal neostriatum (NCM), the caudomedial mesopallium (CMM), and nucleus interfacialis (NIf, Fortune and Margoliash, 1992, Fortune and Margoliash, 1995, Vates et al, 1996). HVc, an area involved in song production in males, receives most of its auditory input from NIf (Coleman and Mooney, 2004). Studies of the avian auditory brainstem and midbrain have generally focussed on how spatial cues are extracted and

processed (Takahashi and Konishi, 1984, Sullivan and Konishi, 1984, Sullivan 1985), while studies of the thalamus and forebrain have focussed on how complex song stimuli may are identified and learned (Margoliash 1983, Lewicki and Arthur, 1996, Stripling etal, 2001, Theunissen et al, 1998, 2000; Sen et al, 2001, Woolley et al, 2005, 2006).

Most studies of the auditory processing of song have taken one of two approaches. One approach has concentrated on a search for "selective" neurons, that respond (with a broad increase in firing rate) only to a single complex stimulus. Neurons selective for a male bird's own song have been found in the song production nucleus HVC (Margoliash, 1983, Margoliash and Fortune, 1992, Theunissen and Doupe, 1998), but lesion studies trying to implicate this area in general auditory perception have produced inconsistent results (Brenowitz, 1991; MacDougall-Schackleton et al., 1998; Gentner et al., 1999). Neurons that showed weaker selectivity for familiar conspecific songs have been found in the high-level auditory areas NCM and CMM (Stripling et al, 2001; Gentner and Margoliash, 2003). On the basis of such studies, the auditory forebrain has been described as a selectivity hierarchy (Margoliash 1986, Lewicki and Arthur, 1996, Sen et al, 2001), with low-level areas such as field L responding well to nearly all natural and synthetic auditory stimuli, while higher areas such as NCM, cHV, and Nif show a larger proportion of neurons selective for aspects of song.

Selectivity has also been examined using genomic responses to song stimuli. Much of the auditory forebrain expresses the immediate early gene ZENK in response to playback of novel conspecific songs, but not heterospecific songs or simple tone sequences (Mello et al. 1992, Bailey et al, 2002). This selective activation is strongest in areas NCM and CMM, and has been taken as evidence that these areas store memories of learned songs (Jarvis et al, 1995; Bolhuis et al, 2001; Stripling et al, 2001; Terpstra et al, 2004, 2006; Gobes et al, 2007). The thalamic input area L2 shows little zenk expression in response to song or other sounds (Mello and Clayton, 1994), although electrophysiological studies demonstrate that it responds robustly to many kinds of sounds (Theunissen et al, 2000, Sen et al, 2001, Woolley et al, 2005, 2006). The biological meaning of this gene activation thus remains unclear.

A second tradition of auditory research has aimed to characterize the auditory response properties of neurons systematically. Many such studies use simple stimuli, such as tones and noise, to map responses as a function of some auditory parameter (Scheich et al. 1979, Muller and Leppelsack, 1985, Heil and Scheich, 1985, Hose et al, 1987, Gehr et al. 1999). More recently, a technique known as reverse correlation (Eggermont 1983) has been used to characterize the responses of auditory neurons to spectrally rich stimuli such as ripples and songs (Kowalski et al, 1996, 1997;

Theunissen et al, 2000; Sen et al, 2001; Miller et al, 2002, Depireux et al, 2001; Linden, 2003). These techniques capture the linear aspects of a neuron's response to a set of complex sounds, and provide a compact model of how the neuron will respond to novel stimuli. Although this technique has great promise, the general organization of the avian forebrain's primary auditory area, field L, remains unknown. Few studies have tried to characterize nonlinearities in neural responses systematically. Finally, very few studies have compared neural response properties to perceptual behavior. While the acoustic parameters that govern how we perceive speech have been quantified (Avedaño et al, 2004; Handel, 1989), those governing birds' responses to song have not. This omission makes it difficult to relate the response properties of neurons to the the bird's perception.

In this dissertation I have tried to approach several of these questions: How are complex sounds represented by neurons in field L? How are auditory responses related to the anatomical and physiological structure of field L? What nonlinearities are present in these neuron's responses and how can we model them mathematically? What acoustic parameters are most critical for identifying an individual by his song? At the end of the dissertation, I discuss how the neural representation in field L may be related to the birds' perceptual behavior, and propose experiments to test these hypotheses.

In chapters 1 and 2, I use reverse correlation to ask what aspects of a complex sound drive different neurons in field L to fire. In chapter 1, I use this technique to quantify temporal response properties--responses to changes in the overall level of a sound with fixed frequency content-- and demonstrate that these temporal response properties depend in systematic ways on the mean and variance of the stimulus. At low mean intensities, cells act more "integrators," responding whenever the stimulus is on. At high mean intensities, cells acts more like "differentiators," responding only when the stimulus intensity changes abruptly. When the stimulus variance increases, neural gain decreases, allowing cells to adjust their range of outputs to their range of inputs. These nonlinearities are closely related to those observed in the retina (Enroth-Cugell and Lennnie, 1975; Chander and Chichilnisky, 2001; Kim and Rieke, 2001; Baccus and Meister, 2002) and suggest that similar computational principles may underlie sensory encoding in the auditory and visual domains.

In chapter 2, I extend the methods of chapter 1 to characterize responses as a function of both frequency (spectrum) and time. I identify three classes of cells in field L: cells tuned to temporal modulations, cells tuned to spectral modulations, and cells tightly tuned in both spectrum and time. I demonstrate that spectrally-tuned cells differ in their physiology and anatomical distribution from the other two types. Finally, I examine how spectro-temporal responses depend on stimulus intensity and

find that similar changes across cells allow each cell type to become more specialized for its preferred modulations when the stimulus is louder. These findings both extend the results of chapter 1 to two dimensions, and suggest organizing principles for the structure of field L.

Several innovations distinguish these two studies from previous ones that used similar techniques to characterize neural responses in both birds (Theunissen et al, 2000, Sen et al, 2001, Woolley et al, 2005, 2006) and mammals (Kowalski et al, 1996, 1997; Miller et al, 2002, Depireux et al, 2001; Linden, 2003). First, I developed novel synthetic stimuli that share many statistical properties with song but have a more uniform and controlled structure. These stimuli drove neurons exceptionally well, and permitted clean and robust estimates of neural response properties. Second, I systematically varied the parameters of this stimulus to characterize nonlinearities in neural responses. Finally, I recorded from chronically-implanted unanesthetized animals, revealing structures in response patterns that may have been absent from previous, anesthetized, recordings.

In the last chapter of my thesis I use an operant behavioral paradigm (Beecher et al, 1994; Gentner and Hulse, 1998) to ask what acoustic parameters zebra finches use to identify songs of different individuals. After training females to categorize a large set of songs from two different males, I test them on songs altered in their pitch, duration, and volume. I find that females correctly classify songs whose duration was significantly altered, but do not generalize well to songs shifted in pitch or played at much louder or softer volumes. These findings place constraints on which aspects of neural firing (such as temporal patterning and spectral selectivity) are used to identify and discriminate between songs.

Together, these three chapters provide a coherent picture of auditory processing in field L and suggest hypotheses for the role of this area in auditory processing and song recognition.

References:

Avedaño, C., Deng, L., Hermansky, H., and Gold, B. (2004) "The analysis and representation of speech." Speech Processing in the Auditory System. New York, NY, Springer-Verlag.

Baccus, S. A. and M. Meister (2002). "Fast and slow contrast adaptation in retinal circuitry." Neuron 36(5): 909-19.

Bailey, D. J., J. C. Rosebush, et al. (2002). "The hippocampus and caudomedial neostriatum show selective responsiveness to conspecific song in the female zebra finch." J Neurobiol 52(1): 43-51.

Beecher, M. D., S. E. Campbell, et al. (1994). "Song Perception in the song sparrow: birds classify by song type but not by singer." Animal Behavior 47: 1343-1351.

Bolhuis, J. J., E. Hetebrij, et al. (2001). "Localized immediate early gene expression related to the strength of song learning in socially reared zebra finches." Eur J Neurosci 13(11): 2165-70.

Brenowitz, E. A. (1991). "Altered perception of species-specific song by female birds after lesions of a forebrain nucleus." Science 251(4991): 303-5.

Chander, D. and E. J. Chichilnisky (2001). "Adaptation to temporal contrast in primate and salamander retina." J Neurosci 21(24): 9904-16.

Coleman, M.J., Mooney, R. (2004) "Synaptic transformations underlying selective auditory representations of learned birdsong." J Neurosci, 24(33):7251-65.

Depireux, D. A., J. Z. Simon, et al. (2001). "Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex." J Neurophysiol 85(3): 1220-34.

Doupe, A. J. and P. K. Kuhl (1999). "Birdsong and human speech: common themes and mechanisms." Annu Rev Neurosci 22: 567-631.

Eggermont, J. J., P. M. Johannesma, and A.M. Aertsen (1983). "Reverse-correlation methods in auditory research." Q Rev Biophys 16(3): 341-414.

Enroth-Cugell, C. and P. Lennie (1975). "The control of retinal ganglion cell discharge by receptive field surrounds." J Physiol 247(3): 551-78.

Fortune, E. S. and D. Margoliash (1992). "Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (Taenopygia guttata)." J Comp Neurol 325(3): 388-404.

Fortune, E. S. and D. Margoliash (1995). "Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (Taeniopygia guttata)." J Comp Neurol 360(3): 413-41.

Gehr, D. D., B. Capsius, P. Grabner, M. Gahr, and H.J. Leppelsack (1999). "Functional organisation of the field-L-complex of adult male zebra finches." Neuroreport 10(2): 375-80.

Gentner, T. Q. and S. H. Hulse (1998). "Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris." Anim Behav 56(3): 579-594.

Gentner, T. Q., S. H. Hulse, et al. (2000). "Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds." J Neurobiol 42(1): 117-33.

Gentner, T. Q. and D. Margoliash (2003). "Neuronal populations and single cells representing learned auditory objects." Nature 424(6949): 669-74.

Gobes, S. M. and J. J. Bolhuis (2007). "Birdsong memory: a neural dissociation between song recognition and production." Curr Biol 17(9): 789-93.

Handel, S. (1989) Listening: An introduction to the perception of auditory events. Cambridge, MA. MIT Press.

Hartmann, G.W. (1946). "Further Evidence on the Unexpected Large Size of Recognition Vocabularies among College Students." Journal of Educational Psychology 37:436 - 439.

Heil, P. and H. Scheich (1985). "Quantitative analysis and two-dimensional reconstruction of the tonotopic organization of the auditory field L in the chick from 2-deoxyglucose data." Exp Brain Res 58(3): 532-43.

Hose, B., G. Langner, and H. Scheich (1987). "Topographic representation of periodicities in the forebrain of the mynah bird: one map for pitch and rhythm?" Brain Res 422(2): 367-73.

Jarvis, E. D., C. V. Mello, et al. (1995). "Associative learning and stimulus novelty influence the song-induced expression of an immediate early gene in the canary forebrain." Learn Mem 2(2): 62-80.

Kim, K. J. and F. Rieke (2001). "Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells." J Neurosci 21(1): 287-99.

Kirkpatrick, E.A. (1891). "The Number of Words in an Ordinary Vocabulary." Science 18:107 - 108.

Konishi, M. (1985) "Birdsong: from behavior to neuron." Annu Rev Neurosci, 8:125-70.

Kowalski, N., D. A. Depireux, et al. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra." J Neurophysiol 76(5): 3524-34.

Kowalski, N., D. A. Depireux, et al. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra." J Neurophysiol 76(5): 3503-23.

Lewicki, M. S. and M. Konishi (1995). "Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order." Proc Natl Acad Sci U S A 92(12): 5582-6.

Lewicki, M. S. and B. J. Arthur (1996). "Hierarchical organization of auditory temporal context sensitivity." J Neurosci 16(21): 6987-98.

Linden, J. F., R. C. Liu, et al. (2003). "Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex." J Neurophysiol 90(4): 2660-75.

MacDougall-Shackleton, S. A., S. H. Hulse, et al. (1998). "Neural bases of song preferences in female zebra finches (Taeniopygia guttata)." Neuroreport 9(13): 3047-52.

Margoliash, D. (1983). "Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow." J Neurosci 3(5): 1039-57.

Margoliash, D. and E. S. Fortune (1992). "Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc." J Neurosci 12(11): 4309-26.

Mello, C. V. and D. F. Clayton (1994). "Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system." J Neurosci 14(11 Pt 1): 6652-66.

Mello, C. V., D. S. Vicario, et al. (1992). "Song presentation induces gene expression in the songbird forebrain." Proc Natl Acad Sci U S A 89(15): 6818-22.

Miller, D. B. (1979). "The acoustic basis of mate recognition by female zebra finches (Taeniopygia guttata)." Animal Behavior 27: 376-380.

Miller, L. M., M. A. Escabi, et al. (2002). "Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex." J Neurophysiol 87(1): 516-27.

Muller, C. M. and H. J. Leppelsack (1985). "Feature extraction and tonotopic organization in the avian auditory forebrain." Exp Brain Res 59(3): 587-99.

Scheich, H., B. A. Bonke, D. Bonke, and G. Langner (1979). "Functional organization of some auditory nuclei in the guinea fowl demonstrated by the 2-deoxyglucose technique." Cell Tissue Res 204(1): 17-27.

Sen, K., F. E. Theunissen, et al. (2001). "Feature analysis of natural sounds in the songbird auditory forebrain." J Neurophysiol 86(3): 1445-58.

Singh, N. C. and F. E. Theunissen (2003). "Modulation spectra of natural sounds and ethological theories of auditory processing." J Acoust Soc Am 114(6 Pt 1): 3394-411.

Stripling, R., A. A. Kruse, et al. (2001). "Development of song responses in the zebra finch caudomedial neostriatum: role of genomic and electrophysiological activities." J Neurobiol 48(3): 163-80.

Stripling, R., L. Milewski, et al. (2003). "Rapidly learned song-discrimination without behavioral reinforcement in adult male zebra finches (Taeniopygia guttata)." Neurobiol Learn Mem 79(1): 41-50.

Sullivan, W. E. and M. Konishi (1984). "Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl." J Neurosci 4(7): 1787-99.

Sullivan, W. E. (1985). "Classification of response patterns in cochlear nucleus of barn owl: correlation with functional response properties." J Neurophysiol 53(1): 201-16.

Takahashi, T., A. Moiseff, et al. (1984). "Time and intensity cues are processed independently in the auditory system of the owl." J Neurosci 4(7): 1781-6.

Takahashi, T. T., A. D. Bala, et al. (2003). "The synthesis and use of the owl's auditory space map." Biol Cybern 89(5): 378-87.

Theunissen, F. E. and A. J. Doupe (1998). "Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches." J Neurosci 18(10): 3786-802.

Theunissen, F. E., K. Sen, et al. (2000). "Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds." J Neurosci 20(6): 2315-31.

Terpstra, N. J., J. J. Bolhuis, et al. (2004). "An analysis of the neural representation of birdsong memory." J Neurosci 24(21): 4971-7.

Terpstra, N. J., J. J. Bolhuis, et al. (2006). "Localized brain activation specific to auditory memory in a female songbird." J Comp Neurol 494(5): 784-91.

Wild, J. M., H. J. Karten, and B.J. Forst (1993). "Connections of the auditory forebrain in the pigeon (Columba livia)." J Comp Neurol 337(1): 32-62.

Woolley, S. M., T. E. Fremouw, et al. (2005). "Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds." Nat Neurosci 8(10): 1371-9.

Woolley, S. M., P. R. Gill, et al. (2006). "Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain." J Neurosci 26(9): 2499-512.

Zahn, R.A. (1996). The Zebra FInch: A Synthesis of Field and Laboratory Studies. Oxford University Press.

Chapter 1: Temporal processing and adaptation in the songbird auditory forebrain

Abstract:

Auditory neurons must encode the dynamics of natural sounds over many intensity ranges. We investigated how songbird auditory forebrain neurons encode amplitude modulations when the distribution of intensities changes. Using reverse-correlation, we modeled neural responses as the output of a linear filter and a nonlinear gain function, then asked how filters and nonlinearities depend on the mean and variance of amplitude modulations. Filter shape depended strongly on mean amplitude (volume): at low mean, most neurons integrated sound over many milliseconds, while at high mean, neurons responded more to local changes in amplitude. Increasing the variance (contrast) of amplitude modulations had less effect on filter shape but decreased the gain of firing in most cells. Both filter and gain changes occurred rapidly after a change in amplitude statistics, suggesting that they represent nonlinearities in processing. These changes may permit neurons to signal effectively over a wider dynamic range and are reminiscent of findings in other sensory systems.

Introduction:

A central problem for all sensory systems is how to represent complex dynamic stimuli over a wide range of intensities.

Songbirds, like humans, recognize their vocalizations whether they are soft or loud (Lohr et al, 2003). How the brain achieves this level-invariant recognition is unclear. At its highest levels, the songbird auditory system contains 'feature detectors' that respond selectively to individual learned songs (Margoliash 1983, Margoliash and Fortune, 1992; Gentner and Margoliash, 2003). These areas receive input from an area called field L, the avian analog of primary auditory cortex (Wild et al, 1993). Field L responds broadly to many classes of auditory stimuli, and shows organized tuning for basic auditory properties such as spectral frequency and temporal modulations (Scheich et al. 1979, Muller and Leppelsack, 1985, Heil and Scheich, 1985, Hose et al, 1987, Lewicki and Arthur 1996, Gehr et al. 1999, Hausberger et al., 2000, Theunissen and Sen 2000, Sen et al. 2001, Grace et al. 2002). Several groups have tried to model how song-selective responses might arise from combinations of field L outputs (Lewicki and Konishi, 1995, Drew and Abbott, 2005). However, few studies have looked at how neural responses in these areas depend on song or stimulus intensity.

In the mammalian auditory brainstem and midbrain, responses to simple tones and noises can depend on intensity in highly non-linear ways (Young and Voigt, 1982; Sachs and Young, 1980; Nelken and Young, 1994; Nelken and Young, 1997). In the cortex, the picture is less clear. Many studies have shown non-monotonic and non-linear responses to tones, intensity modulations, and ripple stimuli (Phillips and Hall, 1987; Phillips et al, 1994; Calhoun and Schreiner, 1998), while other studies have stressed the linearity of cortical responses (Kowalski et al, 1996; Escabi et al, 2003; Barbour and Wang, 2003). How post-thalamic auditory neurons encode stimulus features across intensities remains controversial.

The goal of the present study was to understand how the responses of field L neurons to complex dynamic stimuli depend on the intensity of the stimulus. We focused on coding of amplitude modulations in time, because these are a prominent feature of both song and speech, and can carry a great deal of the information present in these signals (Shannon et al, 1995, Theunissen and Doupe, 1998).

To describe the processing of naturalistic amplitude modulations over different intensity ranges, we developed a set of synthetic stimuli that capture many aspects of song's amplitude modulations, but sample the space of possible modulations more thoroughly. We then used reverse correlation techniques (Eggermont 1983, 1993; Epping and Eggermont, 1986; Kim and Young, 1994; Brenner et al., 2000; Chander and Chichilnisky, 2001; Fairhall et al., 2001; Depireux et al., 2001; Kim and Rieke, 2001, Miller et al., 2002; Escabi and Schreiner 2002; Baccus and Meister, 2002) to extract a linear receptive field (filter) and a nonlinear gain function from the responses to these stimuli. Together, the filter and nonlinearity allowed us to characterize the feature selectivity, threshold, and gain of each cell, and to predict responses to novel sounds.

We found that our stimuli permitted robust estimates of filters and nonlinear gain functions for field L neurons. These models made good predictions of responses to novel synthetic stimuli, and revealed neural sensitivity to a broad range of stimulus time scales and features. We then demonstrated that filters and gain functions depend in systematic and specific ways on the mean and variance of the stimulus amplitude. These changes are very reminiscent of those observed in the early visual system (Enroth-Cugell and Lennie, 1975; Chander and Chichilnisky, 2001; Baccus and Meister, 2002), and shown mathematically to improve coding efficiency (Attick, 1992). Finally, we examined the time course of these changes and found that they occurred on two time scales: a fast change in filter shape and gain, followed by a slower change in threshold.

Our findings suggest that common computational strategies underlie sensory processing in

multiple domains, but challenge current models of song recognition.

Results:

Using Amplitude-Modulated Noise Stimuli to Drive Field L Neurons

To study processing of amplitude modulations in field L, we developed a stimulus composed of two parts: a slowly varying modulation envelope and a rapidly varying noise carrier.

The envelope was designed to capture the temporal frequency and amplitude distributions of natural sounds (Attias and Schreiner, 1998 , Singh and Theunissen, 2003). Because they are dominated by slow changes in amplitude, natural sounds have power spectra that decrease as a function of temporal frequency (Singh and Theunissen, 2003). Natural communication sounds also contain many silent periods, giving rise to an exponential distribution of amplitudes, but a more Gaussian distribution of log (dB) amplitudes (Singh and Theunissen, 2003).

To create a stimulus with these temporal frequency and amplitude properties**,** we first created a Gaussian noise signal with a decreasing exponential distribution of temporal frequencies (figure 1A, lower panel). This modulation signal specifies the loudness of the stimulus in decibels at each point in time (figure 1A, upper panel). We then exponentiated the modulation signal to produce a pressure envelope with an exponential distribution of linear amplitudes (figure 1B). Finally, we multiplied this envelope with a narrow or broadband noise carrier (figure 1C) to generate the full stimulus (figure 1D, see Methods for full details).

Although this stimulus consisted of both fast (carrier) and slow (envelope) fluctuations, we analyzed responses only with respect to the slowly varying modulation signal (figure 1A). Recent studies have suggested that this log envelope is the property most linearly encoded by neurons in field L (Gill et al, 2006). To isolate responses to the slowly varying signal, we repeated the same modulation signal but used different noise segments to form the carrier each time. Figure 1E shows the response raster of a cell to 100 such repeats: the columns of spikes represent reliable spike patterns driven by the slow modulation signal.

Half of the five-second long segments in our stimulus were randomly assigned to be repeats of the same modulation signal; the remaining signals were unique. Unique signals were used to broadly sample the space of possible amplitude modulations, and to estimate model parameters. Repeated trials were used to test the model. Because we used different data to fit and test the model, we ensured that the quality of our predictions was not due to over-fitting.

Responses of Field L Neurons are Well-Modeled Using a Linear-Nonlinear Model

To characterize the temporal response properties of each cell, we modeled its responses using a linear-nonlinear model (Brenner et al., 2000; Chander and Chichilnisky, 2001; Fairhall et al., 2001; Baccus and Meister, 2002).

Figure 2 illustrates this process for two cells. The linear filter (figures 2A and D) tells us what feature of the stimulus best drives the neuron. It is extracted from the data by computing the average modulation signal surrounding a spike, then removing the influence of stimulus correlations from the resulting waveform (Methods). The spike occurs at zero on the x-axis; the filter has structure only to the left of (before) the spike, as the cell responds causally to features of the stimulus. Small error bars (dotted lines, Methods) indicate that our estimates of filter shape were robust.

The shape of the filter indicates its preferred temporal feature. Filters with a single dominant peak behave as low-pass "integrators," and produce responses that look like a smoothed, delayed version of the log stimulus envelope. A response of this type (from the cell in figure 2A) is shown in figure 1E, where a column of spikes follows each peak in the stimulus above (figure 1D). The sign of the filter peak determines whether the cell's firing rate is elevated or depressed by peaks in the stimulus. Filters with a biphasic shape behave as band-pass "differentiators," and respond to either onsets or offsets in the stimulus, depending on the order of the positive and negative components in the filter. Most of our cells behave as combinations of integrators and differentiators, making a strict classification of cell types difficult. The filter shown in figure 2A has a small negative peak followed by a large positive peak; it behaves mostly as a positive integrator, but is sensitive to onsets as well. The filter shown in figure 2B behaves mostly as a negative integrator, but also responds weakly to offsets.

The width of the filter sets an upper limit on the temporal frequencies to which the cell can respond. Because the exact frequency response depends both on the width of the filter and its shape, preferred temporal frequency is best characterized in the frequency domain (figure 4A-C, 2nd column). Figure 2G shows the 50% width of the absolute value of each filter (50% width), versus the peak of each filter's power spectrum (its "best modulation frequency" or BMF). Each parameter has an approximately uniform distribution along its log axis, suggesting a greater than 10-fold spread in temporal frequency sensitivities among cells.

The nonlinearity (figures 2B and E) describes the relationship between filter output and firing rate. It is calculated by passing the mean-subtracted modulation signal through the filter—measuring the similarity of the stimulus to the filter at each point in time-- then comparing the filter output to the neuron's actual firing rate (Methods). Most of the cells we recorded had nonlinearities with a flat, sub-threshold region at negative values, and a linear coding region at positive values, when the stimulus most closely resembled the filter. Some, like that in figure 2B, showed saturation at high positive values.

To assess the quality of our models, we used the filter and nonlinearity pair to generate a prediction of each neuron's response to the repeated modulation signal (figures 2C and 2F). We then compared this prediction (thick gray lines) to the actual PSTH (thin black line). The correlation coefficient between prediction and data for the neuron in figures 2A-2C was 0.89, making this one of our best models. The neuron in figures 2D-2F had a correlation coefficient of 0.67, the median for our population. The distribution of correlation coefficients is shown in figure 2H (n = 33 significant filters from 36 recorded cells). The strength of these correlations suggests that our model captures a significant portion of the behavior of neurons in our population.

Responses of Field L Neurons to Changes in Stimulus Statistics

To examine how coding depends on stimulus statistics, we altered the mean and the variance of amplitude modulations in our stimulus (figure 3A). In our baseline condition (gray), the modulation signal had a mean of 30dB, and a standard deviation of 6dB. In the high mean condition (red), the mean was raised to 63dB, while in high variance (green) the standard deviation was raised to 18dB. The high mean and high variance conditions were designed to have approximately the same overall power (67 dB). Stimulus conditions were presented continuously, and the baseline condition was repeated twice in the series allowed us to characterize the transitions to and from this condition to the other two.

Changes in Stimulus Mean Produce Systematic Changes in Firing Rate and Filter Shape

Most of the cells we recorded responded strongly to a change in stimulus mean, with a rapid

change in firing rate followed by a slow decay (figure 3B). The rates of decay varied across cells from approximately 4 msec to 4 seconds and were generally much slower than would be predicted by the width of the filters. Decay times were poorly correlated with filter width (figure 3D, cc = -0.20 for low to high mean, and 0.04 for high to low mean), suggesting that filter width and decay time represent two separate forms of temporal sensitivity—one on the time scale of syllables, and one on the time scale of motifs or bouts. High and low mean decay times were weakly correlated (supplementary figure 3).

An increase in stimulus mean produced systematic changes in filter shape. Figure 4 shows three examples of filters derived from low mean (gray) and high mean (red) conditions for the same cells. All three show related changes when the mean increases: the filters become narrower, and the sizes of the positive and negative components become more closely matched. These changes in filter shape are also reflected in the power spectrum representation of each filter. The narrowing of the filter in time is reflected in a shift of the power spectrum peak towards higher frequencies, while the increased balance between positive and negative components reduces the response to low frequencies.

The cell shown in figure 4A was typical, with a narrower filter and a larger negative component at high mean, but a similar shape—a negative peak followed by a positive one-- in both conditions. Figure 4D illustrates the consequences of this filter change by comparing a segment of the cell's response under low mean (gray) and high mean (red). Under low mean, the response resembles a smoothed version of the stimulus: larger peaks occur only when the stimulus is above its mean. Under high mean the response contains many more rapid peaks, reflecting the faster frequency tuning and narrower shape of the filter under this condition. The peaks in this condition are of similar size throughout the segment, showing that the neuron has largely filtered out the slow modulations, and responds more to local changes in stimulus intensity.

The cell in figure 4B shows a more dramatic change in filter shape: under low mean conditions, it has a positive integrator shape, while under high mean conditions, it adopts an offset differentiator shape. The neuron shown in figure 4C was the least common. Its filter is almost entirely negative at low mean but gains a small positive component at high mean.

Together, these examples suggest that when the mean sound amplitude is low, neurons integrate over a longer time, while when the mean is high, the same cells respond more to local changes. These different response properties can improve the cells' ability to signal effectively in

different stimulus regimes (Atick, 1992), and can help prevent saturation when sounds are loud. To determine whether these findings were true across the population, we measured two parameters of the filters: the ratio of positive to negative parts of the filter (Pos/Neg), and the best modulation frequency (BMF).

Figure 5A shows the ratio of positive to negative parts of the filter (Pos/Neg) under high mean (y-axis) versus low mean (x-axis; n = 28 cells with significant filters under both conditions). Most points lie below the diagonal, indicating that most filters have larger negative components at high mean. The small number of cells that show an increase in Pos/Neg represent filters like the one shown in figure 4C that are predominantly negative under low mean conditions, and gain a positive component under high mean. Most points lie close to the line y=1, which represents balanced positive and negative components under high mean.

To examine whether neurons as a population became sensitive to faster modulations at high mean, we plotted the best modulation frequency (BMF) for all cells under high mean versus low mean (figure 5B). In this graph, most points lie above the diagonal, indicating that most cells are more sensitive to faster modulations at the high mean. The circled points represent cells that showed a significant increase in best modulation frequency (n=16/28, one-tailed t-test, $p<0.05$); they span the full range of modulation sensitivity seen in our neurons. Together these data suggest that an increase in stimulus mean leads to filters that behave more like differentiators and to a shift in the sensitivity of cells towards high temporal frequencies.

Are the differences we observe in filter shape due to true differences in responses, or do they represent different approximations to the neuron's response? To answer this question, we compared predictions made by filters with the same mean to predictions made by filters with different means. In all cases data used to test the model were distinct from data used to generate the model.

Filters generated under both conditions performed similarly when predicting responses to novel stimuli of the same condition: the mean correlation coefficient between data and prediction was 0.59 +/- 0.18 (sd) under high mean and 0.66 +/- 0.18 under low mean. However, both sets of filters were much worse at predicting responses to the other condition. Figure 5C shows the correlation coefficients between predicted and actual PSTHs using filters with the same mean (self-prediction CC) versus filters with a different mean (cross-prediction CC). Self-predictions are significantly better than cross-predictions for both high and low mean data. This confirms that the changing filter shapes capture a real change in the features to which the neurons respond.

Filters Change Gradually as a Function of Mean

To better understand how changes in filter shape occur--particularly the most dramatic switches from onset to offset selectivity— we varied the stimulus mean randomly among four levels (30, 40, 50, or 60dB), while holding the variance at 5dB. Figures 5D and 5E show filters calculated under these four conditions for two cells – one that retains an onset shape across levels, and one that switches gradually from onset sensitivity at low mean (gray trace), to offset or acceleration sensitivity (indicated by a tri-phasic filter) at high mean (red trace). In both cells, a negative peak that is small at 30dB grows larger and decreases in latency as the mean increases. This suggests that even very striking changes in filter shape might arise through simple mechanisms**,** such as a change in the relative strength and latency of inhibitory inputs to the cell. Population data from 13 cells (supplementary figure 3) support the finding that filters change gradually and systematically as a function of mean stimulus amplitude.

Changes in Stimulus Variance Do Not Produce Systematic Changes in Filter Shape

In contrast to the effect of mean amplitude, changing the stimulus variance had little effect on mean firing rate or filter shape. Figure 6A shows the effects of an increase in standard deviation from 6 to 18dB on the cell shown in figure 4A. Although the change in variance produces the same increase in stimulus power as did the change in mean, the cell shows little significant change in filter shape.

Across our population, correlation coefficients between high variance and low variance filters (mean = 0.84 +/- 0.03, se) were significantly higher than the correlation coefficients between high mean and low mean filters (0.12 +/- 0.10, p = 5.0e-8), confirming that variance had much less effect on filter shape than did mean amplitude. Changing stimulus variance also had fewer effects on the ratio of positive to negative filter components (figure 6B), and on best modulation frequency (figure 6C).

Similar to the filters generated under high mean, filters generated under high variance did well at predicting responses to stimuli with the same statistics. The mean correlation coefficient between data and prediction was 0.54 +/- 0.17 (sd) under high variance and 0.66 +/- 0.18 under low variance. However, predictions made by swapping filters--using high variance filters to predict low variance data and vice versa-- were only slightly worse than predictions made with matched filters (figure 6D). Although an increase in variance caused changes in the shapes of some filters, these

changes were smaller, less common, and less systematic than those observed with an increase in mean.

Effects of Mean and Variance on Nonlinearities

Although variance had only small effects on filter shape, it had a significant effect on neural gain. To examine the effects of stimulus mean and variance on the gain of the neural response, we calculated the nonlinear relationship between the filtered stimulus and the instantaneous firing rate under each condition (Methods). Filters were normalized such that the variance of their output was equal to the variance of their input (Baccus and Meister, 2002) to ensure that changes in gain were not due to changes in filter shape.

Figures 7A and B show the effects of increased mean and variance on the nonlinearities of two cells. An increase in mean (high mean red versus low mean gray) led to a decrease in the gain of the cell in figure 7A, and an increase in the gain of the cell in figure 7B. An increase in variance (high variance green versus low variance gray) led to a decrease in the gain of both cells. These decreases in gain with variance may be considered adaptive, as they allow the cells to maintain a similar range of firing rate fluctuations when the range stimulus fluctuations increases, and so to take better advantage of their signaling capacity.

The trends we observed in these examples continued across our population. An increase in stimulus mean led to both increases and decreases in gain, with no net change in gain across the population. An increase in stimulus variance led to systematic decreases in gain. To quantify changes in gain, we calculated the average slope of each nonlinearity, excluding sub-threshold and saturation regions where the slope fell below 5% of its maximum (Methods). Figure 7C shows slopes under high mean versus low mean: points are distributed on both sides of the diagonal, indicating no systematic change in gain. Figure 7D shows slopes under high variance versus low variance. Here, most points (27/32) lie below the diagonal, indicating a decrease in gain.

The fly H1 neuron has been shown to reduce its gain in proportion to the increase in the standard deviation of the stimulus (Brenner et al., 2000; Fairhall et al., 2001). If our cells adapted proportionally, the increase in mean should produce no change in gain, but the increase in variance should decrease gain by a factor of 1/3 (blue dashed line in figure 7D), to compensate for a 3-fold increase in the standard deviation.

To quantify the amount of variance adaptation in each cell, we calculated an adaptation index (see Methods). The adaptation index is zero if there is not change in gain with variance, and one if

the gain under high variance is exactly 1/3 of the gain under low variance. The distribution of adaptation scores is shown in figure 7E, and is highly skewed (median = 0.74): many of the cells show close to proportional adaptation, while some show less than proportional adaptation. A few cells showed no adaptation, or adaptation in the opposite direction.

Time Course of Changes in Filter Shape and Gain

Our analysis of field L response properties in different statistical conditions indicates that their encoding changes with both mean and variance. Are these changes due to time-dependent mechanisms, or do they represent fixed nonlinearities, for example differences in the relative strength of excitatory and inhibitory inputs at different intensities? To examine this question, we looked at the time course of changes in filter shape and gain.

To ask how fast filter shapes changed, we calculated filters from different epochs of the response before and after the step change in mean. Figure 8A shows filters derived from three epochs surrounding the switch from low to high mean: the last 500 msec of the low mean response (black), the first 100 msec after the switch to high mean (purple), and the last 500 msec of the high mean response (gray). Although the mean firing rate is still changing during the first 100msec after the switch, the filter derived from these data (purple) has the narrower shape and larger negative component characteristic of the fully adapted high mean filter (gray); it is quite different from the wide integrator shape of the filter under low mean (black).

To ask whether filters changed this fast across our population, we calculated correlation coefficients between filters from these three epochs (figure 8B). Correlation coefficients between early (first 100msec) and late (last 500msec) filters from the same condition were significantly greater than correlation coefficients between early filters and filters from the end of the previous condition. Correlation coefficients between early and late filters were not significantly different from correlation coefficients between two fully adapted filters taken 4500msec and 4900msec after the switch. Together these data suggest that filters change shape within 100 msec after a change in mean stimulus level.

To look at the time course of changes in gain, we calculated nonlinearities from different epochs before and after a change in variance. Figure 8C shows nonlinearities calculated during three epochs around the switch from low to high variance. The nonlinearity calculated from the last 500

msec of the low variance condition has a high gain (black), while nonlinearities taken from both the first 100msec (purple) and last 500msec (gray) of the high variance response both show a lower gain.

This rapid change in gain can also be seen in the PSTHs of this cell's response to repeated trials. If the decrease in gain occurred slowly, we would expect firing rate fluctuations under high variance to be initially much larger than firing rate fluctuations under low variance. As shown in figure 8D, firing rate fluctuations under high and low variance are nearly identical by the time of the first large peak, about 50 msec after the switch, indicating that the gain of the cell has decreased by this time.

To quantify the speed of gain change across the population, we compared the gain of nonlinearities taken from the three epochs of the response. The gain of nonlinearities derived from the first 100 msec of data were significantly different from the gain of nonlinearities under the previous condition ($p = 3.7e-5$ for increased variance, $p = 2.5e-5$ for decreased variance) but were not significantly different from the gain measured later under the same condition ($p = 0.75$ for increased variance, $p = 0.43$ for decreased variance). Together, these data indicate that nonlinear gain changes within 100 msec of a change in stimulus variance.

Our analyses indicate that changes in filter shape and nonlinear gain occur quickly, well before the mean firing rate has finished adapting. This suggests that gradual firing rate adaptation corresponds to a change in the threshold or set-point of firing, rather than to a change in gain or in temporal feature selectivity. To ask whether firing rate adaptation is correlated with a change in threshold, we plotted nonlinearities from several epochs around the switch from low to high mean. Figure 8E shows these nonlinearities for a single cell. Following the increase in mean, the nonlinearity undergoes a rapid shift upward and to the left (purple versus thin black), as well as a small change in slope. The shift occurs immediately, and is probably due to the change in filter shape rather than to a change in neural sensitivity. Over the next 500msec, the nonlinearity moves gradually back towards the center (violet and light blue), but does not change slope. The slow decay in firing rate thus seems to determine the shifting position of the nonlinearity.

To confirm that this trend was consistent across our population, we calculated the correlation between the mean firing rate during each 100 msec-long epoch, and the y-intercept of the nonlinearity measured during that epoch. The average correlation coefficient across cells was high-- 0.82 +/- 0.02 (se)-- suggesting that firing rate adaptation is best considered as a change in the position or set-point of the nonlinearity.

Discussion

The goal of our study was to investigate how coding of complex amplitude modulations in a primary auditory area depends on the distribution of stimulus intensities. To characterize coding, we developed a stimulus that mimics several features of natural song stimuli, while still permitting a randomized and systematic search of possible stimuli. From responses to these stimuli, we were able to generate linear-nonlinear models that successfully predict many features of the neurons' responses to novel stimuli.

The filters we extracted from our neurons revealed sensitivity to a range of temporal features and time scales. Peak temporal frequency sensitivities varied from 10-150Hz. Neurons showed sensitivity to many combinations of onsets, offsets, and continuous stimulation. This range of sensitivities may be important for tracking the complex contours of zebra finch syllables. Previous studies in this area found a lower range of temporal frequency sensitivities, and a preponderance of onset units (Sen et al., 2001; Woolley et al. 2005). Our results support findings in other species showing a broader range of temporal frequency sensitivities and response types in un-anesthetized animals (Liang and Wang, 2002; Wang 2005).

To examine how coding changes with the distribution of stimulus amplitudes, we altered the mean and the variance of amplitude modulations in our stimulus. Generating linear-nonlinear models under three different statistical conditions revealed that changes in mean and variance led to distinct types of adaptive changes in coding.

The first adaptive change that we described was a remodeling of filter shape that depends on the mean amplitude of the stimulus. Under low mean conditions neurons act more like low-pass integrators. As the stimulus mean increases, negative components of the neural filter grow stronger and decrease in latency, causing neurons to behave more like band-pass differentiators. This transition could help prevent firing rate saturation at high sound levels. It could also allow cells to signal effectively over a wider range of stimulus levels. The amplitudes of natural sounds tend to be similar at nearby points in time (Attias and Schreiner, 1998 , Singh and Theunissen, 2003), so responding only to changes in amplitude—which are more rare—may be efficient at high signal-to-noise levels. At low signal-to-noise levels, however, integrating over a longer time can improve the chance of detecting a quiet sound.

Several groups (Woolley et al. 2005, Narayan and Sen, 2005) have shown that analyzing songs with band-pass or differentiating filters allows for better theoretical discrimination of different individuals' songs. However, these studies only analyzed songs under high signal-to-noise conditions. It would be interesting to ask whether low-pass filters permit better song discrimination when songs are soft or noise is prevalent.

Changes in filter properties similar to those we describe have also been observed in retinal ganglion cells. At low light intensities, ganglion cells lose their receptive field surrounds and increase their integration times (Enroth-Cugell and Lennie, 1975). Atick (1992) used information theory to show how these changes in receptive field structure can maximize information transmission in bright versus dim conditions. Similar principles should apply to the changes we observe in auditory receptive field structure when the volume of an auditory signal changes.

The second adaptive change we described was a reduction in gain when the stimulus variance increased. This reduction in sensitivity allows the cell to match its range of outputs more closely to its range of inputs. Similar properties have been described in the auditory midbrain (Dean et al, 2005) and in several areas of the visual system, including the retina (Chander and Chichilnisky, 2001; Baccus and Meister, 2002), the LGN (Mante et al., 2005), and the fly H1 neuron (Brenner et al. 2000; Fairhall et al. 2001). In the fly, gain adaptation is proportional to the increase in the standard deviation of a velocity stimulus. In our system, neurons showed a distribution of variance adaptation. Many cells showed close to proportional gain adaptation, while some showed weaker adaptation, and a few showed no adaptation, or adaptation in the opposite direction. The difference between the fly motion detection system and the songbird auditory system may be related to the number of neurons involved. In the fly, a single neuron must encode visual motion under all behavioral conditions, while in the songbird, a population of neurons may allow some specialization for different statistical regimes.

If changes in filter shape and gain arose through some time-dependent process, we would expect them to occur slowly after a step change in stimulus statistics. Instead we found that both these changes occur within 100msec of the switch in stimulus statistics. In a few cells (not shown), we attempted to estimate filters from even shorter time intervals, and found that they changed essentially within the time scale of the filter--as fast as we could measure them. Fast changes in sensitivity have also been described in retinal filters (Baccus and Meister, 2002) and in the fly H1 neuron (Brenner et

al. 2000; Fairhall et al. 2001). This does not rule out the possibility that these changes are due to adaptive processes operating within a few tens of milliseconds--such as fast adaptation measured in cortical cells (Nowak et al., 2003, McCormick et al., 1985), or fast adaptive processes acting in the cochlea (LeMasurier and Gillespie, 2005). However, it suggests that the changes we observed may best be described and modeled as "adaptive nonlinearities" in processing rather than as traditional time-dependent adaptations (Borst et al, 2005).

The fact that these changes occur so quickly can constrain models of how these properties might arise. One hypothesis is that the positive and negative peaks in our filters reflect excitatory and inhibitory inputs, whose latency and relative strength depend differently on the volume of the stimulus. Dramatic changes in filter shape with mean stimulus amplitude will arise if inhibitory inputs have a higher threshold than excitatory inputs, and if their gain rises more steeply as a function of stimulus amplitude. This arrangement of inhibition has been described in several parts of the ascending auditory system, particularly in type IV units of the dorsal cochlear nucleus (Nelken and Young, 1994; Yu and Young, 2000).

Properties related to those we describe have also been seen in the ventral cochlear nucleus. The temporal modulation transfer functions of VCN chopper units show a transition from low-pass to band-pass temporal frequency sensitivity with increased stimulus volume (Frisina et al. 1990). In the inferior colliculus (IC), a similar transition from low-pass to band-pass sensitivity has been observed (Rees and Moller, 1987, Krishna and Semple, 2000). This transition could arise in the cochlear nucleus and be filtered to lower frequencies as it is propagated through the ascending auditory system. Alternatively, it could arise anew at each stage of processing due to inhibitory interactions like those we hypothesize.

Two other studies in the mammalian IC described properties related to those seen here. Dean et al. (2005) showed that IC neurons change their rate-level functions depending on the local distribution of stimulus amplitudes. However**,** because that study calculated rate-level functions by directly comparing stimulus amplitude and firing rate, it could not discriminate changes in neural sensitivity (gain) from changes in feature selectivity (filter shape). Kvale and Schreiner (2003) also saw small changes in neural gain and filter shape of IC units when the stimulus variance changed, but contrast to our results, these changes occurred gradually during adaptation. We think it is likely that the adaptive properties we describe first arise early in the ascending auditory system, but may be augmented by processing at many stages.

In addition to the filter and gain changes, we observed a range of slow decays in firing rate following a change in the mean amplitude of the stimulus, similar to the wide range of decay times described in auditory cortex by Ulanovsky et al (2004) and by Bartlett and Wang (2005). The rate of decay was not related to the width of each cell's filter, nor did this decay arise from a change in filter shape or gain-- because both those changes occurred very quickly. Instead, slow decays appear to be related to a shift in the setpoint or threshold of firing: as time progresses, a fixed amount is subtracted from the neuron's response to the same feature.

A striking feature of our results is their similarity to findings in the visual system (e.g. Enroth-Cugell and Lennie, 1975, Baccus and Meister, 2002)--particularly in the retina, where techniques similar to ours have been most employed. In both systems, neural filters become more differentiating as stimulus intensity increases. In both systems, neural gain decreases with increased stimulus contrast. And in both systems these changes occur rapidly, perhaps within the time scale of filtering, and independent of slower decays in firing rate. Although the neurons we studied lie at a very different stage in the sensory hierarchy from these retinal cells, they seem to share common computational strategies. These may represent solutions to common problems faced by all sensory systems that must represent a wide dynamic range of signals.

The changes in coding that we describe may help neurons to accurately encode stimulus characteristics in very different regimes, but they raise problems for current models of encoding and decoding natural stimuli. For instance, several existing models of primary auditory cortex assume that spectro-temporal features are encoded by a bank of linear filters (Chi and Shamma, 2005). Models of complex song-selective neurons in higher areas of the avian brain also assume that the inputs to these cells are static and largely linear (Margoliash 1983, Lewicki and Konishi, 1995, Drew and Abbott, 2003). In contrast, our study suggests that the inputs to higher-order auditory areas are highly nonlinear, but change in systematic ways that could be effectively modeled. A model that incorporates the nonlinearities we describe should not only make better predictions about how primary neurons respond to complex stimuli like song or faces, but will also be crucial to understanding how the brain constructs high-level feature detectors such as song- or face-selective neurons.

Methods:

Chronic Recording and Electrophysiology

We used chronically implanted microdrives (Hessler and Doupe, 1999) to record single units from 5 adult male zebra finches. A detailed description of microdrive construction and implantation are given in that paper. Electrodes (2-3 tungsten electrodes, 4-5 MOhms, MicroProbe Inc, Gaithersburg, MD) were implanted 1.5 µm lateral (left), and 1.5 µm rostral to the posterior border of the branch point of the central sinus, at an initial depth of 400 µm.

During recording, the bird was attached to a commutator by a flexible lead and op-amp. Electrical traces were digitized, amplified (1000x), filtered (300-5000Hz), and recorded using TDT System 3 hardware (Tucker-Davis Technologies, Alachua, FL) interfaced with custom-written Matlab software. The electrode bundle was advanced manually in small steps (40-160 microns). Putative single units were identified on the oscilloscope by their stable spike waveform and clear refractory period. All spikes were re-sorted offline using a custom-written software window discriminator (Matlab) based on the similarity of overlaid spike waveforms and on clustering of waveform projections in a two-dimensional principal component space. Neural recordings were considered single units if they contained fewer than one violation of 1 msec refractoriness per thousand spikes after sorting (Supplementary figure 1). Units that responded to auditory stimuli were found at depths of 1000-2500 µm. Between each recording session, the electrodes were retracted to a position above where the first auditory units were found.

After the final recordings, histological sections were prepared to confirm that electrode tracks, and in some cases, marker lesions, were located in field L. Birds were lethally anesthetized and perfused with saline followed by 4% paraformaldehyde. Alternate 40 micron sections of fixed brain tissue were Nissl-stained, and labeled for enkephalin (mouse anti-leucine enkephalin monoclonal antibody, Accurate Chemical & Scientific Corporation), a marker for the nucleus interface (NIf) which abuts the anterior end of field L2a. All sites were identified to be in field L layers L1, L2, and very occasionally L3. Although the preferred frequency and linearity of cells varied across these areas as described in Sen et al (2001), we saw no systematic differences in the sensitivity to stimulus statistics across layers.

Stimulus presentation

During recording, the bird was placed inside a small cage (20cm x 20cm floor area) within a sound-attenuating chamber (Acoustic Systems). The chamber lights were kept off to minimize movement and birds were monitored using an infra-red camera. Birds generally sat in one corner of the cage for the duration of the experiment although the commutator permitted free movement within the cage. Auditory stimuli were presented free-field from a small speaker (Bose) located 50 cm from the center of the bird's cage. Using a calibrated microphone (B&K) we verified that 250Hz to 10kHz tones designed to play at 80dB appeared at 79.6+/-2.3dB, and that the highest harmonic distortion peak observed was less than 23.3dB (65.9 +/- 8.7dB SNR).

Stimulus Construction

Although it is likely that field L neurons respond to other parameters of acoustic stimuli (Elhilali et al. 2004), we designed our stimulus to isolate responses to modulations of the log amplitude envelope. Recent studies in field L (Gill et al, 2006) and in mammalian inferior colliculus (Escabi et al, 2003) suggest that modulations of the log envelope drive cells better and are more linearly encoded than modulations of the linear envelope. Our stimulus consisted of two parts: a slowly varying envelope with fixed statistics that was repeated exactly in every experiment, and a rapidly varying carrier that could be adjusted for the frequency preference of each cell.

The log envelope, or "modulation signal" ($n(t)$), consisted of Gaussian noise filtered to have an exponential power spectrum, then normalized to have unit standard deviation and zero mean:

(1) $$P(f) = e^{-f/50Hz}$$

To generate a linear voltage envelope ($E(t)$) from the logarithmic modulation signal ($n(t)$), we exponentiated it according to:

(2) $$E(t) = 1e^{-5} \times 10^{[\mu+\sigma \times n(t)]/20}$$

where m is the mean amplitude of the stimulus in dB, and s is the standard deviation. When multiplied by a noise carrier with unit standard deviation, this produces a sound whose local amplitude in dB is given by

(3) $$A(t) = \mu + \sigma \times n(t)$$

and whose overall RMS amplitude in dB is given by

(4) $$RMS = \mu + \frac{\log(10)}{20}\sigma^2$$

We presented stimuli with three different statistics: low mean/low variance (30 +/- 6dB S.D., RMS = 34dB), low mean/high variance (30 +/- 18dB, RMS = 67dB), and high mean/low variance (63 +/- 6dB, RMS = 67dB). The low mean/high variance and high mean/low variance stimuli were designed to have the same RMS power. Stimulus statistics changed abruptly every 5 seconds, with the low mean/low variance condition appearing in between each presentation of low mean/high variance, or high mean/low variance. A single trial consisted of 400 continuous 5 second segments, half repeats of the same segment, and half unique. Repeated and unique segments were randomly distributed throughout the sequence.

The carrier was a Gaussian noise stimulus, digitally created and filtered online. Upon encountering a cell, we first determined its frequency preference using broad- (500-8000Hz), and narrowband (500Hz-wide, 750-7750Hz center frequencies) noise bursts, presented at 60dB SPL. If broadband noise drove the cell robustly, a broadband carrier was used (n=21/36), otherwise we chose the narrowband carrier that best drove the cell. At several sites (n = 8) we repeated the experiment using both a narrowband and a broadband carrier. Changing the carrier altered the precise shape of the filter, but did not affect our basic findings about changes with mean and variance or the distribution of filter widths. Because the carrier was varied throughout the experiment, responses to repeated trials represent the response to the modulation signal, averaged over many carriers.

We used a variation of this stimulus to characterize filters at 30, 40, 50, and 60dB. This stimulus consisted of unique segments only. All segments had a standard deviation of 5dB, and the order of statistical conditions was completely randomized.

Extraction of Filters and Nonlinearities

Filters were calculated in the Fourier domain according to:

(5) $$F(\omega) = \frac{< s^*(\omega) r(\omega) >}{< s^*(\omega) s(\omega) >}$$

where $F(\omega)$ is the Fourier transform of the filter, $s(\omega)$ is the Fourier transform of the normalized log stimulus envelope $n(t)$, and $r(\omega)$ is the Fourier transform of the spike train. The * indicates the complex conjugate. The numerator in this equation is equal to the spike triggered average, while the

denominator is equal to the power spectrum of the log stimulus envelope, which by design is exponential (eq 1). We verified that deviations of the full stimulus power spectrum from this predicted power spectrum did not significantly impact our calculations (Supplementary figure 2).

In practice, division or decorrelation by the power spectrum results in a noisy estimate of the filter, because it boosts power in high frequencies that are poorly sampled in n(t). To recover meaningful filters, we placed an exponential cutoff on F(w), given by

(6) $c(\omega) = 1$ for $|\omega|$ < cutoff

$c(\omega) = e^{-|\omega - cutoff|/10}$ for $|\omega|$ >= cutoff

The cutoff frequency for each cell was placed where the power spectrum of the raw spike-triggered average fell below 2 standard deviations of the power spectrum of an average of random spike times. The highest cutoff across conditions was used for all calculations involving a single cell. Cutoffs ranged from 23-167Hz.

Filters were considered significant if at least 10msec of the filter lay outside 3 standard deviations of the random-triggered average. Non-significant filters arose exclusively when the firing rate during a particular condition was extremely low. All but one cell recorded produced significant filters under at least two conditions. To estimate the error in our filter calculations, we divided our spikes into 5 random pools, and calculated separate filters from each. The standard deviation of these 5 estimates are shown as error bars in the figures.

The nonlinearity describes the probability of spiking given a value of filter output: $P(spike \mid F \otimes stim)$, and was calculated using a Bayesian formula described in Brenner et al, (2000).

(7) $$P(spike \mid F \otimes stim) = \frac{P(F \otimes stim \mid spike)P(spike)}{P(F \otimes stim)}$$

where $P(spike)$ is the mean firing rate over the stimulus condition. $P(F \otimes stim)$ is the distribution of the mean-subtracted filtered stimulus, and $P(F \otimes stim \mid spike)$ is the distribution of spike-triggered mean-subtracted filtered stimulus segments.

The slope of the nonlinearity depends on the scale of the filter F: if the amplitude of F is increased, the width of the distribution $P(F \otimes stim)$ will also increase, decreasing the slope of the nonlinearity. We normalized each filter so that the variance of its output was equal to the variance of

its input (Baccus and Meister, 2001). This ensured that changes we measured in slope were not due to changes in filter shape. The mean was subtracted from each stimulus prior to calculating the nonlinearity to minimize a leftward shift due to the decrease in the integral of the filter with increased mean. To ensure that the nonlinearity was well-sampled, we restricted our analysis to a region from two standard deviations below the mean of the filtered stimulus distribution, to two standard deviations above. To measure the reliability of the nonlinearity we performed a jackknife operation, where 1/5 of the spikes were excluded in each of 5 estimates of the nonlinearity. Error bars on nonlinearities shown in figures represent the standard deviation of the jackknife estimate (Sen et al, 2001):

$$(8) \qquad std = \sqrt{\frac{n-1}{n}\sum_j (x_j - < x_j >)^2}$$

where n is the number of jackknife estimates, and x_j is the jth estimate of the nonlinearity. To calculate nonlinearities as a function of time for a single condition, we used the filter obtained by pooling all data from that condition, as our analysis showed that filters do not change over this time period.

To predict responses to the repeated modulation signal, we convolved it with the neural filter, quantized the result, then transformed the quantized signal according to the nonlinearity. PSTHs were obtained by smoothing the mean spike count per 1 msec bin with an 8msec wide hanning window. When comparing actual and predicted PSTHs, we omitted the first 500msec when mean firing rates were strongly adapting, as our filters were not designed to capture this response feature. The quality of the prediction depended strongly on the amount of data recorded.

Analysis of Filters and Nonlinearities

Parameters of filters were defined as follows: The 50% width was the duration in msec over which the absolute value of the filter was greater than or equal to half its maximum absolute value. The best modulation frequency (BMF) of a filter was the frequency at which its power spectrum was maximum. To minimize noise, power spectra were calculated from a segment of each filter, from 25 msec before the absolute value of the filter reached 25% of its maximum, to the time of the spike.

The ratio of positive to negative filter components (Pos/Neg) was obtained by dividing the sum of all positive parts of the filter by the absolute value of the sum of all negative parts of the filter. Each parameter was measured on five estimates of the filter to obtain error bars.

To eliminate sub-threshold and saturation regions from our estimates of nonlinear gain, values of the nonlinearity slope less than 5% of the maximum gain under any condition were excluded prior to averaging. To compare gain across conditions, we compared the log ratio of gain under the two conditions to zero.

Analysis of Decay Times

To estimate the decay time following a change in mean, we fit a single exponential of the form

(9a) $$r(t) = A + \Delta R \times e^{(-t/\tau)}$$

to the unsmoothed PSTH. In this formula, r(t) is the PSTH, ΔR is the magnitude of the decay, τ is the decay time constant, and A is the steady state firing rate. Although several decays exhibited complex transients, we focused our analysis on the final slow decay. To avoid mistakes in fitting introduced by transients, we fit only a portion of the response:

(9b) $$r(t_{lat} : end) = A + \Delta R \times e^{(-t/\tau)}$$

where t_{lat} was a latency parameter from 1-500msec. Fits for latency values that did not converge were excluded. Of the remaining fits, we chose the one that resulted in the smallest mean squared error between the actual PSTH and the exponential model. A few PSTHs were fit with only a subset of latency values (n = 4). Decay time constants were considered significant only if they plateaued within our five second trial (all cells but one), and if they decayed over a range of greater than 5Hz (32/35 cells under high mean, 31/35 cells under low mean).

Acknowledgements

This work was supported by grants from HHMI (to KN) and NIH (NS34835 and MH55987 to AJD). We would like to thank T. Sharpee, D. Schoppik, and W. Bialek for very helpful suggestions related to experimental design and analysis, A. Arteseros for performing the histology, and M. Kao, B.D.

Wright, and M.P. Stryker for comments on the manuscript.

Figure Legends

Figure 1

Construction of a stimulus with a power spectrum and amplitude distribution similar to natural sounds.

The modulation signal (A, top panel) specifies the local amplitude of the stimulus in decibels, and consists of filtered Gaussian noise with a power spectrum P(f) = exp(-f/50Hz) (A, bottom panel). Exponentiating the modulation signal creates a linear voltage envelope (B) with an exponential distribution of amplitudes. The envelope modulates a carrier (C) consisting of narrow or broadband noise. The stimulus (D) is generated by multiplying together the envelope (B) and the carrier (C).

(E) shows responses of a single unit to 100 repeats of the modulation signal shown in A, each paired with a different carrier. Each row represents the spike times during one segment repeat. Because the carrier varied across trials, columns of spikes represent responses to the modulation signal alone. Response peaks follow peaks in the aligned stimulus example above.

Figure 2

A linear-nonlinear model successfully predicts a large fraction of the response to our stimulus

A and D) Two examples of linear filters derived from the responses of single units to the non-repeated modulation signals (30dB mean, 6dB variance). The amplitude of each filter is normalized so that the dot product of the filter with itself is one. Filters thus describe the features that drive each cell, but not the magnitude of the cell's response. Dashed lines represent the standard deviation of 5 independent estimates of each filter's shape.

B and E) Nonlinear gain functions show the relationship between the firing rate of the neuron and the output of the filters in (A) and (D). The x-axes show the projection of the mean-subtracted

modulation signal onto each filter and indicate the similarity of the stimulus to the filter. Dotted lines indicate the standard deviation of five jackknife estimates of the gain function. The flat dashed line in each nonlinearity plot represents the average spontaneous firing rate of each neuron recorded prior to stimulus playback.

C and F) Predicted (wide gray) and actual (narrow black) PSTHs of the response to repeated modulation signal segments. The correlation coefficients between data and prediction were 0.89 for the cell in (A-C) and 0.67 for the cell in (D-F).

G) Left panel: filters from four different units show a broad range of temporal frequency preferences. Right panel: distribution of the best modulation frequency (BMF) versus the 50% width of the absolute value of the filter (50% width) for all significant filters (33 from 36 cells). The four cells shown in the left-hand plot are indicated by colored boxes.

H) A histogram of correlation coefficients between predicted and actual repeated segments for all the significant filters (33 out of 36 cells recorded with a 30dB mean and 6dB standard deviation stimulus) indicates that the linear-nonlinear model performed well. The mean correlation coefficient was 0.64 +/- 0.18 (standard deviation).

Figure 3

Changes in stimulus statistics lead to slow changes in firing rate

A) To explore how coding depends on the statistics of the stimulus, we altered the mean and standard deviation of the modulation signal every five seconds. Statistical conditions were presented in a fixed order: low mean/low variance (gray: 30dB +/- 6dB), low mean/high variance (green: 30dB +/- 18dB), low mean/low variance, and high mean/low variance (red: 63dB +/- 6dB). These colors will be used throughout the paper to indicate responses to each condition.

B) Cells adapt at different rates to a change in mean stimulus amplitude. Top: PSTH of a cell showing rapid decays after the switch from low to high mean (red: tau = 28 msec) and after the switch from high to low mean (gray: 56 msec). Bottom: PSTH of a cell showing slow decays. Tau = 574 msec

for the switch from low to high mean, 730 msec for the switch from high to low mean. PSTHs are binned in 5msec windows. Only the first 2.5 seconds of the response after each switch are shown.

C) Filter width under high mean/low variance versus decay time for both low to high (red) and high to low (gray) transitions. Filter time scales and decay time scales are poorly correlated (correlation coefficients = -0.2 for low to high mean and 0.04 for high to low mean.), suggesting that filter width and decay time represent independent forms of temporal sensitivity. The two example cells are shown by a blue circle and red square; open symbols represent low to high mean transitions, and filler symbols represent high to low mean transitions.

Figure 4

Filter shape depends strongly on mean stimulus amplitude.

A-C) Linear filters (left) and their power spectra (right) for three different cells under the low mean/low variance (gray) and high mean/low variance (red) conditions. For all three cells, the filter obtained under high mean is narrower and has positive and negative components of more similar size. Colored dashed lines surrounding each filter represent the standard deviation of 5 different estimates of the filter (Methods). Black dotted lines at zero represent the time of the spike. Filters are normalized so that the dot product of each with itself is one.

D) Responses of the unit shown in A to 120msec of the repeated modulation signal (black, bottom panel) under low mean (gray) and high mean (red) conditions. The stimulus has been shifted forward by 11 msec, the latency of the peak of the low mean filter, to facilitate comparison.

Figure 5

Population analysis of mean effects on filters

A) Ratio of positive to negative areas of filters (Pos/Neg) from high mean (y-axis) versus low mean (x-axis) conditions. Most points lie below the dotted line, indicating that negative components of the filter are larger under the high mean condition. Black circles indicate filters that showed a significant

change in Pos/Neg between the two conditions (n = 21/28). The mean Pos/Neg ratio across the population was significantly lower under high mean (0.97 +/- 0.04, se) than under low mean (1.41 +/- 0.11, p = 5.9e-4), and was not significantly different from one (black line) at high mean, indicating balanced positive and negative components. The three example cells shown in figure 4 are indicated by the pink square (A), blue diamond (B), and yellow triangle (C).

B) BMF under high mean/low variance (y-axis) versus low mean/low variance (x-axis) conditions. Most points lie above the dotted diagonal line, indicating higher BMF under the high mean/low variance condition. Black circles indicate significant changes (n = 16/28). Example cells are indicated by the filled symbols.

C) Filters derived from data with the same mean make much better predictions (y-axis) than filters from data with a different mean (x-axis). Red dots indicate correlation coefficients between data and predictions based on high mean filters (mean correlation coefficient = 0.59 +/- 0.03 (se) for same statistics, 0.03 +/- 0.07 for different statistics, p = 4.0e-9). Gray dots indicate correlation coefficients between data and predictions based on low mean filters (mean correlation coefficient = 0.66 +/- 0.03 for same statistics, 0.08 +/- 0.07 for different statistics, p = 2.3e-10). Data for the example cells are indicated by open (high mean filter predictions) and filled (low mean filter predictions) symbols.

D-E) Linear filters for two cells at four different mean levels: 30dB (gray), 40dB (brown), 50dB (yellow), and 60dB (red), with a standard deviation 5dB. As the stimulus mean increases, the negative component of each filter grows larger and decreases in latency, leading to substantial changes in filter shape.

Figure 6

Unlike a change in mean, a change in variance does not lead to significant changes in filter shape.

A) Linear filters for the cell shown in figure 4A under low mean/low variance (gray) and low mean/high variance (green) conditions. The increase in variance has no significant effect on filter shape.

B) The balance of positive and negative components across the population shows no systematic change with variance. Ratio of positive to negative areas in filters derived under low mean/high variance (y-axis) versus low mean/low variance (x-axis) conditions. Black circles indicate filters that showed a significant change in Pos/Neg between the two conditions (n = 15/32). The mean population ratio of positive to negative areas was not significantly different between high variance (1.53 +/- 0.15 se) and low variance (1.43 +/- 0.10) conditions (p = 0.20). The example cell shown in (A) is indicated by a red square.

C) Best modulation frequencies across the population do not change significantly with variance. BMF under low mean/high variance (y-axis) versus low mean/low variance (x-axis) conditions. Black circles indicate significant changes (n = 4/32). The example cell is indicated by a red square.

D) Filters derived from low variance and high variance conditions make similar predictions. Green dots indicate correlation coefficients between data and predictions based on high variance filters (mean correlation coefficient = 0.54 +/- 0.03 (se) for same statistics, 0.56 +/- 0.03 for different statistics, p = 0.41). Black dots indicate correlation coefficients between data and predictions based on low variance filters (mean correlation coefficient = 0.66 +/- 0.03 for same statistics, 0.52 +/- 0.03 for different statistics, p = 6.5e-8). Data for the example cell are indicated by open (high variance filter prediction) and filled (low variance filter prediction) squares.

Figure 7

Effects of Mean and Variance on Nonlinearities

A-B) Nonlinearities for two cells, calculated under three conditions: low mean/low variance (gray), high mean/low variance (red), and low mean/high variance (green). An increase in mean increases gain in one cell (B) and decreases it in the other (A). An increase in variance decreases gain in both cells. Colored dotted lines represent the standard deviation of 5 jackknife estimates of the nonlinearity (see Methods). The black dashed lines represent the spontaneous firing rates of each cell.

C) An increase in mean produced no net change in gain across the population. Nonlinear gain (Hz/ normalized stimulus projection) under high mean/low variance (y-axis) versus low mean/low variance

(x-axis) conditions. The dotted diagonal line indicates no change. The red line represents the average population ratio of 1.17. It is not significantly different from one (p = 0.32).

D) An increase in variance produced a systematic decrease in gain across the population. Nonlinear gain under low mean/high variance (y-axis) versus low mean/low variance (x-axis) conditions. The green line represents the average population ratio of 0.58. It is significantly less than one (p = 3.7e-5). The dashed blue line represents a decrease in gain proportional to the increase in the standard deviation.

E) Distribution of adaptation scores (see Methods) for all cells comparing low mean/low variance (gray) to low mean/high variance (green, n = 32). An adaptation score of 1 indicates that gain under high variance was exactly 1/3 of gain under low variance. 0 indicates no change in gain between conditions.

Figure 8

Changes in filter shape and gain both occur within 100msec of a change in stimulus statistics.

A) An example of rapid filter change. Top: PSTHs showing the mean firing rate of a single cell around the switch from low to high mean. Black: last 500 msec of low mean. Purple: first 100 msec of high mean. Gray: last 500 msec of the high mean. Bottom: Filters derived from each of the three epochs shown above. The filter derived from the first 100 msec of high mean (purple) resembles the filter from the last 500 msec of high mean (gray) much more than it does the filter from the last 500 msec of low mean(black).

B) Comparisons of filter similarity across the population. First column: average correlation coefficient between filters derived from the first 100msec and last 500msec of the same condition (red: high mean, 0.20 +/- 0.03; gray: low mean, 0.14 +/- 0.04). Second column: average correlation coefficient between filters derived from the first 100msec of one condition, and from the last 500msec of the previous condition (red: low mean to high mean transitions, -0.03 +/- 0.03; gray: high mean to low mean transitions, -0.04 +/- 0.03). Correlation coefficients between early and previous filters were significantly smaller than those between early and late filters (p = 3.0e-5 for high mean, p =

1.3e-3 for low mean). Third column: average correlation coefficient between filters derived from two 100msec epochs at the end of the same condition (red: high mean, 0.22 +/- 0.03; gray: low mean, 0.19 +/- 0.04). Correlation coefficients between early and late filters were not significantly different from those between two late filters ($p = 0.33$ for high mean, $p = 0.15$ for low mean). These data indicate that filters from the first 100msec after the switch were significantly different from filters under the previous condition, but were no more different from the fully adapted filters than two fully adapted filters were from each other.

C) An example of rapid change in nonlinear gain. Top: PSTH showing the mean firing rate of a single cell around the switch from low variance to high variance. Black: last 500msec of low variance. Purple: first 100msec of high variance. Gray: last 500msec of high variance. Bottom: Nonlinearities derived from each epoch after the switch from low variance to high variance. The nonlinearity derived from the first 100msec and last 500msec of high variance (gray) both have a lower gain than the low variance nonlinearity (black).

D) Rapid gain adaptation is evident in responses to repeated trials. PSTH of the response of the neuron in (C) to the first 280 msec after the switch from low to high variance (green), and after the switch from high to low variance (gray). The peaks are approximately the same within 50 msec after the switch, indicating that the cell is already compensating for the difference in the magnitude of stimulus fluctuations.

E) Slow decays in firing rate represent shifts in the nonlinearity. Nonlinearities from different epochs after the switch from low to high mean. Black: 500 msec preceding the switch. Purple: 0-100 msec after the switch. Violet: 200-300 msec after the switch. Light blue: 400-500 msec after the switch. Gray: last 500 msec of the high mean response. Immediately after the switch from low mean to high mean, the nonlinearity shifts up and to the left, then moves slowly back down and to the right.

References:

Attias, H. and C.E. Schreiner (1998) “Coding of Naturalistic Stimuli by Auditory Midbrain Neurons.” Advances in Neural Information Processing Systems 10

Atick, J. J. (1992). "Could Information-Theory Provide an Ecological Theory of Sensory Processing."

Network-Computation in Neural Systems 3(2): 213-251.

Baccus, S. A. and M. Meister (2002). "Fast and slow contrast adaptation in retinal circuitry." Neuron 36(5): 909-19.

Barbour, D. L. and X. Wang (2003). "Auditory cortical responses elicited in awake primates by random spectrum stimuli." J Neurosci 23(18): 7194-206.

Bartlett, E. L. and X. Wang (2005). "Long-lasting modulation by stimulus context in primate auditory cortex." J Neurophysiol 94(1): 83-104.

Borst, A., V. L. Flanagin, and H. Sompolisnky (2005). "Adaptation without parameter change: Dynamic gain control in motion detection." Proc Natl Acad Sci U S A 102(17): 6172-6.

Brenner, N., W. Bialek, and R. de Ruyter van Steveninck (2000). "Adaptive rescaling maximizes information transmission." Neuron 26(3): 695-702.

Calhoun, B. M. and C. E. Schreiner (1998). "Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics." Eur J Neurosci 10(3): 926-40.

Chander, D. and E. J. Chichilnisky (2001). "Adaptation to temporal contrast in primate and salamander retina." J Neurosci 21(24): 9904-16.

Chi, T., P. Ru, et al. (2005). "Multiresolution spectrotemporal analysis of complex sounds." J Acoust Soc Am 118(2): 887-906.

Dean, I., N. S. Harper, and D. McAlpine (2005). "Neural population coding of sound level adapts to stimulus statistics." Nat Neurosci 8(12): 1684-9.

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

Drew, P. J. and L. F. Abbott (2003). "Model of song selectivity and sequence generation in area HVc of the songbird." J Neurophysiol 89(5): 2697-706.

Eggermont, J. J., P. M. Johannesma, and A.M. Aertsen (1983). "Reverse-correlation methods in auditory research." Q Rev Biophys 16(3): 341-414.

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

Enroth-Cugell, C. and P. Lennie (1975). "The control of retinal ganglion cell discharge by receptive field surrounds." J Physiol 247(3): 551-78.

Epping, W. J. and J. J. Eggermont (1986). "Sensitivity of neurons in the auditory midbrain of the grassfrog to temporal characteristics of sound. II. Stimulation with amplitude modulated sound."

Hear Res 24(1): 55-72.

Escabi, M. A. and C. E. Schreiner (2002). "Nonlinear spectrotemporal sound analysis by neurons in the auditory midbrain." J Neurosci 22(10): 4114-31.

Escabi, M. A., L. M. Miller, et al. (2003). "Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus." J Neurosci 23(37): 11489-504.

Fairhall, A. L., G. D. Lewen, W. Bialek, and R. de Ruyter van Steveninck (2001). "Efficiency and ambiguity in an adaptive neural code." Nature 412(6849): 787-92.

Fortune, E. S. and D. Margoliash (1992). "Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (Taenopygia guttata)." J Comp Neurol 325(3): 388-404.

Fortune, E. S. and D. Margoliash (1995). "Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (Taeniopygia guttata)." J Comp Neurol 360(3): 413-41.

Frisina, R. D., R. L. Smith, and S.C. Chamberlain (1990). "Encoding of amplitude modulation in the gerbil cochlear nucleus: I. A hierarchy of enhancement." Hear Res 44(2-3): 99-122.

Gehr, D. D., B. Capsius, P. Grabner, M. Gahr, and H.J. Leppelsack (1999). "Functional organisation of the field-L-complex of adult male zebra finches." Neuroreport 10(2): 375-80.

Gentner, T. Q. and D. Margoliash (2003). "Neuronal populations and single cells representing learned auditory objects." Nature 424(6949): 669-74.

Gill, P., J. Zhang, S.M. Woolley, T. Fremouw, and F.E. Theunissen (2006). "Sound representation methods for spectro-temporal receptive field estimation." J Comput Neurosci. Epub online.

Grace, J. A., N. Amin, N.C. Singh, and F.E. Theunissen (2003). "Selectivity for conspecific song in the zebra finch auditory forebrain." J Neurophysiol 89(1): 472-87.

Hausberger, M., E. Leppelsack, J. Richard, and H.J. Leppelsack (2000). "Neuronal bases of categorization in starling song." Behav Brain Res 114(1-2): 89-95.

Heil, P. and H. Scheich (1985). "Quantitative analysis and two-dimensional reconstruction of the tonotopic organization of the auditory field L in the chick from 2-deoxyglucose data." Exp Brain Res 58(3): 532-43.

Hessler, N. A. and A. J. Doupe (1999). "Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches." J Neurosci 19(23): 10461-81.

Hose, B., G. Langner, and H. Scheich (1987). "Topographic representation of periodicities in the forebrain of the mynah bird: one map for pitch and rhythm?" Brain Res 422(2): 367-73.

Kim, K. J. and F. Rieke (2001). "Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells." J Neurosci 21(1): 287-99.

Kim, P. J. and E. D. Young (1994). "Comparative analysis of spectro-temporal receptive fields, reverse correlation functions, and frequency tuning curves of auditory-nerve fibers." J Acoust Soc Am 95(1): 410-22.

Kowalski, N., D. A. Depireux, et al. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra." J Neurophysiol 76(5): 3524-34.

Kowalski, N., D. A. Depireux, et al. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra." J Neurophysiol 76(5): 3503-23.

Krishna, B. S. and M. N. Semple (2000). "Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus." J Neurophysiol 84(1): 255-73.

Kvale, M. N. and C. E. Schreiner (2004). "Short-term adaptation of auditory receptive fields to dynamic stimuli." J Neurophysiol 91(2): 604-12.

LeMasurier, M. and P. G. Gillespie (2005). "Hair-cell mechanotransduction and cochlear amplification." Neuron 48(3): 403-15.

Lewicki, M. S. and M. Konishi (1995). "Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order." Proc Natl Acad Sci U S A 92(12): 5582-6.

Lewicki, M. S. and B. J. Arthur (1996). "Hierarchical organization of auditory temporal context sensitivity." J Neurosci 16(21): 6987-98.

Liang, L., T. Lu, and X. Wang (2002). "Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates." J Neurophysiol 87(5): 2237-61.

Lohr, B., T. F. Wright, et al. (2003). "Detection and discrimination of natural calls in masking noise by birds: estimating the active space of a signal." Animal Behaviour 65: 763-777.

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

Malone, B. J. and M. N. Semple (2001). "Effects of auditory stimulus context on the representation of frequency in the gerbil inferior colliculus." J Neurophysiol 86(3): 1113-30.

Mante, V., R. A. Frazor, V. Bonin, W.S. Geisler, and M. Carandini (2005). "Independence of luminance and contrast in natural scenes and in the early visual system." Nat Neurosci 8(12): 1690-7.

Margoliash, D. (1983). "Acoustic parameters underlying the responses of song-specific neurons in the white-crowned sparrow." J Neurosci 3(5): 1039-57.

Margoliash, D. and E. S. Fortune (1992). "Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc." J Neurosci 12(11): 4309-26.

McCormick, D. A., B. W. Connors, J.W. Lighthall, and D.A. Prince (1985). "Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex." J Neurophysiol 54(4): 782-806.

Miller, L. M., M. A. Escabi, H.L. Read, and C.E. Schreiner (2002). "Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex." J Neurophysiol 87(1): 516-27.

Muller, C. M. and H. J. Leppelsack (1985). "Feature extraction and tonotopic organization in the avian auditory forebrain." Exp Brain Res 59(3): 587-99.

Nelken, I. and E. D. Young (1994). "Two separate inhibitory mechanisms shape the responses of dorsal cochlear nucleus type IV units to narrowband and wideband stimuli." J Neurophysiol 71(6): 2446-62.

Nelken, I., P. J. Kim, et al. (1997). "Linear and nonlinear spectral integration in type IV neurons of the dorsal cochlear nucleus. II. Predicting responses with the use of nonlinear models." J Neurophysiol 78(2): 800-11.

Nowak, L. G., R. Azouz, M.V. Sanchez-Vives, C.M. Gray, and D.A. McCormick (2003). "Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses." J Neurophysiol 89(3): 1541-66.

Phillips, D. P. and S. E. Hall (1987). "Responses of single neurons in cat auditory cortex to time-varying stimuli: linear amplitude modulations." Exp Brain Res 67(3): 479-92.

Phillips, D. P., M. N. Semple, M.B. Calford, and L.M. Kitzes (1994). "Level-dependent representation of stimulus frequency in cat primary auditory cortex." Exp Brain Res 102(2): 210-26.

Sachs, M. B. and E. D. Young (1980). "Effects of nonlinearities on speech encoding in the auditory nerve." J Acoust Soc Am 68(3): 858-75.

Sachs, M. B., H. F. Voigt, et al. (1983). "Auditory nerve representation of vowels in background noise." J Neurophysiol 50(1): 27-45.

Scheich, H., B. A. Bonke, D. Bonke, and G. Langner (1979). "Functional organization of some auditory nuclei in the guinea fowl demonstrated by the 2-deoxyglucose technique." Cell Tissue Res 204(1): 17-27.

Sen, K., F. E. Theunissen, and A.J. Doupe (2001). "Feature analysis of natural sounds in the songbird auditory forebrain." J Neurophysiol 86(3): 1445-58.

Shannon, R. V., F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid (1995). "Speech recognition with primarily temporal cues." Science 270(5234): 303-4.

Singh, N. C. and F. E. Theunissen (2003). "Modulation spectra of natural sounds and ethological

theories of auditory processing." J Acoust Soc Am 114(6 Pt 1): 3394-411.

Theunissen, F. E., K. Sen, and A.J. Doupe (2000). "Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds." J Neurosci 20(6): 2315-31.

Ulanovsky, N., L. Las, D. Farkas, and I. Nelken (2004). "Multiple time scales of adaptation in auditory cortex neurons." J Neurosci 24(46): 10440-53.

Voigt, H. F. and E. D. Young (1980). "Evidence of inhibitory interactions between neurons in dorsal cochlear nucleus." J Neurophysiol 44(1): 76-96.

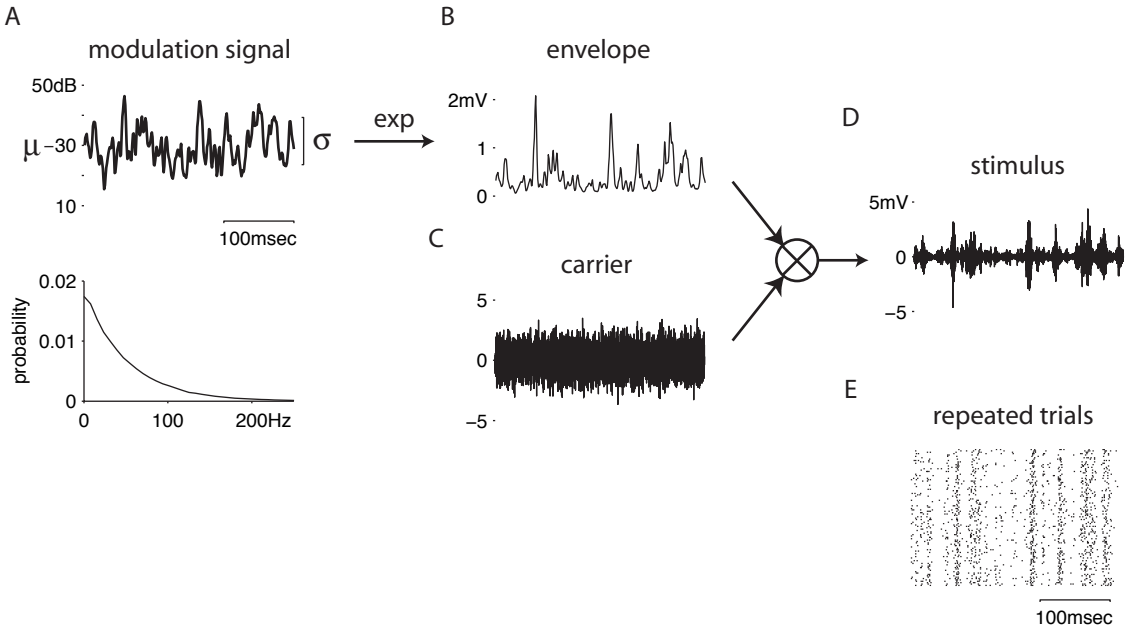
Wang, X., T. Lu, R.K. Snider (2005). "Sustained firing in auditory cortex evoked by preferred stimuli." Nature 435(7040): 341-6.

Wild, J. M., H. J. Karten, and B.J. Forst (1993). "Connections of the auditory forebrain in the pigeon (Columba livia)." J Comp Neurol 337(1): 32-62.

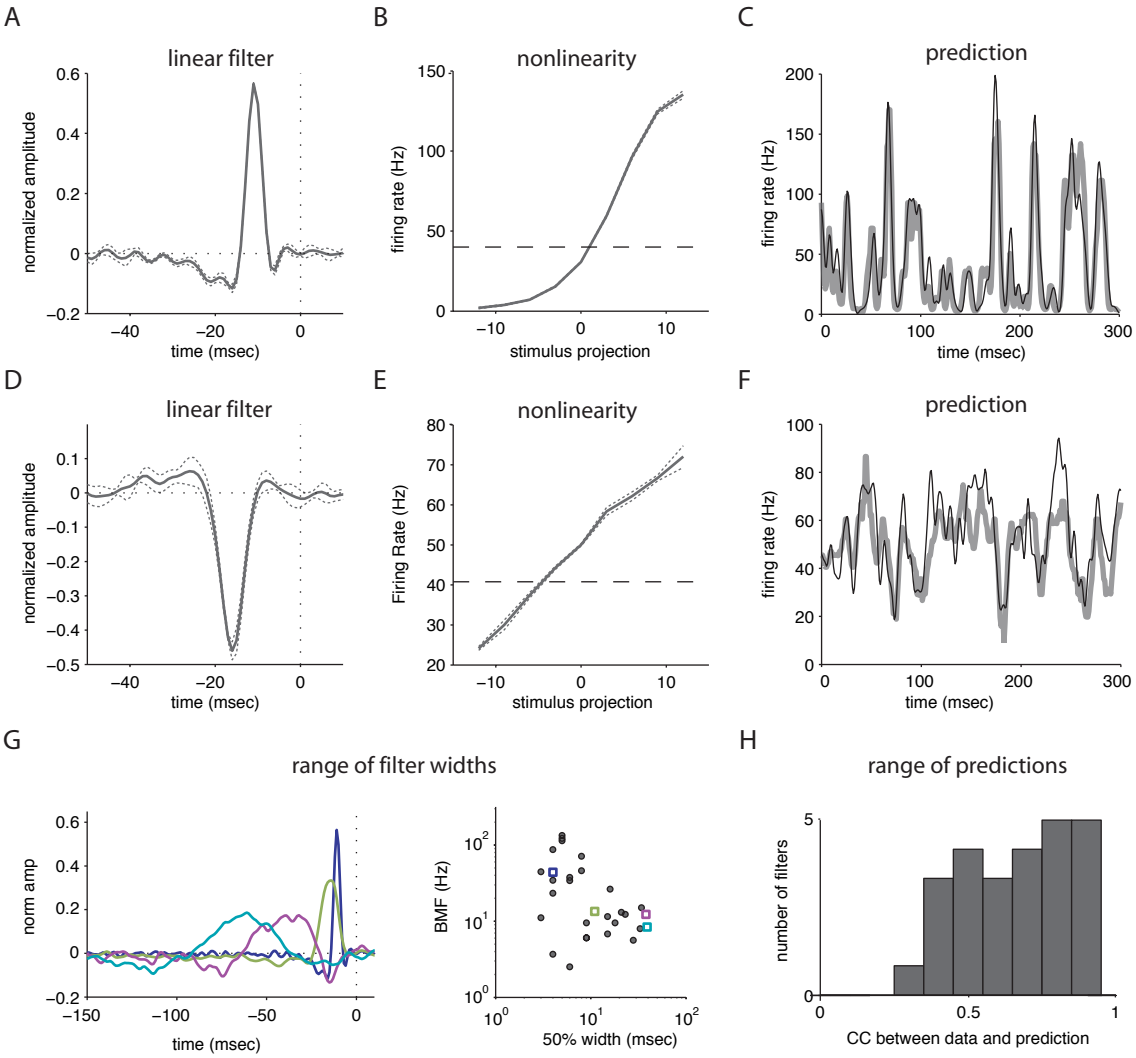
Woolley, S. M., T. E. Fremouw, A. Hsu, and F.E. Theunissen (2005). "Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds." Nat Neurosci 8(10): 1371-9.

Young, E. D. and H. F. Voigt (1982). "Response properties of type II and type III units in dorsal cochlear nucleus." Hear Res 6(2): 153-69.

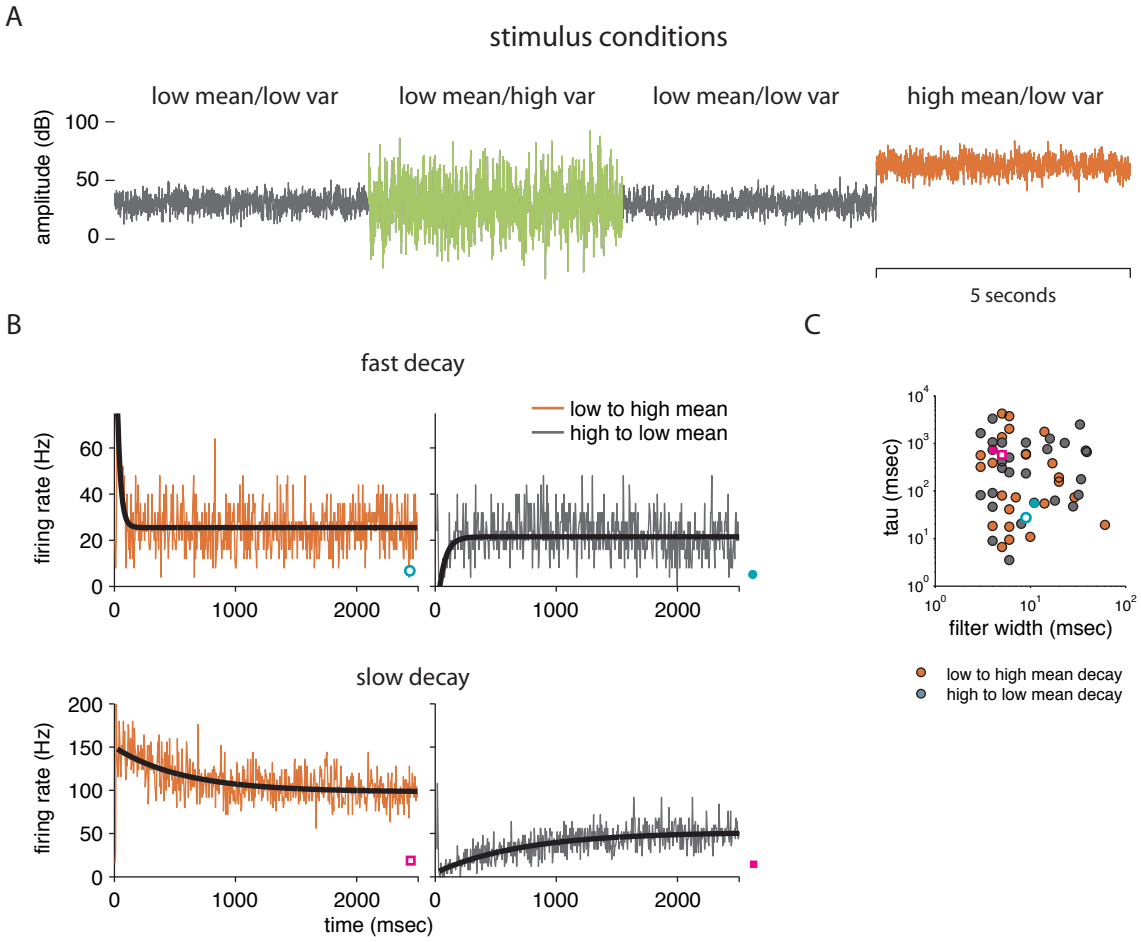
Yu, J. J. and E. D. Young (2000). "Linear and nonlinear pathways of spectral information transmission in the cochlear nucleus." Proc Natl Acad Sci U S A 97(22): 11780-6.



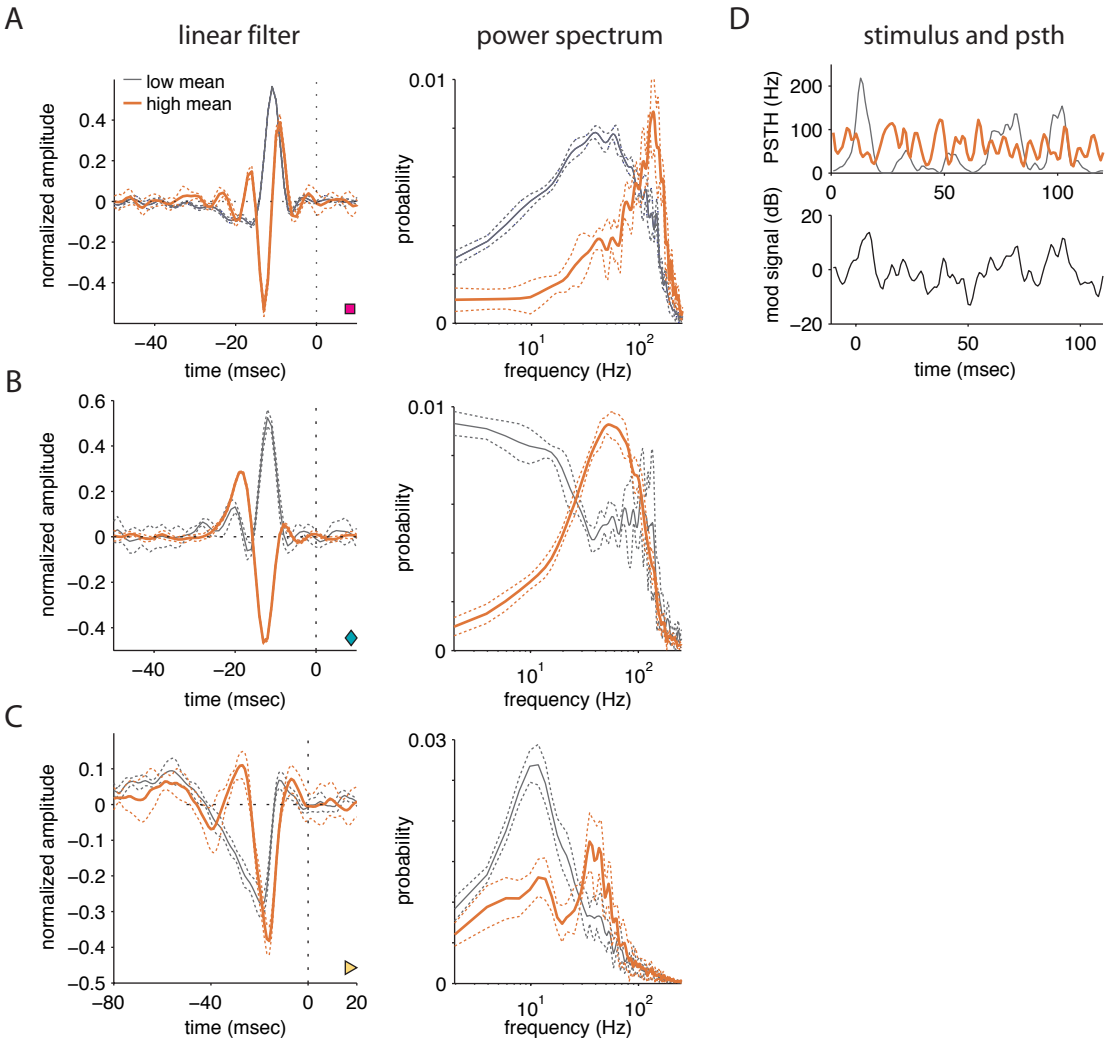
chapter 1, figure 1: stimulus construction



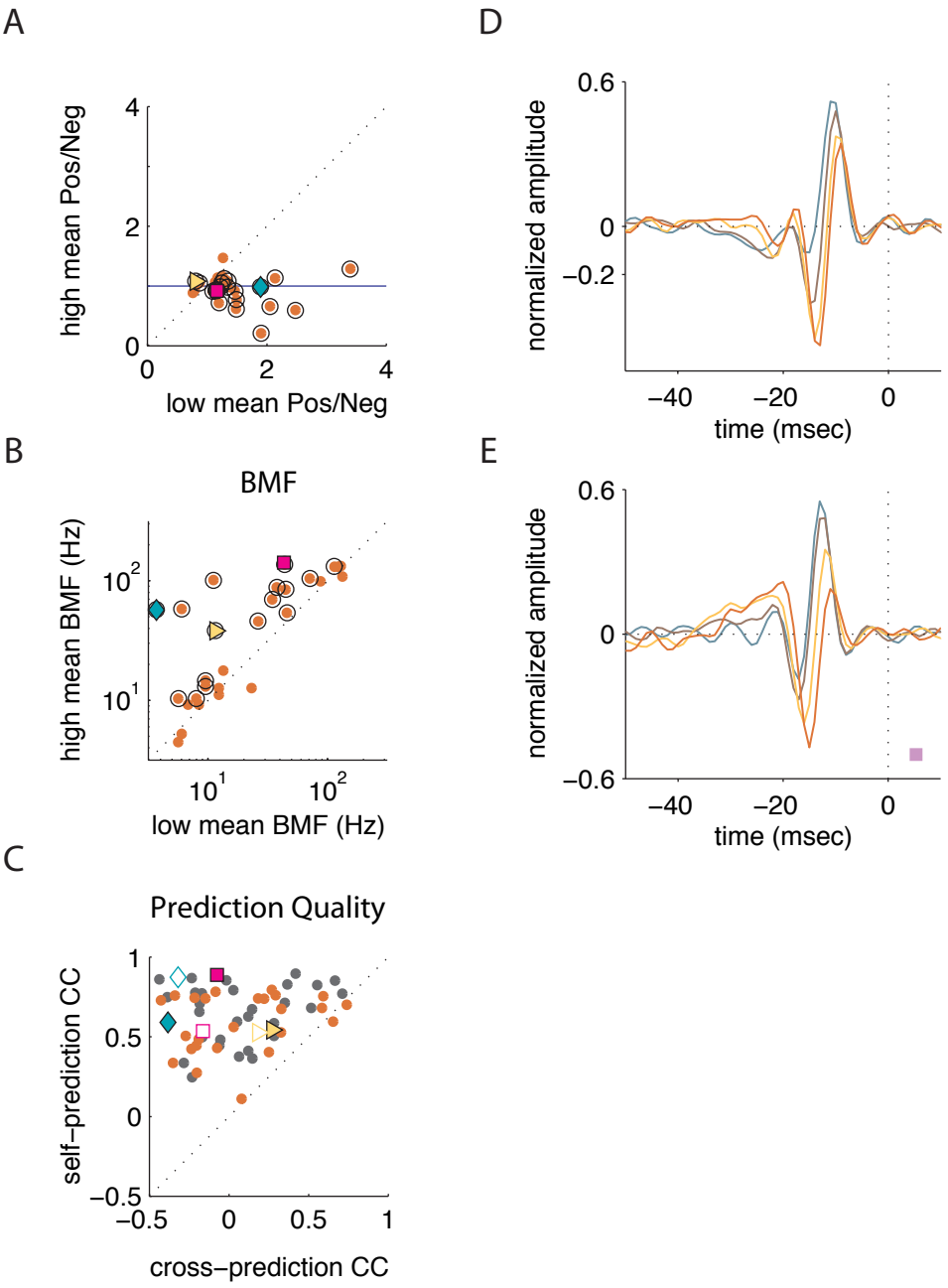
chapter 1, figure 2: linear non-linear model



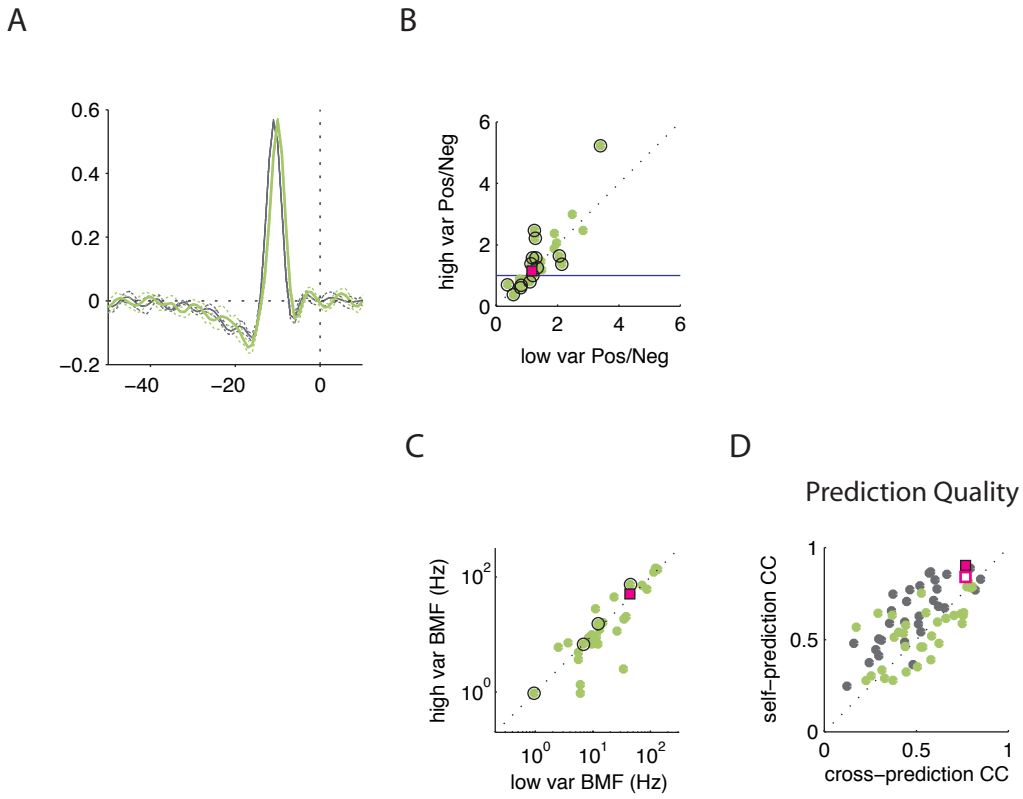
chapter 1, figure 3: stimulus conditions



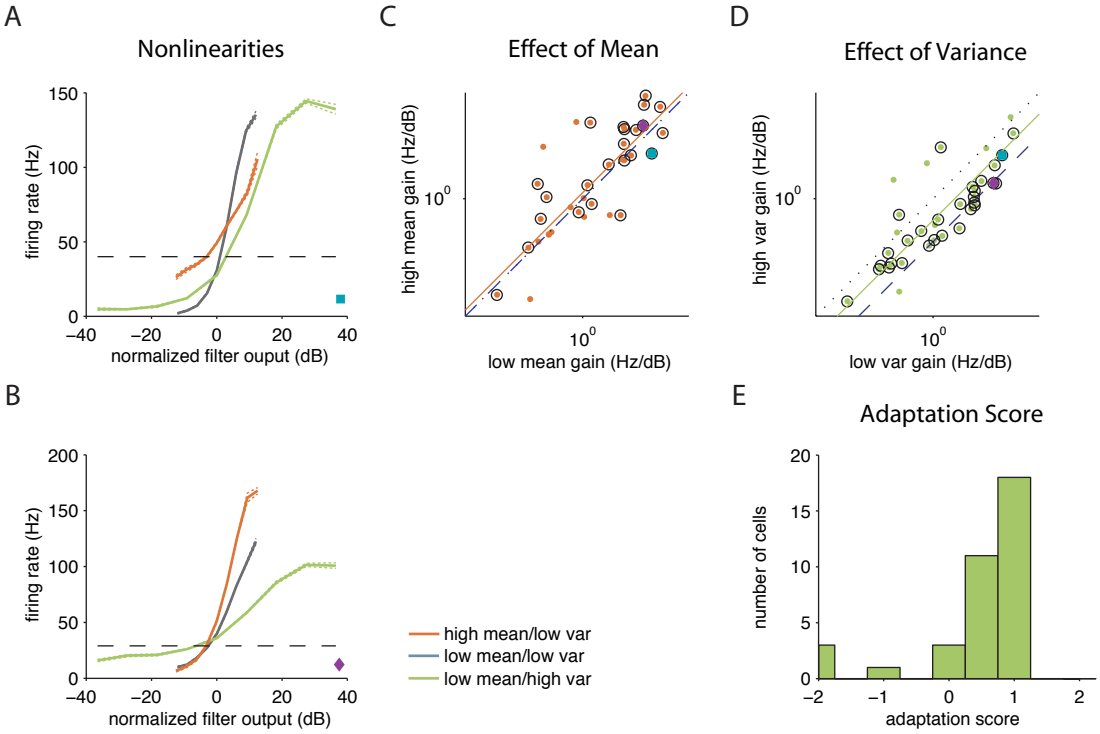
chapter 1, figure 4: effects of stimulus mean on filter shape



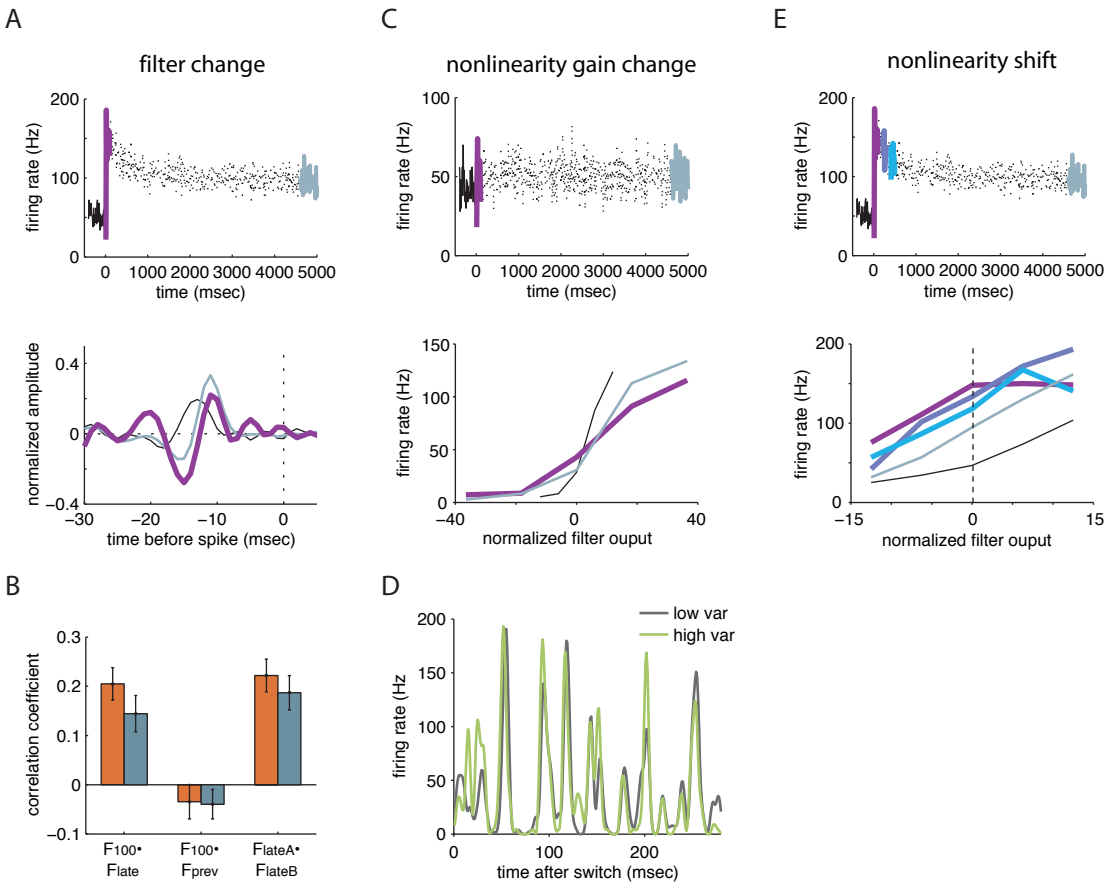
chapter 1, figure 5: population effects of stimulus mean on filter shape



chapter 1, figure 6: effects of stimulus variance on filter shape



chapter 1, figure 7: effects of mean and variance on nonlinearities



chapter 1, figure 8: time course of changes in filters and nonlinearities

Chapter 2: Spectro-Temporal Encoding in the Songbird Auditory Forebrain

Abstract:

How are complex sounds decomposed by the auditory system and what cellular properties underlie this decomposition? Using a synthetic broadband stimulus that shares many properties of natural sounds, we measured spectro-temporal receptive fields (STRFs) of auditory neurons in the avian primary auditory area field L. We found three classes of STRFs: those narrowly tuned for spectral modulations, those narrowly tuned for temporal modulations, and those narrowly tuned in both domains. Cells tuned for fast and slow temporal modulations differed in their firing rate, spike shape, and anatomical distribution, suggesting that cells with distinct electrical or morphological properties may give rise to different response types. Finally, we studied how the shape of each STRF depended on the intensity of the stimulus. At low intensities, STRFs were dominated by single positive peaks, while at high intensities, stronger negative regions gave each STRF type greater sensitivity to specific spectral, temporal, or spectro-temporal modulations respectively. These data link the response properties of auditory neurons to the distribution of energy in natural sounds, and to biophysical differences between cell types. They suggest that response properties depend in systematic ways on the amplitude of the stimulus.

Introduction:

Understanding how complex stimuli are decomposed and represented by populations of neurons is a central goal of sensory neuroscience. In visual cortex, cells are tuned to orientation, and have response properties that depend on their spatial position within a column (Hubel and Wiesel, 1962, 1963, 1974). In the auditory brainstem, parallel pathways encode the timing, intensity, and frequency of incoming sounds in ways that depend on the electrical, synaptic, and morphological properties of different cell populations (Rhode and Smith, 1983, 1987; Spirou et al, 1990; Smith et al, 1991; Oertel 1991, Takahashi and Konishi, 1984, Sullivan and Konishi, 1984, Sullivan 1985).

At higher levels of the auditory system, the organizing principles are less clear. Until recently, responses to different sound frequencies and responses to slow amplitude modulations in time were generally studied separately (Phillips and Irvine, 1981; Phillips and Hall, 1987; Schreiner et al, 1992; Achreiner et al, 1997; Lu et al, 2001; Liang et al, 2002; Barbour and Wang, 2003). More

recently, reverse correlation approaches have made it possible to measure spectral and temporal response properties together (Miller et al, 2002; Depireux et al, 2001; Kowalski et al, 1996, 1997, Theunissen et al, 2000; Sen et al, 2001; Woolley et al, 2005, 2006). While such studies have begun to identify differences in auditory selectivity between areas (Sen et al, 2001; Miller et al, 2002; Linden et al, 2003), they have chiefly revealed a diversity of response types, and few clear links between response properties, cellular properties, and anatomy have emerged.

The songbird provides an excellent model system to study the neural representation of complex sounds. Songbirds produce and perceive complex learned sounds that share many acoustic features with human speech (Singh and Theunissen, 2002). The songbird forebrain contains a primary auditory area known as field L that is analogous to the primary auditory cortex of mammals, and forms the gateway for auditory information to reach forebrain areas involved in song production and recognition (Wild et al, 1993; Fortune and Margoliash, 1995). Studies using simple tone stimuli have identified multiple tonotopic maps in the field L complex (Scheich et al, 1979; Heil et al, 1985; Gehr et al, 1999, Terleph et al, 1996), while anatomical studies have identified 3 layers (L1,2, and 3) that differ in their connectivity and cyto-architecture (Fortune and Margoliash, 1992). Recent studies have compared the aggregate response properties across field L layers (Sen et al, 2001), and between field L and surrounding auditory regions (Woolley et al, 2005, 2006), but have not described the distribution of single cell responses in detail, nor linked these response properties to different cell types.

In this study, we used a rich synthetic stimulus to measure spectro-temporal receptive fields (STRFs) in field L of unanesthetized animals. We found three classes of STRFs--spectral, temporal, and spectro-temporal-- that tile a space of modulations similar to those found in natural sounds. Cells giving rise to fast temporal and spectro-temporal STRFs had distinct physiology from those giving rise to spectral STRFs, and were localized to a different region of field L. Finally, we measured STRFs from the same cells at different stimulus intensities, and found that common nonlinearities--increases in the magnitude and decreases in the latency of negative or inhibitory regions-- allowed STRFs of different types to become more specialized for their preferred type of modulation at higher stimulus intensities. These data suggest the the response properties of field L neurons are organized around the types of modulations prevalent in natural stimuli, and provide a unifying account of how those response properties depend on stimulus intensity.

Results:

Stimulus Design

We designed our stimulus to sample a range of natural time and frequency combinations without imposing the higher order structure of song. The stimulus consisted of 32 logarithmically spaced frequency bands (figure 1A, column 1), each modulated by a different time-varying amplitude envelope (column 2), then summed to produce the final signal (column 3). The envelopes were designed such that the log amplitude of each band was a random Gaussian noise signal with an exponential distribution of frequencies. Envelopes were statistically identical to those used in a previous experiment (Nagel and Doupe, 2006). Frequency bands were gaussian in log-frequency and overlapped by one standard deviation. Full details of the stimulus construction are given in the Methods.

As a result of its construction, the frequency content of our stimulus varied randomly and smoothly in time. Nearby frequencies were generally elevated or depressed together, and remained high or low over a period of a few milliseconds. This can be seen in the spectrogram of the stimulus (3rd column, 2nd panel) which shows the intensity of sound at each frequency as a function of time. Frequency peaks in the stimulus are broader and last longer than those in the white noise spectrogram below it. Local correlations like those in our stimulus are found in most natural sounds, including song, speech, and environmental noise (Singh and Theunissen, 2003). They enabled our stimulus to drive our neurons more effectively and reliably than white noise (data not shown).

The correlations in the stimulus can be quantified by plotting its modulation spectrum (figure2B, left). This heat map shows the energy in the stimulus as a function of temporal and spectral frequency. Energy along the x-axis represents temporal modulations, such as sharp onsets or offsets in the overall sound level. Energy along the y-axis represents spectral modulations such as harmonic frequency combinations. Energy off the axes represents joint spectro-temporal modulations, such as upward and downward frequency sweeps. Asymmetry between the two halves would indicate that upward or downward sweeps predominated in the stimulus, while the symmetric distribution seen here indicates that they were equally likely. Although our stimulus is dominated by low spectral and temporal modulation frequencies, it contains small amounts of energy at higher frequencies. These high frequency tails can be seen most clearly in marginal distributions (figure 1C) that show energy in the stimulus as a function of temporal or spectral frequency alone.

The distribution of energy in our synthetic stimulus shared many features with the statistics of natural sounds. Like our stimulus, most natural sounds have the most energy at low spectral and temporal modulation frequencies, with a long tail of higher frequencies. Temporal and spectral modulation spectra for these sounds can be approximated by a power law (Singh and Theunissen, 2003). In contrast to our stimulus, the energy in zebra finch song and speech is more tightly concentrated along the x and y axes of the 2-dimensional modulation spectrum plot (Singh and Theunissen, 2003), indicating that they consist largely of pure spectral modulations (such as the harmonic combinations found in vowels) and pure temporal modulations (such as the broadband amplitude modulations found in consonants).

In contrast to song or speech, our stimulus contained no higher order structure or correlations. This ensured that the influence of stimulus correlations on STRFs could be removed by "decorrelating" them (see Methods), and that our STRF estimates were not biased by higher order stimulus correlations (Sen et al, 2001; Machens et al, 2004; Ringach et al, 2002; Sharpee et al 2004, 2006).

We played our stimulus at two average intensities: 63dB and 30dB, which alternated continuously every five seconds for 33 minutes. We discuss results obtained with the higher intensity stimulus first, and compare results from the two different intensity conditions at the end of the paper.

Types of STRFs

Using our 63dB stimulus, we were able to obtain robust estimates of spectro-temporal receptive fields for neurons in field L. Although the shapes of these STRFs were diverse, several patterns emerged repeatedly. Examples of the three most common patterns are shown in figure 2.

Figure 2A shows a cell that is primarily selective for temporal modulations. Its positive and negative subfields are arranged sequentially in time, and the highest positive peak extends over 0.6 octaves in frequency (50% width). Due to this arrangement of subfields, the power in this filter is concentrated at high temporal frequencies and low spectral frequencies. This is reflected in its modulation spectrum (2nd column) where energy is concentrated along the x-axis. The slightly asymmetric distribution of energy between the two halves of the modulation spectrum indicates that the STRF is slightly oriented in time-frequency space, and responds more vigorously to downward than to upward sweeps.

In keeping with its sensitivity to fast temporal modulations, the PSTH of this cell's response to repeated trials (third row, black line) shows rapid fluctuations over a 100 millisecond interval. Many of these peaks are captured by the STRF model (red line). The correlation coefficient between PSTH and prediction for this cell was 0.48, just below the population mean of 0.51 +/- 0.14 (sd). Data used to fit the STRF were kept separate from the data used to generate the PSTH, ensuring that these correlations were not due to overfitting. As shown in supplementary figure 2A, the quality of the prediction was highly dependent on the amount of data collected.

Figure 2B shows an example of a cell that is primarily sensitive to spectral modulations. It has a single long positive peak flanked by negative sidebands, and extends over 13.4 msec in time, but is constrained to less than 0.3 octaves in frequency. Its modulation spectrum shows energy concentrated along the y-axis, at high spectral frequencies and low temporal frequencies. It responded much more sparsely to the stimulus than the first cell, with a single broad burst of spikes after a long silent interval.

Figure 2C shows a cell that is sensitive to both spectral and temporal modulations. It has a single central positive peak, 3.3 msec wide in time, and 0.3 octaves wide in frequency. This central peak is flanked by prominent negative sidebands in frequency, giving the cell strong sensitivity to spectral modulations. However, the STRF also contains small but significant positive regions flanking the negative sidebands in time, which give the cell some sensitivity to temporal modulations.

This sensitivity to both spectral and temporal modulations is reflected in the cell's modulation spectrum, which shows energy along diagonals away from the axes. The energy peaks along the x-axis arise because of the very profound negative sidebands in frequency, which give rise to a DC component in spectral frequency. (See supplementary figure 2B for a schematic of the relationship between a filter waveform and its power spectrum.) A second cell of the same type (figure 2D) is more spectrally oriented, and has energy concentrated closer to the y-axis. Both cells have relatively symmetric modulation spectra reflecting relatively un-oriented STRFs. Like the first cell, these two also responded to repeated stimulus segments with fast fluctuations in firing rate. Most such cells had similar short latencies and asymmetric negative sidebands that were stronger and narrower on the high frequency side and broader and shallower on the low frequency side (supplementary figure 2C).

Distribution of STRF shapes

To examine whether these three patterns --temporal, spectro-temporal, and spectral--reflect the distribution of cell types in our population, we looked at the distribution of cells' preferred spectral and temporal modulation frequencies. To calculate preferred modulation frequencies, we first folded and averaged together the two sides of the modulation spectrum, as shown in the top panel of figure 3A for an example cell. We then averaged across spectral frequencies, or across temporal frequencies, to compute the marginal distribution of STRF power as a function of temporal (2nd panel) or spectral (3rd panel) frequency alone. The peaks of these two marginal distributions were taken as the best temporal modulation frequency (temporal BMF) or best spectral modulation frequency (spectral BMF). Plotting the best spectral modulation frequency against the best temporal modulation frequency for each cell in our population (figure 3B, n=72) revealed three classes, corresponding to three classes of example cells in figure 2. "Temporal" cells along the x axis are sensitive to high temporal frequencies and low spectral frequencies. These include the example from figure 2A, indicated by a green square. "Spectral" cells, including the example from figure 2B (blue square) are distributed along the y-axis and are sensitive to high spectral frequencies but low temporal frequencies. A third group of points fall along a diagonal between the two axes, and correspond to cells sensitive to both high spectral and high temporal modulation frequencies. These include the example from figure 2D indicated by a red square. Cells are relatively uniformly distributed along each axis, suggesting that the population evenly samples or 'tiles' the space of pure spectral and pure temporal modulation frequencies. This can also be seen in the marginal distributions at the sides of the plot which both show a peak of cells with the lowest modulation frequencies, and a flat distribution of higher modulation frequencies over the range we sampled.

A limitation of this analysis of modulation spectrum peaks is that it does not capture the behavior of cells like example 2C, which are highly sensitive to spectral modulations, but whose spectral power spectra peak at 0 cycles per octave (red circle in figure 3B). A principal components analysis of the folded modulation spectra allows us to characterize the population distribution of STRFs based on the overall shape of the spectrum, not just its peak. The first two principal components are shown in figure 4A. The first principal component (top panel) has a large positive region along the x-axis and a large negative region along the y-axis. It resembles the difference between the spectra of a strongly temporal STRF and a strongly spectral STRF, and indicates how

sensitive the STRF is to temporal versus spectral modulations. The second principal component resembles the spectrum a spectro-temporal STRF, with a negative region near the origin and a positive region along the diagonal between the axes. A large projection onto this component indicates that the STRF is sensitive to both spectral and temporal modulations.

Plotting the projection of each cell's modulation spectrum onto principal component 2 versus principal component 1 illustrates the distribution of STRF shapes found in our population (figure 4B). Spectral cells (blue) are concentrated at the left side, with large negative projections onto PC1, and small projections onto PC2. Spectro-temporal cells (red) are spread across the top of the triangular distribution. They have large positive projections onto PC2, and a range of projections onto PC1. This range of projections corresponds to our observation that spectro-temporal cells range from more spectral (like the example in figure 2D, red square) to more temporal (like the example in figure 2C, red circle). Temporal cells (green) are distributed along the right face of the triangle and have a range of positive projections onto PC1. The PCA plot suggests that cells in field L form a continuum ranging from purely spectral, through spectro-temporal, to spectral. We divided the PCA plot into colored regions to help describe and analyze regions that behaved differently on average. Placement of linear divisions was based on a consideration of both the PCA plot and the location of example cells (Methods) and is not intended as a strict classification of cell types.

Cells whose modulation spectra fell in different regions of the PCA plot differed in their latency and distribution of orientations. As shown in figure 4C, spectro-temporal cells had the shortest latencies to peak, followed by temporal cells, and spectral cells. Temporal cells often had two significant positive peaks, and their longer latencies arose because the dominant peak was usually the second. Spectral cells had a wider distribution of latencies.

To examine whether cells showed selectivity for oriented spectro-temporal sweeps, we calculated the symmetry index (Methods) of the two sides of the modulation spectrum. As shown in figure 4D, most cells of all types had symmetry indices near zero, indicating that they were largely un-oriented and justifying our use of folded modulation spectra in the above analyses. However, the STRF types differed in their range of symmetries. Temporal cells showed the highest spread in symmetry values (std 0.21), spectro-temporal cells showed intermediate values (std 0.15), while spectral cells showed the least spread (std 0.07). These data indicate that some cells in field L show sensitivity to sweeps oriented just off vertical, but that cells sensitive to spectral modulations were generally unoriented.

These analyses indicate that 3 patterns we observed are characteristic of 3 ranges of STRF shapes seen in our population. The population can be loosely divided into cells sensitive to spectral, spectro-temporal, and spectral modulations, although these may form a continuous distribution, as suggested by the principal components analysis. Cells sample a range of pure spectral modulation frequencies, pure temporal modulation frequencies, and symmetric spectro-temporal modulations. This highly structured distribution of STRF types may be related to the distribution of energy in natural sounds--including song-- which have most of their energy in pure temporal and pure spectral modulations. Our finding contrasts with previous studies in both field L (Woolley et al, 2005) and auditory cortex (Miller et al, 2002), which have suggested that these areas are selective for temporal modulations but low-pass in the spectral domain.

Physiological Correlates of STRF characteristics

Cells whose STRFs had different temporal response properties showed differences in their firing rate and spike waveform shape. Figure 5A shows spike waveforms and the distribution of inter-spike intervals (ISI) for a typical cell with a fast STRF (temporal best modulation frequency = 90.5Hz). It has a narrow spike and a short (1 msec) refractory period, and fired at high average rates both spontaneously (35 Hz) and when driven by our stimulus (104Hz). Figure 5B shows the same data for a cell with a slow STRF (temporal BMF = 5Hz). Its spike has a much wider second peak, and the ISI histogram peaks at a longer latency. This cell fired at 6 Hz spontaneously, and at 10Hz on average when driven by the stimulus.

To examine whether these differences in firing rate and spike shape held at the population level, we plotted each cell's spontaneous rate and spike width against a measure of its temporal response speed, the half-width of its STRF in time. Temporal half widths were measured by finding the peak of the STRF, and measuring the half-width of the temporal cross-section that intersected the peak (Figure 5C, top panel).

The spontaneous firing rate of a cell was negatively correlated with the temporal half-width of its STRF (figure 5D, correlation coefficient = -0.64, p = 1.3e-9), indicating that cells with high spontaneous firing rates gave rise to STRFs with fast time courses, while cells with low spontaneous firing rates gave rise to STRFs with slow time courses. The correlation between spontaneous firing

rate and best temporal modulation frequency was also strongly significant (cc = 0.58, p = 1.07e-7) but less linear.

To examine this relationship in more detail, we plotted the spontaneous firing rate of each cell as a function of its position in the PCA space of figure 4B (figure 5E). Spectral cells, at the left hand corner, had uniformly low spontaneous firing rates (< 20Hz). The remaining temporal and spectro-temporal cells had a range of higher spontaneous firing rates (20-60Hz). The highest spontaneous firing rates (>50Hz) were seen in cells with large positive projections onto both of the first two principal components (top right corner) which were sensitive to both high temporal and high spectral modulation frequencies.

The width of a cell's spike was positively correlated with the temporal half-width of its STRF (correlation coefficient = 0.66, p = 1.0e-10), indicating that cells with narrow spikes gave rise to fast STRFs and cells with wide spikes gave rise to slow STRFs. Spike width was measured on the mean spike over the recording session (5900-374241 waveforms), from the first negative peak to the subsequent positive peak (figure 5C, bottom panel). The correlation between spike width and temporal best modulation frequency was -0.54, p = 5.02e-6).

Because the shape of an extracellular spike depends on the distance between the recording electrode and the cell (), we worried that this relationship could be due to differences in recording conditions. There was a weak correlation between spike width and spike amplitude in our data (cc = 0.26, p = 0.03), arising from a few units with large narrow or small wide spikes. To verify that the relationship between spike width and STRF width was not due to recording distance, we plotted only cells with mean spike amplitudes between 200 and 600μV (n=62/72). Spike and STRF width were still strongly correlated In this sub-population (figure 5F, cc = 0.71, p = 4.8e-11), but spike width and spike amplitude were not (cc = 0.09, p =0.49). There was no significant relationship between spike amplitude and STRF width (correlation coefficient = -0.14, p = 0.27) within this sub-population.

Finally, we noticed that there was a gap in our distribution of spike widths around 350 microseconds, suggesting two clusters of spike widths. In figure 5G, we plotted the location in PCA space of all cells with spikes wider than 350 microseconds. These are mostly concentrated at the left side, and correspond to spectral STRFs.

Together, these data indicate that cells sensitive to fast temporal modulations (including both temporal and spectro-temporal cells) have high firing rates and narrow spikes, while cells sensitive to

slower modulations (spectral cells) have lower firing rates and wider spikes. These data suggest that fast and slow responses arise from distinct classes of cells.

Anatomical distribution of STRF types

Field L is composed of several sub-regions (Fortune and Margoliash) defined by their anatomical location and different distributions of cell morphologies. From dorsal to ventral they are numbered L1, L2, and L3. L2 receives thalamic input and is reciprocally connected to areas L1 and L3. L2 can be further subdivided into L2a, a thin sheet that bisects field L from ventro-rostral to dorso-caudal, and L2b, a more diffuse area located dorsal and posterior to L2a.

To examine whether different STRF types were localized to different regions of field L we performed multi-unit mapping studies in 3 head fixed birds sedated with diazepam. In each experiment, we advanced a four-electrode linear array in steps of 100 microns through the field L complex, and recorded single or multi-unit responses to our stimulus on all channels at each depth. At the end of the recording session, we made marker lesions and sacrificed the bird. We prepared sagital sections of each brain, and stained alternate sections with nissl and with an antibody against the cannabinoid receptor CB1 that selectively labels the input area L2 (Soderstrom et al, 2004).

Figure 6A shows CB1 and nissl-stained sections through one bird. Area L2 can be seen as a dark-staining area in the CB1 section, and as a region of densely-packed cells surrounding the lighter-staining input fibers in the nissl section (white arrows). The tracks of the four electrodes can be seen intersecting L2 obliquely as they descend from the dorsal surface of the brain. Marker lesions on electrode 3 were made at 1000, 1300, and 2200 microns from the surface.

Figure 6B shows the best modulation frequency of raw spike-triggered averages calculated from sorted single- or multi-unit activity at each depth. Cells with high best modulation frequencies (temporal and spectro-temporal cells) were constrained to a narrow region of each penetration, that-- like area L2-- traverses the penetration field from ventro-rostral to dorso-caudal, and is more diffuse on the posterior side. Pink arrow show the depths of the marker lesions on electrode 3, and support the localization of faster cells to area L2. Cells with low best modulation frequencies (spectral cells) were found throughout the penetration but were more common above and below the region of fast cells.

Localization of cells with fast response properties to L2 was seen in three additional mapping sessions from two birds. Figure 6C shows BMF as a function of distance from the CB1-defined center

of L2 in each of the three experiments. Data from 3-4 electrodes are plotted on each axis, and aligned so that zero represents the center of L2. Cells with higher best modulation frequencies are concentrated near L2, while those sensitive to slower features are located above and below it. Raw spike-triggered averages at the far right are taken from the last mapping experiment. They illustrate that robust tuning and changes with depth could be observed using multi-unit recording without decorrelation. Data from the chronically recorded birds show a similar localization pattern, with narrow-spiking cells sensitive to fast modulations generally concentrated in one region of the penetration (supplementary figure 3).

In these experiments we used two different stimulus configurations to try to estimate STRFs well in a short amount of time. One stimulus--also used in the chronic experiments-- had modulation envelopes filtered so power was proportional to 1/50Hz. A second stimulus was identical except that modulation envelopes were filtered so power was proportional to 1/20Hz. We found this slower stimulus to be useful in obtaining robust estimates of slower STRFs from a smaller number of spikes. Examples of STRFs from the same site estimated with both stimuli are shown in supplementary figure 4. STRFs estimated with the slower stimulus were generally smoother and less noisy but otherwise did not differ significantly from those estimated with the faster stimulus. In figures 6B-6E, blue dots represent STRFs estimated with the 50Hz stimulus, and gray dots represent STRFs estimated with the 20Hz stimulus. In three of these experiments (figures 6B, 6D and 6E), a single stimulus condition was used; each experiment shows that fast (high best modulation frequency) STRFs are localized to L2. In a single experiment (figure 6C), we used different stimulus conditions at different depths. However, STRFs recorded with the 50Hz stimulus alone (blue dots) are sufficient to show that fast STRFs are localized to field L2.

Together, these data indicate that cells with long integration times are concentrated in the L1 and L3 regions of field L. Cells with short integration times are mostly localized to field L2a.

STRF changes with stimulus intensity

In a previous study of temporal receptive fields (Nagel and Doupe, 2006), we found that cells in field L showed systematic changes in the shapes of their receptive fields in response to a change in the mean stimulus amplitude. Specifically, at higher mean intensities, the negative parts of the receptive fields grew larger and decreased in latency, causing the cells to become more sensitive to

changes in stimulus amplitude. These changes occurred rapidly after a change in stimulus statistics, suggesting that they represent nonlinearities in the neural response rather than time-dependent adaptations. To ask whether spectro-temporal receptive fields would show related nonlinearities, we compared STRFs obtained from responses to the same stimulus played in an interleaved fashion at 63dB and at 30dB.

Figure 7A and B shows STRFs obtained from the same spectro-temporal cell with two different stimulus intensities. At 30dB (figure 7A), the STRF has a single large positive peak flanked by shallow negative regions on all sides. The negative regions that flank the positive peak in spectrum lag the peak in time. At 63dB (figure 7B), the negative spectral sidebands have grown deeper and shifted forward in time relative to the positive peak. In addition, small positive peaks have appeared flanking these negative sidebands in time.

These changes can be seen more clearly in temporal and spectral cross-sections through the peak of the low intensity STRF (figures 7C and 7D). The temporal cross-section (figure 7C) illustrates that the 63dB STRF (red) is narrower and shifted forward in time relative to the 30dB STRF (blue). The spectral cross-section (figure 7D) reveals that the negative sidebands are much more prominent in the 63dB STRF (red), than in the 30dB STRF (blue). Dashed lines in these two plots represent the standard deviation of five jackknife estimates of the STRF (see Methods).

The consequences of these changes in STRF shape can be seen in rasters of the neuron's response to repeated trials of the same stimulus at the two intensities (figure 6E and 6F). The cell responds robustly in both conditions, with peaks of equal magnitude, indicating that the difference in the two STRFs is not simply due to reduced spiking at 30B. In addition, although some peaks--such as the first two-- occur in both responses, they occur at a shorter latency in response to the 63dB stimulus, while other peaks are entirely different between the two conditions. These data indicate that subtle changes in the strength and relative latency of STRF peaks are associated with dramatic changes in the neural response to the same stimulus at different intensities. Similar changes were observed for most spectro-temporal cells, as quantified in the next section.

Figure 7F shows changes typical of a temporal cell. At 30dB (top left panel), this cell's STRF has a single positive peak that is narrow in time but elongated in frequency. A large weak negative region follows this ridge in time. At 63dB (bottom left panel), it has two significant positive peaks divided by a strong negative region. This dramatic temporal phase change can also be seen in the

temporal cross-sections (top right panel) through the peak of the 30dB STRF (top right panel). In a previous experiment, we observed very similar phase changes in purely temporal receptive fields (Nagel and Doupe, 2006). There we used several different stimulus intensities (30 40 50 and 60dB) to show that the negative part of the filter gradually increased in magnitude and decreased in latency with increasing stimulus intensity. Similar mechanisms are likely to underlie the nonlinearity observed here.

While most temporally organized STRFs showed temporal phase changes with intensity, changes in the spectral domain were inconsistent across these cells. Some, like this example, became somewhat more narrowly tuned at high mean. Others became more broadly tuned, or showed a shift in their preferred frequency.

Figure 7G shows changes in a cell tuned for spectral modulations. At 30dB (top left panel), the STRF consists mostly of a positive peak that is narrow in frequency but extended in time. It is flanked by a weak negative region that is more prominent on the high frequency side. At 63dB (bottom left panel), the main positive peak grows narrower in time (top right panel), the negative sidebands increase in depth, and an additional positive peak appears above the preferred frequency (bottom panels).

The error bars on the estimates of spectral STRFs were generally much larger than those for other STRF types, due to their significantly lower firing rates (figure 5D). Despite recording them for longer (2-4 runs through the entire stimulus), we were able to obtain fewer spikes from these cells (mean 2500 spikes per condition for spectral cells, 11000 for temporal cells, and 17000 for spectro-temporal cells). Changes in these STRFs therefore appear to be less significant (figure G, right hand panels), although they are broadly similar to changes observed in other STRF types (stronger negative regions, narrower temporal responses, more significant positive peaks). To quantify the amount of nonlinearity in a way that does not depend on estimating the STRF, we compared the PSTH of the cell's response to the same stimulus segment played at the two intensities. If the cell behaves linearly, this correlation coefficient should be high, since the responses should differ only by a scaling factor. If the cell behaves nonlinearly, the patterns of response to the same stimulus will differ and the correlation coefficients will be low. The mean correlation between PSTHs for spectral cells (0.34 +/- 0.19) was not significantly greater than the mean for spectro-temporal cells (0.35 +/- 0.11), which showed highly significant changes in STRF shape. These data suggest that all cell types showed significant nonlinearities with stimulus intensity.

Population Analysis of Intensity Nonlinearities

To quantify changes in STRF shape across our population, we fit the STRFs in each condition to a bi-variate mexican hat model (Methods). This model generates a synthetic STRF composed of a Gaussian function multiplied by a quadratic function in time and spectrum. The model has seven parameters: an overall scale, a latency, a best frequency, a temporal and a spectral width term, and two parameters that we call alpha and beta, which control the depth of the sidebands in time and in frequency respectively. We chose this model because it is flexible enough to fit STRFs sensitive to purely temporal and purely spectral modulations, as well as STRFs sensitive to both (figure 8A). In addition, sensitivity to spectral and temporal modulations are each governed by a single parameter, the alpha and beta coefficients on the quadratic terms. By comparing the fits of these two parameters to the low and high intensity STRFs, we can ask whether cells show changes in sensitivity to spectral and/or temporal modulations.

Figure 8B shows examples of fits of the model to the example STRFs. Qualitatively, the model is able to capture several aspects of each STRF, including its preferred frequency and latency, its widths in spectrum and time, and its selectivity for spectral,temporal, or both spectral and temporal modulations. Supplementary figure 5 shows the distribution of these parameters obtained from model fits, along with estimates of these parameters obtain by directly measuring the STRFs. The two distributions show good agreement. These examples also illustrate some limits of the model: because the model can produce only three peaks in each dimension, it cannot capture the additional peaks in the spectral STRF. The temporal STRF with two positive peaks can only be fit by changing the sign of the scaling term. The version of the model we used contains no orientation or asymmetry terms, making it unable to capture the subtle orientation of the temporal STRF, or the asymmetries of the spectro-temporal STRF. These terms can be added to the model to produce more accurate fits. However, because these parameters did not show consistent changes between conditions, we chose to omit them to simplify our fitting procedure.

Figure 8C shows the distribution of the parameter beta controlling the depth of spectral sidebands at 30 versus 63dB. Many cells have values of alpha close to zero under one or more conditions, indicating that they do not show much sensitivity to spectral modulations. Of the

remaining cells, 25 cells show a significant increase in beta with an increase in stimulus intensity, while 4 show a significant decrease. Overall, values of beta were higher at 63dB than at 30dB ($p = 6.6e-7$). Figure 8D shows the PCA location of cells that showed a significant increase in beta. They include most of the spectro-temporal cells and a fraction of the spectral cells.

Figure 8E shows the distribution of the parameter alpha controlling the depth of temporal sidebands, at 30 versus 63dB. 20 showed a significant increase in alpha at 63dB versus 30dB, while only 5 showed a significant decrease. Across the population, values of alpha were significantly higher at 63dB than at 30dB ($p = 3.5e-8$). These data suggest that as a population, cells tend to become more sensitive to temporal modulations at higher intensities Figure 8F shows the location of cells that showed a significant increase in alpha in PCA space (dark green squares). These cells include most of the spectro-temporal cells, and a substantial fraction of the temporal cells.

As described above, many temporal cells showed a dramatic change in temporal phase with stimulus intensity, with one significant positive peak at 30dB, and two significant positive peaks at 63dB. In the context of our model, such phase changes appear as a change in the sign of the scaling term from positive (producing one positive peak and two negative peaks) to negative (producing one negative peak and two positives). Cells that show such a change in sign are indicated by light green circles in figure 8F. These data indicate that nearly all temporal cells show either an increase in the depth of temporal sidebands, or a phase change, when the stimulus intensity increases.

Together these data indicate that STRFs, like purely temporal receptive fields, show significant nonlinearities with stimulus intensity. Spectro-temporal cells show significant deepening of both spectral and temporal sidebands at 63dB versus 30dB. Temporal cells show either deeper temporal sidebands, or a phase change, or both at 63dB. Many spectral cells show deeper spectral sidebands at 63dB. At low intensities, STRFs are dominated by a single peak and sideband peaks are reduced in magnitude. At higher mean intensities, sidebands become more prominent, making cells more selective for particular spectral and temporal modulations. These data suggest that at the low stimulus intensities, cells act more like detectors, responding to the presence of their preferred frequencies, regardless of what other sounds are nearby. At higher intensities cells become more specialized for their particular kind of modulations.

Discussion:

Using a stimulus that shares many properties of natural sounds, we mapped the spectro-temporal receptive fields (STRFs) of neurons in the primary auditory area field L of awake zebra finches. We found a highly structured distribution of STRFs, with most cells specializing in either temporal or spectral processing; a third population was highly tuned in both these domains. We related these different response properties to differences in the spike shape, spontaneous firing rate, and anatomical distribution of the cells that produced them, suggesting that cells are morphologically and biophysically specialized for spectral versus temporal processing, and that these two types of processing are anatomically segregated in the zebra finch forebrain. Finally, we explored how STRF shapes depended on the intensity of the stimulus. We found that each STRF type became more sensitive to its preferred type of modulation when the stimulus intensity was high, and behaved more like a low-pass detector when the stimulus was soft. We discuss each of these three findings in detail below.

Distribution of STRF Types

We found that STRFs of neurons in field L can be roughly divided into three classes: STRFs that are primarily sensitive to temporal modulations (temporal cells), STRFs that are primarily tuned for spectral modulations (spectral cells), and STRFs tightly tuned in both spectrum and time (spectro-temporal cells). Within each class neurons showed a distribution of properties. Temporal cells were insensitive to spectral modulations but evenly spanned the range of temporal modulations (0-150Hz). Spectral cells were insensitive to temporal modulations but evenly spanned the space of possible spectral modulations (0-2 cycles per octave). Spectro-temporal cells were centered at many best frequencies, but shared a similar latency and receptive field structure, with a compact central positive peak flanked by asymmetric negative regions. Spectro-temporal cells varied from mostly temporal-- with energy concentrated near the temporal axis of the modulation spectrum plot-- to mostly spectral, but were generally symmetric in time-frequency space, indicating that they did not differentially respond to upward versus downward frequency sweeps. Strongly oriented sweep-selective cells, and broadly tuned cells that would act as overall sound level detectors, were largely absent from our population. This distribution of response properties may be related to the structure of many natural sounds--including both zebra finch song and speech-- which are dominated by purely temporal and

purely spectral modulations, and contain comparatively few strongly oriented spectro-temporal modulations (Singh and Theunissen, 2002, Woolley et al, 2005).

While our data agree broadly with previous studies of the distribution of STRF types in field L or mammalian auditory cortex, they differ in several important respects. Woolley et al (2005) calculated the average modulation spectra across field L neurons, and compared this ensemble modulation spectrum to the distribution of energy found in song. They concluded that field L was band-pass in the temporal domain but low-pass in the spectral domain. By looking at the distribution of individual cell types, rather than aggregate measures, we found that separate populations of temporal and spectral cells were tuned to modulations in each of these domains, while a third population was tuned to high frequency modulations in both domains. As in our study, Miller et al (2002) found that most neurons in anesthetized cat A1 have fairly symmetric modulation spectra, indicating that they were not selective for oriented frequency sweeps. However, that study found no systematic relationship between the spectral and temporal tuning properties of A1 neurons, while we found a strong trade-off between temporal and spectral selectivity. These differences may arise from the structure of the avian forebrain, which contains many fast-firing cells able to follow rapid modulations in the stimulus, while the mammalian auditory cortex has generally slower responses (Miller et al, 2002, Depireux et al, 2001, Lu et al, 2001). They may also arise from differences in recording conditions. Most previous studies have measured STRFs under pentobarbital anesthesia, while we recorded from unanesthetized animals. Anesthesia can profoundly influence the temporal dynamics of cortical auditory responses (Wang et al, 2005). Chi and Shamma (2005) have proposed that A1 neurons evenly span a range of time-frequency orientations, similar to the tiling of 2D spatial orientations in visual cortex. However, the same group reported that A1 neurons show mostly symmetric tuning properties (Depireux et al, 2000). Together with data from many other studies, our findings suggest that tiling of pure spectral and temporal modulations, not orientation in time-frequency space, may be the organizing principle of forebrain auditory sensitivity.

Correlations with Physiology and Anatomy

The second main finding of our study was that cells with different response properties also differ in their physiology and anatomical distribution. Cells sensitive to fast stimulus modulations, including both temporal and spectro-temporal cells, had high spontaneous and driven rates, narrow

spike waveforms, and were located in a restricted region of each penetration. Our multi-unit mapping study indicated that this restricted region corresponds to the input layers, L2a and L2b. Spectral cells sensitive to slower stimulus modulations were found throughout the penetration, but were more prevalent in the flanking regions L1 and L3. These cells had much lower spontaneous and driven rates, and wider spike waveforms. These data suggest that fast and slow response types may arise from cells with different morphological or electrical properties, concentrated in different anatomical regions.

Biophysically, narrow spikes and high firing rates can arise from small or electrotonically compact cells that can repolarize quickly and track rapid fluctuations in their synaptic input. Such cells may have specialized potassium channels that give them faster kinetics (Martina 1997, 1998; Chow et al, 1999). Conversely, a large or electrotonically extended cell will have a wider spike and longer refractory period, leading to lower firing rates, and more low-pass filtering of its input. The subregions of field L have been shown to differ in their distribution of cell sizes and morphologies (Fortune and Margoliash, 1992): L2a contains more small and medium bodied cells with compact dendritic fields. L1 and L3 have more large cell bodies and more medium-bodied cells with extensive dendritic fields. We therefore hypothesize that temporal and spectro-temporal responses may arise from smaller cells with more compact dendrites, while spectral responses may arise from larger cells with dendritic conductances that allow them to integrate inputs over time.

In the auditory brainstem of both mammals and birds, distinct temporal response patterns have been linked to the distinct morphologies and electrical properties of bushy and stellate cells (Rhode and Smith, 1983, 1986, Sullivan and Konishi, 1984, Oertel, 1991). These different response properties are in turn related to different functional roles in encoding phase and intensity information (Sullivan, 1985; Takahashi et al, 1984, Yin and Chan). Although intracellular recordings and labeling experiments will be required to definitively link cellular properties to auditory responses in field L, our data suggest that similar relationships between structure and function exist for auditory neurons in the avian forebrain.

In the mammalian hippocampus and cortex, narrow extracellular spike waveforms have been linked to inhibitory interneurons (Henze et al, 2000, Buzsaki and Eidelberg 1982), which have narrower spike waveforms in intracellular recordings (Buhl et al. 1996; Scharfman 1995; Hasenstaub, 2005) . About 30-40% of neurons in the auditory forebrain are estimated to be GABAergic (Pinaud et al, 2004, 2007), but physiological differences between inhibitory and excitatory cells in these areas have not been explored. Due to the density of units with fast responses and narrow spikes in regions

of our penetration, we think it is unlikely that fast responses arise exclusively from inhibitory neurons, although this remains an open possibility.

Our data fit with previous studies showing that auditory neurons become tuned for progressively slower temporal modulations at successive levels of the auditory hierarchy (Creutzfeldt, O. 1980; Sen et al, 2001; Miller et al, 2002; Linden et al, 2003). In contrast to mammalian systems, where a large difference in temporal following is seen between IC and cortex (Langer and Schreiner, 1988, Miller et al, 2002), but not between successive cortical stages, we see a sharp distinction between the input and output layers of the avian primary auditory area. This may indicate that L2a shares more properties with the mammalian thalamus, as suggested by Las et al (2005).

Several further questions are raised by our finding about the localization of distinct cell types. First, how can slow spectral cells be built from fast spectro-temporal or temporal cells? If tightly tuned spectro-temporal cells were found at many latencies, then spectral cells could be built by simply summing the responses of many fast spectro-temporal cells. Fast cells, however, show uniformly short latencies (figure 3C), indicating that spectral cells cannot arise from summation alone. An interesting alternative is that the morphology and electrical properties of spectral cells allows them to low-pass filter their inputs, and so integrate information over much longer time scales.

A second question raised by this finding is what happens to the fast temporal information represented by temporal and spectro-temporal cells? Anatomical evidence suggests that higher-order areas involved in song recognition and production receive input from L1 and L3, but not directly from L2. Estimating the auditory response properties of these higher-order areas using linear reverse correlation techniques has proved difficult (Sen et al, 2001; Gentner et al, 2003). It would be interesting to ask whether sensitivity to fast temporal modulations is preserved in nonlinear responses of L1/L3 or higher order neurons. This could be done using more complex techniques such as spike-triggered covariance (STCM, Rust et al, 2005; Touryan et al, 2002, 2005) and the method of most-informative dimensions (MID, Sharpee et al, 2004,2006).

Nonlinearities in spectro-temporal processing

Our third main finding was that cells with different response properties showed different nonlinearities with stimulus intensity. At high intensities, spectro-temporal cells showed increases in

the magnitude of both spectrally-oriented and temporally-oriented negative STRF regions, leading to increased sensitivity to both spectral and temporal modulations. Temporal cells showed dramatic changes in temporal phase, with stronger temporally-oriented inhibition, and often two significant excitatory peaks. Spectral cells showed less dramatic changes between low and high intensity conditions, but did become more spectrally bandpass at the higher intensity.

In a previous study of temporal response properties, we found that most cells in field L integrate over longer times at low intensities and become more temporally differentiating at high intensities (Nagel and Doupe, 2006). We argued that such changes could be adaptive, allowing cells to become selective for particular temporal modulations when signal-to-noise levels were high, and to behave more like detectors when signal quality was low. Similar nonlinearities have been described in retinal ganglion cells, which respond to differences in local light intensity at high contrast, but integrate over larger spatial regions at low contrast (Enroth-Cugell and Lennie, 1975). The results of the current study extend these findings, suggesting that cells in field L become more specialized for particular kinds of temporal and spectral modulations at high intensities, while at low intensities all behave more like low-pass detectors.

While different cell types show distinct nonlinearities with intensity, many of these changes could arise through similar mechanisms. Most nonlinearities we observed seem to arise from an increase in the relative strength, and decrease in the relative latency, of negative regions of the STRF. This leads to subtle changes in the shape of spectral and spectro-temporal STRFs, whose negative regions are generally more spectrally oriented, and dramatic phase changes in temporal STRFs. These data suggest a model, in which each cell receives at least two inputs, one excitatory and one inhibitory. As the stimulus intensity increases, both inputs increase in magnitude and decrease in latency. The excitatory input has a lower threshold and saturation point, while the inhibitory input has a higher threshold and decreases its latency more in proportion to the input level. If the underlying excitatory and inhibitory inputs had different spectro-temporal receptive fields, this simple model could account for the nonlinear response properties of many different auditory neuron types. Such circuitry could be present in the forebrain or cortex (Wehr and Zador, 2003; Tan and Schreiner, 2004), or could shape responses of more peripheral neurons that pass their response properties on to higher areas (Nelken and Young, 1994; Yu and Young, 2000).

Previous studies in auditory cortex have measured response strength as a function of the intensity of tones or amplitude modulated noise (Phillips and Irvine, 1981; Phillips et al, 1994, Calhoun and Schreiner, 1998; Polley et al, 2004). These studies have often seen an array of response types--

from those that monotonically increase with intensity, to those that are strongly non-monotonic. By considering responses as a function of both frequency and time, our data suggest that simple and unified processes may underlie this diversity of responses: at higher stimulus intensities, negative inputs increase in magnitude and decrease in latency relative to positive inputs, causing cells to respond more selectively to particular kinds of modulations. Thus, studies of joint spectro-temporal response properties can reveal unifying principles underlying the apparent diversity of response types.

Methods:

Chronic Electrophysiology

We used chronically implanted microdrives (Hessler and Doupe, 1999) to record single units (n = 72) from 5 adult male zebra finches. A detailed description of microdrive construction and implantation are given in that paper. Electrodes (2-3 tungsten electrodes, 4-5 MOhms, MicroProbe Inc, Gaithersburg, MD) were implanted 1.5 µm lateral (left), and 1.5 µm rostral to the posterior border of the branch point of the central sinus, at an initial depth of 400 µm.

During recording, the bird was attached to a commutator by a flexible lead and op-amp. Electrical traces were digitized, amplified (1000x), filtered (300-5000Hz), and recorded using TDT System 3 hardware (Tucker-Davis Technologies, Alachua, FL) interfaced with custom-written Matlab software. The electrode bundle was advanced manually in small steps (40-160 microns). Putative single units were identified on the oscilloscope by their stable spike waveform and clear refractory period. All spikes were re-sorted offline using either a custom-written software window discriminator (Matlab) or a commercial spike sorter (Plexon). In both cases spikes were sorted based on the similarity of overlaid spike waveforms and on clustering of waveform projections in a two-dimensional principal component space. Neural recordings were considered single units if they contained fewer than one violation of 1 msec refractoriness per thousand spikes after sorting. Units that responded to auditory stimuli were found at depths of 1000-2500 µm. Between each recording session, the electrodes were retracted to a position above where the first auditory units were found.

During recording, the bird was placed inside a small cage (20cm x 20cm floor area) within a sound-attenuating chamber (Acoustic Systems). The chamber lights were kept off to minimize movement and birds were monitored using an infra-red camera. Birds generally sat in one corner of

the cage for the duration of the experiment although the commutator permitted free movement within the cage. Auditory stimuli were presented free-field from a small speaker (Bose) located 50 cm from the center of the bird's cage. Using a calibrated microphone (B&K) we verified that 250Hz to 10kHz tones designed to play at 80dB appeared at 79.6+/-2.3dB, and that the highest harmonic distortion peak observed was less than 23.3dB (65.9 +/- 8.7dB SNR).

Acute Electrophysiology

We performed acute mapping experiments in three additional adult male birds. A few days prior to the experiment, we prepared the bird for acute physiology under equithesin (3.5µl/g, Hessler and Doupe, 1999) anesthesia by removing the scalp, making an opening through the skull over the location of field L, and affixing a stereotaxic metal pin to the skull with dental cement. The skull opening was then covered with a silicone elastomer (World Precision Instruments) and the bird was allowed to recover fully.

On the day of the experiment, the bird was lightly sedated with diazapam (30µl), wrapped in a cloth jacket and placed in a stereotaxic device. The bird's head was held in place using the previously implanted pin. The silicone covering was then removed from the skull, a hooked ground electrode was placed under the skull, at the corner of the opening, and a linear array of four extracellular electrodes was lowered into the brain.

For mapping experiments, we lowered the electrode array in 100 microns steps and recorded activity on all channels if auditory activity was present on any channel. Single and multi-unit activity was sorted using the Plexon offline sorter. Multi-unit clusters were separated into multiple clusters if this separation increased the signal quality of the STRFs produced from each cluster.

After the final recordings, we made electrolytic lesions at several depths along the electrode penetrations and prepared histological sections to identify the location of recording sites. Birds were lethally anesthetized and perfused with saline followed by 4% paraformaldehyde. Alternate 40 micron sections of fixed brain tissue were Nissl-stained, and labeled for CB1, a marker for the input region L2 (Soderston et al, 2004).

Stimulus

To estimate spectro-temporal receptive fields (STRFs) for our cells, we created a stimulus that shared some properties with natural song but which sampled the space of possible frequency combinations more evenly, and contained only local 2nd order correlations. Like song, our stimulus was broadband, and had
more energy concentrated at low temporal and spectral frequencies than at higher ones (Singh and Theunissen, 2003)

The stimulus was composed of 30-32 logarithmically spaced narrowband noise signals, each created by passing a white noise signal through a filter that was gaussian in log frequency (Figure 1A). The center frequencies of the filters cf_n were given by

$$cf_n = \exp(\log(400) + 0.1(n-1))$$

where cf_n is the center frequency of the nth band. Filters overlapped by one standard deviation so that the summed narrowband noises had close to a flat envelope (Theunissen et al, 1998). Each narrowband signal was modulated by a different envelope (Figure 2B), designed such that the amplitude of each band in dB fluctuated randomly and smoothly in time. Specifically, the log envelope was given by a gaussian noise signal smoothed with an exponential filter:

$$P(f) = \exp(-f/50\text{Hz})$$

where P is the power in the log envelope and f is frequency. These envelopes were statistically identical to those used in a previous experiment (Nagel and Doupe, 2006). The independently modulated noise bands were then summed to create the final stimulus.

By adjusting the mean and the variance of the log envelope, we could manipulate the intensity of our stimulus. We used two stimulus conditions, 63dB mean, 6dB standard deviation, and 30dB mean 6 dB standard deviation, that alternated smoothly every five seconds. A third condition, with a mean of 30dB and a standard deviation of 18dB, was also presented in interleaved segments but is not discussed in this paper.

Due to the overlapping frequency bands, and to smoothing of the envelopes, our stimulus had local 2nd-order correlations in both frequency and time. Figure 1B illustrates these correlations by

plotting the two dimensional modulation power spectrum of the stimulus spectrogram. Closely related to a normal power spectrum, which shows energy as a function of frequency, this modulation spectrum shows how energy is distributed as a function of temporal and spectral modulation frequency. Energy along the x-axis represents temporal modulations, and energy along the y-axis represents spectral modulations. Energy off the axes represents joint spectro-temporal modulations, such as frequency sweeps. Like many natural stimuli, including song and human speech, our stimulus had energy concentrated at low spectral and temporal modulation frequencies. This can be seen in the two dimensional heat map, where energy is concentrated near the origin, and in the two marginal plots showing energy as a function of spectral or temporal frequency alone. Our stimulus contained no higher order correlations, ensuring that the influence of stimulus correlations on STRFs could be removed by "decorrelating" them as described in the following section.

Half of our stimulus was composed of repeats of the same stimulus segment while the remainder were unique segments. Unique segments were used to generate the STRF models, while their predictions were tested on the repeated segments. This segregation of model generation data from test data ensured that the quality of predictions was not due to over-fitting.

STRF estimation

Our analysis attempted to relate the firing of field L neurons to a linear filter of the stimulus spectrogram. If this spectrogram were uncorrelated in both frequency and time, we could calculate this filter, or spectro-temporal receptive field (STRF), by simply cross-correlating the spike train with each row of the spectrogram, yielding the spike-triggered average, or STA. Because the spectrogram of our stimulus contains local correlations, this STA will be equal to the true neural filter, convolved with the autocorrelation of the spectrogram. The influence of these stimulus correlations can be removed from the STA by inverting the autocorrelation matrix. Preparing our stimulus for analysis thus required two steps: first calculating its spectrogram, and then calculating the autocorrelation matrix of the spectrogram envelopes.

To calculate the spectrogram $S(cf,t)$, we first filtered the stimulus through the same bank of 32 log-gaussian filters used to create the narrowband noise signals. We then extracted the envelope of each narrowband signal by passing its absolute value through the exponential filter used to generate the stimulus envelopes. We down-sampled the resulting low-pass signal to 1kHz to speed analysis, took its logarithm, and normalized it to have zero mean and unit variance. Because of the overlap in

our frequency bands, the log envelopes of several of the original noise bands contributed to each of these extracted log envelopes and the extracted envelopes *S(cf,t)* were not exactly equal to the envelopes used to create the stimulus.

The autocorrelation matrix of the spectrogram consisted of 128 complex matrices $A_\omega(cf, cf)$, representing the correlations between the log envelopes *S(cf,t)* at each of 128 temporal frequencies ω. The diagonals of each matrix $A_\omega(cf, cf)$ represent the power in each log envelope at the temporal frequency ω, while the off-diagonal terms represent the correlation between log envelopes at that temporal frequency. To calculate these complex matrices, we broke *S(cf,t)* into 256 sample-long chunks *S(cf,t:t+255),* applied a hanning window to each chunk, then took its Fourier transform, yielding a complex matrix $\tilde{S}(cf,\omega)$. The complex matrix $A_\omega(cf, cf)$ was then equal to the average cross-product across chunks:

$$\left\langle \tilde{S}(cf,\omega)' \times \tilde{S}(cf,\omega) \right\rangle$$

To estimate each STRF, we first calculated the raw spike-triggered average $C_n(cf,\tau)$ by cross-correlating the spike train ρ(t) with each log envelope *S(cf,t)* in the spectrogram:

$$C(cf,\tau) = \frac{1}{n}\sum_t \rho(t) S(cf, t-\tau)$$

where n is the total number of spikes. We obtained spike-triggered averages over an interval from 100msec before the spike to 100msec after. The portion of this average that followed the spike was used as an estimate of the noise in our averaging process.

To decorrelate the spike-triggered average, we first took its Fourier transform, yielding a complex matrix $\tilde{C}(cf,\omega)$. Each column of this matrix $\tilde{C}(cf,\omega)$ corresponding to a particular temporal frequency ω, was then decorrelated by applying the pseudo-inverse of the autocorrelation matrix for that temporal frequency $A_\omega(cf, cf)$. To apply the pseudo-inverse, we first took the singular value decomposition of the matrix $A_\omega(cf, cf)$:

$$U\Sigma V' = A_\omega(cf,cf)$$

where U and V are square matrices and Σ is a matrix of singular values that has non-zero values only along the diagonal. We then calculated a matrix Σ_{inv}, by taking the inverse of the values of Σ along the diagonal:

$$\Sigma_{inv}(s,s) = \frac{1}{\Sigma(s,s)}$$

We used Σ_{inv},with U and V to decorrelate the spike-triggered average:

$$\tilde{D}(cf,\omega) = V\Sigma_{inv}U'\tilde{C}(cf,\omega)$$

The inverse Fourier transform of this "decorrelated" matrix $\tilde{D}(cf,\omega)$ is equal to the decorrelated spike-triggered average $D(cf,\tau)$. This method is very similar to the one used by Theunissen et al (2000).

A limit of this decorrelation technique is that it tends to boost the power in high temporal frequencies that are poorly sampled by the stimulus, leading to very noisy estimates of the STRF. This is because the inversion increases the magnitude of the smallest singular values. To limit this noise we systematically reduced the amplitudes of the largest inverse singular values. We first obtained the singular values of all the autocorrelation matrices and sorted them in descending order (largest to smallest). We then found the largest singular value of the autocorrelation matrix at a cutoff frequency $A_{\omega cut}$, whose choice is described below. We call this cutoff singular value s_{cut}. Singular values greater than this cutoff value were inverted as described above. Singular values below this cutoff were reduced in amplitude according to their distance from the cutoff value

$$\Sigma_{inv}(s,s) = \frac{1}{\Sigma(s,s)} \times \exp(-k/250)$$

where k is the number of singular values between the current value and the cutoff. By smoothly reducing the amplitude of large inverse singular values, this method reduces "ringing" in the STRF that can arise from using a hard cutoff on the decorrelation.

We tried several methods for choosing an appropriate cutoff frequency. The first was to find the frequency at which the power in the spike-triggered average fell below 3 times the mean power in the acausal part of the the spike-triggered average. This "noise cutoff" yielded reasonable results for some cells, but unreasonably high cutoffs (150-250 Hz) for others. STRFs decorrelated using a cutoff frequency greater than 150Hz were uniformly noisy and impossible to interpret. We therefore tried an alternate method, finding a cutoff frequency that contained 99% of the power in the raw spike-triggered average. This "99% cutoff" yielded much better results for STAs with excessively high noise cutoffs, but was too high for those with low noise cutoffs. We therefore chose one of these two cutoffs for each STRF depending on whether they yielded a significant STRF. If the 99% cutoff produced an STRF with at least 20 points above the maximum absolute value of the acausal STRF, we used that cutoff (n=49/97). Otherwise we used the lower noise cutoff (n=23). We observed no significant differences between the groups of cells decorrelated with one cutoff or the other and results for the two groups are pooled. Un-decorrelated spike triggered averages for the example cells in figure 2A-C are shown in supplementary figure 1. Both differences in tuning between cells, and differences in the same cell's STRF with intensity, are evident in these raw averages.

Both the causal and acausal parts of the STA were decorrelated together. STRFs were considered significant if the causal part of the STRF contained at least 20 points that were greater than the maximum absolute value in the acausal part of the STRF. To obtain error bars on our STRF estimates we divided our spikes randomly into 5 groups. Separate STAs and decorrelated STRFs were calculated from each group of spikes although a uniform cutoff was used for all decorrelations. Error bars shown in the figure represent the standard deviation of these five STRF estimates.

Measurements of STRFs

We measured several parameters of the STRFs to characterize their properties. The latency and best frequency of each STRF were defined as the time and frequency of the maximum of the STRF. To calculate the 50% width in time we looked at the temporal cross-section of the STRF that passed through this best frequency. The 50% width was defined as the interval over which this cross-section had an amplitude greater than or equal to half the maximum amplitude. Cross-sections were linearly interpolated to a resolution of 100kHz before measuring their half-width. The 50% bandwidth was similarly defined as the interval over which the spectral cross-section through the peak was greater than half its maximum. Spectral cross-sections were linearly interpolated to 100 times

their resolution before measuring their half width. Bandwidth measurement were converted to octaves according to:

$$\text{bw50 (octaves)} = \text{samplerate}/\ln(2) * \text{nsamples}$$

Modulation Spectra

Modulation spectra of STRFs were obtained by taking the two dimensional Fourier transform of the two dimensional autocorrelation function of the STRF (Singh and Theunissen, 2003). The x axis of this plot represents the temporal frequencies in the STRF (modulations along its time axis) while the y-axis represents the spectral frequencies (modulations along its spectral axis). The left side of the spectrum represents downward frequencies sweeps while the right side represents upward sweeps. More of our STRFs had modulation spectra that were fairly symmetric around the y axis. We calculated a symmetry index (Miller et al, 2002) by dividing the modulation spectrum into left and right halves (M_{left} and M_{right}, respectively), then computing:

$$\frac{\Sigma M_{right} - \Sigma M_{left}}{\Sigma M_{right} + \Sigma M_{left}}$$

This index ranges from -1 to 1, and is zero for perfectly symmetric STRFs. Our STRFs had a mean symmetry index of 0.02, and a standard deviation of 0.15. Because of this symmetry, we folded the spectra about the y-axis and averaged them before performing other analyses (see Principal Components Analysis, below). The spectral or temporal best modulation frequency (BMF) of a STRF was obtained by averaging this folded spectrum across columns or rows then finding the location of its maximum.

Principal Components Analysis

To describe the distribution of STRF types in our population we performed a principal components analysis on the folded STRF modulation spectra. Very similar results were obtained using the two-sided modulation spectra rather than the folded spectra. To obtain the principal components of the spectra, we normalized all spectra to have a unit norm, vectorized each spectrum,

subtracted the mean vector, and assembled the vectors into a large matrix M_{pop}. We then performed a singular value decomposition on this matrix:

$$U\Sigma V' = M_{pop}$$

The columns of V represent the principal components of the spectra, and could be reshaped into matrices for direct comparison with the modulation spectra. We examined the projections of normalized modulation spectra onto the first two principal components (figure 3C), which together accounted for 14 percent of the variance in the spectra.

The distribution of projections in this two-dimensional space appeared to be a fairly continuous triangle, with STRFs in different regions of this distribution showing very different response properties, physiology, anatomy, and nonlinearities with intensity. Consideration of additional components did not break this continuity. To facilitate our discussion of these differences we loosely classified STRFs by linearly dividing the two-dimensional PCA space (figure 3D). STRFs whose projections satisfied the inequality PC2 < 0.91*PC1-0.25 were classified as "temporal." STRFs whose projections satisfied the inequality PC2 > 28*PC1+9 were classified as spectral, and STRFs that satisfied the two inequalities PC2 > -0.29 + 0.17 and PC2 > 0.91*PC1-0.25 were classified as spectro-temporal. The choice of these inequalities was motivated by examination of STRFs whose projections fell in different regions of the distribution, and by the locations of example STRFs. These classifications are intended only to denote different regions of the distribution, and are not intended to indicate that STRFs cluster into discrete groups.

Bi-Variate Mexican Hat Model

To quantify changes in STRF shape with intensity, we fit each STRF under each stimulus condition to a bi-variate mexican hat model:

$$x' = x - \mu$$
$$y' = y - \nu$$

$$G(x,y) = Ae^{-\sigma^2 x'^2 - \gamma^2 y'^2}(1-\alpha\sigma^2 x'^2)(1-\beta\gamma^2 y'^2)$$

This model has seven parameters: A, the magnitude and sign of the filter; μ, the latency; ν, the center frequency; σ, the temporal width; γ, the spectral width; α, the depth of temporal sidebands, and β, the depth of spectral sidebands. To fit the model, we used a constrained least-squares fitting algorithm based on the Newton method (lsqcurvefit.m, Matlab). To choose starting values for each of the parameters, we first created a matrix of initial parameter values. For A, μ, and ν, there were two initial values, equal to the magnitude, and x- and y-coordinates, of the maximum and minimum point on each STRF. Initial values of α and β were [0 0.25 0.5 1 2 4]. The parameters σ and γ were fit by fitting their inverses 1/σ and 1/γ, to avoid the possibility of having a zero in the denominator of the function. Initial values for these two inverse parameters were [0.1:0.1:0.8].

We calculated the mean squared error between the actual STRF and the model produced by each of the 18432 combinations of initial parameter values. The combination of parameters that produced the lowest error was chosen as the starting point for the algorithm. The algorithm was constrained to find values of μ and ν that fell within the temporal and spectral dimensions of our STRF. The two width parameters σ and γ were constrained to be positive, and the two depth parameters α and β were constrained to be greater than or equal to zero. Confidence intervals for parameters estimates were made using the function nlparci.m in Matlab.

Nonlinearities and Predictions

The nonlinear relationship between each STRF's output and the cell's actual response was calculate following the method of Brenner et al (2001). Predicted responses to repeated stimulus segments were made by first applying the STRF to the spectrogram of this repeated segment, summing across bands, then transforming the output according to the nonlinearity. The first 500 milliseconds of both the predicted and actual PSTH were omitted from the comparison because this often contained a slowly adapting component not predicted by the STRF model. Correlation coefficients between the remaining 4.5 seconds are shown in supplementary figure 1A.

Figure Legends

Figure 1

Construction of a stimulus with a naturalistic distribution of spectral and temporal modulation frequencies.

A) Our stimulus was composed of 32 frequency bands that were gaussian in log-frequency, and overlapped by one standard deviation. Filters for each band are shown in the right-hand column. Each frequency band was modulated by an independent amplitude modulation signal (2nd column) whose log amplitude was gaussian and had a power spectrum that decreased exponentially as a function of temporal frequency. The complete stimulus was formed by summing all the bands. A short segment of the complete 33 minute stimulus is shown as both an oscillogram (top panel) and as a spectrogram (2nd panel) in the left hand column. The spectrogram shows the frequency content of the stimulus as a function of time. Because of the overlapping frequency bands, and the exponential smoothing of the modulating waveforms, nearby frequencies rise and fall together, and remain elevated or depressed for a longer time than in a pure white noise stimulus (bottom panel). These correlated modulations are more similar to those found in natural sounds and drove our neurons more effectively than white noise.

B) Quantitative analysis of correlations in the stimulus. The two-dimensional modulation spectrum of the stimulus (left-hand panel) shows the distribution of energy in the stimulus as a function of slow temporal modulation frequencies (1-250Hz, x-axis) and fast spectral modulation frequencies (0-3 cycles per octave, y-axis). Energy away from the axes represents joint spectro-temporal modulations such as frequency sweeps.

C) Marginal distributions show energy as a function of temporal frequency (top panel) or as a function of spectral frequency (bottom panel). These indicate that energy in the stimulus is concentrated at low temporal and spectral modulation frequencies, but has tails that reach to higher frequencies.

Figure 2

Examples of three common types of spectro-temporal receptive fields (STRFs): temporal, spectro-temporal, and spectral.

A) A cell sensitive to temporal modulations has an STRF (column 1) with positive and negative subfields arranged sequentially in time. This makes it sensitive to high temporal frequencies but low spectral frequencies, reflected in the energy concentrated along the x-axis of its modulation spectrum (column 2). The firing rate of the cell was modulated rapidly by our stimulus, as shown in a PSTH of its responses to 50 repeated stimulus segments (column 3, black). The response predicted by the STRF model (red) captures most but not all of these peaks. The mean driven firing rate of this cell was 116 Hz, and the correlation coefficient between actual and predicted PSTHs was 0.48.

B) A cell sensitive to spectral modulations. Its STRF has a temporally elongated positive region flanked by negative spectral sidebands. The energy in the STRF is concentrated at high spectral frequencies and low temporal frequencies. This cell fired at an average of 10Hz in the presence of the stimulus, but did respond robustly to distinct stimulus features. These sparse bursts were well captured by the STRF model, with a correlation coefficient between actual and predicted PSTHs of 0.56.

C) A cell sensitive to both spectral and temporal modulations has an STRF with a central positive peak flanked by negative regions in both spectrum and time. Its modulation spectrum has energy in two places: away from the axes and along the x-axis. Energy off the axes arises because the cell is sensitive to both high temporal and high spectral modulation frequencies. The energy along the x-axis arises because the very profound spectral sidebands give the cell additional sensitivity to low spectral frequencies (see supplementary figure 1). This cell was also driven very robustly by the stimulus, with a mean driven rate of 104Hz. The correlation between actual and predicted PSTHs was 0.70.

D) A second cell sensitive to both spectral and temporal modulations, but centered at a different best frequency, and with energy concentrated closer to the spectral axis. Mean driven rate was 60Hz and the correlation between actual and predicted PSTHs was 0.71.

Figure 3

The population distribution of preferred modulation frequencies reveals three classes of cells

A) Schematic of the calculation of best temporal and spectral modulation frequencies. Top panel: a modulation spectrum that has been folded along its y-axis and averaged. 2nd panel: marginal distribution of temporal frequencies calculated by averaging the folded spectrum across rows. The best temporal modulation frequency (temporal BMF) is indicated by a red arrow. 3rd panel: marginal distribution of spectral frequencies calculated by averaging the folded spectrum across columns, with the best spectral modulation frequency (spectral BMF) indicated by a red arrow.

B) Distribution of preferred spectral versus preferred temporal modulation frequencies. Most points are distributed along the two axes, corresponding to cells sensitive to pure temporal (x-axis) or pure spectral (y axis) modulations. A third group of cells are sensitive to both types of modulations. Marginal distributions for each axis indicate that cells sensitive spectral (or temporal) modulations sample the range of spectral (or temporal) modulations fairly uniformly. Example cells from figure 2A-D are indicated by green (2A), blue (2B), and red (2D squares), and by a red circle (2C).

Figure 4

A principal components analysis (PCA) of STRF modulation spectra supports the division of our population into three classes.

A) Principal components describing the main axes of variation in the folded modulation spectra. Top panel: first principal component, accounting for 7 percent of the total variance. It resembles the difference between a strongly temporal STRF, with energy along the x axis, and a strongly spectral STRF, with energy along the y axis. Bottom panel: second principal component, accounting for 6 percent of the variance. It resembles a spectro-temporal STRF with energy concentrated away from the origin.

B) Distribution of population projections onto principal components 1 and 2. Colored dots indicate our division of this fairly continuous distribution into spectral (blue), spectro-temporal (red), and

temporal (green) STRFs, based on the magnitude of their projections onto the two principal components, and the locations of example STRFs. White dots indicate STRFs that had low projection values onto both axes and showed low or contradictory tuning. Example cells from figure 2A-D are indicated by green (2A), blue (2B), and red (2D squares), and by a red circle (2C).

C) Cells found in different regions of the PCA plot showed different average latencies. Spectro-temporal cells (red) has the shortest latencies, followed by temporal cells (green). Spectral cells (blue) had both the longest latencies and the widest range of latencies.

D) The asymmetry of the modulation spectrum reflects the degree to which an STRF is selective for oriented time-frequency sweeps. Although most cells of all types showed very little orientation (symmetry values near zero), temporal cells (green) showed the widest distribution of orientations, while spectral cells (blue) showed very little.

Figure 5

STRFs with different temporal response properties arise from cells with distinct physiology

A) Panel 1: An STRF with fast temporal response properties. Panels 2 and 3: raw voltage trace and narrow mean spike waveform of the cell that produced this STRF. Dashed lines in panel 3 indicate the standard deviation of the waveform. Panel 4: inter-spike interval (ISI) distribution for this cell, showing its short 1 msec refractory period. The mean spontaneous firing rate of this cell was 34Hz, while its mean stimulus-driven rate was 104Hz.

B) Equivalent data for an STRF with slow temporal response properties. This cell had a much wider spike, with a rounded and elongated second positive peak (panels 2 and 3). It also had a much longer relative refractory period, with the peak of its ISI distribution occurring at 5 milliseconds. Both the spontaneous and mean driven firing rates were lower for this cell (6 Hz spontaneous, 10Hz average during the stimulus).

C) Top: schematic of temporal half-width measurement. The temporal half-width of an STRF was equal to the duration over which the temporal cross-section through its peak was greater than half of

the peak value. Bottom: schematic of spike width measurement. The width of the mean spike for each cell was measured from its trough to the subsequent peak.

D) Temporal half-width of STRFs versus spontaneous firing rate for all single units that produced significant STRFs. Half-width and spontaneous firing rate were significantly correlated (correlation coefficient = -0.64, p = 1.3e-9).

E) Spontaneous firing rate as a function of projections onto the first two principal components. Spectral cells, at the left hand corner, have lower spontaneous firing rates (< 20Hz) than temporal and spectro-temporal cells.

F) Temporal half-width of STRFs versus spike width from trough to peak for all cells with spike amplitudes between 200 and 600 μV. STRF and spike width were positively correlated (cc = 0.71, p = 4.8e-11), while STRF width and spike amplitude were not (cc = 0.09, p =0.49).

G) Cells with the widest spike waveforms (>350 microseconds) were almost all sensitive to spectral modulations, as indicated by their position on the left of the PCA plot.

Figure 6

STRFs with different temporal response properties are localized to different regions of field L

A) Histological slides showing the regions of field L and the locations of electrode penetrations in a head-fixed mapping experiment. Top panel: slide stained with an antibody to CB1, the cannabinoid receptor, which selectively labels the input area L2 (white arrow). Above the stained area is area L1, below it is area L3. Tracks of four electrodes can be seen crossing the three layers of field L. Electrode 1 is lesioned at 2100 microns; electrode 2 at 2200 microns; and electrode 3 and 1000, 1300, and 2200 microns (pink arrows). Bottom panel: nissl-stained slide adjacent to above showing the lamina that define the borders of the field L complex (white arrow). The input fibers to area L2 appear in the nissl preparation as a lighter-staining diagonal stripe. Terminal lesions on all four electrodes are visible.

B) Temporal best modulation frequency of raw multi-unit spike-triggered averages as a function of recording depth on each of four electrodes. Pink arrows mark the depths of the lesions shown in the top panel of (A). Sites with higher best modulation frequencies are found within a restricted range of depths that is deeper and narrower for the anterior electrodes and shallower and wider for the posterior electrodes. The location of these faster sites corresponds well to the location of the dark-staining area in the CB1 slide, suggesting that these faster sites are localized to area L2.

C-E) Temporal best modulation frequency of multi-unit sites as a function of recording depth for three additional experiments. Each axis represents one penetration with four electrodes; data from different electrodes are plotted on the same graph. Data from different electrodes have been aligned so that zero marks the center of the field L2, as defined anatomically. In each experiment, cells with higher best modulation frequencies are found only at the center, surrounding L2. Examples of STRFs recorded at different depths in the last experiment are shown at the right, and indicate that STRFs in all regions could show strong spectral tuning. Blue dots represent STRFs recorded with a stimulus whose modulation envelopes had power proportional to 1/50Hz. Gray dots represent STRFs from a stimulus with power proportional to 1/20Hz. As we discuss in the text, localization of faster STRFs to L2 can be seen witheither stimulus configuration alone.

F) Examples of raw spike-triggered averages recorded at different depths in the experiment summarized in figure 6E. Spike-triggered averages at all depths show good frequency tuning; those located near the center of field L2 show much narrower temporal tuning than those located above and below them.

Figure 7

The three types of STRFs show distinct changes with stimulus intensity

A-D) STRFs measured from the same cell with two different stimulus intensities. (A) STRF at 30dB. (B) STRF at 63dB. The 63dB STRF has shorter latency, more prominent negative regions, and additional positive regions. (C) Temporal cross sections at 2.7 kHz through the 30dB (blue) and 63dB (red) STRFs. Dashed lines indicate the standard deviation from five jackknife estimates of the STRF.

The 63dB STRF has a significantly shorter latency than the 30dB STRF. (D) Spectral cross sections at -12 msec through both STRFs, colored as above. Negative spectral sidebands are much more prominent in the 63dB STRF. (E) rasters (top two panels) and PSTHs (bottom panel) of the cell responding to same stimulus segment played at 30dB and at 63dB. Although the mean firing rate of the cell was slightly higher at 63dB (104Hz at 63dB versus 76Hz at 30dB), the firing rate is modulated over the same range in both conditions (0-300Hz, bottom panel). Some peaks in the response, like the first two, occur earlier in the response to 63dB, while other peaks differ completely between the two responses.

F) STRFs measured from a temporal cell at 30dB and and 63dB. Top left panel: 30dB STRF. Bottom left panel: 63dB STRF. Top right panel: temporal cross-sections through both STRFs at 3.0 kHz. Bottom right panel: spectral cross sections through both STRFs at -14 msec. This cell shows a dramatic temporal phase change between stimulus intensities.

G) STRFs measured from a spectral cell at 30dB and 63dB. Top left panel: 30dB STRF. Bottom left panel: 63dB STRF. Top right panel: temporal cross-sections through both STRFs at 1.8 kHz. Bottom right panel: spectral cross sections through both STRFs at -32 msec. At 63dB the negative regions of the STRF are a bit deeper, and a second positive peak has appeared at a higher frequency. The 63dB STRF (red) is slightly narrower in time than the 30dB STRF (blue). The error bars on the estimates of spectral STRFs were generally much larger due to their low firing rates. The significance of changes in spectral STRFs is discussed in the text.

Figure 8

A bi-variate mexican hat model can quantify changes in STRF shape with stimulus intensity

A) Schematic of different STRF shapes that can be created from the bivariate mexican hat model by manipulating the parameters alpha and beta on the quadratic terms. Top left: when the alpha and beta are zero, the model reduces to a two-dimensional gaussian. Top right: when the parameter alpha is positive, and beta is zero, the model generates a STRF sensitive to temporal modulations. Bottom left: when beta is positive, and alpha is zero, the model generates a STRF sensitive to spectral modulations. Bottom right: when alpha and beta are both positive, the model generates a STRF

sensitive to both spectral and temporal modulations. The model is thus capable of capturing each of the main STRF shapes seen in our population.

B) Examples of model fits to real data. Top panels: STRFs shown as examples in previous figures. Bottom panels: STRFs generated by model fits to those examples.

C) Values of the parameter beta when fit to 63dB STRFs (y-axis) versus 30dB STRFs (x-axis). Blue squares represent cells that show a significant increase in beta when the stimulus intensity increases from 30dB to 63dB (n=25). Gray circles represent cells that show a significant decrease (n=4). Across the population, values of beta were significantly higher at 63dB than at 30dB ($p = 6.6e-7$), indicating that cells became more sensitive to spectral modulations at the higher intensity.

D) Location of cells showing a significant increase in beta on the PCA plot. These include most of the spectro-temporal cells along the top face of the triangle, and a large fraction of the spectral cells at its left-hand corner. These data indicate that cells sensitive to spectral modulations tend to become more sensitive to spectral modulations at the higher intensity.

E) Values of the parameter alpha fit to 63dB STRFs (y-axis) versus 30dB STRFs (x-axis). Green squares represent cells that showed a significant increase in alpha when the stimulus intensity increases from 30dB to 63dB (n=25), while gray squares represent cells that show a significant decrease in alpha (n=5). Across the population, values of alpha were significantly higher at 63dB than at 30dB ($p = 3.5e-8$), indicating greater sensitivity to temporal modulations at the higher intensity.

F) Location of cells showing a significant increase in alpha (green squares) or a sign inversion (change of phase, pale green circles) in PCA space. Most spectro-temporal cells (top of the triangle), and some temporal cells (right hand face), showed an increase in alpha. Nearly all temporal cells showed either a significant increase in alpha or a sign change. These data indicate that both spectro-temporal and temporal cells show changes in the temporal domain with stimulus intensity.

References:

Barbour, D. L. and X. Wang (2003). "Auditory cortical responses elicited in awake primates by random spectrum stimuli." J Neurosci 23(18): 7194-206.

Brenner, N., W. Bialek, and R. de Ruyter van Steveninck (2000). "Adaptive rescaling maximizes information transmission." Neuron 26(3): 695-702.

Buhl, E. H., T. Szilagyi, et al. (1996). "Physiological properties of anatomically identified basket and bistratified cells in the CA1 area of the rat hippocampus in vitro." Hippocampus 6(3): 294-305.

Buzsaki, G. and E. Eidelberg (1982). "Direct afferent excitation and long-term potentiation of hippocampal interneurons." J Neurophysiol 48(3): 597-607.

Calhoun, B. M. and C. E. Schreiner (1998). "Spectral envelope coding in cat primary auditory cortex: linear and non-linear effects of stimulus characteristics." Eur J Neurosci 10(3): 926-40.

Cheung, S. W., S. S. Nagarajan, et al. (2001). "Auditory cortical neuron response differences under isoflurane versus pentobarbital anesthesia." Hear Res 156(1-2): 115-27.

Chi, T., P. Ru, et al. (2005). "Multiresolution spectrotemporal analysis of complex sounds." J Acoust Soc Am 118(2): 887-906.

Chow, A., A. Erisir, et al. (1999). "K(+) channel expression distinguishes subpopulations of parvalbumin- and somatostatin-containing neocortical interneurons." J Neurosci 19(21): 9332-45.

Creutzfeldt, O., F. C. Hellweg, et al. (1980). "Thalamocortical transformation of responses to complex auditory stimuli." Exp Brain Res 39(1): 87-104.

Depireux, D. A., J. Z. Simon, et al. (2001). "Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex." J Neurophysiol 85(3): 1220-34.

DeWeese, M. R., M. Wehr, et al. (2003). "Binary spiking in auditory cortex." J Neurosci 23(21): 7940-9.

Eggermont, J. J., P. M. Johannesma, and A.M. Aertsen (1983). "Reverse-correlation methods in auditory research." Q Rev Biophys 16(3): 341-414.

Escabi, M. A., L. M. Miller, et al. (2003). "Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus." J Neurosci 23(37): 11489-504.

Fortune, E. S. and D. Margoliash (1992). "Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (Taenopygia guttata)." J Comp Neurol 325(3): 388-404.

Fortune, E. S. and D. Margoliash (1995). "Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (Taeniopygia guttata)." J Comp Neurol 360(3): 413-41.

Frisina, R. D. (2001). "Subcortical neural coding mechanisms for auditory temporal processing." Hear Res 158(1-2): 1-27.

Gehr, D. D., B. Capsius, P. Grabner, M. Gahr, and H.J. Leppelsack (1999). "Functional organisation of the field-L-complex of adult male zebra finches." Neuroreport 10(2): 375-80.

Gentner, T. Q. and D. Margoliash (2003). "Neuronal populations and single cells representing learned auditory objects." Nature 424(6949): 669-74.

Gill, P., J. Zhang, et al. (2006). "Sound representation methods for spectro-temporal receptive field estimation." J Comput Neurosci 21(1): 5-20.

Hasenstaub, A., Y. Shu, et al. (2005). "Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks." Neuron 47(3): 423-35.

Heil, P. and H. Scheich (1985). "Quantitative analysis and two-dimensional reconstruction of the tonotopic organization of the auditory field L in the chick from 2-deoxyglucose data." Exp Brain Res 58(3): 532-43.

Henze, D. A., Z. Borhegyi, et al. (2000). "Intracellular features predicted by extracellular recordings in the hippocampus in vivo." J Neurophysiol 84(1): 390-400.

Hessler, N. A. and A. J. Doupe (1999). "Singing-related neural activity in a dorsal forebrain-basal ganglia circuit of adult zebra finches." J Neurosci 19(23): 10461-81.

Hubel, D. H. and T. N. Wiesel (1962). "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex." J Physiol 160: 106-54.

Hubel, D. H. and T. N. Wiesel (1963). "Shape and arrangement of columns in cat's striate cortex." J Physiol 165: 559-68.

Kowalski, N., D. A. Depireux, et al. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra." J Neurophysiol 76(5): 3524-34.

Kowalski, N., D. A. Depireux, et al. (1996). "Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra." J Neurophysiol 76(5): 3503-23.

Lacaille, J. C., A. L. Mueller, et al. (1987). "Local circuit interactions between oriens/alveus interneurons and CA1 pyramidal cells in hippocampal slices: electrophysiology and morphology." J Neurosci 7(7): 1979-93.

Las, L., E. A. Stern, et al. (2005). "Representation of tone in fluctuating maskers in the ascending auditory system." J Neurosci 25(6): 1503-13.

Lewicki, M. S. and M. Konishi (1995). "Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order." Proc Natl Acad Sci U S A 92(12): 5582-6.

Lewicki, M. S. and B. J. Arthur (1996). "Hierarchical organization of auditory temporal context sensitivity." J Neurosci 16(21): 6987-98.

Liang, L., T. Lu, et al. (2002). "Neural representations of sinusoidal amplitude and frequency modulations in the primary auditory cortex of awake primates." J Neurophysiol 87(5): 2237-61.

Linden, J. F., R. C. Liu, et al. (2003). "Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex." J Neurophysiol 90(4): 2660-75.

Lu, T., L. Liang, et al. (2001). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates." Nat Neurosci 4(11): 1131-8.

Machens, C. K., M. S. Wehr, et al. (2004). "Linearity of cortical receptive fields measured with natural sounds." J Neurosci 24(5): 1089-100.

Martina, M. and P. Jonas (1997). "Functional differences in Na+ channel gating between fast-spiking interneurones and principal neurones of rat hippocampus." J Physiol 505 (Pt 3): 593-603.

Martina, M., J. H. Schultz, et al. (1998). "Functional and molecular differences between voltage-gated K+ channels of fast-spiking interneurons and pyramidal neurons of rat hippocampus." J Neurosci 18(20): 8111-25.

Miller, L. M., M. A. Escabi, et al. (2002). "Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex." J Neurophysiol 87(1): 516-27.

Muller, C. M. and H. J. Leppelsack (1985). "Feature extraction and tonotopic organization in the avian auditory forebrain." Exp Brain Res 59(3): 587-99.

Nagel, K. I. and A. J. Doupe (2006). "Temporal processing and adaptation in the songbird auditory forebrain." Neuron 51(6): 845-59.

Nelken, I., A. Fishbach, et al. (2003). "Primary auditory cortex of cats: feature detection or something else?" Biol Cybern 89(5): 397-406.

Nelken, I., J. K. Bizley, et al. (2004). "Large-scale organization of ferret auditory cortex revealed using continuous acquisition of intrinsic optical signals." J Neurophysiol 92(4): 2574-88.

Oertel, D. (1991). "The role of intrinsic neuronal properties in the encoding of auditory information in the cochlear nuclei." Curr Opin Neurobiol 1(2): 221-8.

Phillips, D. P. and D. R. Irvine (1981). "Responses of single neurons in physiologically defined primary auditory cortex (AI) of the cat: frequency tuning and responses to intensity." J Neurophysiol 45(1): 48-58.

Phillips, D. P. and S. E. Hall (1987). "Responses of single neurons in cat auditory cortex to time-varying stimuli: linear amplitude modulations." Exp Brain Res 67(3): 479-92.

Phillips, D. P., M. N. Semple, M.B. Calford, and L.M. Kitzes (1994). "Level-dependent representation of stimulus frequency in cat primary auditory cortex." Exp Brain Res 102(2): 210-26.

Pinaud, R., T. A. Velho, et al. (2004). "GABAergic neurons participate in the brain's response to birdsong auditory stimulation." Eur J Neurosci 20(5): 1318-30.

Pinaud, R. and C. V. Mello (2007). "GABA immunoreactivity in auditory and song control brain areas of zebra finches." J Chem Neuroanat.

Polley, D. B., M. A. Heiser, et al. (2004). "Associative learning shapes the neural code for stimulus magnitude in primary auditory cortex." Proc Natl Acad Sci U S A 101(46): 16351-6.

Qiu, A., C. E. Schreiner, et al. (2003). "Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition." J Neurophysiol 90(1): 456-76.

Rhode, W. S., D. Oertel, et al. (1983). "Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus." J Comp Neurol 213(4): 448-63.

Rhode, W. S., P. H. Smith, et al. (1983). "Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat dorsal cochlear nucleus." J Comp Neurol 213(4): 426-47.

Rhode, W. S. and P. H. Smith (1986). "Encoding timing and intensity in the ventral cochlear nucleus of the cat." J Neurophysiol 56(2): 261-86.

Ringach, D. L., M. J. Hawken, et al. (2002). "Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences." J Vis 2(1): 12-24.

Rust, N. C., O. Schwartz, et al. (2005). "Spatiotemporal elements of macaque v1 receptive fields." Neuron 46(6): 945-56.

Scharfman, H. E. (1995). "Electrophysiological diversity of pyramidal-shaped neurons at the granule cell layer/hilus border of the rat dentate gyrus recorded in vitro." Hippocampus 5(4): 287-305.

Scheich, H., B. A. Bonke, D. Bonke, and G. Langner (1979). "Functional organization of some auditory nuclei in the guinea fowl demonstrated by the 2-deoxyglucose technique." Cell Tissue Res 204(1): 17-27.

Schreiner, C. E., J. R. Mendelson, et al. (1992). "Functional topography of cat primary auditory cortex: representation of tone intensity." Exp Brain Res 92(1): 105-22.

Schreiner, C. E. and M. L. Sutter (1992). "Topography of excitatory bandwidth in cat primary auditory cortex: single-neuron versus multiple-neuron recordings." J Neurophysiol 68(5): 1487-502.

Schreiner, C. E., J. Mendelson, et al. (1997). "Temporal processing in cat primary auditory cortex." Acta Otolaryngol Suppl 532: 54-60.

Sen, K., F. E. Theunissen, et al. (2001). "Feature analysis of natural sounds in the songbird auditory forebrain." J Neurophysiol 86(3): 1445-58.

Sharpee, T., N. C. Rust, et al. (2004). "Analyzing neural responses to natural signals: maximally informative dimensions." Neural Comput 16(2): 223-50.

Sharpee, T. O., H. Sugihara, et al. (2006). "Adaptive filtering enhances information transmission in visual cortex." Nature 439(7079): 936-42.

Singh, N. C. and F. E. Theunissen (2003). "Modulation spectra of natural sounds and ethological theories of auditory processing." J Acoust Soc Am 114(6 Pt 1): 3394-411.

Smith, P. H. and W. S. Rhode (1987). "Characterization of HRP-labeled globular bushy cells in the cat anteroventral cochlear nucleus." J Comp Neurol 266(3): 360-75.

Smith, P. H., P. X. Joris, et al. (1991). "Projections of physiologically characterized globular bushy cell axons from the cochlear nucleus of the cat." J Comp Neurol 304(3): 387-407.

Soderstrom, K., Tian, Q., Valenti, M., and Di Marzo, V. (2004) "Endocannabinoids link feeding stats and auditory perception-related gene expression." J Neurosci 24(44):10013-21.

Spirou, G. A., W. E. Brownell, et al. (1990). "Recordings from cat trapezoid body and HRP labeling of globular bushy cell axons." J Neurophysiol 63(5): 1169-90.

Sullivan, W. E. and M. Konishi (1984). "Segregation of stimulus phase and intensity coding in the cochlear nucleus of the barn owl." J Neurosci 4(7): 1787-99.

Sullivan, W. E. (1985). "Classification of response patterns in cochlear nucleus of barn owl: correlation with functional response properties." J Neurophysiol 53(1): 201-16.

Sutter, M. L. and C. E. Schreiner (1995). "Topography of intensity tuning in cat primary auditory cortex: single-neuron versus multiple-neuron recordings." J Neurophysiol 73(1): 190-204.

Takahashi, T., A. Moiseff, et al. (1984). "Time and intensity cues are processed independently in the auditory system of the owl." J Neurosci 4(7): 1781-6.

Terleph, T. A., C. V. Mello, et al. (2006). "Auditory topography and temporal response dynamics of canary caudal telencephalon." J Neurobiol 66(3): 281-92.

Theunissen, F. E. and A. J. Doupe (1998). "Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches." J Neurosci 18(10): 3786-802.

Theunissen, F. E., K. Sen, et al. (2000). "Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds." J Neurosci 20(6): 2315-31.

Touryan, J., B. Lau, et al. (2002). "Isolation of relevant visual features from random stimuli for cortical complex cells." J Neurosci 22(24): 10811-8.

Touryan, J., G. Felsen, et al. (2005). "Spatial structure of complex cell receptive fields measured with natural images." Neuron 45(5): 781-91.

Ulanovsky, N., L. Las, et al. (2004). "Multiple time scales of adaptation in auditory cortex neurons." J Neurosci 24(46): 10440-53.

Wang, X., T. Lu, et al. (2005). "Sustained firing in auditory cortex evoked by preferred stimuli." Nature 435(7040): 341-6.

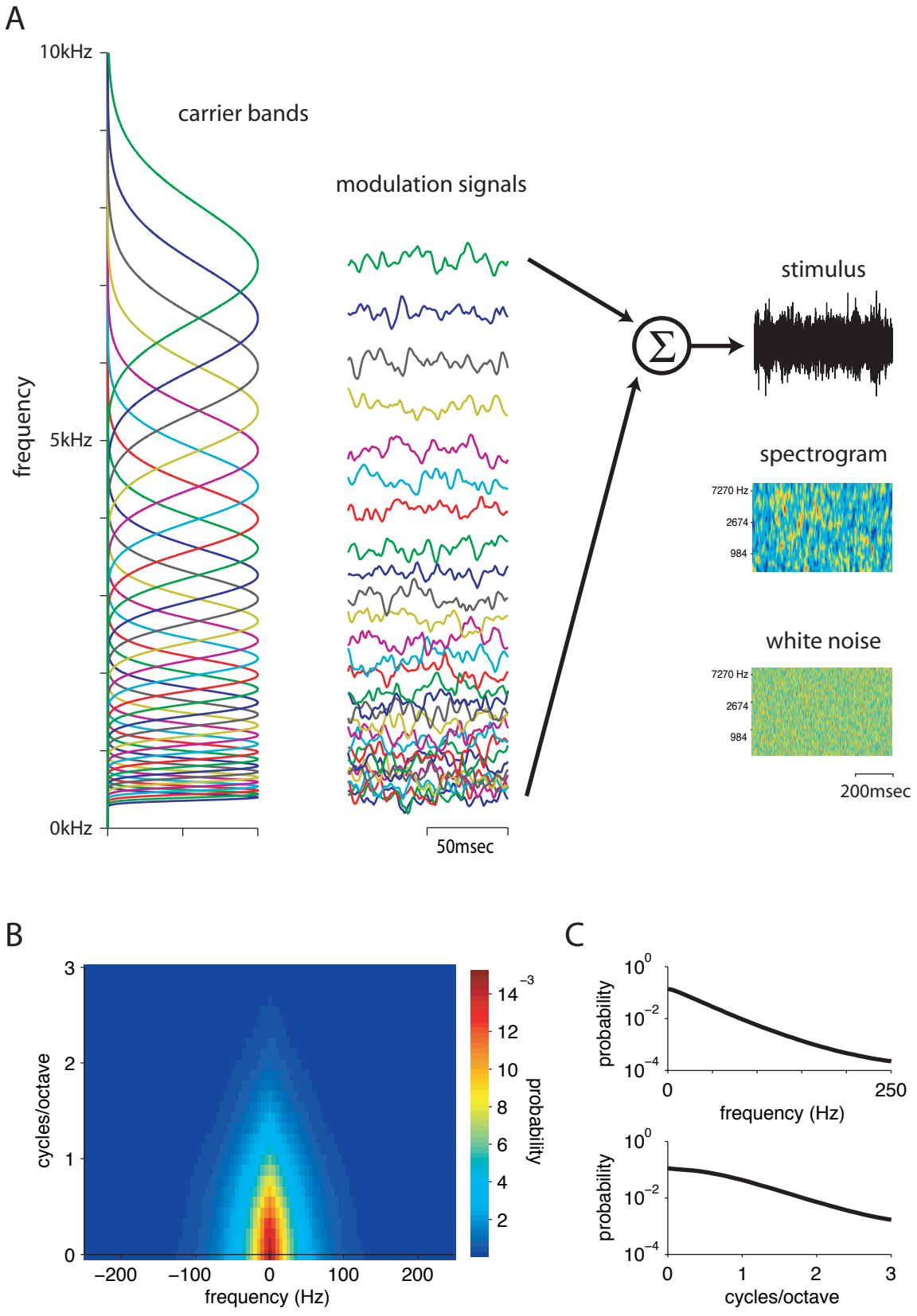
Wehr, M. and A. M. Zador (2003). "Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex." Nature 426(6965): 442-6.

Wiesel, T. N. and D. H. Hubel (1974). "Ordered arrangement of orientation columns in monkeys lacking visual experience." J Comp Neurol 158(3): 307-18.

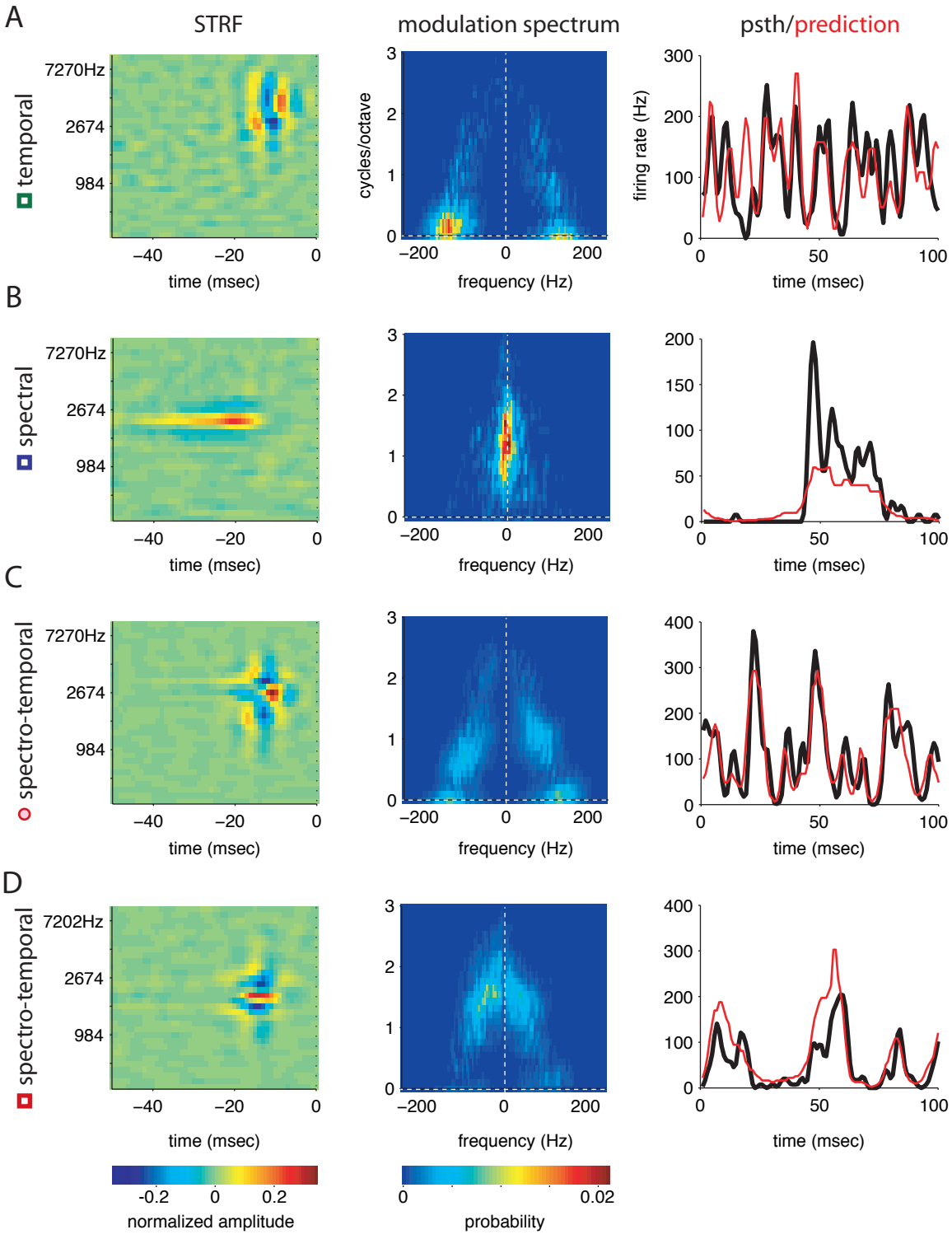
Wild, J. M., H. J. Karten, and B.J. Forst (1993). "Connections of the auditory forebrain in the pigeon (Columba livia)." J Comp Neurol 337(1): 32-62.

Woolley, S. M., T. E. Fremouw, et al. (2005). "Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds." Nat Neurosci 8(10): 1371-9.

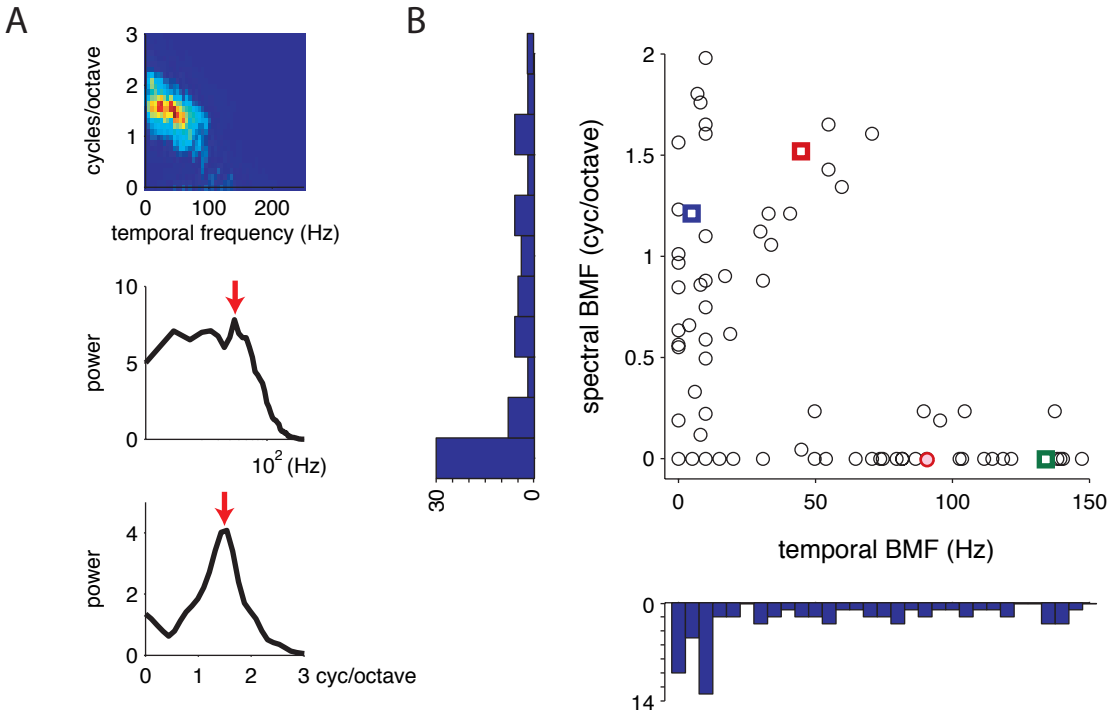
Woolley, S. M., P. R. Gill, et al. (2006). "Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain." J Neurosci 26(9): 2499-512.



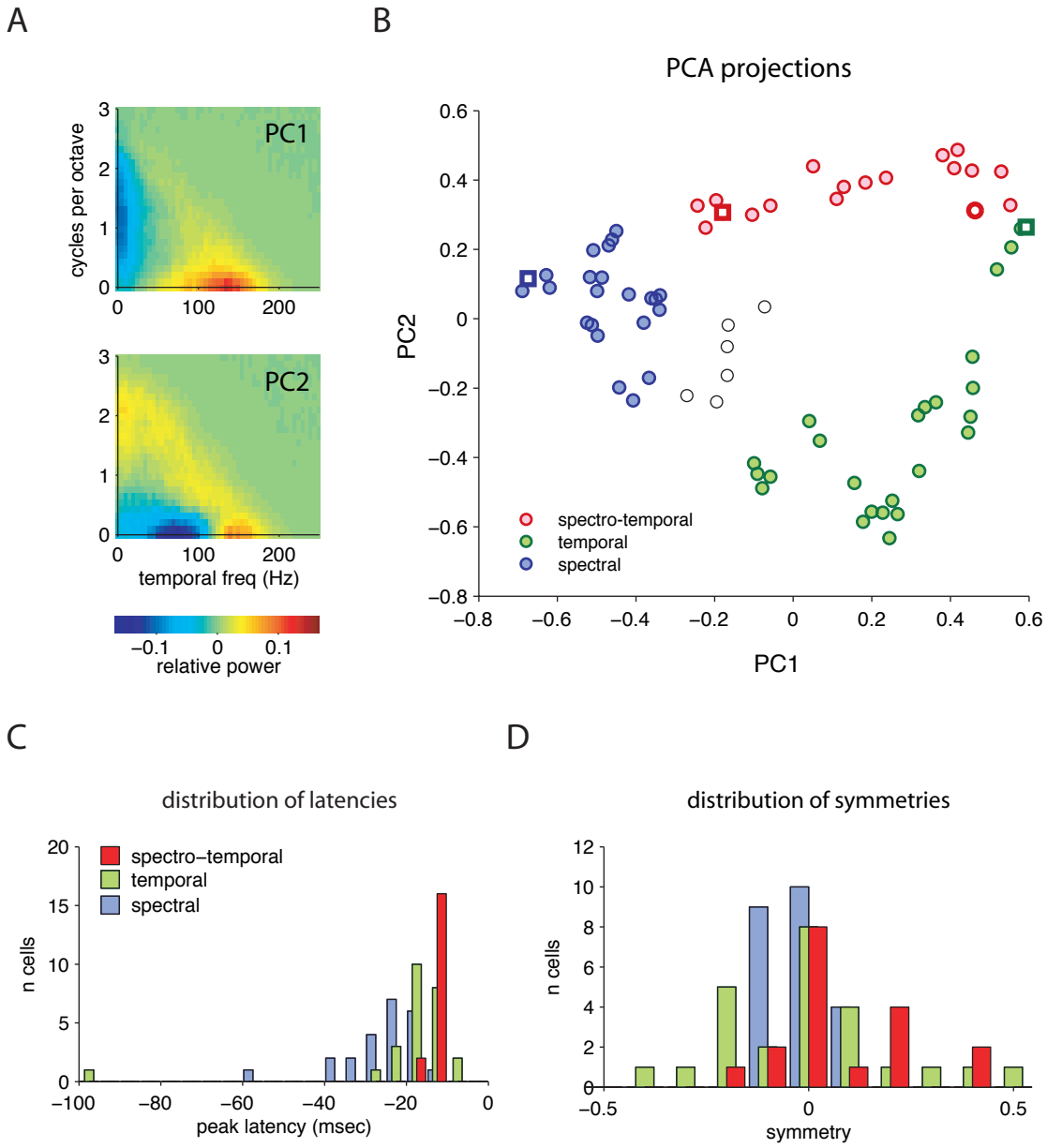
chapter 2, figure 1: stimulus construction



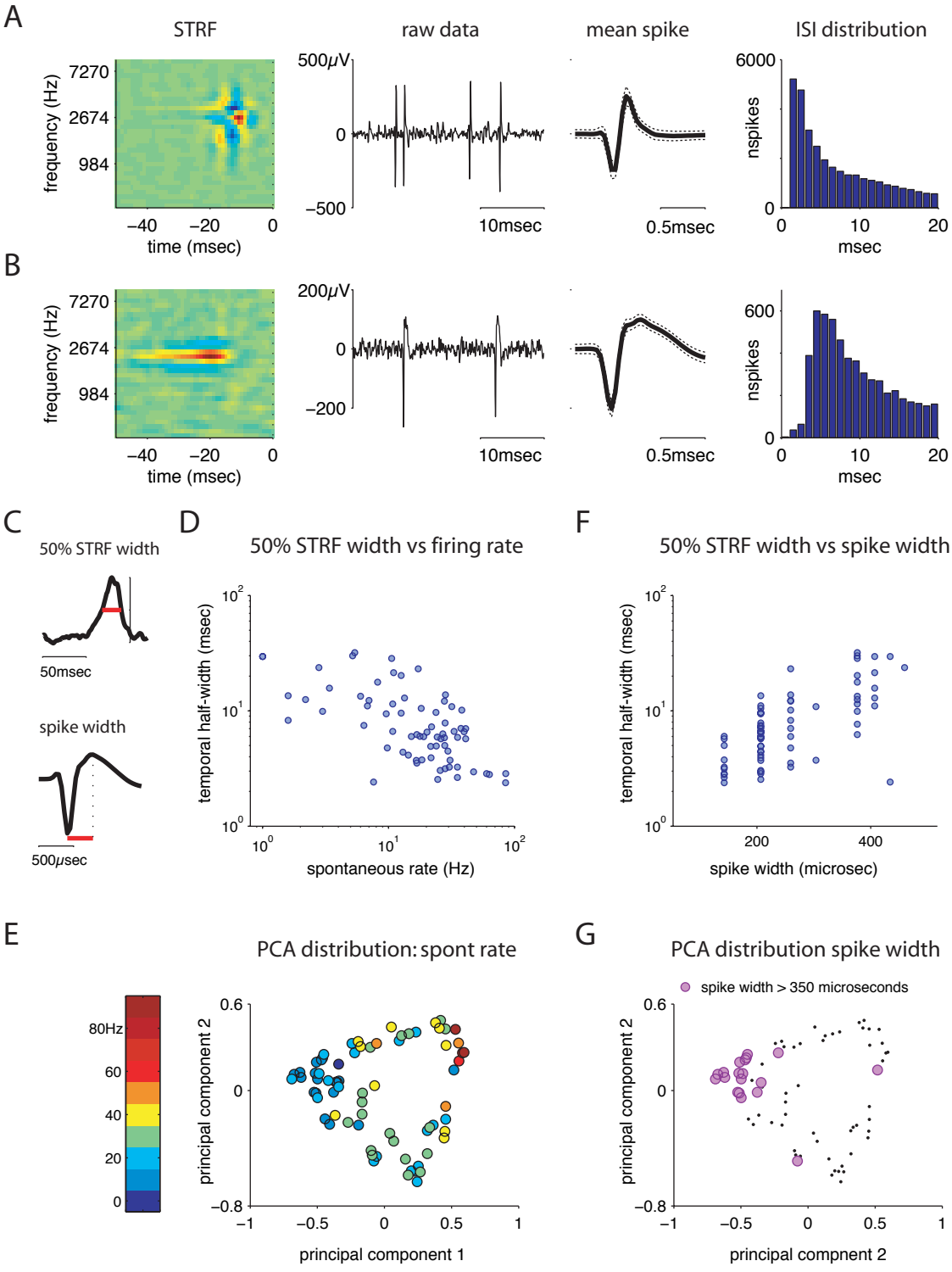
chapter 2, figure 2: examples of STRFs



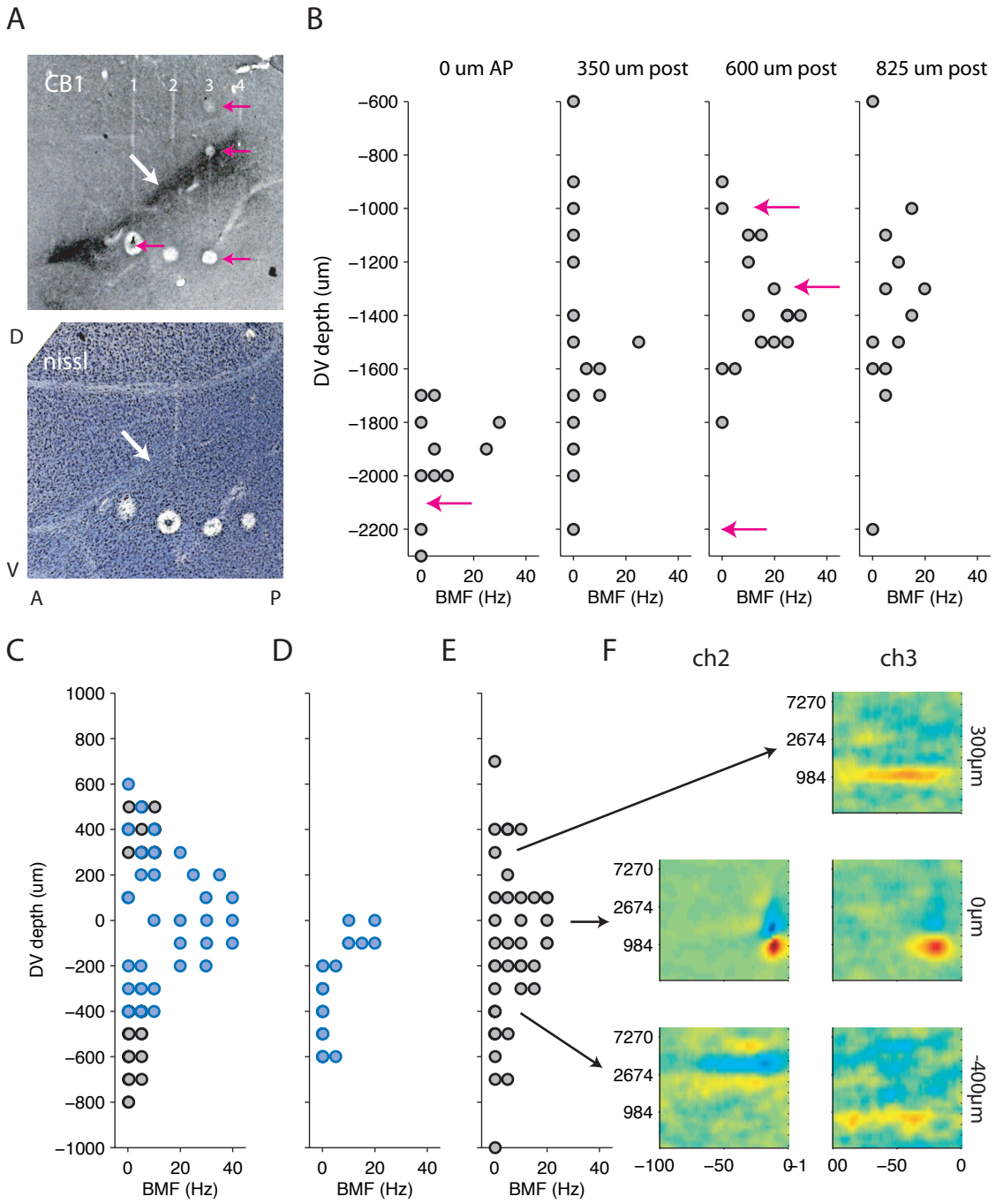
chapter 2, figure 3: distribution of best modulation frequencies



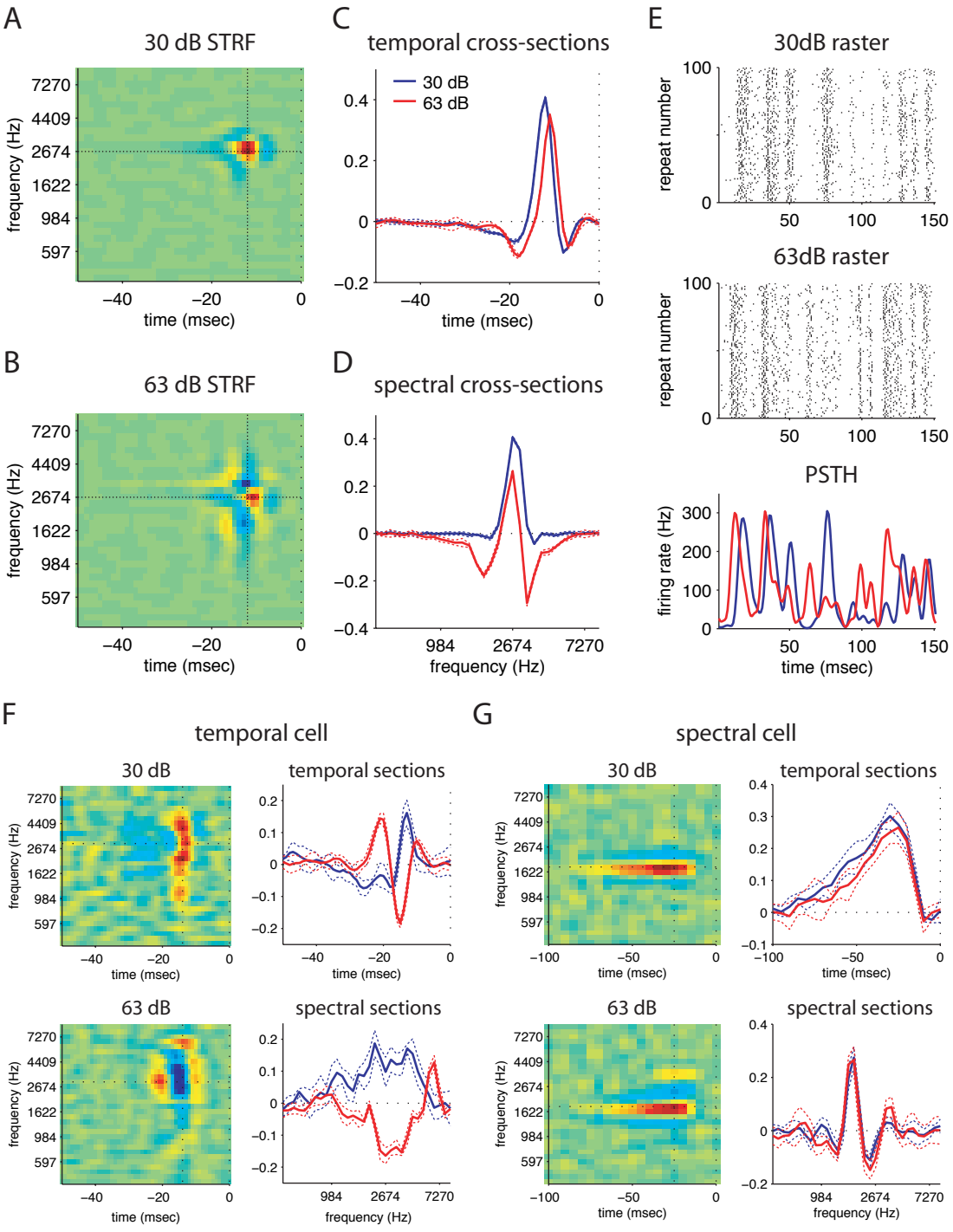
chapter 2, figure 4: principal components analysis of STRF modulation spectra



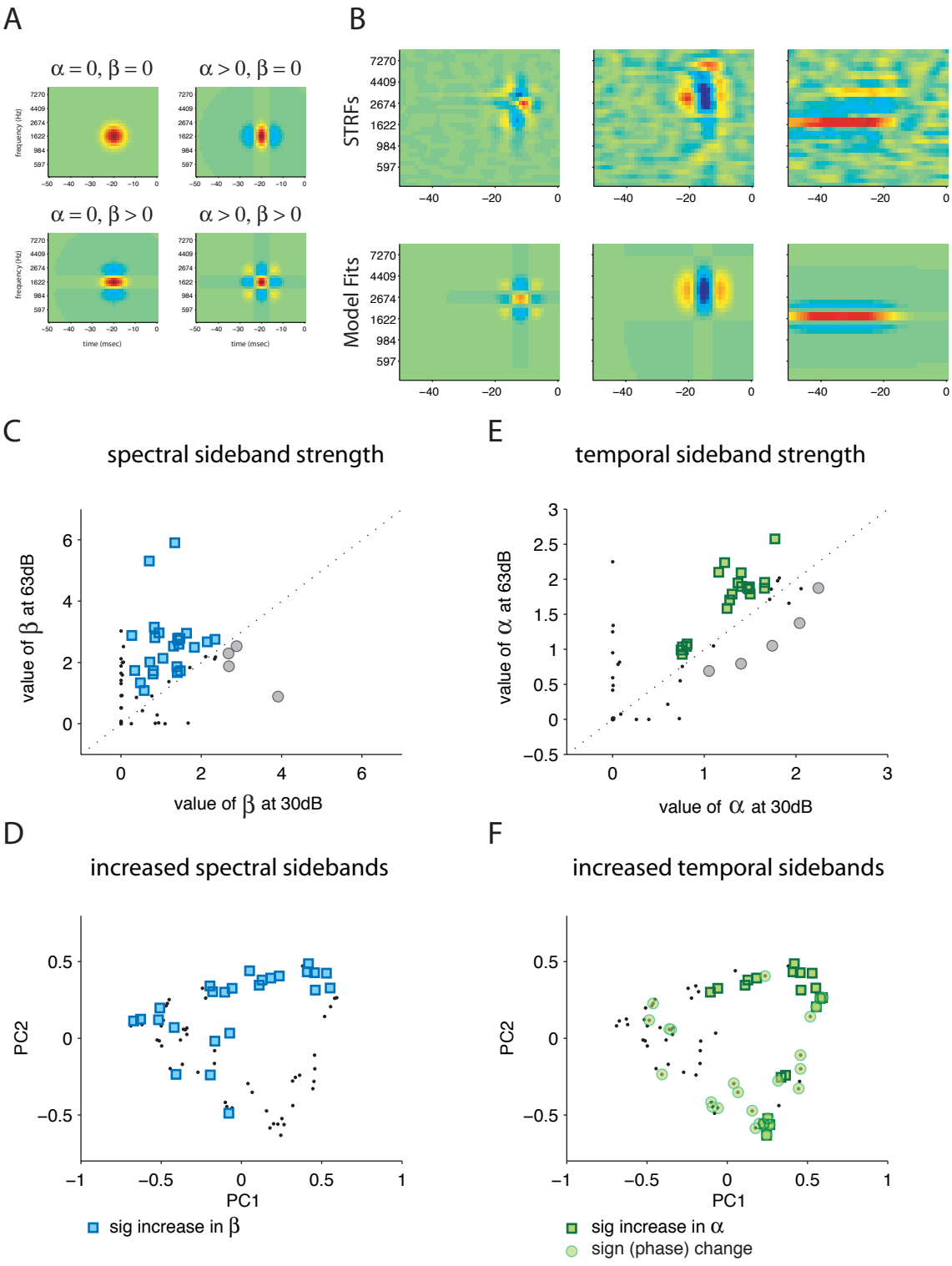
chapter 2, figure 5: correlations between STRF width and cellular physiology 98



chapter 2, figure 6: correlations between STRF width and anatomy



chapter 2, figure 7: examples of STRF dependence on stimulus intensity



chapter 2, figure 8: population analysis of intensity nonlinearities

Chapter 3: Acoustic Parameters Underlying Generalization of a Learned Song Discrimination Task

Abstract:

Songbirds are often used to study the neural mechanisms underlying learning and perception of complex sounds, yet the behavioral relevance of different song parameters is unknown. In this study we used an operant discrimination task to examine the role of various acoustic parameters in song discrimination. Female zebra finches were first trained to discriminate songs from two different individuals. We then asked whether they could generalize what they had learned to songs that were altered in some way. We found that pitch cues played a much more important role than duration cues in our behavioral task: females generalized poorly to songs whose pitch fell outside the range of natural variation, but generalized correctly to songs with very extended or compressed durations. Females could learn to correctly discriminate between songs over a wide range of volumes (13-73dB rms). However, when trained at a single intensity, they could only generalize to intensities within 10-20dB of the training volume. These results place important constraints on how perceptual behavior may be related to neural responses in the primary auditory area field L, suggesting that songs are identified by the spatial pattern of neurons they activate, rather than the temporal pattern of firing they elicit.

Introduction:

Songbirds, like humans, communicate using complex learned acoustic signals, and provide a rich biological model for the neural mechanisms underlying the perception and recognition of complex sounds (Doupe and Kuhl, 1999). In the zebra finch, males learn to produce a single sequence of complex syllables termed a motif (Zahn, 1996). Females do not sing, but are able to recognize individuals--such as their mates-- from their songs alone (Miller, 1979; Stripling et al, 2003; Woolley and Doupe, submitted).

Bird songs, like human words, are high-dimensional stimuli, with structure that can be parameterized in many ways. Yet each song rendition is unique: its volume will depend on the distance between singer and listener; background noises--often other songs-- will corrupt the signal (Zahn, 1996); and the pitch and duration of individual notes vary slightly across renditions (Kao and Brainard, 2006; Glaze and Troyer, 2006). To recognize a bird by his song, then, females must be able

both to discriminate between the songs of different individuals, and to generalize across variations in a single bird's song.

Electrophysiological studies have begun to clarify what features of complex song stimuli are encoded in different regions of the songbird auditory system (Theunissen, et al, 2000; Sen et al, 2001, Woolley et al, 2005; Nagel and Doupe, 2006, in preparation), and to identify areas that respond selectively to learned songs (Gentner and Margoliash, 2003). Song-induced expression of the immediate early gene zenk has revealed brain areas that are selectively activated by novel conspecific songs (Mello et al, 1992, Mello and Clayton, 1994; Stripling et al, 2001; Bailey et al, 2002; Gentner et al, 2004). Lesion studies have implicated many areas that may be necessary for recognition learning (Brenowitz, 1991; MacDougall-Schackleton et al, 1998; Scharff and Nottebohm, 1998; Gentner et al, 2000). Yet the behavioral limits of song recognition have received comparatively little attention. How much variation in pitch or duration can birds tolerate before a song becomes unrecognizable? What range of song volumes are birds able to discriminate correctly? These questions are critical to understanding how the representation of song features or parameters by neurons in the zebra finch brain are related to their perception of song.

We used generalization of a learned discrimination, or "discrimination transfer" (Beecher et al, 1994; Gentner and Hulse, 1998) as a model paradigm for studying song recognition. In this operant task, a bird is taught to categorize a large number of stimuli, then tested on novel probe stimuli. Probe stimuli are rewarded without respect to the bird's choice, ensuring that the bird cannot not learn the "correct" answer for these stimuli from the reward pattern. Performance significantly above chance on a probe stimulus is considered evidence for generalization to that probe. This method has several specific advantages: it allows one to present many test stimuli in a short period of time, and it makes it possible to distinguish incorrect responses (which signify a failure to generalize) from lack of response (which could signify tiredness, indifference, or lack of attention).

To investigate the acoustic parameters underlying learned song recognition, we trained female zebra finches to discriminate a set of one male's songs from those of a second male, then tested their performance on systematically altered songs. We found that female zebra finches showed little tolerance for variations in song pitch, generalizing only to stimuli that were close to the natural range of pitch variation found in zebra finch song. In contrast, females showed great tolerance for changes in song and syllable duration, responding well above chance to a 2-fold range of syllable lengths. Finally we investigated the females' ability both to discriminate songs at different volumes, and to generalize a discrimination learned at one volume to another. Females were able to

discriminate songs over a wide range of volumes, from 23-73dB. However, when trained at a single volume, they generalized only to volumes within 10dB of the training set. These results place important constraints on what aspects of neural responses to song may be important for behavioral recognition.

Methods:

Behavioral task and training:

We trained adult female zebra finches (n=8) to discriminate the songs of two unfamiliar conspecific males using a two-alternative forced choice paradigm. Training took place in custom-built operant cages (Lazlo Bocskai, Ken McGary) and was controlled using a custom-written Matlab program that interfaced with a TDT board.

A discrimination trial began when the female hopped on a central "song perch" that faced a speaker (figure 1A). Following this hop, the computer selected a stimulus from a database of songs from two individuals, described in detail below. The female had to decide which male the song belonged to, and to report her choice by hopping on one of two response perches, located to the right (for male A) and to the left (for male B) of the song perch. If her response was correct, a feeder located under the song perch was raised for 2-5 seconds, allowing her access to food. If it was incorrect no new trials could be initiated for a time-out period of 20-30 seconds. Females had up to 6 seconds after song playback ended to make a response and were allowed to respond at any point during song playback; playback was halted as soon as a response was made. If no response was made within 6 seconds of the song ending, the trial was scored as having no response, and a new trial could be initiated by hopping again on the song perch.

Birds required one to two weeks of training to acquire this task. Training consisted of 4 stages. In "song mode," the female was habituated to the operant cage. Food was freely available. Hopping on the song perch produced a song drawn at random from a separate training database, that contained no songs from the two individuals used for discrimination trials. In "food mode," food was restricted, and the female had to hop on one of the two response perches to raise the feeder. The song perch continued to perform the same function. Food and song could be procured independently. In "sequence mode," the female had to hop first on the song perch, then on either of the two side perches within 6 seconds to receive a reward. Finally, the female began discrimination

trials, the training song set was replaced with the discrimination database, and trials were rewarded (with food) or punished (with a time-out) depending on which response perch the female chose.

Birds spent one or two days on each of the first three training stages (song, food, and sequence). Birds were moved to the next stage of training when they performed more than 200 hops per day (for song mode), or earned more than 200 rewards per day (for food and sequence mode). Females generally learned the discrimination task (greater than 75% correct) within 4-5 days of beginning discrimination trials (Figure 2A and 2C). One female did not perform significantly above chance after three weeks of training and was removed from the study.

Females occasionally developed a bias, in which they classified most of one individual's songs correctly (>90% correct), but performed more poorly (70-90% correct) on songs from the other individual. When this happened, we altered the percent of correct trials that received a reward so that the preferred perch was rewarded at a lower rate. This procedure was usually successful at correcting large biases in performance, and reward rates were equalized over 1-3 subsequent days. Data from days on which the two perches were rewarded at different rates were not used for analysis. In addition, the computer automatically adjusted the rate of song presentation to present songs associated with the non-preferred perch. At the beginning of a session, the computer formed a list of all songs in the database. When a song was played and got a response, it was removed from this list. Songs that did not receive a response remained on the list. Songs were drawn from this list until all songs in the database had elicited a response. At the end of such a "block" the computer calculated the percent correct for "A" songs and "B songs. If these percentages were unequal, it removed A or B songs from the next block so that the percent of A (or B) songs presented was inversely proportional to the fraction of correct responses to A (or B) songs:

P(present A) = P(correct on B)/(P(correct on A)+P(correct on B))

According to this formula, if the bird made the same number of correct responses to A and B songs, both songs would be presented equally. However, if she got only one song correct (for example, if she chose the A perch on all trials and so got all A's correct and all B's incorrect), the computer would adjust song presentation rates so that mostly songs associated with the non-preferred perch were presented. Together with manual adjustment of reward rates, this adaptive song presentation scheme helped ensure that females correctly classified both sets of songs and did not display too strong a bias towards one perch or the other.

Females were allowed to work continuously for the duration of their 14 hour day. In general, females received all of their food by performing this task. However, bird weight was monitored closely, and birds received supplemental food when their weight decreased. Birds had free access to water throughout the experiment. After birds reached criterion in their discrimination behavior, the overall reward rate for correct trials was lowered to 75-95%. This generally increased the number of trials the birds performed each day. All experimental procedures were approved by UCSF's Institutional Animal Care and Use Committee.

Probe Trials:

To test the trained females ability to generalize the learned discrimination to new and altered stimuli, we used a probe trial configuration. Probe trials were randomly interspersed between training trials at a low rate (10-15% of trials), and were rewarded at a fixed rate equal to the overall reward rate (75-95%) regardless of which response perch the female chose. This ensured that no information about the probe stimulus identity was contained in its reward rate, and that females could not easily discriminate probe trials from normal trials on the basis of reward. Probe trials that did not receive a response were not rewarded and remained on the probe stimulus list until they elicited a response.

Song stimuli:

The training database consisted of 100 songs, 50 recordings of one male's song ("male A"), and 50 recordings of another ("male B", figure 1B). We used this large database to ensure that females learned to associate each perch with an individual, and not with a single rendition of his song. The song renditions were chosen to have similar distributions of overall duration, RMS volume, and overall power spectrum across the two individuals (figure 1B, lower panels), but had quite distinct syllable structure and timing (figure 1B, upper panels). Human listeners could easily discriminate songs from one individual from those of the other. Songs contained 1 to 6 motifs and were drawn from both directed (sung to a female) and undirected (sung alone) recording sessions. When necessary, low levels of white noise were added to the recordings to ensure that the background noise level prior to song playback was the same for all stimuli.

To test that females had learned to discriminate individuals, rather than memorizing the classification of all 100 songs in the database, we tested their responses to 10 additional song renditions that were not included in the original training set.

Song Manipulations:

The pitch and duration of individual zebra finch notes are known to vary between song renditions. Analyses in other papers indicate that pitch can vary by 1-3% (Kao), while duration can vary by 1-4% (Troyer). We used a phase vocoder algorithm () to independently manipulate the overall pitch and duration of training songs. To alter the duration of a song without changing its pitch, the vocoder broke it into short overlapping segments, and took the Fourier transform of each segment, yielding a magnitude and phase for each frequency. We used segments of 256 samples, or 10.5 msec at 25kHz, overlapping by 128 samples. This produced a function of time and frequency similar to the spectrogram (such as those shown in the upper panels of figure 1B) often used to visualize an auditory stimulus but with a phase value as well as a magnitude for each point

$$S(t,f) = (A,phi)$$

We then chose a new time scale for the song, and interpolated the magnitude and phase of the signal at the new time points based on the magnitude and phase at nearby times.

$$S'(t',f) = (A',phi')$$

Finally, the vocoder reassembled the stretched or compressed song from the interpolated spectrogram. This process yielded a song that was slower or faster than the original but whose frequency information was largely unchanged.

To alter the pitch of a song without changing its duration, we first used the above algorithm to stretch or compress the song in time. We then resampled it back to its original length yielding a song with a higher (for stretched songs) or lower (for compressed songs) pitch but the same duration as the original. We made songs with durations and pitches increased and decreased by 0, 1, 2, 4, 8, 16, 32, and 64%. Pitch-shifted and time-dilated songs were generated from 20 different base stimuli: 10 from the A training set and 10 from the B training stimuli. We computed the averaged responses to

each manipulation of the 20 base stimuli to calculate performance as a function of pitch shift or duration change. Songs with 0% change were presented as probe stimuli to ensure that nothing about our probe stimulus presentation caused the birds to respond differently.

The loudness or volume of song stimuli was adjusted by scaling the digital representation of the song before outputting it to the speaker. White noise was added to song by creating a random noise vector in Matlab, scaling it to have a specified RMS volume, then adding it to the original song stimulus. Training songs had a distribution of volumes centered around 73dB SPL, as illustrated in figure 1B. In pitch and duration experiments, all songs, including training and probe stimuli, were set to have an RMS volume of 63dB.

Three experiments looked at discrimination and generalization as an explicit function of song volume or noise level. To look at discrimination performance as a function of song volume (figure 6), we created a stimulus set in which 10 A training songs and 10 B training songs were set to 8 volumes (3 to 73dB in steps of 10). To look at discrimination performance as a function of noise level, we set all 100 training stimuli to two volumes (33dB and 63dB) with varying levels of background noise (0-50dB for 33dB songs, 0-80dB for 63dB songs). Noise was generated de novo on each trial insuring that birds never heard the same noise signal twice. In both of these two experiments, all stimuli were presented as "discrimination" trials, meaning that birds were rewarded for getting the answer right, and punished with a time-out for getting the answer wrong. They could thus learn the correct association for each stimulus from the reward patterns.

In a third experiment (figure 8), birds were trained on 100 stimuli with an average intensity of 73 or 30dB. We then used probe trials to test whether they could generalize to songs at other volumes (3-73dB). Probe stimuli were drawn from the same stimulus set used for discrimination training in figure 6, and added an additional silent probe stimulus to verify that birds could not perform the discrimination when there was no signal.

Data Analysis:

Each trial initiated by hopping on the song perch could be scored as correct, incorrect, or no response. The percent of correct responses ("percent correct") was calculated by dividing the total number of correct responses, by the total number of trials to which the bird responded. The percent of responses ("percent response") was calculated by dividing the total number of responses by the

total number of trials. Reaction time was calculated from the song perch hop to the subsequent response hop. 95% confidence intervals on percent correct were calculated by fitting the total number of correct responses, and the total number of responses, to a binomial model (binofit.m, Matlab). The same algorithm was used to estimate confidence intervals on percent response. Error bars on reaction time measurements represent the mean standard deviation across several days. For estimates of behavior on control trials, we show the mean and standard deviation of percent correct and percent response across days. We use this estimate rather than the binomial estimate because there are so many more control trials (9 control trials for every probe) that the binomial estimate of error on control trials is exceedingly small.

Results:

Training and Generalization:

8 out of 9 females trained on the discrimination acquired the task within 4-5 days of beginning discrimination trials (figures 2A and 2C). Performance was very good, with 79 to 95 percent of songs classified correctly. Mean response times ranged from 1.66 to 2.38 seconds (figures 2B and 2D), long enough to hear 1-2 motifs of each song but generally before the song bout ended.

To test that females had learned to discriminate individuals, rather than memorizing the classification of all 100 songs in the database, we tested the responses of four females to 10 additional song renditions that were not included in the original training set. These additional song renditions were presented as probe stimuli: they occurred randomly on 10% of trials, and were rewarded if the bird responded, regardless of which perch she chose. Responses to probe songs were statistically indistinguishable from responses to training songs, in percent of correct responses (figure 2E, 1st panel), percent of songs responded to (figure2E, 2nd panel), and reaction time (figure 2E, 3rd panel). These data indicate that discrimination was robust to trial-to-trial variations in song structure, and suggest that these females had learned to classify songs from an individual, rather than the exact song variants used in the training.

Pitch and Duration:

The pitch and duration of individual notes vary across song renditions: pitch by 1-3% of the note's fundamental frequency (Kao and Doupe), duration by 1-4% (Troyer, Kao). Females identifying a male by his song must therefore be robust to small variations in the pitch and duration of his song.

To examine the importance of pitch for identifying a song, we tested trained females on probe songs that were shifted in pitch (see Methods). As illustrated in figure 3A, the frequency content of these stimuli was altered but their temporal pattern--including the sequence of harmonic and noisy notes-- was not. As a result, human listeners could readily distinguish strongly pitch-shifted songs of one individual from those of the other. If females responded correctly to strongly pitch-shifted songs, it would indicate that pitch is not an important parameter in determining song identity. Incorrect responses would indicate that pitch plays an important role. Probe stimuli were constructed from 10 song variants used in training. Each song variant was presented at 14 different shifts (+/- 0, 0.01, 0.03, 0.06, 0.11, 0.21, 0.40, and 0.71 octaves, equal to changes of 0, 1, 2, 4, 8, 16, 32, and 64%), designed to sample a range of shifts inside and outside the range of natural variability. Responses to different variants were pooled to obtain responses as a function of shift.

Figure 3B (top panel) shows one female's response to pitch-shifted probe songs. The blue bar represents the mean +/- one standard deviation across days of her performance on control discrimination trials. For small pitch shifts less than 0.5 octaves, her performance on probes overlapped with her control performance. However, her performance fell off rapidly for shifts larger than those naturally observed (+/-3%, indicated by dashed lines), and she performed at chance for pitch shifts of 0.2 octaves or greater. The lower panels of figure 3B show this bird's percent response (lower left panel) and reaction time for probes (black lines) and control training trials (blue bar). Percent response and reaction times for probes were statistically indistinguishable from those for control trials.

Similar results were seen in all birds tested on these probes (n=5). As shown in figure 5A (top panel), all birds performed close to their control levels for probe stimuli within the range of natural variation, but performed poorly for shifts outside this range. All birds performed at chance for the largest pitch shifts (+/-0.71 octaves). Percent response and reaction time did not change significantly with pitch (figure 5A, lower two panels). Together these data suggest that pitch within a narrow range is critical for discriminating the songs of individuals.

To examine the importance of syllable duration for song recognition, we used probe songs that were stretched or compressed in time (figure 4A, Methods). The frequency content of these

songs was unchanged, as were the relative durations of different syllables and intervals. Probe stimuli with altered duration were constructed from the same 5 song variants as the pitch-shifted probes. Duration was altered by the same amounts (+/- 0 1, 2, 4, 8, 16, 32, and 64%).

In contrast to their responses to pitch-shifted songs, females showed great tolerance for changes in song duration. Figure 4B (top panel) shows responses as a function of song duration for the example bird shown in figure 3B. 100% duration indicates no change. The blue bar represents the bird's performance on normally rewarded control trials, and the dashed vertical lines represent the range of natural variation (+/- 4%). In contrast to her behavior on pitch-shifted probes, this female responded well above chance to all duration probes, including those compressed or expanded well beyond the range of natural variability. For duration changes up to 32%, her performance overlapped with her control performance. For a 64% change in duration, her performance dropped but remained above chance. Percent response (lower left panel) was similar to control for all probes except the most compressed (61% duration). Reaction times (lower right panel) were indistinguishable from control for all probes.

Data from 5 birds (figure 5B) confirmed this result. Birds behaved well above chance on all duration probes, generalizing over a 2.6-fold range of syllable durations. Performance dropped slightly for large duration changes (61 and 164%), but remained significantly over chance ($p<0.05$) for all birds. Most birds showed a small drop in response rates for the most compressed songs (61% duration) but responded equally well to all other probe stimuli. Reaction times for one bird increased with song duration, but remained similar for the other birds. Together these data suggest that the discrimination learned in this task does not depend heavily on syllable or interval duration.

Volume and Noise:

Females recognizing a male in a natural setting must contend with variations in the intensity of the song and of background noise. We therefore asked whether females were able to discriminate songs played at different volumes, or in the presence of background noise. Finally, we asked whether birds trained to discriminate songs at one volume could generalize to other volumes.

To test birds' ability to discriminate soft songs, we trained females (n=7) to discriminate songs played at 7 different intensities. RMS song volumes ranged from 3-73dB in steps of 10dB. These

stimuli were presented as discrimination trials, meaning that females were rewarded (with access to food) for correctly classifying a song, and punished (with a time-out) for an incorrect classification. All song intensities occurred with equal probability.

Figure 6A (top panel) shows the performance of one female as a function of song intensity. Her performance falls off gradually as the song intensity decreases, but remains high (>80% correct) down to 23dB RMS. At intensities below 23dB, she responded to fewer songs and had longer average reaction times (figure 6A, lower panels).

Data from all birds showed the same pattern. As shown in figure 6B, performance for all birds remained high down to at least 23dB. Below this intensity, performance fell off. 6 out of 7 birds performed significantly better than chance at the lowest volume tested (3dB RMS). Most birds showed a decrease in percent response, and an increase in mean reaction time for songs 23dB or softer.

To test birds' ability to discriminate songs in the presence of noise, we trained them to discriminate songs at two volumes (33 and 63dB) in the presence of varying levels of background noise (0-50dB for 33dB songs and 0-80dB for 63dB songs). Noise was generated electronically on each trial and was unique in each presentation. As in the previous experiment, birds were differentially rewarded or punished depending on whether their answer was correct, and all volume and noise combinations occurred with equal probability.

Figure 7A shows the performance of one female as a function of noise level for two different song volumes. Vertical lines represent the RMS amplitude of the songs. As expected, performance depends on the ratio of song volume to noise volume. Her performance began to fall off as the noise level approached the RMS level of the song and reached chance when the noise level was 17dB greater than the RMS level of the song. Because song is composed of distinct syllables, its amplitude varies strongly in time. Peak amplitudes could reach 7-14dB louder than the RMS song volume. On average, the bird had a slightly longer reaction time for 33dB songs, even when noise levels were low and her discrimination performance on these songs was very good (figure 7A, lower right panel). In this experiment, she responded equally well to all stimuli (figure 7A, lower left panel).

Data for all birds showed the same pattern (figure 7B). All birds' performance began to fall off when the noise level was 3-8dB below the rms volume of the song, and reached chance when the noise level was 7-17dB louder than the rms volume of the song, close to where the noise level was equal

to the peak amplitude of the song. Most birds showed slightly longer average reaction times for 33dB versus 63dB songs, and responded equally well to all songs.

The experiments shown in figures 6 and 7 test birds' ability to discriminate songs at many volumes but do not address whether they can generalize a task learned at one volume to another intensity. Because birds were rewarded or punished for their responses in these experiments, they could learn to associate loud and soft versions of a song with the same perch based on reward cues, rather than on the acoustic similarity between these stimuli. To examine whether females trained at one volume would generalize automatically to other volumes, we trained birds at a single mean volume level (73dB or 30dB), then tested them with probe stimuli at many volumes.

Surprisingly, we found that birds performed significantly above chance only for intensities within 10-30dB of the intensity at which they were trained. Figure 8A shows responses of two birds trained at 73dB to probe stimuli at different volumes. In contrast to the discrimination performance illustrated in figure 6A, these birds responded above chance only to songs presented at 43dB (pink) or 53dB (red) and above. To check that this steep decline in performance was not due to differences between these birds and others, we later trained these birds at all volumes, using the stimulus set from figure 6 (figure 8B). Within 1 day of training, performance on volume probes improved markedly, with both birds responding above chance to stimuli down to 13dB.

Birds trained only with loud songs (73dB) might not attend to softer songs, and the decrease in their performance at soft volumes could be due to a lack of attention rather than an inability to generalize. To differentiate these two options, we trained two birds on 30dB songs, then tested them with the same set of volume probes (figure 8C). Both birds responded above chance only to probe stimuli between 13-43dB (red), or 13-53dB (blue). In response to the loudest probe stimuli (73dB), both birds performed at chance. When trained on all volumes, both birds quickly learned to correctly discriminate songs from 13-73dB. These data indicate that the birds only generalize to intensities within 10-30dB of the training song.

Discussion:

Songbirds, like humans, learn to recognize complex species-specific vocalizations. Because these vocalizations vary from presentation to presentation, this recognition must involve both

discrimination of sounds corresponding to different individuals or words, and generalization across non-meaningful variations in these sounds. Because of their small size, rapid learning, and amenability to electro-physiological recordings, songbirds provide an excellent model for understanding how the human brain may solve these two problems. A crucial step for linking neural activity to recognition is mapping the space of stimuli that a bird considers to be one song.

To map this space quantitatively, we trained female zebra finches to discriminate a large set of songs produced by one individual from an equally large set produced by another. We then used probe stimuli--infrequent stimuli rewarded without respect to the female's response-- to test whether the birds could generalize the learned discrimination to novel or altered stimuli. The percent of correct responses to systematically altered probe stimuli fell off gradually as a function of the acoustic parameter altered (pitch, duration, or volume), indicating that generalization is a graded phenomenon. As a convention, in this paper we state that a female generalized to a probe stimulus when she classified it correctly at a level significantly greater than chance.

In our first experiment, we tested whether birds relied more on pitch or duration cues to distinguish the songs of two individuals. While pitch and duration are not the only cues present in zebra finch song, they are easy to measure, and their level of variability has been quantified in several previous studies. Thus, manipulating pitch and duration allows us to compare the tolerance of females' perceptual behavior to the variation in male vocalizations.

Although pitch and duration vary by similar amounts (1-3% for pitch, 1-4% for duration), females showed much narrower tuning for pitch than for duration. Females generalized over a wider range than the 0.5% just noticeable difference for pitch measured in most avian species (Dooling et al, 2000). However, performance on pitch-shifted stimuli fell to chance rapidly outside the range of natural variation. In contrast, females performed significantly above chance over a more than 2-fold range of syllable durations. The acoustic parameters used by these songbirds to identify individuals may be closer to those used by human being to identify speakers rather than words. Humans depend heavily on pitch cues to identify speakers (O'Shaughnessy, 1986), but not to identify phonemes (Avendaño et al, 2004). Word duration can be an important cue for speaker identity (Walden, 1978), but is a less reliable cue than spectral features of speech (O'Shaughnessy. 1986).

This finding has important consequences for understanding how neural responses to song may be related to perceptual behavior. Neurons in field L, the avian analog of primary auditory

cortex, respond to song with precisely time-locked spikes (Woolley et al, 2006), whose temporal pattern provides significant information about what song was heard (Narayan et al, 2006, Wang et al, 2007). Several papers have therefore suggested that these temporal patterns of spiking are critical for birds discrimination and recognition of songs.

Temporal patterns in field L, however, arise largely because spikes lock to the onsets or offsets of syllables (Nagel and Doupe, 2006, Woolley et al, 2006). They would thus be expected to change if the song is expanded or contracted in time. The ability of birds to generalize across large changes in duration calls into question the importance of field L's temporal spiking patterns for song recognition.

In contrast, the female's tolerance for pitch-shifted songs agrees approximately with the measured bandwidths of field L neurons. The half-width of the pitch tuning curves was 0.11-0.22 octaves (8-16%), while the half-widths of field L neurons have a minimum of 0.13 octaves, and a median of 0.33 octaves (chapter 2).

In a previous study (chapter 2), we found that temporally tuned cells predominated in the input layer of field L, while cells in the output layer were tightly tuned in frequency (median bandwidth 0.22 octaves) but integrate broadly in time. We therefore expect that--like the females' generalization behavior-- the responses of these output cells to song would change markedly with pitch but be relatively invariant to changes in duration. Recordings from auditory neurons with these altered song stimuli will be necessary to test whether the responses of output cells predict females' generalization behavior.

In our second series of experiments we explored females' ability to discriminate songs at different volumes or with noisy backgrounds. We found that females were able to discriminate songs above chance at volumes from 13-73dB and in the presence of noise up to 7dB louder than the mean amplitude of the song. These data indicate that the bird's auditory system can extract sufficient information from these stimuli to perform the discrimination, and form a benchmark to compare the discrimination performance of birds to the discrimination performance of their neurons.

In our last experiment, we asked birds whether they could generalize a task learned at one stimulus intensity to another. The ability to generalize across intensities is considered fundamental to pattern recognition tasks, in both the auditory and visual domains. Yet most auditory neurons show strong nonlinearities in their response as a function of sound level: at the periphery, neurons lose their

frequency selectivity at high volumes (Geissler, 1998), while in the forebrain, neurons become more sensitive to local differences at high volume (Nagel and Doupe, 2006). An outstanding question has therefore been where in the auditory system invariance to volume emerges (Drew and Abbott, 2003; Wang et al, 2007).

Surprisingly, we found that females did not generalize to songs more than 20dB louder or softer than the training stimuli. This held true even when the probe songs were much louder than the training stimuli. These data suggest that the females do not automatically generate a volume-invariant representation of the song, and instead learn to generalize across a wide volume range only when both loud and soft songs are associated with a reward. They open the possibility that intrinsically volume-invariant areas may not be found in the bird's brain. Instead, behavior and neural invariance for particular songs may be learned by associating songs at many intensities with other cues from the same individual. Invariance in other sensory systems, such as to position and scale in the visual system (Cox et al, 2005; Kortzi and DiCarlo, 2006), or to odor concentration in the olfactory system (Gross-Isserhof and Lancet, 1988), might be similarly learned.

We used an operant paradigm-- generalization of a learned discrimination-- to investigate the range of acoustic signals that a bird responds to as one individual's song. This paradigm allows us to test many song modifications in a controlled manner, and allows us to distinguish an inability to generalize from a disinclination to respond. Simpler assays, such as measuring approach behavior (Miller, 1979), calls, (Vicario et al, 2002; Vicario, 2004), singing (Stripling et al, 2003), or displays in response to song presentation have also been used to gauge mate recognition and song familiarity, but have no such positive control.

A disadvantage of our operant task is that the birds' behavior may reflect the task constraints more than their innate inclinations. For example, in our task, birds could receive their reward sooner by responding during the on-going song. This may bias them towards making decisions based on small amounts of information, and trusting pitch information over temporal structure. We think this is unlikely, as most natural female responses to song such as approach, calls and displays can similarly occur before a song bout has ended. However, we could test the effect of the temporal structure of the task on performance by training females to wait until a song bout had ended before responding. In addition, we could test whether their decision was based on only the first few syllables by using probe stimuli with the first syllables systematically removed.

Our paradigm could also be used to examine the conditions under which birds can learn to generalize across volumes. If birds were hardwired to generalize across song volumes, or learned this early in development, we would have expected this to show up in the experiments we performed in figure 8. However, we can also test whether prior exposure to an individual's song improves generalization by testing females on familiar songs, such as those of their mates or brothers. Finally, we could train females on one set of songs at many volumes, and see if this improves their generalization on songs of unfamiliar individuals.

The ability of humans and animals to recognize complex patterns is justly celebrated. Understanding how the brain achieves such recognition is a fundamental goal of neuroscience. Often, however, physiological studies assume that we know what tasks the brain solves-- such as building invariance to position and scale and luminance in the visual system, concentration in the olfactory system, and volume in the auditory system. Our data suggest that animals may group stimuli in unexpected ways. Relating neural activity to perceptual behavior will require detailed and quantitative assessments of behavior as well as of neural activity.

Figure Legends

Figure 1

An operant discrimination task to study song recognition

A) Diagram of the operant cage seen from above. Three perches are located on three sides of the cage. After a hop on the central perch, the computer plays a song from the speaker located directly in front of it. The bird can respond by hopping on either of the two response perches. After a correct answer, a feeder beneath the cage is raised for 2-5 seconds, allowing the bird access to seed. After an incorrect answer, all perches cease to function for a 20-30 second time-out.

B) Songs used for training. The goal of our study was to train birds to associate each response perch with the songs of one individual. Example song from the two individuals chosen are shown here both as oscillograms (top panels) and spectrograms (2nd row). Each song is composed of several repeated motifs, indicated by the blue bars over the spectrograms. The two individuals differ in the temporal

pattern and frequency content of their syllables. 50 songs from each individual were used in training, to avoid over-training on a single song rendition. The distribution of overall song durations, mean (rms) amplitudes, and power spectra are similar for the two individuals, as shown in the lower panels.

Figure 2

Learning and generalization of the song discrimination task

A) Learning curve for one individual showing percent of correct responses on A (black) and B (white) songs as a function of the number of days of discrimination training. By the second day of training, this bird performed well above chance on both stimulus types. This bird had 9 days of pre-training (song, food, and sequence modes, as described in Methods) before beginning discrimination trials.

B) Distribution of reaction times (after training) for the bird in figure 2A. The mean reaction times were 1.92 seconds for A songs, and 2 seconds for B songs. This indicates that the bird generally heard at least one full motif of the song but usually responded before the stimulus had played to completion. As shown in figure 1B, overall stimulus durations ranged from 2-6 seconds for both A and B songs.

C) Learning curves for 7 other birds. The dotted line at 0.5 represents chance behavior. All show greater than chance performance by the second day of training. Black curves represent birds trained with songs played at an average intensity of 73dB SPL. Green curves represent birds trained with songs with an average intensity of 30dB.

D) Mean and standard deviation of reaction time for all A (black) and B (white) songs for all birds.

E) Responses to novel renditions of the same two birds' songs were used to test whether birds had learned to associate the perch with an individual rather than a particular set of acoustic signals. Novel renditions were presented as probes, meaning that they occurred with low probability, and were rewarded at a fixed rate regardless of the birds' response (see Methods). The first panel shows the mean and 95% confidence intervals on the percent of correct responses for novel probe songs (y-axis) versus training songs (x-axis). The second panel shows the mean and 95% confidence intervals for the percent of responses to these two types of stimuli. The third panel show the mean and standard

deviation of reaction time. In all cases, responses to novel probes were statistically indistinguishable from responses to training songs.

Figure 3

Responses to pitch-shifted songs

A) Examples of pitch-shifted songs. Center panel: unshifted song. Top panel: same song shifted up by 0.71 octaves (64%). Bottom panel: same song shifted down by 0.71 octaves (61%). In this manipulation the frequency content of the song is changed but the temporal pattern remains the same. A shift in pitch leads to expansion or compression of notes with harmonic structure. Pitch-shifted probe stimuli were presented at low probability between unaltered training trials, and were rewarded without regard to the female's response.

B) Performance of one bird on pitch-shifted probe stimuli. Top panel: percent correct as a function of pitch shift (black line). Error bars represent 95% confidence intervals on percent correct obtained by fitting data to a binomial distribution. The blue bar at the top represents the mean +/- one standard deviation of her performance on control training trials, averaged across days (n=16). Dashed lines at 97% and 103% represent +/- one standard deviation of the range of natural pitch variation. Her performance overlaps with control performance within this range and falls off rapidly outside it. Bottom left panel: percent response as a function of pitch shift (black line), with 95% confidence intervals. The blue bar represents the mean +/- one standard deviation of percent response for control trials. Response rates do not change with pitch and overlap with response rates for control trials. Bottom right panel: Reaction time as a function of pitch shift (black line) with standard deviation. The blue bar represents the mean +/- one standard deviation of reaction time for control trials. Reaction times for probes trials fall within the range of reaction times for control trials.

Figure 4

Responses to songs with altered duration

A) Examples of songs with altered duration. Center panel: unaltered song. Top panel: same song expanded by 64%. Bottom panel: same song compressed by 64% (duration = 1/1.64 or 61%). Expansion or compression alters the duration of both syllables and intervals, but does not change their relative durations, nor the frequency content of the song. Duration probes were presented at low probability between unaltered training trials, and were rewarded without regard to the female's response.

B) Performance of one bird on duration probe stimuli. Top panel: percent correct as a function of song duration (black line). Error bars and blue bar as described in the legend for figure 3B. Dashed lines representing the range of natural variability are at 96% and 104%. From 76 to 132% duration her performance on probes overlaps with her performance on control trials, and she performs significantly above chance on all probes. Bottom left panel: percent response as a function of song duration. Response rates for probes were comparable to response rates for control trials except for the most compressed songs. Bottom right panel: Reaction time as a function of duration. Reaction time generally did not increase with song duration.

Figure 5

Population data for pitch and duration experiments

A) Responses of 5 birds (including the example in figure 3B) to pitch-shifted probe stimuli. Top panel: colored dots represent the overall percent correct for each bird at each condition. The black line represents the mean across birds. Error bars at the top left represent the mean +/- one standard deviation of each birds' performance on control trials. Data from the same bird are in the same color. The dashed lines represent the range of natural variability (+/-3%), and the dotted line represents chance behavior. All birds show a decrease in performance for songs outside the range of natural variability and chance behavior for the largest pitch shifts. Bottom left panel: Percent response for all birds as a function of pitch shift, with the mean across birds shown in black. Bottom right panel: Mean reaction time for all birds with the mean across birds shown in black.

B) Responses of the same 5 birds to duration probe stimuli. All colors and symbols as in (A). Top panel: Percent correct as a function of probe song duration. All birds performed significantly above

chance for all duration probe stimuli, although their performance decreased below control levels for the largest changes 61% and 164% duration). Bottom left panel: Percent response as a function of song duration. Many birds showed a small decrease in response rate to the most compressed (61% duration) stimuli. Bottom right panel: Reaction time as a function of song duration. Most birds showed no change in reaction time with changes in song duration.

Figure 6

Discrimination of songs across intensities

A) Performance of one bird on discrimination trials with songs of different intensities. In this experiment responses to songs at all levels were differentially rewarded or punished depending on the bird's response. Top panel: Percent correct as a function of RMS song volume. Error bars represent 95% confidence intervals obtained from a binomial fit to the data. The performance of this bird decreased gradually with decreasing song intensity, but remained significantly above chance for all volumes tested. Bottom left panel: Percent response as a function of song volume. This bird responded at a much-reduced rate to songs below 23dB. Bottom right panel: Mean and standard error of reaction time as a function of song volume. Mean reaction time increased significantly for song played at 23dB or softer.

B) Performance of 7 birds, including the example in (A), on discrimination trials with song of different intensities. Colored dots represent mean values for each bird. Black lines represent the mean across birds. Top panel: Performance rates fell off gradually, remaining high for songs played at intensities down to 23dB. All but one bird performed above chance at the lowest intensity tested (3dB). Bottom left panel: Percent response as a function of song intensity. Many birds responded at decreased rates to songs played below 23dB. Bottom right panel: Mean and standard deviation of reaction time as a function of song intensity. Many birds showed a gradual increase in reaction time as songs became softer. This increase is most noticeable for songs played at 23dB and below.

Figure 7

Discrimination of songs in the presence of noise

A) Performance of one bird on discrimination trials with different levels of background noise. Top panel: Performance as a function of noise level for 63dB songs (light blue) and 33dB songs (dark blue). The intensities of the songs are indicated by the vertical lines. Error bars represent 95% confidence intervals obtained from a binomial fit to the data. Performance was significantly better than chance for noise levels up to 12dB greater than the RMS amplitude of the song, close to the songs' peak amplitude. Bottom left panel: Percent response as a function of noise level for the two song intensities. This birds responded at high rates to all stimuli. Bottom right panel: Mean and standard error of reaction time as a function of noise level for the two song intensities. Mean response times to the softer song (33dB, dark blue) were significantly longer than those to the louder song (63dB, light blue) at all noise levels tested, but did not change significantly with added noise.

B) Performance of 4 birds on discrimination trials with different noise levels. Solid lines represent the mean across birds. Patterns of behavior for all four birds were similar to the example in A: behavior fell to chance when the noise level was 7-17dB louder than the rms song volume. Response rates did not change with stimulus condition (bottom left panel), and reaction times for the soft song were longer than for the loud song (bottom right panel). Reaction time did not change with noise intensity.

Figure 8

Testing generalization from one volume to another

A) Performance (mean and 95% confidence intervals on percent correct) of two birds initially trained with 73dB RMS songs on probe songs at many volumes. Performance by both birds fell off rapidly for songs softer than 73dB. One bird performed significantly above chance for only songs played at RMS intensities between 53-73dB (red), while the other performed above chance for songs played at 43-73dB (pink). The blue vertical line indicates the RMS volume of the training songs.

B) Performance of the same two birds on identical probe trials after 1-2 days training with songs at many volumes. After training, both birds performed above chance for songs down to 13dB rms.

C) Performance of two different birds trained with 30dB rms songs on the same volume probes. Both birds performed significantly above chance for songs from 13-43dB, and performed at chance level for songs played at 63 and 73dB. The blue vertical line indicates the RMS volume of the training songs.

D) Performance of the same two birds on identical probes after 1-2 days of training with songs at many volumes. After training, both birds performed above chance for songs between 13-73dB.

References:

Avedaño, C., Deng, L., Hermansky, H., and Gold, B. (2004) "The analysis and representation of speech." Speech Processing in the Auditory System. New York, NY, Springer-Verlag.

Bailey, D. J., J. C. Rosebush, et al. (2002). "The hippocampus and caudomedial neostriatum show selective responsiveness to conspecific song in the female zebra finch." J Neurobiol 52(1): 43-51.

Beecher, M. D., S. E. Campbell, et al. (1994). "Song Perception in the song sparrow: birds classify by song type but not by singer." Animal Behavior 47: 1343-1351.

Brenowitz, E. A. (1991). "Altered perception of species-specific song by female birds after lesions of a forebrain nucleus." Science 251(4991): 303-5.

Cox, D.D., Meier, P., Oertelt, N., and DiCarlo, J.J. (2005) "Breaking position-invariant object recognition." Nat Neurosci 8(9): 1145-47.

Dooling, R. J., B. Lohr, et al. (2000). Hearing in Birds and Reptiles. Comparative Hearing. Birds and Reptiles. R. J. Dooling and R. R. Fay. New York, NY, Springer-Verlag. 13: 308-359.

Doupe, A. J. and M. Konishi (1991). "Song-selective auditory circuits in the vocal control system of the zebra finch." Proc Natl Acad Sci U S A 88(24): 11339-43.

Doupe, A. J. and P. K. Kuhl (1999). "Birdsong and human speech: common themes and mechanisms." Annu Rev Neurosci 22: 567-631.

Drew, P. J. and L. F. Abbott (2003). "Model of song selectivity and sequence generation in area HVc of the songbird." J Neurophysiol 89(5): 2697-706.

Eggermont, J. J. (2001). "Between sound and perception: reviewing the search for a neural code." Hear Res 157(1-2): 1-42.

Fortune, E. S. and D. Margoliash (1992). "Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (Taenopygia guttata)." J Comp Neurol 325(3): 388-404.

Fortune, E. S. and D. Margoliash (1995). "Parallel pathways and convergence onto HVc and adjacent neostriatum of adult zebra finches (Taeniopygia guttata)." J Comp Neurol 360(3): 413-41.

Geisler, C.D. (1998) From Sound to Synapse: Physiology of the Mammalian Ear. Oxford University Press.

Gentner, T. Q. and S. H. Hulse (1998). "Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris." Anim Behav 56(3): 579-594.

Gentner, T. Q., S. H. Hulse, et al. (2000). "Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds." J Neurobiol 42(1): 117-33.

Gentner, T. Q. and D. Margoliash (2003). "Neuronal populations and single cells representing learned auditory objects." Nature 424(6949): 669-74.

Glaze, C. M. and T. W. Troyer (2006). "Temporal structure in zebra finch song: implications for motor coding." J Neurosci 26(3): 991-1005.

Gross-Isserhof, R. and Lancet, D. (1988) "Concentration-dependent changes of perceived odor quality." Chemical Senses 13(2):191-204.

Kao, M. H. and M. S. Brainard (2006). "Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability." J Neurophysiol 96(3): 1441-55.

Kourtzi, Z., and DiCarlo, J.J. (2006) "Learning and neural plasticity in visual object recognition." Curr Op Neurobiology 16:152-158.

MacDougall-Shackleton, S. A., S. H. Hulse, et al. (1998). "Neural bases of song preferences in female zebra finches (Taeniopygia guttata)." Neuroreport 9(13): 3047-52.

Mello, C. V. and D. F. Clayton (1994). "Song-induced ZENK gene expression in auditory pathways of songbird brain and its relation to the song control system." J Neurosci 14(11 Pt 1): 6652-66.

Mello, C. V., D. S. Vicario, et al. (1992). "Song presentation induces gene expression in the songbird forebrain." Proc Natl Acad Sci U S A 89(15): 6818-22.

Miller, D. B. (1979). "The acoustic basis of mate recognition by female zebra finches (Taeniopygia guttata)." Animal Behavior 27: 376-380.

Nagel, K. I. and A. J. Doupe (2006). "Temporal processing and adaptation in the songbird auditory forebrain." Neuron 51(6): 845-59.

Narayan, R., G. Grana, et al. (2006). "Distinct time scales in cortical discrimination of natural sounds in songbirds." J Neurophysiol 96(1): 252-8.

O'Shaughnessy, D. (1986). "Speaker recognition." IEEE Acoustics, Speech, and Signal Processing 3(4):4-17.

Scharff, C., F. Nottebohm, et al. (1998). "Conspecific and heterospecific song discrimination in male zebra finches with lesions in the anterior forebrain pathway." J Neurobiol 36(1): 81-90.

Sen, K., F. E. Theunissen, et al. (2001). "Feature analysis of natural sounds in the songbird auditory forebrain." J Neurophysiol 86(3): 1445-58.

Stripling, R., A. A. Kruse, et al. (2001). "Development of song responses in the zebra finch caudomedial neostriatum: role of genomic and electrophysiological activities." J Neurobiol 48(3): 163-80.

Stripling, R., L. Milewski, et al. (2003). "Rapidly learned song-discrimination without behavioral reinforcement in adult male zebra finches (Taeniopygia guttata)." Neurobiol Learn Mem 79(1): 41-50.

Theunissen, F. E., K. Sen, et al. (2000). "Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds." J Neurosci 20(6): 2315-31.

Vates, G. E., B. M. Broome, et al. (1996). "Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches." J Comp Neurol 366(4): 613-42.

Vicario, D. S. (2004). "Using learned calls to study sensory-motor integration in songbirds." Ann N Y Acad Sci 1016: 246-62.

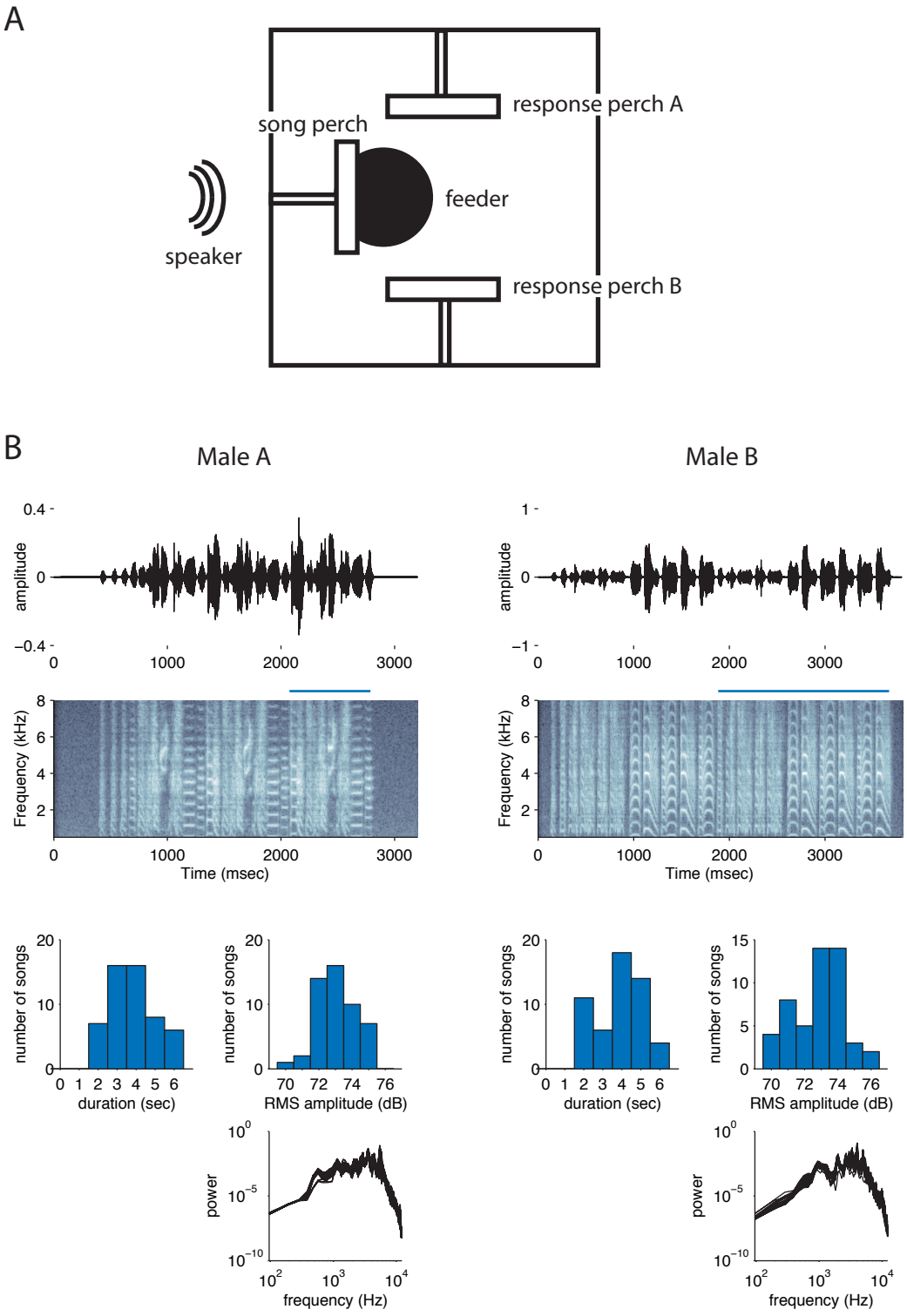
Vicario, D. S., J. N. Raksin, et al. (2002). "The relationship between perception and production in songbird vocal imitation: what learned calls can teach us." J Comp Physiol A Neuroethol Sens Neural Behav Physiol 188(11-12): 897-908.

Walden, B.E., Montgomery, A.A., Gibeily, G.J., Prosek, R.A., and Schwartz, D.M. (1978) "Correlates of psychological dimensions in talker similarity." Journal of Speech and Hearing Research, 21(2): 265-275.

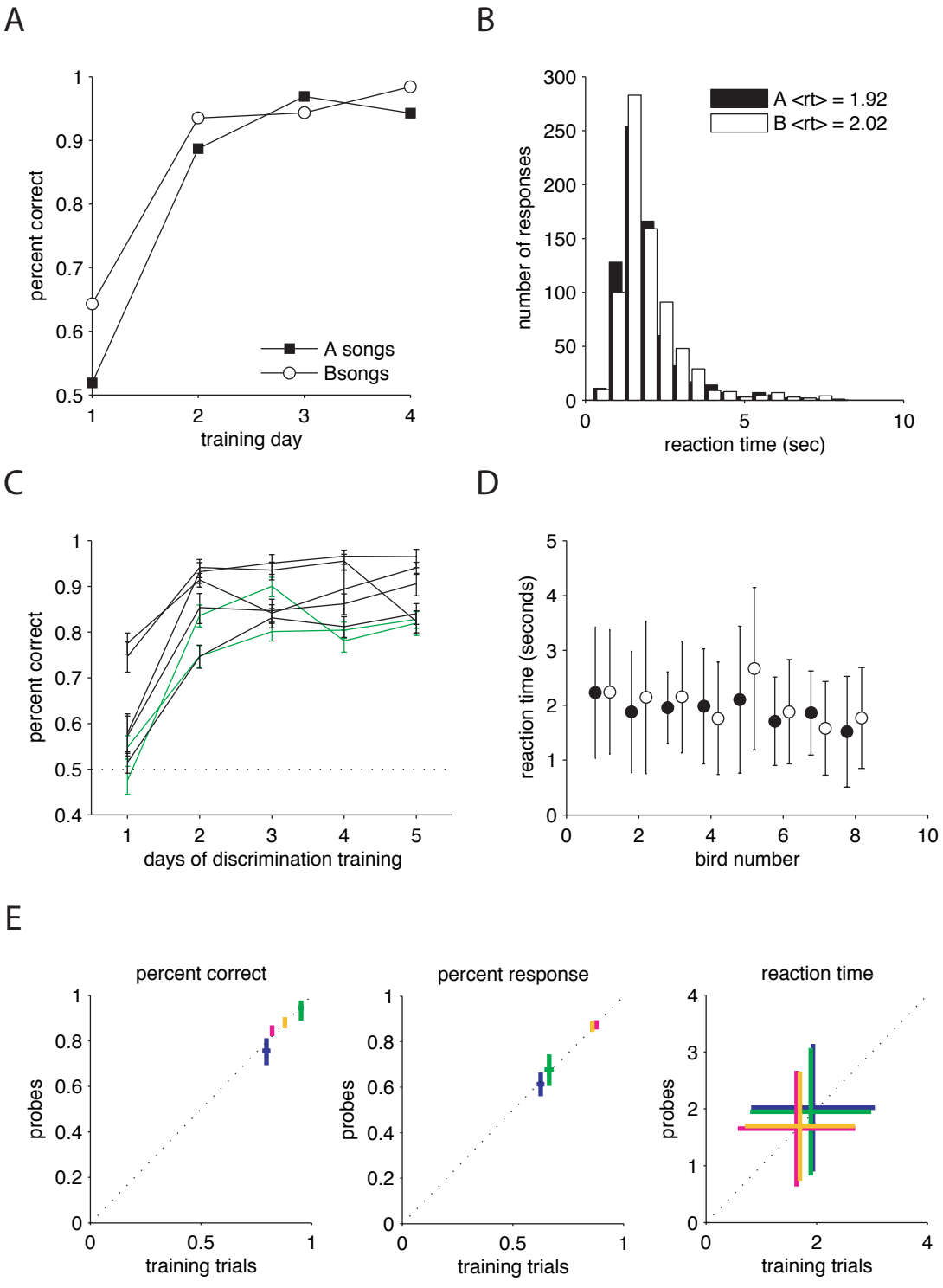
Wang, L., R. Narayan, et al. (2007). "Cortical discrimination of complex natural stimuli: can single neurons match behavior?" J Neurosci 27(3): 582-9.

Woolley, S. M., T. E. Fremouw, et al. (2005). "Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds." Nat Neurosci 8(10): 1371-9.

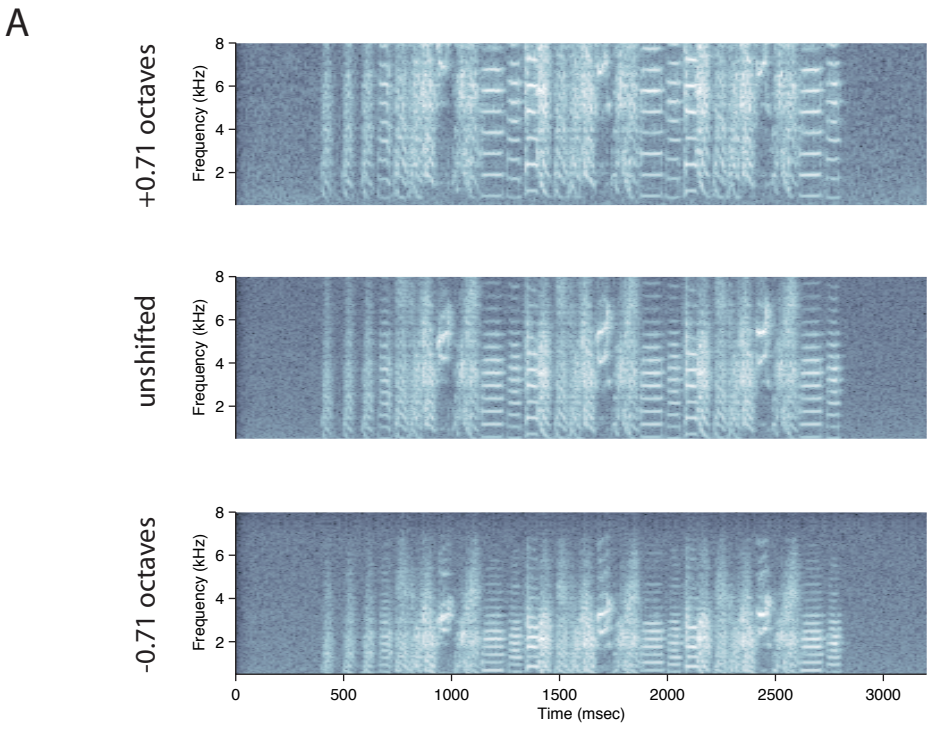
Woolley, S. M., P. R. Gill, et al. (2006). "Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain." J Neurosci 26(9): 2499-512.
Zahn, R.A. (1996). The Zebra Finch: A Synthesis of Field and Laboratory Studies. Oxford University Press.

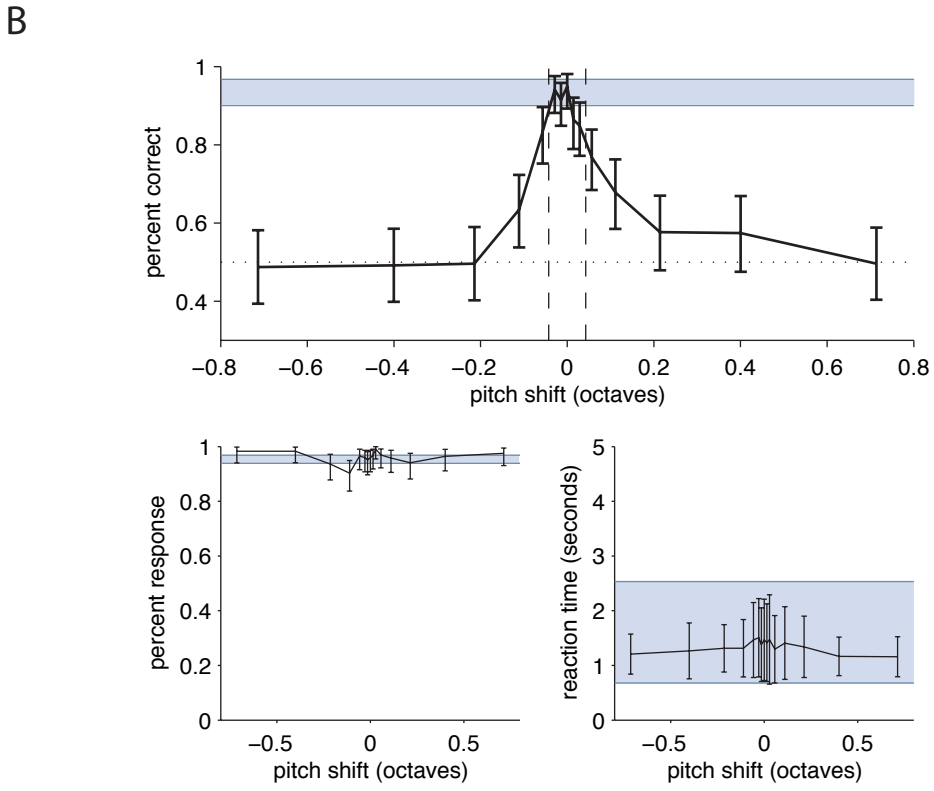


chapter 3, figure 1: operant cage and training stimuli

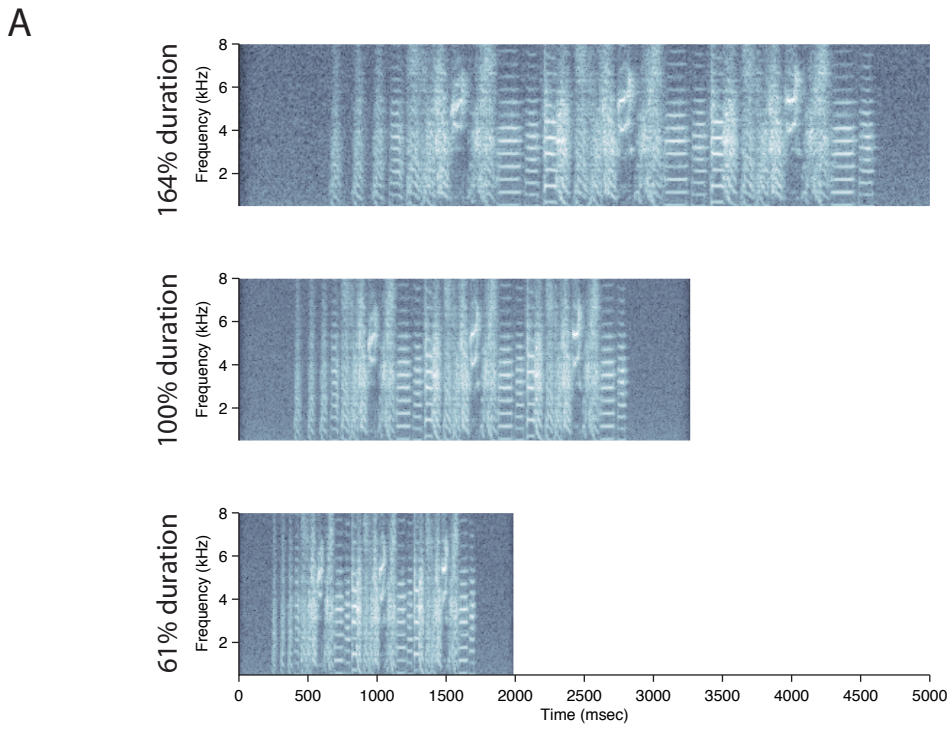


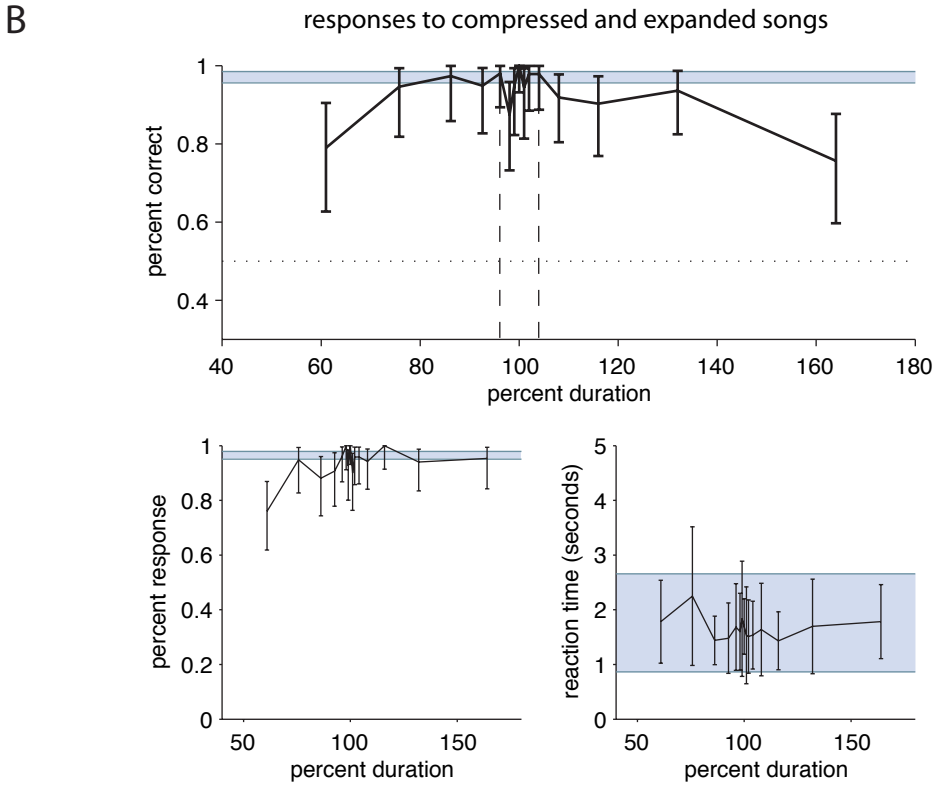
chapter 3, figure 2: learning and performance of the song discrimination task



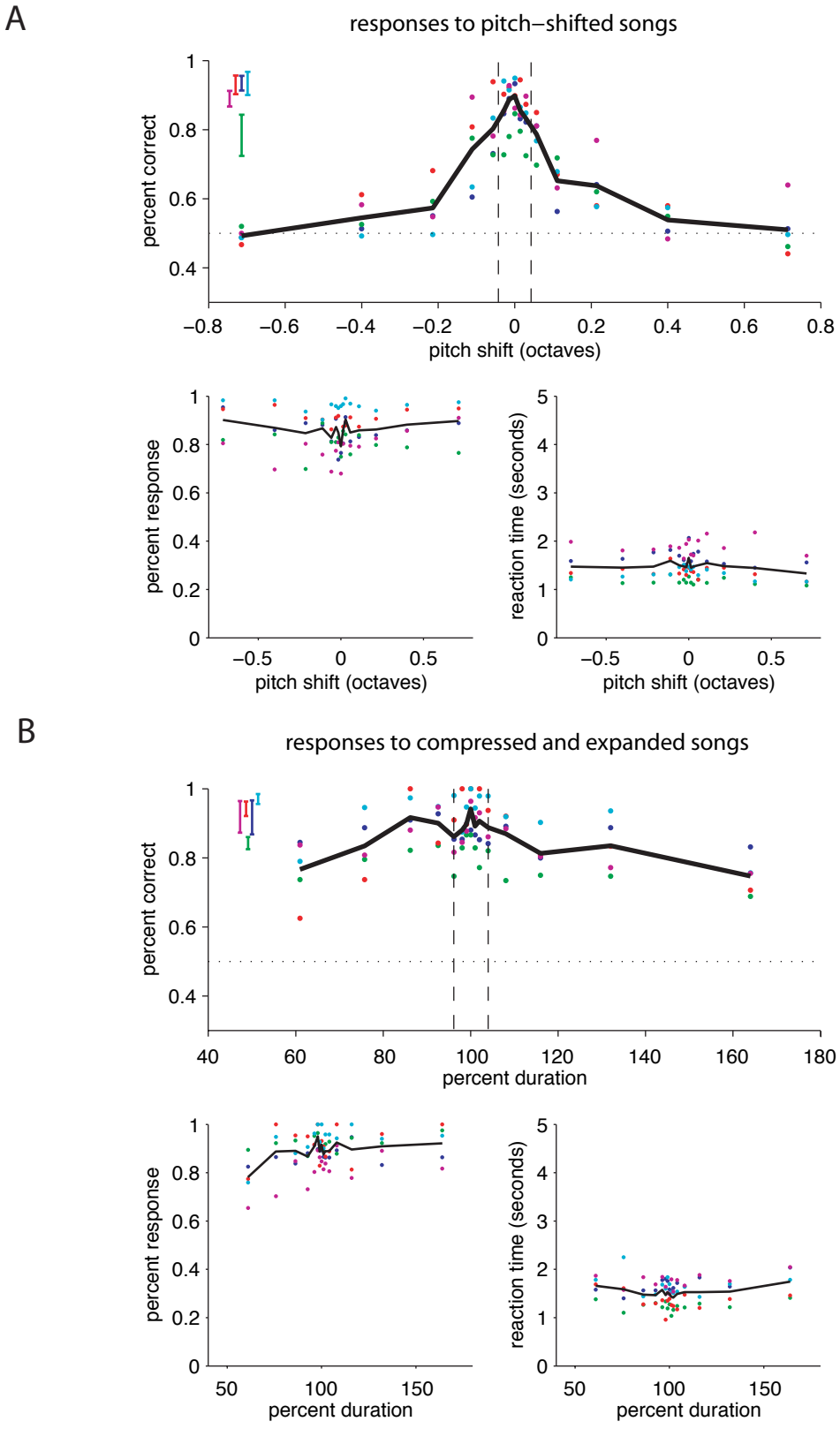


chapter 3, figure 3: responses to pitch-shifted probe songs

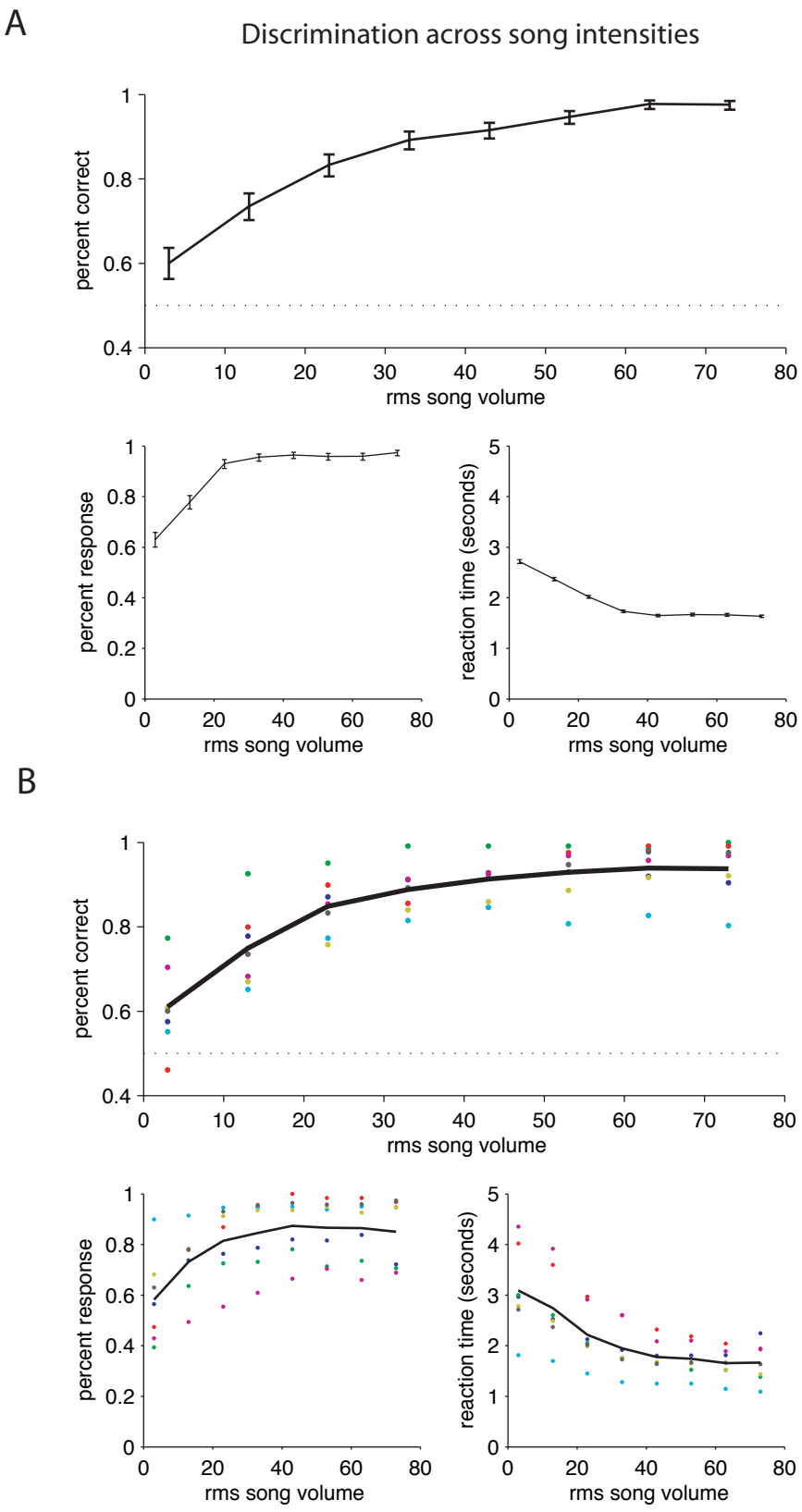




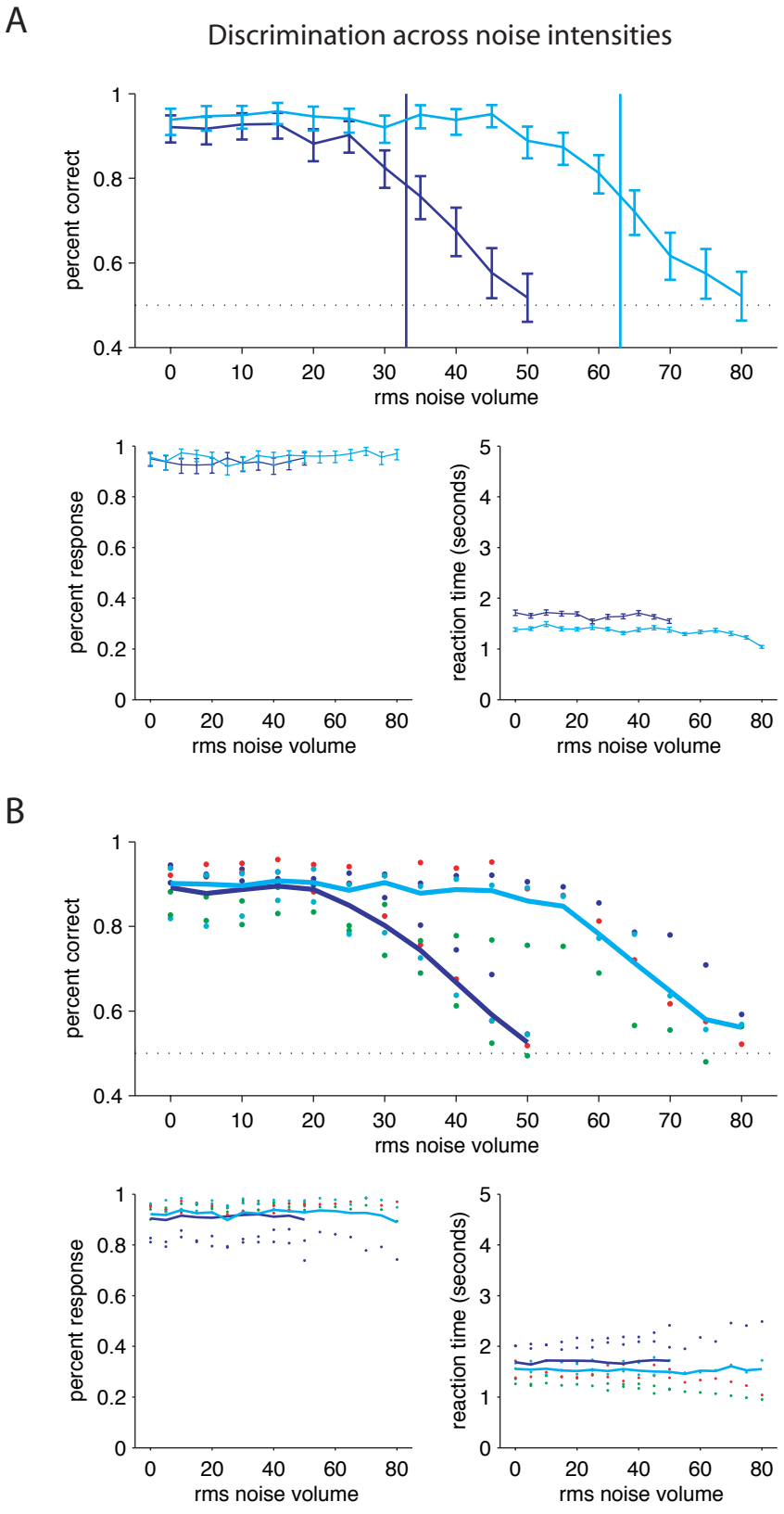
chapter 3, figure 4: responses to probe songs with altered duration



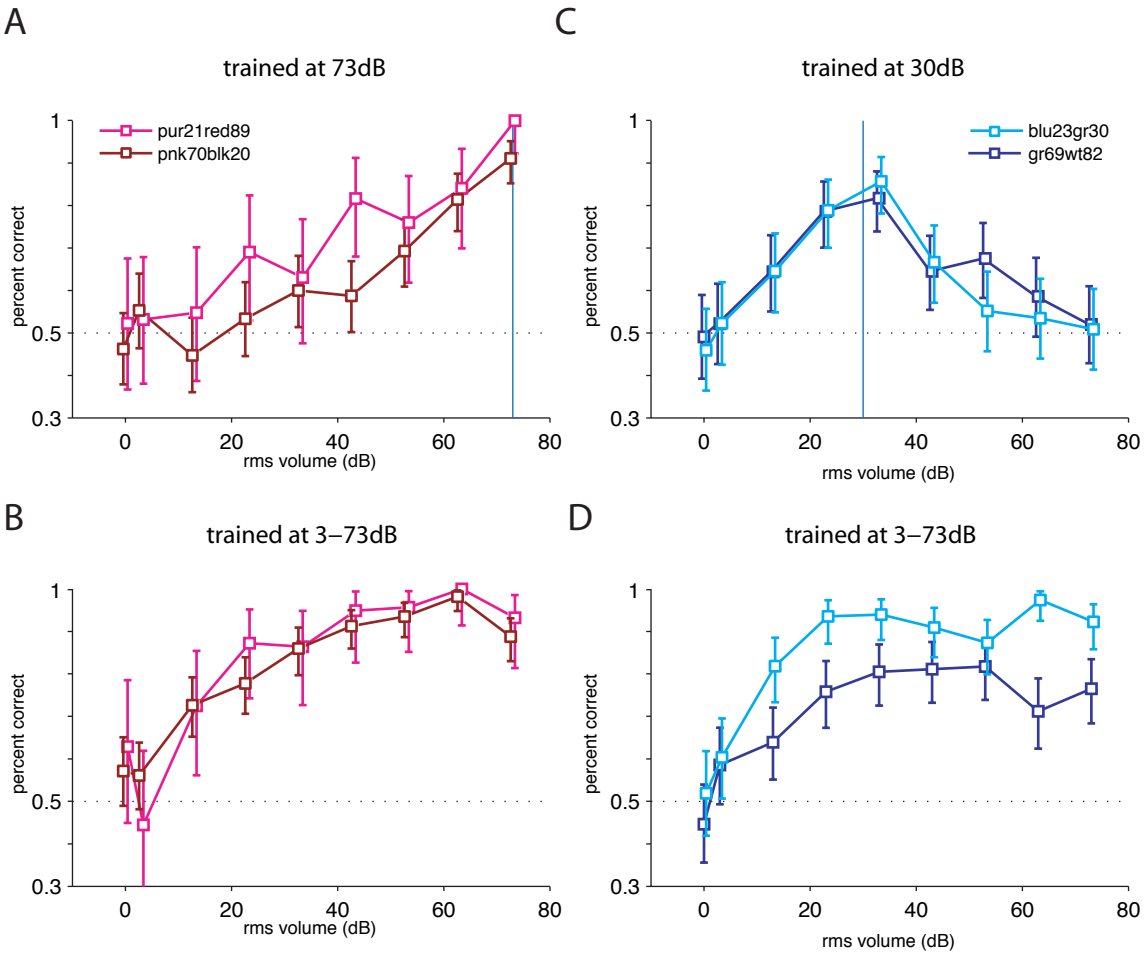
chapter 3, figure 5: population data for pitch and duration experiments



chapter 3, figure 6: discrimination of songs across intensities



chapter 3, figure 7: discrimination of songs in the presence of noise



chapter 3, figure 8: generalization to novel song intensities

General Discussion

I set out to characterize the auditory response properties of field L, and to test the role of various acoustic parameters in song recognition. I presented three main sets of findings.

First, I found that the layers of field L differ in both their physiology and their response properties. The input layer, L2, contains mostly narrow-spiking, fast-firing cells, tuned to fast temporal or spectro-temporal modulations. The output layers, L1 and L3, contain broad-spiking, slow-firing cells, that integrate over long time periods and are tuned for spectral modulations. Narrow-spiking cells evenly sample temporal modulations up to 150Hz, while broad-spiking cells evenly sample spectral modulations up to 2 cycles per octave. Narrow spiking spectro-temporal cells span a range of high frequency modulations from mostly temporal to mostly spectral. This distribution of tuning preferences may reflect the statistical structure of zebra finch song, which has most of its energy in pure spectral and temporal modulations, or it may reflect a very general strategy for processing natural sounds, which share similar statistics (Singh and Theunissen, 2003; Attias and Schreiner, 1998). The difference in spike shape and firing rate between fast temporal and spectro-temporal, and slow spectral cells may reflect biophysical differences in their ion channels and morphology that allow them to track fast stimulus changes, or to integrate over longer time periods, respectively.

Second, I found that the linear response properties of most field L cells depended strongly on the mean intensity of the stimulus. At low mean intensities, most spectro-temporal receptive fields (STRFs) were dominated by a single positive peak, indicating that under these conditions, cells behaved largely as integrators or detectors of sound in a certain frequency range. At high mean intensities, negative components of STRFs grew in magnitude and decreased in latency, while positive components showed smaller decreases in latency. Depending on the frequency range and timing of these negative and positive regions, these nonlinearities allowed different cells to become more specialized for spectral, temporal, or spectro-temporal modulations at high stimulus intensities.

These data suggest a more unified way of looking at the dependence of response properties on stimulus amplitude than the traditional categorization of auditory cells as monotonic or non-monotonic. Instead, different cells are specialized for extracting different kinds of spectro-temporal features, and all cells undergo a transition from detection at low intensities to discrimination at higher intensities. The shift from detection at low intensities to discrimination at high intensities has also been seen in the retina (Enroth-Cugell and Lennie, 1975), where it has been shown computationally to

improve coding efficiency in high and low signal-to-low regimes (Attick, 1992). It may represent a common computational principle shared by many sensory systems.

Finally, I looked at what acoustic features females used to perform a discrimination task between two individuals. I found that the duration of syllables and intervals did not have much impact on discrimination performance, but that pitch and volume did. These data suggest that the spatial pattern of cells activated by a song is more important for identifying the song than the temporal pattern of firing in individual cells.

These findings suggest several lines of future questions. What ion conductances or morphological properties underlie the differences between classes of field L cells? What role do these different cell types and their nonlinearities play in forming a population-level representation of song? How does this representation shape behavioral responses to song? In this discussion, I will outline experiments and models that might be used to address each of these questions in turn.

Relating auditory responses to cellular biophysics:

What biophysical properties underlie the difference between narrow-spiking temporal and spectro-temporal cells, and broad-spiking spectral cells? How do those biophysical properties constrain response properties?

In many mammalian systems, narrow spikes and high firing rates are associated with the expression of potassium channels from the Kv3 family, whose unique kinetics allow cells to rapidly repolarize. In the cortex and hippocampus, most such Kv3-expressing cells are inhibitory interneurons, although Kv3 channels are also expressed in fast-firing GABA-ergic Purkinje cells, glutamatergic cells of the subthalamic nucleus, and glutamatergic terminals in the calyx of Held (Bean, 2007). Kv3-expressing inhibitory interneurons can be identified by their expression of parvalbumin (Kawaguchi and Kondo, 2002).

To ask what cell types might give rise to the different response types we saw, a first approach might be to stain slices of zebra finch forebrain tissue for Kv3 and parvalbumin. Because narrow-spiking cells are found only in the input layer, L2, we can ask whether this region also shows expression of Kv3 and/or parvalbumin. Additionally, we could see if Kv3 co-localized with GAD-65, a marker for GABAergic cells, or with CB1, a marker for field L2. A GAD-65 antibody for birds has

been developed and used successfully by Pinaud et al. (2004). Parvalbumin-positive interneurons have been found in the song control nuclei HVc and RA (Wild et al, 2001, 2005) but have not been examined closely in field L. These experiments would allow us to ask if putative fast-firing neurons of L2 express Kv3, and whether such neurons are inhibitory.

To link such cell types to auditory response properties more directly, it would be useful to record intracellularly from in vitro slice preparations of field L. Several studies have demonstrated that spike shapes recorded extracellularly in vivo are related closely to the time derivative of spike waveforms recorded intracellularly in slice (Buzsaki and Eidelberg 1982; Henze et al, 2000). Although we could not measure auditory response properties of neurons in a slice preparation, we could record their responses to dynamic injections of current. Using this paradigm, we should be able to identify cells with fast and slow integration times. By filling such cells and staining slices as described above, we can ask if cells with fast and slow integration times express different potassium channels, or different neuro-transmitters.

If we are able to identify channels specific to narrow-spiking cells, such as Kv3, we could block these channels phamacologically (using TEA, Goldberg et al, 2005) or manipulate their expression using lentivirus (Wada et al, 2006; Scharf, unpublished data) to probe these cells' behavior in vivo. For example, we could ask whether cells with fast but not slow auditory response properties are affected by a pharmacological blockade of Kv3 channels. We could also ask how these cells' auditory responses change in the absence of functional Kv3 channels.

Modeling population coding in field L:

Is the distribution of response properties that we described particularly suited for encoding zebra finch song or other natural sounds? Do the non-linearities we described with intensity allow for more efficient coding in different regimes? These questions can only be answered through modeling.

Our data suggest that field L could be modeled as a bank of spectro-temporal filters, or synthetic STRFs, with built-in nonlinearities. Such filter banks have been used to model responses of both the auditory (Lewicki 2002; Qiu et al, 2003; Chi et al, 2005), and the visual (Jones and Palmer, 1987; Olshausen and Field, 1996) systems. We could generate synthetic STRFs using the bivariate-mexican hat models from chapter 2. Parameters of the STRFs would be drawn from fits to the distributions of parameters we measured (chapter 2, supplementary figure 5). We could also fit our

data and generate synthetic STRFs using Gabor functions, which would replace the quadratic terms in the mexican hat model with cosine functions.

Nonlinear receptive fields could be added to this model by making each neuron's response the sum of two linear-nonlinear filters: one purely excitatory and one purely inhibitory. The inhibitory filter would have a higher threshold and gain. This causes the linear receptive field of the neuron to depend on the mean of the stimulus, but does not require an explicit representation of the stimulus mean (Figure 1). To make a more realistic model, we could make the temporal width and latency of both inputs depend on the mean intensity of the stimulus. We could also add gain normalization to the model using the function model of Mante and Carandini (2005), and add a range of exponential decays to different cells that approximate the range of adaptation times described in chapter 1. This nonlinear filter bank could be used to synthesize population responses to zebra finch songs. We could then use spike count (Shadlen and Newsome, 2001), and spike timing measures (Van Rossum, 2001; Narayan et al, 2006) to ask whether we can discriminate between the representations of two individuals' songs. By looking at synthesized responses to songs altered in pitch, duration, and volume, we can ask which of these measures, and which classes of cells, most closely approximate the birds' behavior in chapter 3. We can also use information measures (Strong et al, 1998; Hsu et al, 2001) to ask how much information the filter bank captures about different classes of natural and synthetic stimuli.

The strength of this model is that it will allow us to test how particular distributions of cell types, and particular nonlinearities, contribute to song discrimination. By drawing the parameters of our STRFs from different distributions, we can test whether the distributions we observed provide a particularly "informative" representation of zebra finch song. By adding nonlinearities one at a time, we can test what impact each nonlinearity has on the representation of songs at different volumes, and on the representation of songs with noisy backgrounds. For example we could ask whether nonlinear receptive fields allow better discrimination of different songs at low volume, but make generalization from one volume to another more difficult.

Linking neural responses to behavior:

What neural responses determine how birds respond to and categorize altered songs? To link the behavioral results of chapter 3 to patterns of neural firing, we could start by recording from field L

during performance on pitch, duration, and volume probe trials. One hypothesis I proposed in that chapter was that responses of L1/L3 spectral cells would more closely track the birds' behavior than responses of L2 cells with their tightly time-locked responses. By comparing spike counts or PSTHs of responses to altered song with responses to training songs, we can construct neural "tuning curves" for song that we can compare directly with the behavior tuning curves measured in chapter 3 (figures 3, 4, 5, and 8). We can then ask which cell types (spectral, temporal, or spectro-temporal) and what aspects of their firing (spike count over some time interval, PSTH at some time-resolution) most closely matches the birds' behavior. This analysis will allow us to ask how the representation in field L may be "read-out" by neurons downstream.

If we find differences between cells in different layers, we could use pharmacology to test the role of various cell classes more directly. Using reverse dialysis of lidocaine, a post-doc in our lab has successfully and reversibly inactivated neurons in area X of awake, chronically-implanted zebra finches. By recording during the reverse dialysis, she was able to confirm that she was in the right area and precisely measure the time-course of inactivation. We could use this technique to reversibly inactivate neurons in different layers of field L and look at whether this inactivation impacts birds behavior on control or probe trials of the behavior task. If we find specific channels expressed by fast versus slow cells, as described above, we could phamacologically target these channels to alter the response properties of L2 cells and look at their effects on behavior.

Together, the proposed experiments and models will allow us to link biophysical differences between cells to neural representations of sounds, and neural representations to behavior. By taking advantage of the many levels of analysis possible in the songbird, such experiments should help elucidate the neural basis of acoustic pattern recognition.

References:

Attias, H. and C.E. Schreiner (1998) “Coding of Naturalistic Stimuli by Auditory Midbrain Neurons.” Advances in Neural Information Processing Systems 10

Atick, J. J. (1992). "Could Information-Theory Provide an Ecological Theory of Sensory Processing." Network-Computation in Neural Systems 3(2): 213-251.

Bean, BP (2007). "The action potential in mammalian central neurons." Nat Rev Neurosci, 8:451-465.

Buzsaki, G. and E. Eidelberg (1982). "Direct afferent excitation and long-term potentiation of hippocampal interneurons." J Neurophysiol 48(3): 597-607.

Chi, T., P. Ru, et al. (2005). "Multiresolution spectrotemporal analysis of complex sounds." J Acoust Soc Am 118(2): 887-906.

Goldberg, E. M., S. Watanabe, et al. (2005). "Specific functions of synaptically localized potassium channels in synaptic transmission at the neocortical GABAergic fast-spiking cell synapse." J Neurosci 25(21): 5230-5.

Henze, D. A., Z. Borhegyi, et al. (2000). "Intracellular features predicted by extracellular recordings in the hippocampus in vivo." J Neurophysiol 84(1): 390-400.

Hsu, A., S. M. Woolley, et al. (2004). "Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons." J Neurosci 24(41): 9201-11.

Enroth-Cugell, C. and P. Lennie (1975). "The control of retinal ganglion cell discharge by receptive field surrounds." J Physiol 247(3): 551-78.

Kawaguchi, Y. and S. Kondo (2002). "Parvalbumin, somatostatin and cholecystokinin as chemical markers for specific GABAergic interneuron types in the rat frontal cortex." J Neurocytol 31(3-5): 277-87.

Mante, V., R. A. Frazor, et al. (2005). "Independence of luminance and contrast in natural scenes and in the early visual system." Nat Neurosci 8(12): 1690-7.

Narayan, R., G. Grana, et al. (2006). "Distinct time scales in cortical discrimination of natural sounds in songbirds." J Neurophysiol 96(1): 252-8.

Olshausen, B. A. and D. J. Field (1996). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images." Nature 381(6583): 607-9.

Pinaud, R., T. A. Velho, et al. (2004). "GABAergic neurons participate in the brain's response to birdsong auditory stimulation." Eur J Neurosci 20(5): 1318-30.

Qiu, A., C. E. Schreiner, et al. (2003). "Gabor analysis of auditory midbrain receptive fields: spectro-temporal and binaural composition." J Neurophysiol 90(1): 456-76.

Singh, N. C. and F. E. Theunissen (2003). "Modulation spectra of natural sounds and ethological theories of auditory processing." J Acoust Soc Am 114(6 Pt 1): 3394-411.

Shadlen, M. N. and W. T. Newsome (2001). "Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey." J Neurophysiol 86(4): 1916-36.

Strong, S. P., R. R. de Ruyter van Steveninck, et al. (1998). "On the application of information theory to neural spike trains." Pac Symp Biocomput: 621-32.

van Rossum MC. A novel spike distance. Neural Comput 13: 751–763, 2001.

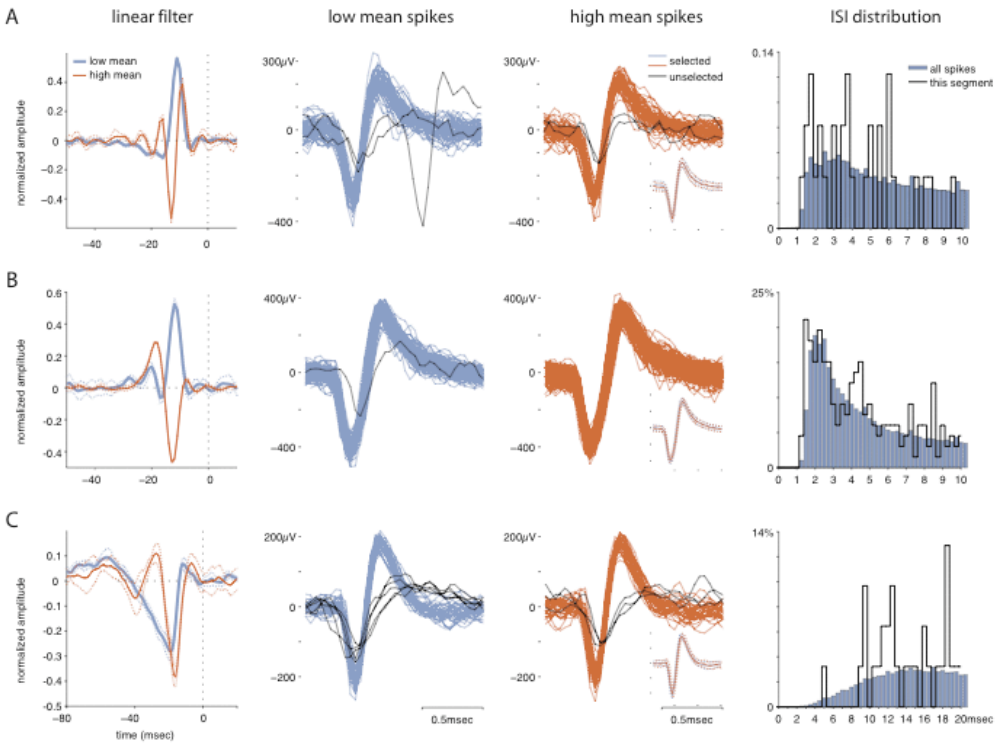
Wada, K., J. T. Howard, et al. (2006). "A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes." Proc Natl Acad Sci U S A 103(41): 15212-7.

Wild, J. M., M. N. Williams, et al. (2001). "Parvalbumin-positive projection neurons characterise the vocal premotor pathway in male, but not female, zebra finches." Brain Res 917(2): 235-52.

Wild, J. M., M. N. Williams, et al. (2005). "Calcium-binding proteins define interneurons in HVC of the zebra finch (Taeniopygia guttata)." J Comp Neurol 483(1): 76-90.

Chapter 1 Supplementary Data

supplementary figure 1



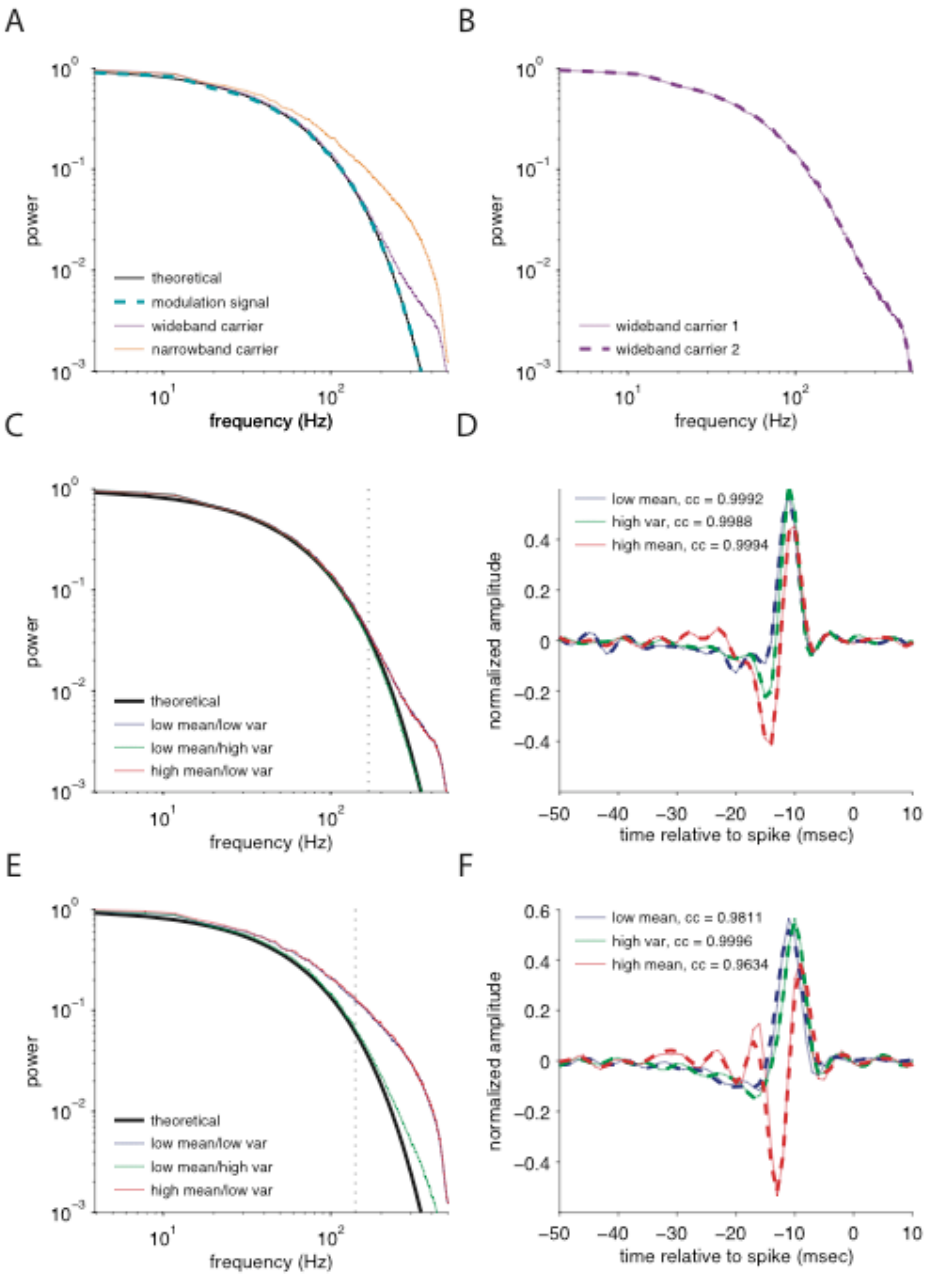
Supplementary Figure 1

Spike waveforms from example cells

The left-hand column shows filter pairs obtained under low and high mean from the three example cells in figure 4A-C of the paper. In the center two columns are thresholded waveforms from each of these three recording sites. The waveforms are taken from adjacent 1 second segments just before (column 2, blue) and after (column 3, orange) the switch from low to high mean. Waveforms shown in black were excluded during sorting due to their smaller size and different shape. Waveforms shown in blue and orange were kept for analysis. These selected waveforms have a very similar shape under both statistical conditions, suggesting that they represent the same unit. The small insets in the lower right hand corner of column 3 show an overlay of the mean +/- one standard deviation of the selected waveforms from these two segments, and reinforce our observation that the waveforms have the same

shape in the two conditions. The column on the far right shows inter-spike interval (ISI) distributions of selected waveforms for the two one-second segments shown for each cell (thick black lines, n= 99, 272, and 63 spikes respectively) and for the full recording (blue bars, n=97555, 211252, and 96943 spikes). The distributions from the short segments are similar to the overall distributions, and all distributions show clear refractory periods, suggesting that the selected waveforms represent the spikes of a single unit. In general, the shape of a presumptive unit's waveform stayed consistent over the course of a recording (30 minutes to 2 hours), although the amplitude of the waveform could change slowly over time. Based on these observations and similar observations at all other sites , we believe that it is unlikely that our results could be due to the presence of different units during different statistical conditions.

supplementary figure 2



Supplementary Figure 2

Decorrelation using the theoretical power spectrum of the modulation signal, versus the actual power

spectrum

In this figure we compare the power spectrum of the full log stimulus envelope to the theoretical power spectrum we used in our analysis, and examine the effects of using these two power spectra to decorrelate the filters.

Supplementary figure 2A shows the effect of the carrier on the frequency spectrum of the log envelope. The black line represents the theoretical distribution of frequencies: P(f) = exp(-f/50Hz). The blue dashed line shows the distribution of frequencies in the modulation signal that we designed and used for our analysis. This distribution is very close to the theoretical distribution.

The frequency distribution of the actual log envelope of our stimulus will depend on the method we use to estimate the envelope. Here, we estimate the envelope by low-pass filtering the rectified stimulus below 500Hz. We also tried estimates using a lower cutoff frequency and found that the frequency spectra of these estimates was dominated by the shape of the low-pass filter.

The frequency spectrum of the log envelope also depends on the frequency content of the carrier. The purple line represents the frequency spectrum of a stimulus with a wideband (500Hz-8000Hz) carrier. It deviates strongly from the theoretical frequency spectrum only at frequencies above about 250Hz. The orange line represents the frequency spectrum of a stimulus with a narrowband (2500-3000Hz) carrier. It deviates from the theoretical prediction more strongly and at lower frequencies than does the stimulus with the wideband carrier. During decorrelation, we divide the Fourier representation of the spike-triggered average by the power spectrum, then low-pass filter the result. As discussed in our Methods, the low-pass filtering is necessary because decorrelation tends to amplify high frequencies, which are poorly sampled in the spike-triggered average. The filter cutoff was determined for each cell individually by looking for the frequency at which the power in the spike-triggered average dropped below the power in an estimate of the noise--the random-triggered average. Therefore, we expect that including the contribution of the carrier in decorrelation will have the most impact on cells with high cutoff frequencies. We examine these cases below.

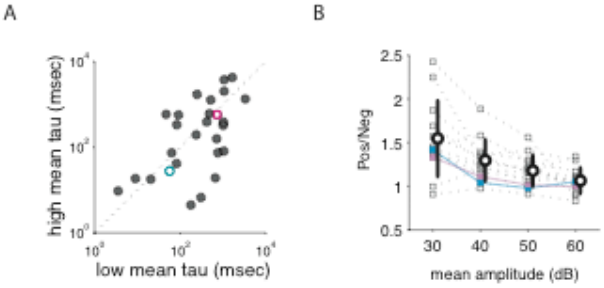
Supplementary figure 2B examines the reliability of the power spectrum estimate. The solid and dashed curves represent two estimates of the log stimulus envelope power spectrum, using different wideband carriers. As can be seen, the two estimates are essentially identical due to the long sampling time (2.4e7 samples per trial). We therefore concluded that the power spectrum depends only on the statistics of the carrier, and not on the precise carrier used.

Supplementary figures 2C and 2D show the effect of using the actual power spectrum to decorrelate filters recorded with a wideband stimulus. A wideband carrier was used to estimate most of our filters (21/36). Here we study the worst case scenario: the effects of the carrier on the estimate of the filter with the highest cutoff frequency (167 Hz). Supplementary figure 2C shows the theoretical frequency spectrum in black, and the spectra of log envelopes for each of our three statistical conditions in blue, green, and red. The frequency cutoff for this cell is indicated by the dotted black line. Because the actual power spectra are very similar to the theoretical power spectrum below this line, we do not expect that using the actual power spectrum will have much impact on the filters.

Supplementary figure 2D confirms this expectation directly: it shows filters obtained by decorrelating with the theoretical power spectrum (thin solid lines) and with the actual power spectrum (thick dashed lines). As we expected, the two estimates are very similar, and the correlations between the two estimates are greater than 0.99 for all conditions.

Supplementary figures 2E and 2F show the effect of using the actual power spectrum to decorrelate filters recorded with a narrowband stimulus, in this case 2500-3000 Hz. Again, we looked at the cell with the highest cutoff frequency, reasoning that it should show the largest possible effect of the difference between actual and theoretical power spectra. Supplementary figure 2E shows the theoretical and actual power spectra with the cutoff frequency for this cell (140 Hz) indicated by the black dotted line. Supplementary figure 2F shows filters obtained by decorrelating with the theoretical power spectrum (thin solid lines) and with the actual power spectrum (thick dashed lines). Although there are small discrepancies between the two filters (for example the second positive peak in the red filter is smaller when using the actual power spectrum), the overall shapes of the filters are very similar. The correlation coefficients between the two filter estimates are high, and the qualitative differences between the filters obtained under different statistical conditions remain. Based on this analysis of worst case scenarios, we concluded that that using the theoretical power spectrum to decorrelate our filters did not significantly impact our results, and is a reasonable approximation.

Supplementary Figure 3



Supplementary Figure 3

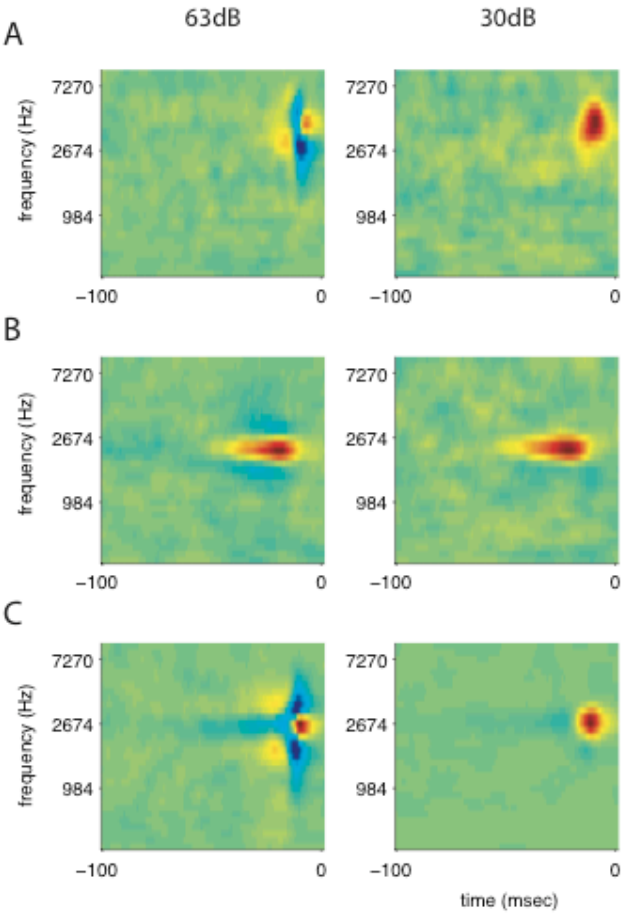
A Low to high mean decay times were correlated with high to low mean decay times (all cells with significant decays , cc= 0.47). The two example cells from figure 3B are indicated by a blue circle and red square. There is a great deal of scatter in this relationship, suggesting that different mechanisms may be involved in adaptation to high and to low mean (Abbott et al, 1997, Fairhall et al, 2001, Kvale and Schreiner, 2004).

Reference: Abbott, L.F., J.A. Varela, K. Sen, and S.B. Nelson (1997). "Synaptic depression and cortical gain control." Science 275(5297):220-4.

B) Changes in filter shape occur gradually as a function of mean stimulus amplitude. Pos/Neg ratios in filters from 13 cells recorded under the four different mean conditions. Dotted lines connect ratios from the same cell. Both the mean (black circles) and the variance (black lines) of the Pos/Neg ratio decrease across the population as the stimulus mean increases. At 60dB, the highest mean condition we examined, the ratios are all closer to one. Together these results suggest that filters change gradually and systematically as a function of mean stimulus amplitude, and that these changes may be related to changes in inhibition. Blue and purple symbols indicate Pos/Neg for the example cells from figure 5D and E.

Chapter 2 Supplementary Data

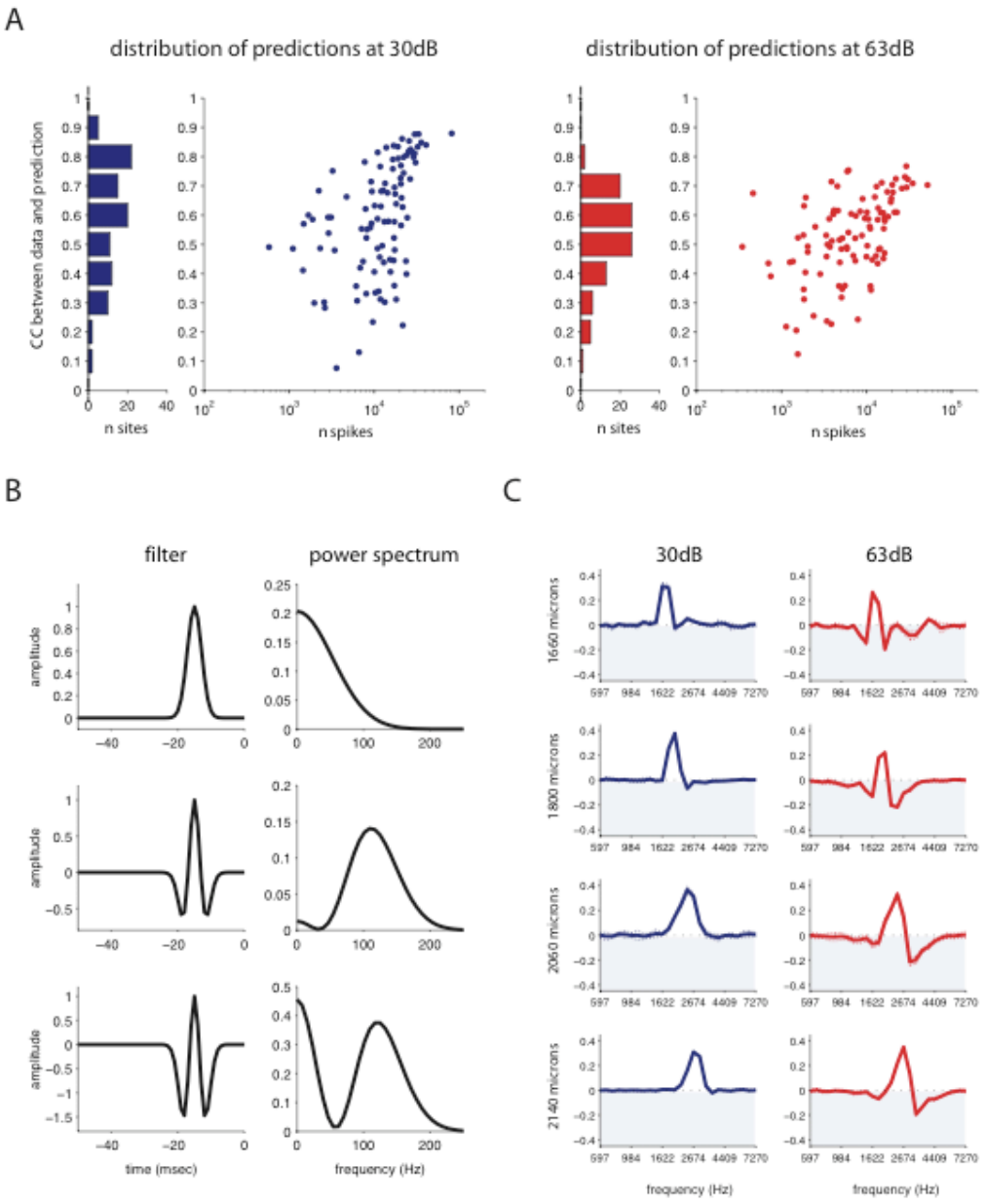
Supplementary Figure 1: 3 raw STAs for 3 example STRFs



Supplementary Figure 1

Raw (un-decorrelated) spike-triggered averages (STAs) for the example cells shown in figure 2A-C. At 63dB, the differences in tuning between a temporal (A), spectral (B), and spectro-temporal (C) cell are evident without decorrelation. At 30dB all three cells have much simpler STAs dominated by a single large positive peak.

Supplementary Figure 2



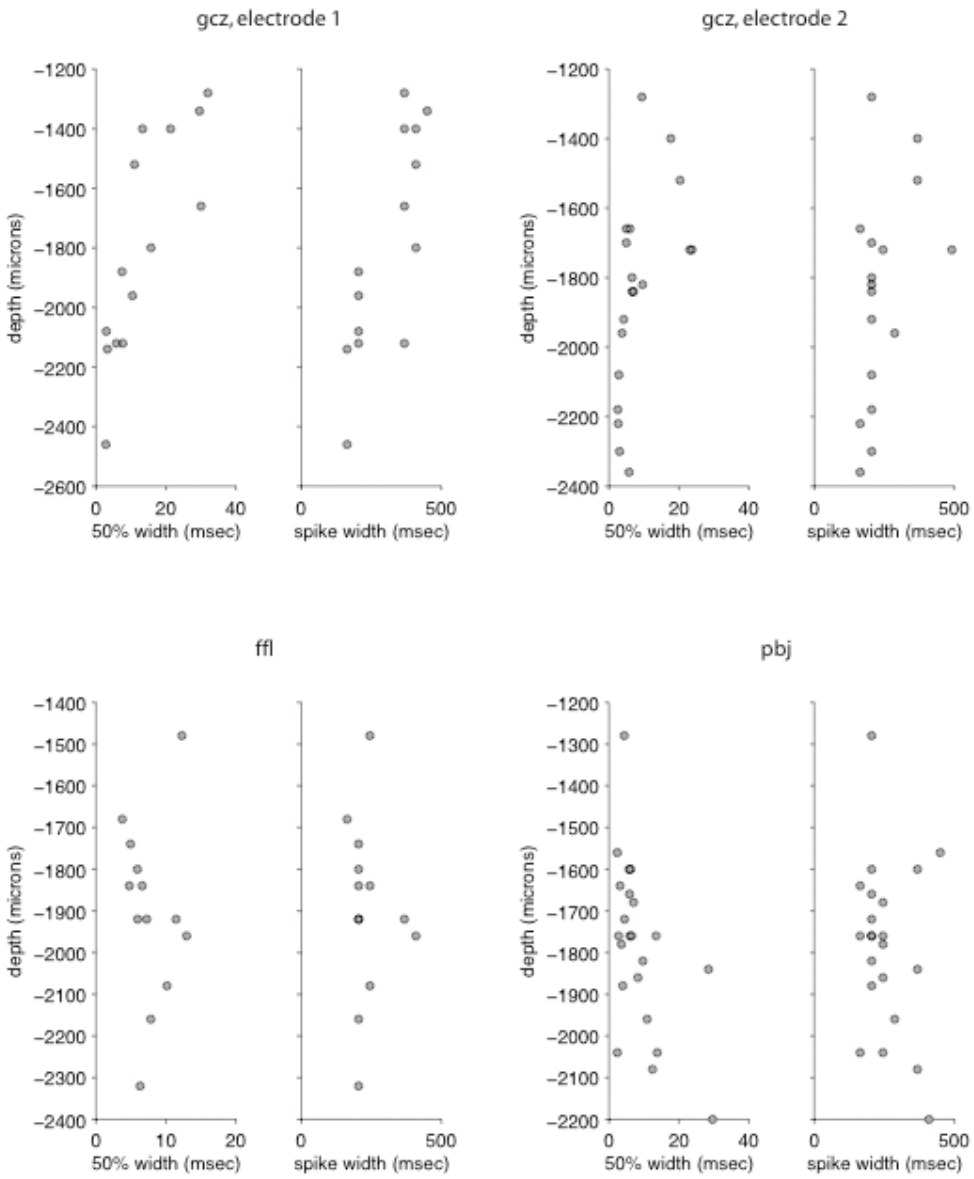
Supplementary Figure 2

A) The distribution of correlation coefficients between predicted and actual PSTHs for stimuli played at 30dB (left) and 63dB (right). PSTHs were obtained from 13 to 200 repeats of an identical 5 second stimulus segment by smoothing the average spike count per millisecond with a 8 msec-wide hanning window. Predictions were made using STRFs and nonlinearities derived from a non-overlapping set of data as described in the Methods. The first 500 msec of both the predicted and actual PSTH were omitted to reduce the effects of slow adaptation on the comparison. The quality of the prediction depended strongly on the number of spikes used to estimate the STRF, but plateaued at lower values for the 63dB stimulus condition. The mean correlation coefficient was 0.57+/- 0.19 (standard deviation) for the 30dB stimulus condition, and 0.52 +/- 0.14 for the 63dB stimulus condition.

B) Different filter shapes give rise to different modulation spectra. The filter shapes shown represent either temporal or spectral-cross sections through the STRF, and the power spectra represent corresponding cross-sections through the two-dimensions modulation spectrum. If a filter has a single peak (top row), its power spectrum will be low-pass, meaning that the cell will respond to all modulations in that domain up to some cutoff determined by the width of the STRF peak. A filter with a single negative peak will give rise to a similar power spectrum. If the filter has balanced positive and negative regions (second row), its power spectrum will be band-pass, meaning that it will only respond to modulation frequencies over a narrow range. Finally, a filter with very profound negative sidebands (3rd row), whose negative regions are larger than its positive regions, will have a power spectrum with two peaks. The alternating positive and negative regions will give rise to a high frequency band-pass peak, while the overall negative sum will give the power spectrum a peak at low frequencies. This property of power spectra complicates our analysis of some strongly spectrally-tuned cells, such as the example in figure 2B, whose STRFs show clear sensitivity to spectral modulations, but whose spectral power spectra peak at 0 cycles per octave.

C) Spectral cross-sections through several spectro-temporal STRFs illustrating their similar structure. Most such cells had a single positive peak at 30dB, while at 63dB they showed strong negative sidebands. These sidebands were generally stronger and narrower on the high frequency side, and shallower and broader on the low-frequency side. The STRFs shown here were collect at the depths indicated in a single bird and suggest a systematic progression of preferred frequencies.

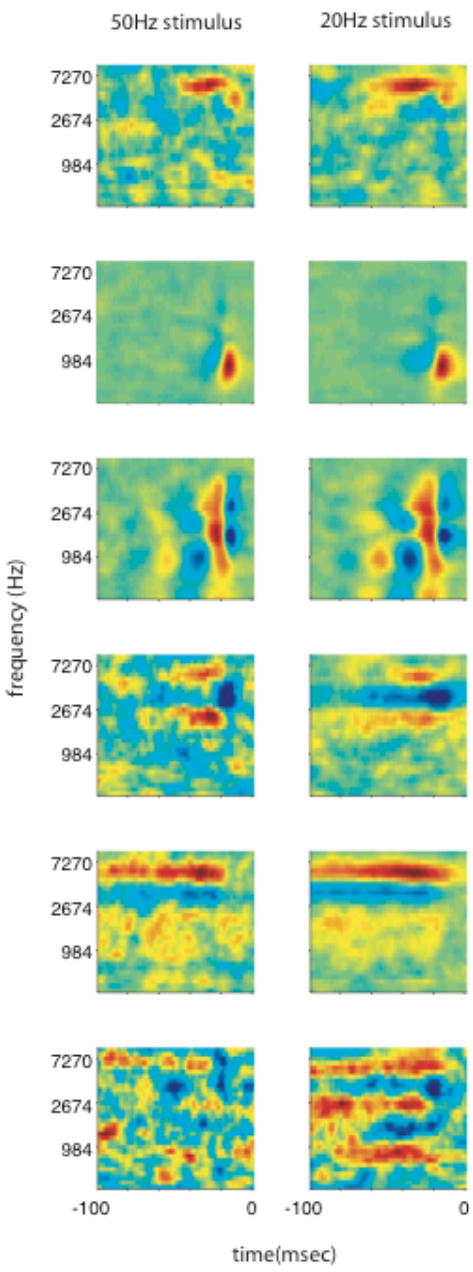
Supplementary Figure 3



Supplementary Figure 3

The anatomical distribution of STRF and spike properties from chronic birds supports our findings in restrained sedated birds (figure 6). This figure shows temporal half-widths of STRFs and widths of spikes as a function of recording depth for four chronic electrode penetrations. Fast STRFs with short temporal half-widths and narrow spikes are mostly found in a restricted region of each penetration. The strong correlation between STRF width and spike width is also clear.

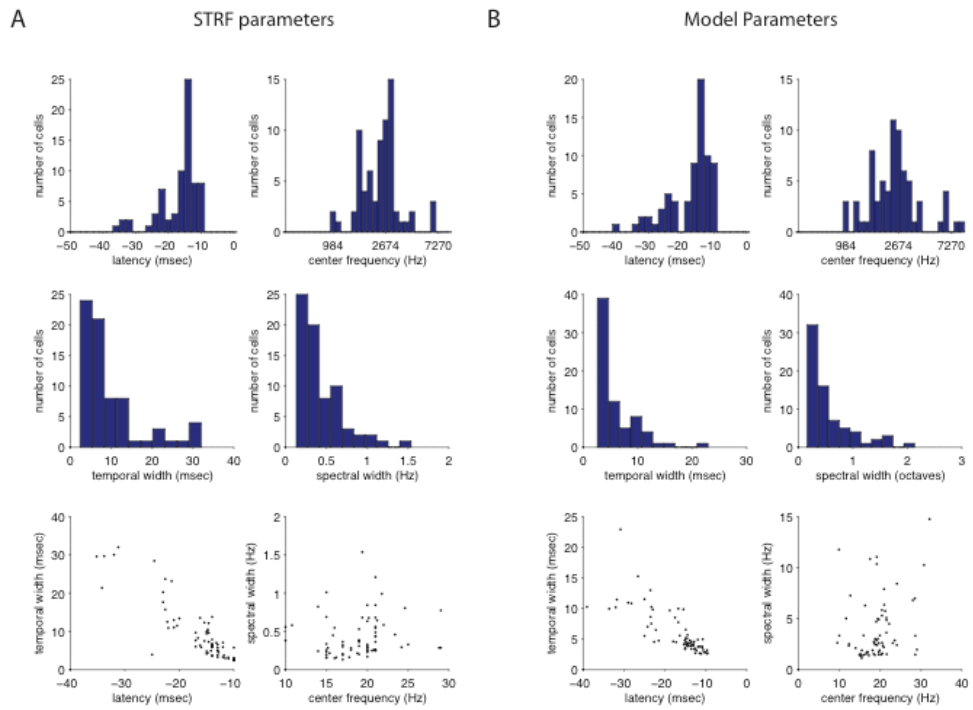
Supplementary Figure 4



Supplementary Figure 4

Spike-triggered averages recorded from the same site under two stimulus conditions, one with faster amplitude modulations (50Hz, left) and one with slower amplitude modulations (20 Hz, right). In general the STRFs obtained under the two conditions were very similar, although the STRFs obtained with the slower, 20Hz stimulus are smoother and slightly broader in both frequency and time. Compared to the striking effects of stimulus intensity on STRF shape (supplementary figure 1), the effects of stimulus frequency are very mild. Occasionally, as at the last site shown, sensitivity to very slow stimulus features was evident with the 20Hz stimulus but not with the 50 Hz stimulus.

Supplementary Figure 5



Supplementary Figure 5

Distribution of STRF parameters obtained by measuring STRFs directly (A) and by fitting a bivariate mexican-hat model to each STRF (B). The measured latency and center frequency of the STRF were defined to be its maximum in time and frequency. The measured temporal half-width, and bandwidth were obtained from cross-sections through the peak. Model parameters were obtained by fitting as described in detail in the Methods. In general, the shape of the two distributions showed a good

agreement. Latencies showed a skewed distribution with more cells having shorter latencies. Center frequencies showed an approximately normal distribution centered around 2.5 kHz. Temporal and spectral widths had more exponential distributions, with more cells having short widths. Latency and temporal width were negatively correlated, while center frequency and spectral width (in octaves) were not. Correlation coefficient between measured and fitted parameters were 0.97 for latency, 0.86 for center freuqency, 0.89 for temporal widths, and 0.57 for spectral widths).

.

Publishing Agreement
It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Katherine Wright 6.26.07

Author Signature Date