

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Auditory Processing and Perception in Songbirds

Permalink

<https://escholarship.org/uc/item/1jv9910r>

Author

McLendon, Helen Marguerite

Publication Date

2015

Peer reviewed|Thesis/dissertation

Auditory Processing and Perception in Songbirds

by

Helen Marguerite McLendon

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Neuroscience

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Copyright 2014

by

Helen Marguerite McLendon

Acknowledgments

This dissertation is a testament to the tireless support and encouragement offered by my advisor, Allison Doupe. Allison's passion for scientific inquiry, incredible breadth of knowledge, and creative insights made every meeting with her exciting. In addition, her boundless optimism and faith in her students were constant sources of inspiration to me.

I am greatly indebted to Michael Stryker, who, in the wake of Allison's passing, stepped in to see me through the last months of my graduate career. He rapidly oriented himself to my unfamiliar data and pushed me to throw together a coherent story. His dogged pursuit kept me from throwing in the towel, and his incisive comments have (hopefully) made this thesis readable. I am also grateful to my other committee members, current and past: Michael Brainard, Christoph Schreiner, Steve Lisberger, and Frédéric Theunissen. Their many thoughtful questions and suggestions have will greatly improve the forthcoming journal articles based on this dissertation.

I've been fortunate to have two additional mentors during my graduate career. Sarah C. Woolley took me under her wing when I joined the lab. Much of the work described in chapter one was accomplished with the aid of her clever problem-solving skills and insights into animal behavior. Chapter two of this dissertation would not exist without the apparently limitless patience and wisdom of Gunsoo Kim. I will strive to emulate his ability to think clearly, calmly, and critically. The Doupe Lab as a whole has been a continual source of constructive feedback and encouragement. I am indebted to all the talented, honest, and dedicated post-docs Allison managed to gather, including Mimi Kao, Hamish Mehaffey, Satoshi Kojima, Laurie Stepanek, Derek Zaraza, Raghav Rajan, and Smitha Jagadish. Thanks to Adria Arteseros for excellent histology and impeccable laboratory management.

I'm grateful to Kathy Nagel and David Schoppik for teaching me MatLab and reverse correlation, and for their intellectual and emotional support through my first years in graduate school. I also owe much of my development as a scientist to my other officemates in HSE-808, including Jenn Li, Ritu Kapur, Etienne de Villiers-Sidani, Mati Joshua, Yan Yang, Sonja Hohl, and Javier Medina.

This thesis would not exist without Tom Clandinin, who hired me as a laboratory technician (even though I knew nothing about neuroscience), encouraged me to apply to graduate school, and helped me with my first grant applications.

Thanks to Darrell Floyd for designing and building the operant cages and feeders used in chapter one and Ken McGary for designing their electrical wiring. I'm grateful to Dirk Kleinhesselink for being the most responsive and considerate IT analyst I could imagine.

I'm deeply appreciative of all the friends who have put up with me for the last eight years. My classmates in the neuroscience program have supported me beyond all reason, and cheered me up countless times. Thanks especially to Erica Korb and Talia Lerner for all of the long bike rides that brought me back from the brink of insanity. I would never have finished this document without the relentless encouragement and nagging of Anika Kinkhabwala and Alex Williams.

I'm especially grateful to have had the love and support of Will Van Trump. His unflagging faith in me has kept me going on many days when I was ready to give up. His presence in my life has made it immeasurably richer and more joyous.

Finally, I thank my brother, J, for constantly challenging me to think deeply and strengthen my arguments, and my parents, Dixie and Jim, for always encouraging my curiosity in the world around me.

Abstract

Auditory processing and perception in songbirds

Helen Marguerite McLendon

The ability to perceive and categorize spoken words is a remarkable capability of the human auditory system. Songbirds are one of the few animal orders that face auditory tasks of similar complexity. In this dissertation, I analyze auditory responses in the zebra finch at both behavioral (chapter one) and neuronal levels (chapter two). In chapter one, I use an operant condition paradigm to show that female songbirds are able to identify the social context in which a male's song was sung (alone, or directed towards a female). Females require only a short segment of recorded song (a single "motif") to perform this task. I also show that, given only simple temporal information about the stimuli, a machine-learning algorithm can classify most males' motifs according to social context. However, the model's behavior is not consistent with that of the females on individual stimuli, indicating that spectral and temporal cues beyond those tested by the model influence the birds' behavior. Finally, lesions of a nucleus required for social context-dependent differences in spectral variability cause most males to produce songs whose social context was still detectable to females performing the task. Chapter two describes the results of a series of acute electrophysiological recordings in anesthetized female zebra finches. I analyze the responses of single neurons in the songbird auditory forebrain to two types of stimuli: birdsong and an artificially generated stimulus. Using a relatively unbiased mutual-information-based technique, I show that the responses of these neurons change dramatically depending on the stimulus. Across different stages of the ascending auditory pathway, song stimuli give rise to more complex receptive fields than the artificial stimulus. Receptive fields calculated in response to the song stimuli also have excellent predictive value, far surpassing that

of the receptive fields calculated from artificial stimuli. Our results indicate that for many neurons in the songbird auditory forebrain, receptive field structure is highly dependent on stimulus statistics, and that receptive fields constructed in response to different stimulus classes contain surprisingly little information regarding responses to other sounds.

Table of Contents

ACKNOWLEDGMENTS	III
ABSTRACT	V
TABLE OF CONTENTS.....	VII
LIST OF FIGURES	VIII
INTRODUCTION.....	1
CHAPTER 1: FEMALES DETECT RAPID SOCIAL CONTEXT-INDUCED CHANGES IN SONG IN AN ACOUSTIC CATEGORIZATION TASK.....	6
ABSTRACT.....	6
INTRODUCTION	7
RESULTS.....	9
<i>Single song motifs contain subtle differences across social context.....</i>	<i>9</i>
<i>Female zebra finches can learn to classify stimuli by social context.....</i>	<i>13</i>
<i>Females use both categorization and memorization to identify social context.....</i>	<i>14</i>
<i>Different females use common strategies to classify individual motifs.....</i>	<i>19</i>
<i>Simple temporal features are not sufficient for social context categorization.....</i>	<i>20</i>
<i>Females can identify the social context of motifs sung after male LMAN lesion.....</i>	<i>26</i>
DISCUSSION.....	32
<i>Evolutionary significance.....</i>	<i>32</i>
<i>Potential mechanisms underlying social context differences.....</i>	<i>33</i>
<i>Broader significance</i>	<i>35</i>
METHODS.....	35
<i>Behavioral task and training.....</i>	<i>35</i>
<i>Classification of novel stimuli.....</i>	<i>38</i>
<i>Song stimuli.....</i>	<i>38</i>
<i>Data analysis.....</i>	<i>40</i>
CHAPTER 2: SPECTROTEMPORAL FEATURE DEPENDENCE ON NATURALISTIC STIMULUS STRUCTURE IN SONGBIRD AUDITORY FOREBRAIN.....	42
ABSTRACT.....	42
INTRODUCTION	43
RESULTS.....	48
<i>Spectrotemporal receptive fields in field L and CM.....</i>	<i>48</i>
<i>Filter shapes differ in response to different stimuli</i>	<i>52</i>
<i>Receptive fields derived from song responses have high predictive value.....</i>	<i>53</i>
<i>Receptive fields fail to predict responses to the across stimulus type.....</i>	<i>59</i>
<i>Narrow- and wide-spiking neurons have similar filter shapes</i>	<i>59</i>
DISCUSSION.....	64
METHODS.....	66
<i>Electrophysiology.....</i>	<i>66</i>
<i>Spike sorting.....</i>	<i>67</i>
<i>Histology</i>	<i>68</i>
<i>Stimulus generation.....</i>	<i>68</i>
<i>Receptive field estimation.....</i>	<i>70</i>
DISCUSSION	72
BIBLIOGRAPHY	76

List of Figures

CHAPTER 1

FIGURE 1: BEHAVIORAL STIMULI	11
FIGURE 2: BEHAVIOR RIG AND LEARNING CURVES.....	16
FIGURE 3: PERFORMANCE ON TRAINING AND PROBE STIMULI.....	18
FIGURE 4: PERFORMANCE ON INDIVIDUAL TRAINING AND PROBE STIMULI	22
FIGURE 5: SOCIAL CONTEXT AND TEMPO.....	24
FIGURE 6: SOCIAL CONTEXT CLASSIFICATION OF MOTIFS SUNG AFTER MALE LMAN LESION.	30

CHAPTER 2

FIGURE 1: SONGBIRD AUDITORY FOREBRAIN DIAGRAM.	47
FIGURE 2: STIMULI USED DURING RECORDINGS.	50
FIGURE 3. EXAMPLE NEURAL RESPONSES.	51
FIGURE 4. FIRST MID FILTER SHAPES.	55
FIGURE 5. EXAMPLE PREDICTIONS.	56
FIGURE 6: SIMILARITY BETWEEN SONG AND RIPPLE-NOISE FILTER SHAPES.	58
FIGURE 7: CORRELATION COEFFICIENT BETWEEN RESPONSE AND PREDICTED RESPONSE.	61
FIGURE 8: CROSS PREDICTIONS.	62
FIGURE 9: BIMODAL DISTRIBUTION OF SPIKE WIDTHS.	63

Introduction

The ability to perceive and categorize spoken words is a remarkable capability of the human auditory system. Songbirds are one of the few animal orders that face auditory tasks of similar complexity (Doupe and Kuhl 1999). Male songbirds learn to sing by listening to songs of others and their own vocalizations; in addition, both male and female songbirds show a lifelong capacity to memorize other birds' songs (Miller 1979, Stripling et al. 2003, Woolley and Doupe 2008). A network of forebrain auditory areas is thought to subserve these complex recognition tasks (Vates et al. 1996), including the equivalent of primary auditory cortex and several higher-order auditory areas. In male songbirds, the ascending auditory pathway ultimately feeds into a region of the brain containing some of the most exquisitely selective neurons known: HVC (acronym used as a proper name) neurons respond nearly exclusively to the sound of the bird's own song (Margoliash 1983, Margoliash and Fortune 1992, Theunissen and Doupe 1998). The songbird thus provides an ideal system for studying the neural mechanisms underlying the learning and recognition of complex sounds.

Auditory information first enters the brain through the cochlear nuclei. From there, it travels to two other nuclei within the brainstem, the lateral lemniscus and superior olive. In the midbrain, the lateral dorsal mesencephalon (MLd; homolog of the mammalian inferior colliculus) receives projections from all three of these nuclei and projects, in turn, to the thalamic region ovoidalis (Ov; medial geniculate nucleus homolog). Field L is the main thalamorecipient region in the songbird auditory forebrain, thus giving rise to comparisons with the mammalian primary auditory cortex (Wang et al. 2010). L2a, the main input and output region of field L, receives topographically organized input from Ov and sends a large projection to the mediocaudal neostriatum (NCM), maintaining topographical organization within this projection.

A number of other ascending pathways also exist, however. Field L's three other sub-regions, L1, L2b, and L3 all receive smaller projections from Ov and Ov itself sends a direct projection to NCM. L3 also sends a projection to NCM, and all four of field L's sub-regions project to the lateral portion of the caudal mesopallium (CM), the second of the two regions studied in chapter two. Within field L, L1, L2a, and L3 are all interconnected, and L2b is reciprocally interconnected with L2a. Lateral CM sends a projection to a small nidopallial region called nucleus interfacialis (NIf). NCM and lateral CM are both reciprocally connected to the medial portion of CM. In males, information flows into HVC via lateral CM and NIf.

Researchers have taken two main approaches towards understanding the ascending auditory pathway in songbirds. One approach is to start with the known properties of song-selective neurons in HVC and move “backwards,” in an attempt to understand how these responses emerge. This is typically done by characterizing the distributions of neuronal responses over a broad range of song stimuli, such as birds' own song, familiar conspecific songs, unfamiliar conspecific songs, heterospecific songs, and various manipulations thereof. Neurons selective for familiar conspecific songs have been found in CMM and NCM (Stripling et al. 2001, Gentner and Margoliash 2003, Meliza et al. 2010, Jeanne et al. 2011, Meliza and Margoliash 2012, Jeanne et al. 2013). Neurons in field L, the first of the two regions studied, on the other hand, appear to respond robustly to nearly all naturalistic auditory stimuli (Margoliash 1986, Lewicki and Arthur 1996, Sen et al. 2001). More recent work has begun to delineate a hierarchy of response types within field L, in which neurons in the thalamorecipient input layer (L2b) and L1 are less selective than those in L3 (Meliza and Margoliash 2012). These electrophysiological investigations have been aided by experiments analyzing the activation of immediate early genes in response to song stimuli. Unlike areas earlier in the auditory pathway,

CMM and NCM express the immediate early gene ZENK in many more cells in response to novel conspecific song than in response to heterospecific song or other auditory stimuli (Mello et al. 1992). Familiar and behaviorally relevant stimuli, such as father's song and mate's song, induce ZENK responses in a greater proportion of CMM cells than NCM cells (Terpstra et al. 2006, Woolley and Doupe 2008). Although the relationship between ZENK response and electrophysiological activity is not well understood, these experiments have served to identify brain regions of interest for further investigation.

The other main approach in auditory neuroscience has been to map systematically the response patterns of neurons. Traditionally, researchers have taken a simple stimulus, such as a tone, and modified it along one or two parameters, such as frequency and intensity, and then used the neural responses to these stimuli to construct a stimulus response function (Scheich et al. 1979, Müller and Leppelsack 1985). Over the past 30 years, researchers have developed another tool to allow characterization of responses to more complex stimuli. Reverse correlation (Eggermont et al. 1983) essentially works by capturing the average stimulus that occurred preceding each spike. The result provides a "snapshot" of the features that drove the neuron to fire during stimulus presentation. This technique has been used to map the auditory responses of neurons throughout the auditory midbrain and forebrain, in mammals and birds (Kowalski et al. 1996a,b, Theunissen et al. 2000, Sen et al. 2001, Miller et al. 2002, Depireux et al. 2001, Linden et al. 2003, Nagel and Doupe 2008, Woolley et al. 2009, Amin et al. 2010, Kim and Doupe 2011). Using stimuli other than random noise creates a problem for reverse correlation, because some frequencies and latencies appear more often than others. This problem has been mitigated by taking into account the correlations in the stimulus to produce a "linear-nonlinear" model that can be used, in principle, to predict how a neuron will respond to any novel stimulus (Theunissen

et al. 2000). Another approach, referred to as maximally informative dimensions (MID), explained in chapter two, can also be used to generate linear-nonlinear modeled receptive fields (Sharpee et al. 2004). Recent work has shown that these models can be highly specific to the statistics of the stimulus ensemble (Nagel and Doupe 2006, Sharpee et al. 2006) as well as the behavioral state of the animal (Fritz et al. 2003, 2005a, 2005b, Elhilali et al. 2007). While scientists have begun to map the stimulus and behavioral dependencies of receptive fields, very little is known about how these properties actually relate to perceptual behavior, especially in songbirds. The acoustic parameters governing birds' responses to song are not well understood, making precise perceptual predictions difficult.

In this thesis, I have tried to address some of the following questions: How are natural sounds represented by neurons in CM and field L? How do these response properties differ from those elicited by artificial stimuli? What subtle auditory cues might female birds use to distinguish social context? At the end of the dissertation, I discuss how the properties of neurons in field L and CM may be related to the auditory cues females use to determine social context, and propose experiments to test these hypotheses.

In chapter one, I show that songbirds can use subtle, social context-dependent cues to classify short excerpts of song. Using an operant behavioral paradigm (Beecher et al. 1994, Gentner and Hulse 1998, Nagel et al. 2010), I train female zebra finches to categorize single brief, characteristic elements (called motifs) of a male's song on the basis of the social context in which they were sung. Male song is known to be less variable in pitch across motifs when it is directed towards a female than when the male sings alone. Here, I show that females can detect a difference when they are asked to classify the song using only a single motif. Additionally, females can classify novel exemplars from familiar males, indicating that they use acoustic

properties common to motifs from each social context to classify the stimuli, rather than memorizing the correct response associated with each rendition. I also demonstrate that, at least for some males, bilateral lesions of a nucleus required for context-dependent modulation of song pitch variability do not prevent females' from identifying the social context in which the motif was sung.

In chapter two, I explore how song-selectivity emerges in the ascending auditory pathway by studying neurons in both CM and field L of zebra finches using both artificial acoustic stimuli and natural stimuli composed of elements of birdsong. By comparing responses to the natural and artificial stimuli, I discover that responses to the natural stimuli are more reliable, and that receptive fields calculated from the natural stimuli give a much better prediction of neuronal responses. Surprisingly, this stimulus dependence is similar in the two brain areas, contrary to our expectation that neurons in CM would be more selective for the natural stimuli than neurons in field L. We characterized neuronal responses with a linear filter followed by a nonlinear response function. In both field L and CM, and for both putative excitatory and inhibitory neurons, natural stimuli gave rise to more complex filter shapes than the artificial stimulus. Receptive fields calculated in response to the natural, song stimuli also had excellent predictive value, far surpassing that of the receptive fields calculated from artificial stimuli. Our results indicate that for many neurons in the songbird auditory forebrain, receptive field structure is highly dependent on stimulus statistics, and that receptive fields constructed in response to different stimulus classes bear surprisingly little information regarding responses to other sounds.

Chapter 1: Females detect rapid social context-induced changes in song in an acoustic categorization task

Abstract

The goal of this study was to determine whether female zebra finches could classify single motifs of male song on the basis of the social context in which they were sung. Social context alters the vocal signals of many species, including humans. Zebra finch (*Taeniopygia guttata*) males use song as a courtship signal. Each male's song contains a brief characteristic element (called a motif), which he repeats to form a bout. Several aspects of overall bout structure change when a male sings to a female (directed social context) rather than singing alone (undirected context). Female zebra finches strongly prefer recordings of directed over undirected song bouts (Woolley and Doupe 2008). Here we show that female zebra finches can distinguish single motifs extracted from directed bouts from those extracted from undirected bouts. Thus, female zebra finches may use the subtle cues found within a single motif to establish song preference. Additionally, females can classify novel motif exemplars from familiar males, indicating that they use acoustic properties common to motifs from each social context to classify the stimuli, rather than memorizing the correct response associated with each rendition. Using a naïve Bayes classifier, we find that the slight changes in motif speed across social contexts does provide sufficient information to allow classification of most males' directed and undirected motifs. However, it does not explain the pattern of behavior exhibited by females. Finally, we demonstrate that, at least for some males, bilateral lesions of a nucleus

required for context-dependent modulation of song pitch variability do not prevent females' from identifying the social context in which the motif was sung.

Introduction

In zebra finches, male song serves as a courtship signal. A male's song contains a single, learned motif (Figure 1A) which he repeats a variable number of times to form a 'bout' of song. Bouts are usually preceded by a number of introductory notes, which are similar to calls. To the casual listener, each motif the bird sings sounds nearly identical.

Males do not restrict their singing to courtship, however; they also sing when they are alone. Although the function of these 'undirected' song bouts is not entirely understood, many hypothesize that undirected song may serve as 'practice', or 'exploration' used to maintain and fine-tune the motor system (Kao et al. 2005). Previous studies have shown that several aspects of the male's song structure change when he sings 'directed' song to a female: he sings more introductory notes before beginning a bout, repeats his motif more times, sings faster, and sings syllables with more consistent spectral structure (Sossinka and Böhner 1980; Kao et al. 2005; Ölveczky et al. 2005; Kao and Brainard 2006). A specialized cortical-basal ganglia loop is known to underlie some, but not all of these social context-dependent effects. Damaging the output nucleus of this loop removes social context differences in song speed and spectral variability, while differences in number of introductory notes and motif repeats remain (Kao and Brainard 2006).

Females strongly prefer to hear bouts of female-directed song to undirected song (Woolley and Doupe 2008). However, it remains unclear whether this preference is caused by aspects of the male's performance controlled by the cortical-basal ganglia loop, or by those features that are independent of its presence. In addition, the spectral variability of song is

usually measured as the variability in the fundamental frequency of a single syllable across renditions of the motif (Kao et al. 2005). It is not known whether changes in the variability of spectral structure within the motif are present, or if they contribute to female preference. Thus, we began with the hypothesis that females were using spectral variability across motifs to classify directed and undirected bouts of song.

Here, however, we use an operant paradigm to show that female zebra finches are highly sensitive to social context-driven differences in single motifs of male song. We used generalization of a learned discrimination (Beecher et al. 1994; Gentner and Hulse 1998) as a paradigm for studying social context recognition. In this operant task, a bird learns to discriminate two stimulus sets; in this case, one set was recorded while the male directed his song to a female, and the other while he was alone. Birds are free to use any of the parameters available in the training stimuli to classify them. After learning, birds continue to perform the discrimination task, but are intermittently tested using novel “probe” stimuli. How the bird categorizes these novel renditions can thus reveal which aspects of the original songs are used for categorization. Probe stimuli are rewarded without respect to the bird’s choice, ensuring that the bird cannot 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, i.e., the novel rendition still falls within the perceptual category defined by the bird in the task.

Using this behavioral paradigm, we found that female zebra finches can distinguish directed and undirected song segments containing only a single rendition of the male’s signature motif. Ten of the 11 females tested were able to classify novel directed and undirected probe motifs correctly, indicating that they had formed perceptual categories in order to classify the

training stimuli. Although they performed well above chance on these probe motifs, the perceptual categories they had formed were not sufficient to generalize across all renditions; most birds performed significantly worse on the probes than they had on the training stimuli. In addition, females were able to classify directed and undirected probes from most male singers even when recordings were made after the output nucleus of the cortical-basal ganglia loop had been damaged. Thus, the differences across social contexts perceived by females in this task can only be partly explained by the addition of variability in motor performance added by the cortical-basal ganglia loop.

Results

Single song motifs contain subtle differences across social context

We recorded song bouts from five male zebra finches as they sang to females (directed song) and as they sang alone (undirected song). Each bout of song contains a number of introductory notes followed by the male's characteristic motif, which he repeats a variable number of times. For the stimuli used in the discrimination task (see below), we selected only the first motif in a bout, as well as the two preceding introductory notes; thus, in the course of these experiments, females were limited to the spectral and temporal cues present within a single motif. From the song bouts of each male, we selected 23-29 directed stimuli and the same number of undirected stimuli. Figure 1A contains example directed and undirected stimuli from one male. From the oscillograms and spectrograms depicted, the striking similarity between directed and undirected stimuli is apparent, especially as compared to the differences between song bouts produced by two different males (Figure 1B). The stimulus amplitudes were set to 80

dB RMS (Figure 1C). The overall power spectra were similar for the directed and undirected stimuli (Figure 1D).

To the casual listener, the most obvious differences between directed and undirected song appear at the level of a bout. When singing to a female, the male sings more introductory notes and repeats his motif more times (Sossinka and Böhner 1980). Also, the male sings a more consistent version of the motif across repeats in the directed condition, as indicated by the reduced fundamental frequency variability of syllables across renditions (Kao et al. 2005). This difference in variability between directed and undirected song is apparent in our stimulus sets. Across all motifs in the directed set depicted in Figure 1, the fundamental frequency of syllable ‘c’ varies less than it does across the motifs from the undirected set (coefficient of variation for directed syllables = 0.0069; undirected = 0.016, $p < 0.001$). However, none of the information above is available to the female performing the discrimination task, as she must make her decisions based solely on a single motif per trial.

Other temporal and spectral cues exist within the stimulus sets that may be used to aid discrimination. On average, males sing slightly faster during directed song than when they sing alone. In the stimulus set depicted in Figure 1, the directed song is shorter than the undirected song (mean directed: 1052 +/- 24ms; mean undirected: 1068 +/- 53ms), but the distributions are highly overlapping, thus this cue alone cannot be used to determine the social context in which the motif was sung (Figure 1E). Recent work in our laboratory has shown pitch varies more in the undirected case even within single syllables (Kojima et al., *in preparation*). This effect holds true for our stimulus set as well, although the distributions are again highly overlapping.

Figure 1: Behavioral stimuli.

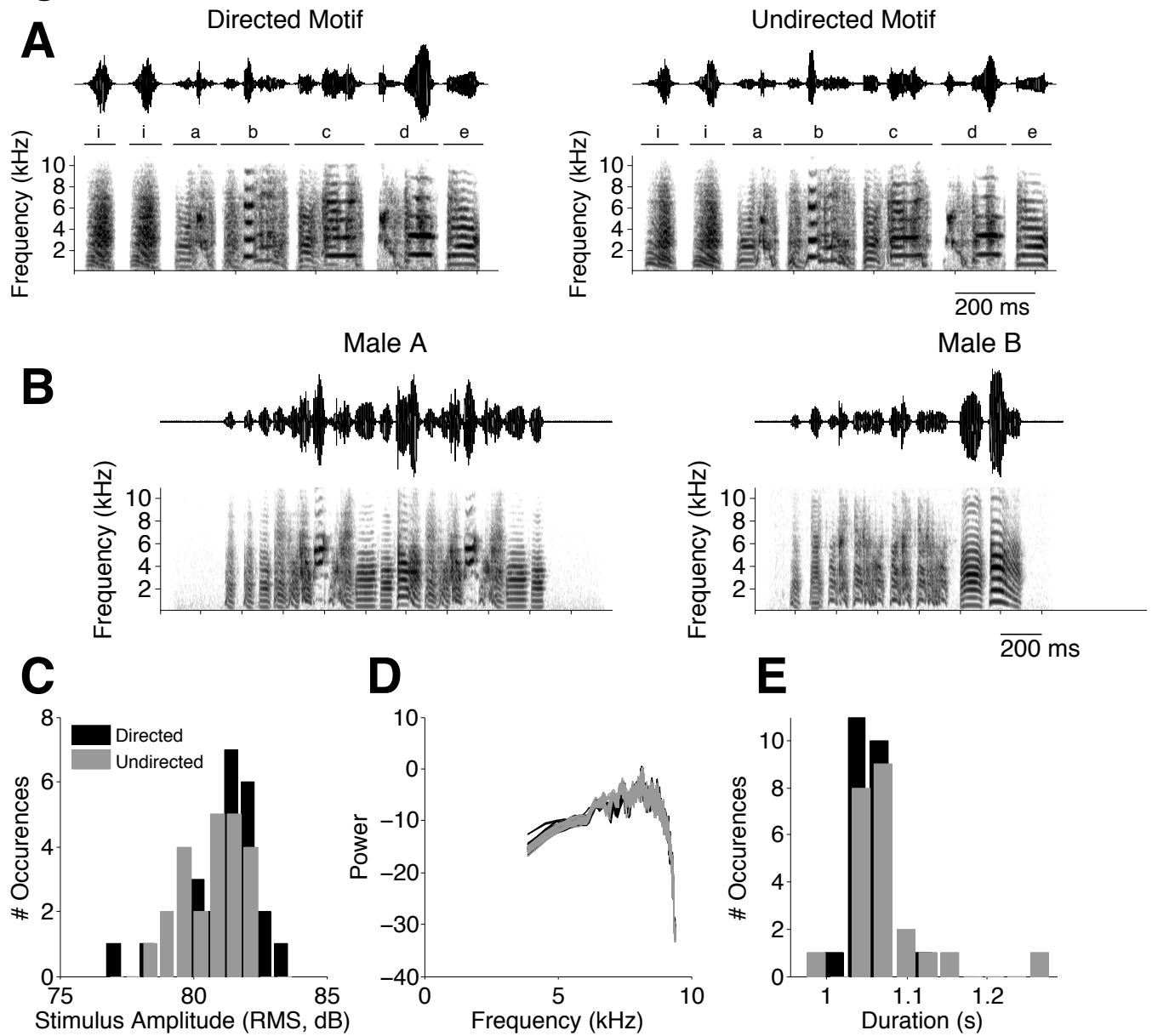


Figure 1: Behavioral stimuli. **A)** Example directed (left) and undirected (right) motif stimuli from Male 1. Syllables are indicated by lines and labeled with letters. Spectrograms (lower part of figure) correspond to the waveform in the upper part of the figure. **B)** Full song bouts from two different males. Stimulus durations shown are between 1 and 2 seconds (see 200ms scale bar at bottom right). **C)** Stimulus amplitudes. Mean amplitudes for all directed (black) and

undirected (gray) training stimuli from Male 1. **D) Power spectra.** Power spectra for stimuli from the two different social contexts were very similar. **E) Stimulus durations.** Distributions of stimulus durations were also similar for the two stimulus sets.

Female zebra finches can learn to classify stimuli by social context

Previous work in our laboratory (Nagel et al. 2010) has shown that female zebra finches can quickly learn to classify song bouts from two different males using an operant procedure. Example stimuli from this task are shown in Figure 1B. Each bird is housed in an isolated sound box containing an operant conditioning rig with three perches (Figure 2A). After learning a series of simpler tasks designed to acclimate the bird to the rig and the procedure, the bird is introduced to the two stimulus sets. The bird hops on the central song perch to initiate a trial. The speaker, located directly behind the song perch, then plays a randomly selected stimulus from one of the two sets. The bird then hops on to one of the two response perches flanking the song perch. If the bird hops on to the perch associated with the same set as the stimulus that just played, the feeder, located below the song perch, comes up for four to five seconds, allowing the bird to feed. If the bird hops on to the other perch, the feeder does not come up and the song perch becomes inactive, preventing her from initiating another trial for 30-90 s.

Female zebra finches can also learn to classify directed and undirected motifs recorded from a single male. The open symbols in Figure 2A depict the performance of a bird as she learned to discriminate between songs recorded from two different males. She rapidly mastered the task, achieving near perfect performance after 3-4 days of training. The closed symbols in Figure 2B represent the behavior of a bird as she learned to classify a male's song motifs as either directed or undirected. Across the first few days, the bird quickly rose above chance (dotted line, 50%) for both directed and undirected stimuli. The bird's performance continued to improve over the first 7-10 days, stabilizing at about 80% correct for both stimulus types. Although the bird learned to classify stimuli as either directed or undirected, she did not do so

nearly as quickly as birds learning to discriminate between song bouts from two different males. Her performance also stabilized at a lower level than the bird learning the two-male task. This lower performance and slower learning curve suggest that the directed and undirected single motifs are much more challenging to classify than the songs of two different male birds. Despite the apparent difficulty of this task, all 6 of the birds tested learned to classify the directed and undirected motif stimuli better than chance (Figure 2C).

In order to ensure that this classification ability was not specific to one male's directed and undirected song, we trained a total of 12 females on overlapping subsets of songs from these five males. Each bird was tested on one to four of the males' songs. All 12 females learned to classify all the directed and undirected motifs that they were tested on (Figure 2D).

One male's directed and undirected motifs were more difficult to learn than motifs from the other three (Figure 2D, Male 3). This male also had the shortest motif (4 syllables, mean motif duration 767 +/- 33ms) of the five males (mean motif durations 1060 +/- 41ms, 840 +/- 14ms, 1144 +/- 31ms, and 1005 +/- 33ms). These results indicate that the ability to classify motifs as directed or undirected is not specific to a single male's song, but that some males' motifs are more difficult to classify than others.

Females use both categorization and memorization to identify social context

A female bird could use two different strategies to complete the classification task. First, she could simply memorize the rewarded response for each motif in the training set. Alternatively, she could learn to categorize the stimuli based on the acoustic properties that distinguish directed motifs from undirected motifs and respond accordingly. In order to determine which approach the females used, we tested the responses of 12 birds to novel motifs that were not included in the original training sets. These new motifs from familiar males were

presented as probe stimuli (see *Methods*), meaning that they were randomly interspersed between training trials at a low rate (10–20% of trials), and were rewarded at a fixed rate equal to the overall reward rate regardless of which response perch the female chose. Since no information about the “correct” category of the probe was given to birds performing the task, their choices indicated the category they would assign to probes on the basis of their experience with the training set.

All 12 females performed above chance on all the directed and undirected probes on which they were tested. Figure 3A depicts the behavior of the females on all five of the different males’ probes. Each cross represents the performance of one female on one male’s directed and undirected motifs. Performance on the probe trials is plotted on the ordinate. Performance on the training trials in which these probes were embedded is on the abscissa. All of the points are above the horizontal dotted line, indicating that all of the birds performed above chance on all of the probe stimuli. This demonstrates that the females learned to associate each perch with motifs sung within a particular social context, rather than with the specific motif exemplars used in training. All of the points are also below the line of unity, meaning that the birds’ performance on the probes was not as high as it was on the training stimuli. This indicates that familiarity with exact motif variants aided behavioral performance on the classification task. Each bird’s probe response rates and reaction times matched her performance on training trials quite well (Figures 3B and 3C), consistent with the idea that the randomly interspersed probe trials were seamlessly integrated into the ongoing task structure, and that the bird could use familiar categorical acoustic cues to classify the novel probes. Thus, all birds used a combination of categorization based on common cues as well as memorization of specific training stimuli to perform the classification task.

Figure 2: Behavior rig and learning curves.

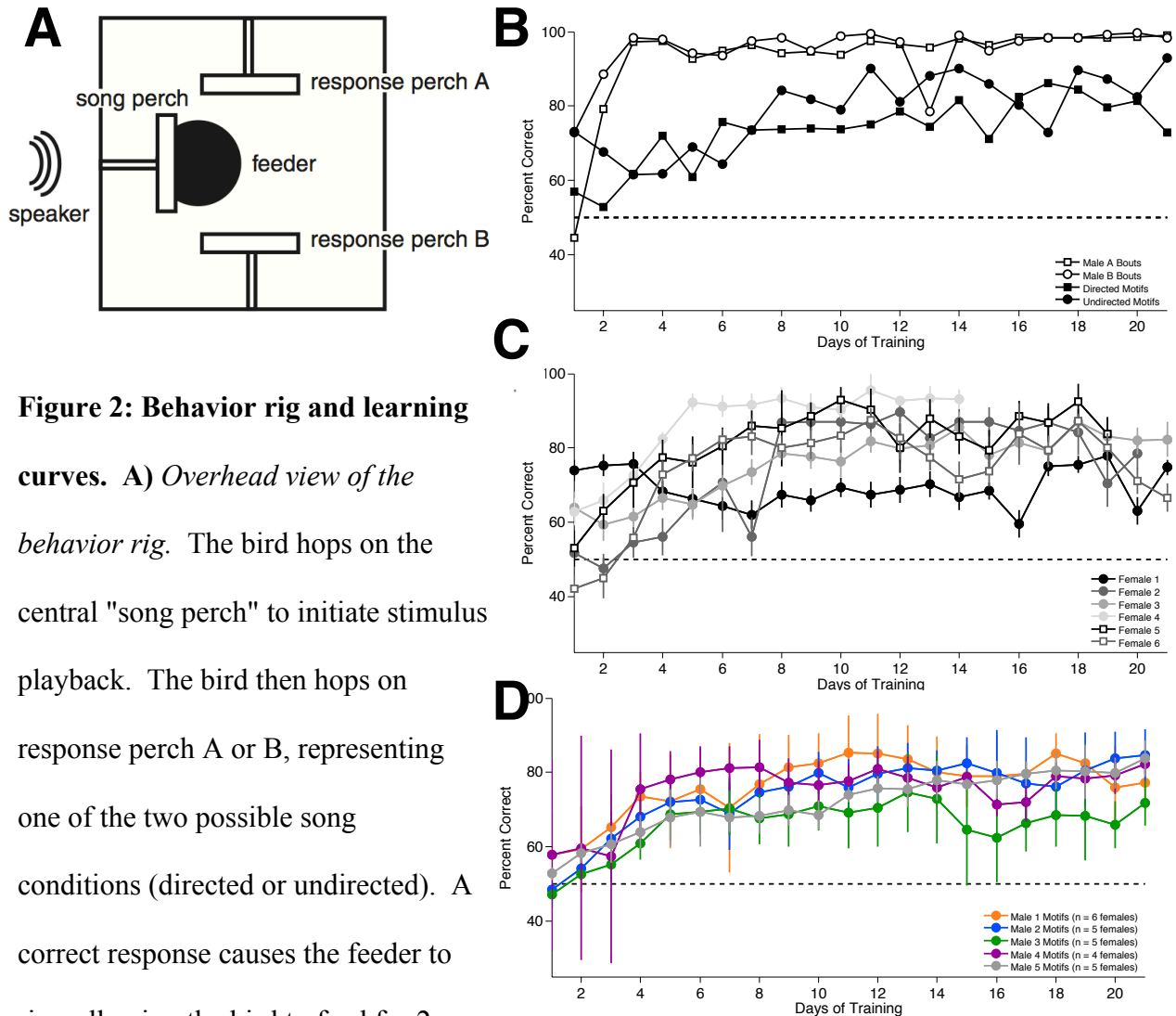


Figure 2: Behavior rig and learning

curves. **A)** *Overhead view of the behavior rig.* The bird hops on the central "song perch" to initiate stimulus playback. The bird then hops on response perch A or B, representing one of the two possible song conditions (directed or undirected). A correct response causes the feeder to rise, allowing the bird to feed for 2-

5s, while an incorrect response results in a brief time-out. **B)** *Example learning curves for two birds.* Classification performance improved as birds learned across days of training. The dashed line represents chance performance (50% correct). Open symbols represent the performance of a bird learning to classify song bouts from two different males (Example stimuli presented in Figure 1B). Closed symbols indicate performance as a bird learned to classify directed and undirected motifs of a single male (Example stimuli presented in Figure 1A). Birds took longer

to learn the social context discrimination, and performance stabilized at a lower rate, indicating the difficulty of this task. **C)** *Learning curves for six females learning to discriminate stimuli from Male 1.* Classification accuracy and learning rates varied across females learning to classify the same stimuli. **D)** *Mean learning curves for all females as they learned to classify stimuli from each of the five males.* Across females, the social context of some males' motifs was harder to classify than others. For example, Male 3's motifs (green) were misclassified at a higher rate than the other males' motifs.

Figure 3: Performance on training and probe stimuli.

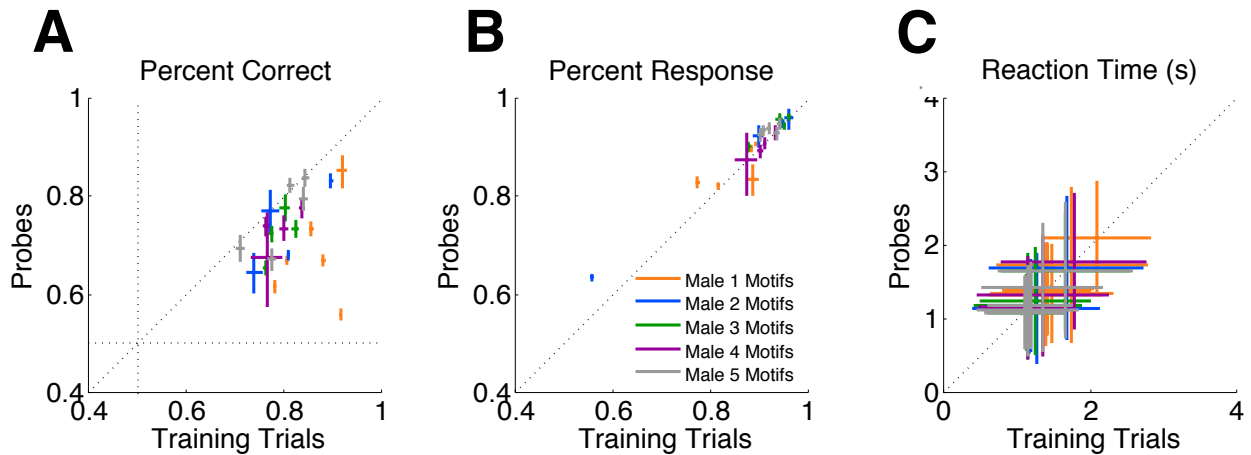


Figure 3: Performance on training and probe stimuli. Comparison of training trials (ordinate) vs. probe trials (abscissa) for all females and all stimulus sets. Stimuli from the five males are depicted in different colors. After a bird reached criterion on the training trials, probe trials were randomly interleaved at a low rate (10-20% of trials). Probe trials were rewarded at a constant rate, regardless of response. These trials served to determine whether the bird could classify novel motifs, or if she had simply memorized the correct response for each stimulus. **A)** *Percent correct.* Although probe trial performance was generally lower than training accuracy, both were substantially above chance for all five stimulus sets. **B)** *Percent response.* For all five stimulus sets, response rates were similar for probe and training trials. **C)** *Reaction times.* Mean time to response for stimuli from each male (in seconds). Reaction times were similar for training and probe stimuli.

Different females use common strategies to classify individual motifs

A closer look at the birds' classification of individual stimuli within the training sets further supports the notion that birds rely on categorical rather than memorization-based techniques to master the task. Figure 4A depicts the responses of four females to all of the directed and undirected training stimuli from Male 1. For each female, the percent correct is remarkably similar across most of the stimuli. This suggests that the females are using similar techniques to classify all of the training stimuli, rather than memorizing a subset of the exemplars and failing to recognize others. For the stimulus set depicted in Figure 4A, there are two notable exceptions to this pattern; directed motifs one and 15 are often incorrectly classified as undirected stimuli by all four females. Despite the fact that all of the training stimuli are reinforced equally, all of the females tested had a tendency to persistently classify the same stimuli incorrectly. These shared 'failures' suggest that, not only are the birds using categorical rather than individual acoustic properties to classify the stimuli, all females are using similar categorical cues to distinguish directed and undirected motifs.

Classification patterns in response to probe stimuli also suggest that females use common categorical cues to perform the task. Probe stimuli represented 10-20% of behavioral trials, and females were rewarded at a constant rate regardless of which response perch they chose. Figure 4B shows the responses of the same four females from Figure 4A to all of Male 1's directed and undirected probe stimuli. Each female's percent correct is less consistent across stimuli than it was in response to the training trials, indicating that the categorical cues gleaned by the females during the training trials were more successful on some probe motifs than on others. Although there is also more variability across females in the probe trials than there is in the training trials, the birds still share some behavioral patterns. In particular, directed motifs 10 and 22 (Figure

4B) are consistently classified as undirected motifs by all females tested. Thus, even when the birds are not receiving any feedback on the ‘correct’ response, they display similar patterns of behavior. This suggests that, although many subtle acoustic cues may differentiate directed and undirected motifs, females use a similar set of acoustic cues to classify these motifs.

Simple temporal features are not sufficient for social context categorization

As a male zebra finch sings, he typically produces 1-2 syllables with each expiration and pauses to breathe in between syllables (Suthers et al. 1999). During directed song, males tend to exhale more quickly, producing shorter syllables, while maintaining similar inspiratory durations (Cooper and Goller 2006). This difference in timing is evident in our training stimuli. Figure 5a depicts the durations of all syllables used in the training sets from all five males. Each male’s syllables are plotted using a different pair of colors; a slightly darker shade is used to distinguish directed from undirected syllable durations. The increase in syllable speed during directed song is especially evident in Male 1’s motifs (plotted in orange). This striking difference in tempo suggests that syllable durations alone may provide sufficient information for classification of directed and undirected motifs.

In order to determine whether changes in syllable duration across social contexts could be responsible for females’ ability classify directed and undirected motifs, we trained a naive Bayes classifier using the durations of each syllable in each motif from Male 1’s training set as the only input parameters. We then used the model to place these training stimuli, as well as the probe stimuli, into one of two clusters. The model’s clusters corresponded quite well to the directed and undirected conditions (Figure 5B). On the training stimulus set, the (over-trained) model performed even better than any of the female zebra finches had performed (Figure 5B, orange symbols). The same classifier performed above chance on the probe stimuli, but

classified fewer probes correctly than all but one of the female birds. When we applied the same approach to the stimulus sets collected from Males 2, 3, and 4, we found that the classifier did quite well overall. Male 5's motifs, however, were not classified very well using this approach (Figure 5B, gray symbols). Model performance was barely above chance on both the training and probe stimulus sets, unlike the birds' performance. These results suggest that, although syllable durations may help female zebra finches perform the task, they do not provide sufficient information to allow the birds to classify all of the stimulus sets.

The naive Bayes classifier also clustered individual training and probe stimuli differently than the birds. Although the model's performance was similar to the birds' behavior overall, the patterns of misclassification observed across all females in both the training and probe stimulus sets were completely different from the model's clusters. Figure 5C depicts the behavioral responses of 4 female birds to the individual training stimuli, as well as the predictions made by the naive Bayes classifier. The model misclassified directed stimulus 12, while the birds misclassified stimuli 1 and 15. Testing the model on the individual probes and the other males' motifs yielded similar results. These results suggest that, although syllable durations provide sufficient information for a bird (or a computer) to classify them as directed or undirected, this is not the only information that females actually use when performing the task. In addition, this approach entirely failed to cluster stimuli from one of the five males according to social context. This further supports the notion that simple temporal cues are not sufficient to allow social context classification of single motifs.

Figure 4: Performance on individual training and probe stimuli.

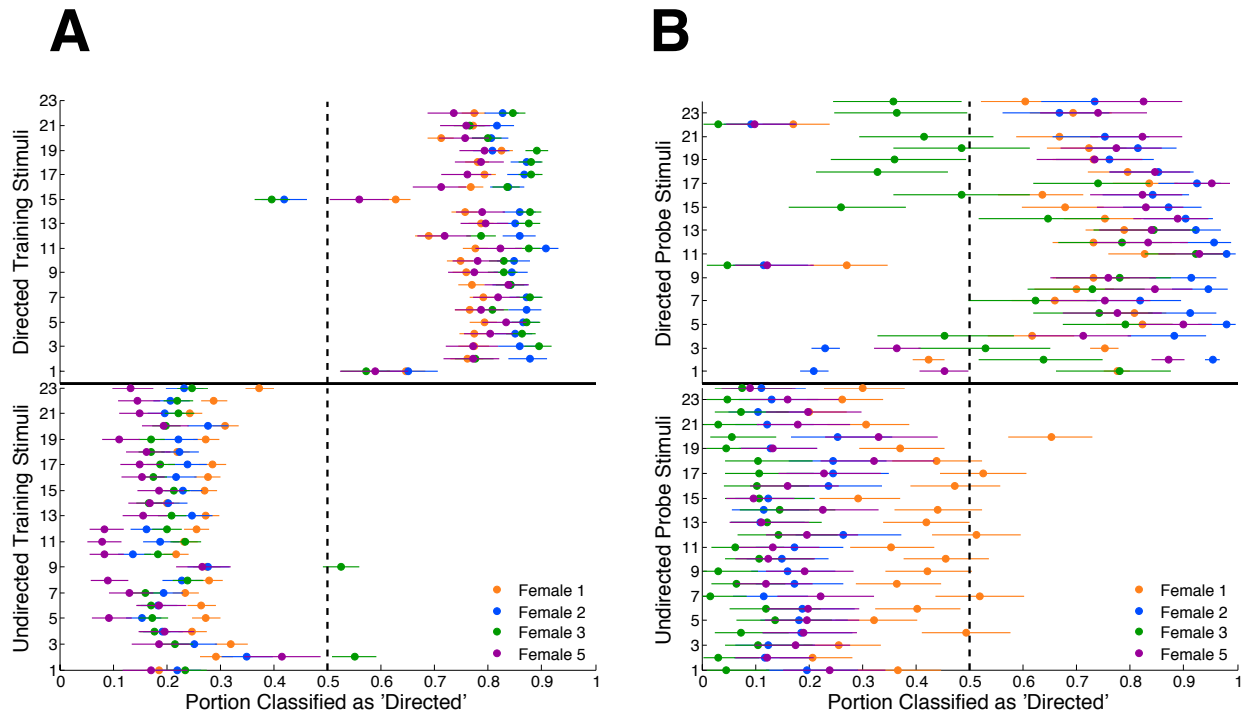


Figure 4: Performance on individual training and probe stimuli. A) *Performance of four females on all training stimuli from Male 1.* Each female’s responses are depicted in a different color. Individual training stimuli appear on the ordinate with directed stimuli above the heavy black line and undirected stimuli below the line. The proportion of trials in which the female classified each stimulus as directed is plotted on the abscissa. Responses that fall to the right of the vertical dashed line indicate that the bird classified the stimulus as directed, while responses to the left of the dotted line correspond to stimuli classified as undirected. All four birds correctly classified most directed and undirected training stimuli, with a few clear exceptions, such as directed stimuli 1 and 15. **B)** *Performance of the same 4 females on individual probe stimuli from Male 1.* Plot colors and symbols are the same as in A. Birds answered a lower portion of probe trials correctly, and were presented with many fewer probe trials than training

trials (10-20%); however, their overall performance was above chance. As with the training stimuli depicted in *A*, all four females persistently misclassified a few probe stimuli, most notably directed stimuli 10 and 22.

Figure 5: Social context and tempo.

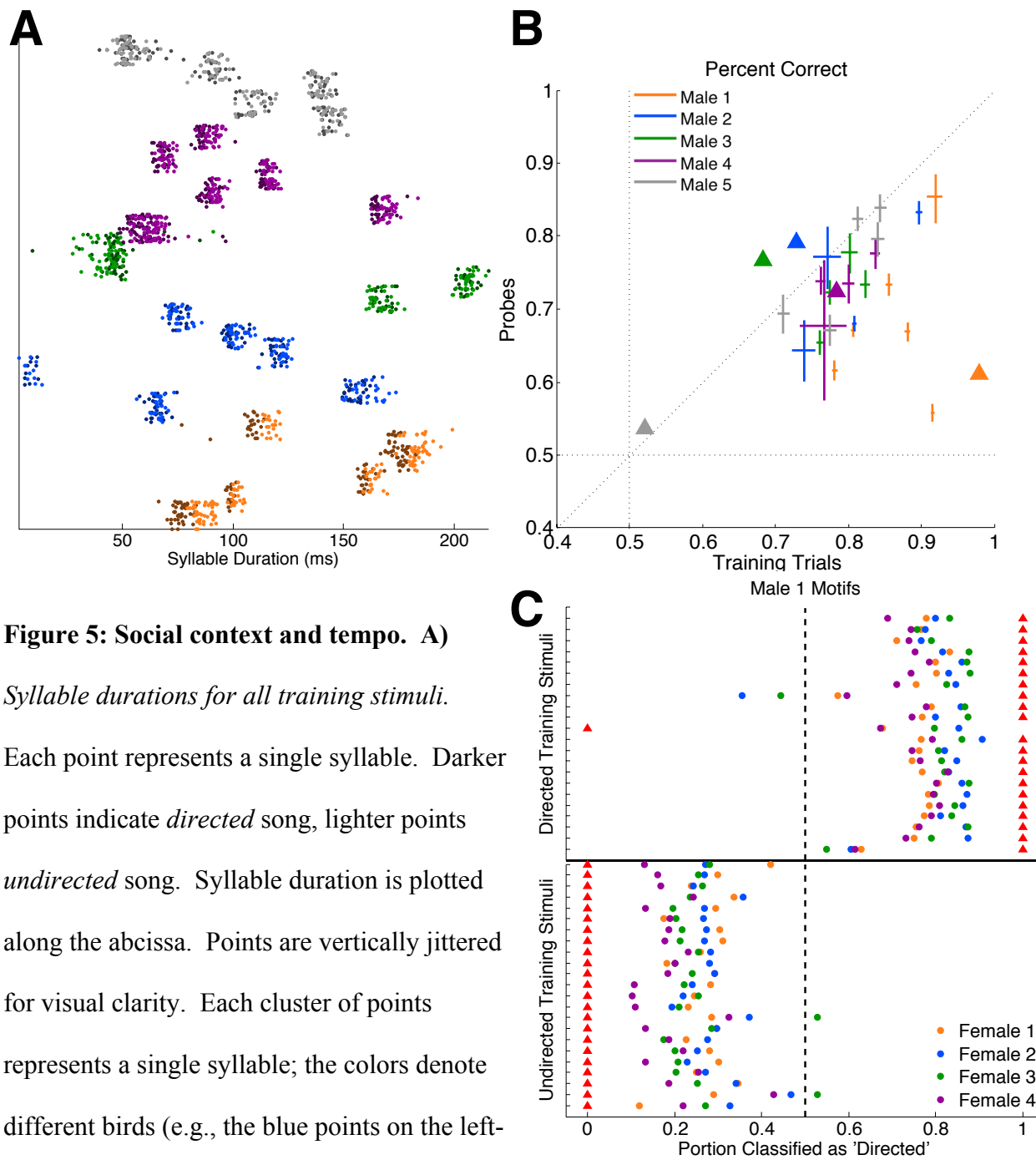


Figure 5: Social context and tempo. A)

Syllable durations for all training stimuli.

Each point represents a single syllable. Darker points indicate *directed* song, lighter points *undirected* song. Syllable duration is plotted along the abscissa. Points are vertically jittered for visual clarity. Each cluster of points represents a single syllable; the colors denote different birds (e.g., the blue points on the left-hand side represent a Male 2's shortest syllable).

B) Performance of a naïve Bayes classifier trained on syllable durations. Female performance on training and probe stimuli is plotted here as in Figure 3A. The added triangles indicate the performance of a naïve Bayes classifier trained

only on the syllable durations of the training stimuli. The color of the triangle indicates the stimuli on which the classifier was trained. For four of the five males, the overall performance of the classifier was a reasonable approximation of the female birds' performance. For Male 5 (gray), however, a classifier trained on syllable durations failed to classify the both the stimuli on which it was trained as well as the novel probe stimuli. **C) Performance of the naïve Bayes classifier on individual training stimuli.** The performance of four females on training stimuli from Male 1 is plotted as in Figure 4A. The added red triangles indicate the performance of the (over)-trained classifier on these same stimuli. The algorithm correctly classifies all but one directed stimulus. Females consistently misclassify a different subset of the stimuli than the algorithm.

Females can identify the social context of motifs sung after male LMAN lesion

Destruction of a male zebra finch's LMAN (lateral magnocellular nucleus of the anterior nidopallium), the output nucleus of a basal ganglia-thalamo-cortical loop specialized for song production, prevents social context differences in song spectral variability across motifs (Kao and Brainard 2006) and within single syllables (Kojima et al., *in preparation*). In particular, the variability of the fundamental frequency across renditions, which is normally elevated during undirected singing, is reduced to the same level as that found in directed song after LMAN removal (Kao and Brainard 2006).

In order to determine whether LMAN-dependent cues influence females' ability to categorize directed and undirected motifs, we analyzed the responses of the females to songs of three of the five males after LMAN lesions. After the males recovered from surgery, we again recorded their directed and undirected songs and created new probe stimuli similar to those we had made before lesion. On each day in which probe trials were interspersed with the training trials, the female randomly heard either pre- or post-lesion probes. Given that undirected song becomes more 'directed-like' after LMAN destruction, one might predict that females would over-classify the undirected motifs from a male after his LMAN was destroyed as directed, thus reducing her overall performance below chance.

We found that female zebra finches retained the ability to classify some male motifs as directed or undirected even after the males' LMANs had been destroyed. Figures 6A-C directly compare the females' performance on pre- and post-lesion probes. Each plot shows the responses of all females tested to probes from one of the three lesioned males; each cross represents the performance of a single female. On the abscissa, the performance of females on the pre-lesion probes is plotted; the ordinate shows their performance on the post-lesion probes. All

five females trained on the songs of Male 2 performed better than chance on his post-lesion probes (Figure 6B), and three of the five females trained on Male 1's motifs also performed above chance on his post-lesion probes. However, for both Males 1 and 2, four of the five females were significantly less able to identify the social context of their motifs after lesion. These results suggest that, although destruction of a male's LMAN significantly reduced the females' ability to identify the social context in which his motifs were sung, some categorical cues remained.

The females' classification performance on post-lesion probes varied widely across males. All five females tested on Male 3's post-lesion probes performed at or near chance (Figure 6C) on his stimuli, suggesting that LMAN destruction had removed all cues previously used to identify social context within the motif. Stimuli from all three males were tested on overlapping subsets of the females, confirming that differences in performance across stimuli from different males reflected true differences in the motifs themselves and not the individual females' abilities to classify them. These results also reinforce the notion that females used some common strategies to accomplish the task.

Female responses, however, never dipped below chance on any post-lesion probe stimuli. This suggests that, while removing LMAN reduced the cues available to females for social context classification, it did not make undirected songs sound more 'directed-like' to them.

Post-lesion probes differed from pre-lesion probes in a number of ways. In addition to lacking signals from LMAN, males' post-lesion songs were recorded 3-12 weeks after his pre-lesion training-set songs. Pre-lesion probe stimuli were randomly chosen (without replacement, of course) from the same recording sets that were used to produce the training stimuli. Post-lesion probes, were, by necessity, taken from completely different recording sessions. It is

possible that conditions specific to the recording session, such as temperature, humidity, and/or microphone settings could be responsible for the decrement in female performance, rather than the loss of LMAN's contribution to the motor pathway. It is also possible that the passage of a few weeks or months might have altered the males' songs significantly.

In order to determine if the changes in female performance were due solely to the loss of the male's LMAN, or if recording conditions or the passage of time had also contributed to the change in male song observed by females, we constructed two control probe sets from Male 5 (whose LMAN was not ablated). The directed and undirected song bouts that were used to construct the training and original probe stimulus sets from Male 5 were sung over the course of 5 days. We extracted the training and original probe stimuli from days 1, 2, 4, and 5. We created new 'mid-probe' stimuli from songs recorded on day 3. Figure 6D clearly shows that females classified the new mid-probe stimuli just as well as they did the original probe set from Male 5; the results for all five females fall near the line of unity. This indicates that the precise conditions under which songs were recorded did not result in differences among the probes large enough to alter female performance. We then constructed a third probe stimulus set from Male 5. The 'late-probe' stimulus set was constructed from songs recorded 7 weeks after the songs used to create the training, original probe, and mid-probe stimulus sets. Unlike the mid-probe stimuli, the late-probe stimuli were significantly more difficult than the original probe set for the same five females to classify correctly. All five females performed significantly above chance on the late-probes, but their decrease in performance with respect to the original probe set demonstrates that the mere passage of time (on the male's part) could be responsible for a portion of the effect observed after LMAN removal in Males 1-3. What remains clear, however,

is that removing a male's LMAN does not necessarily eliminate a female's ability to determine the social context under which the male's song was recorded.

Figure 6: Social context classification of motifs sung after male LMAN lesion.

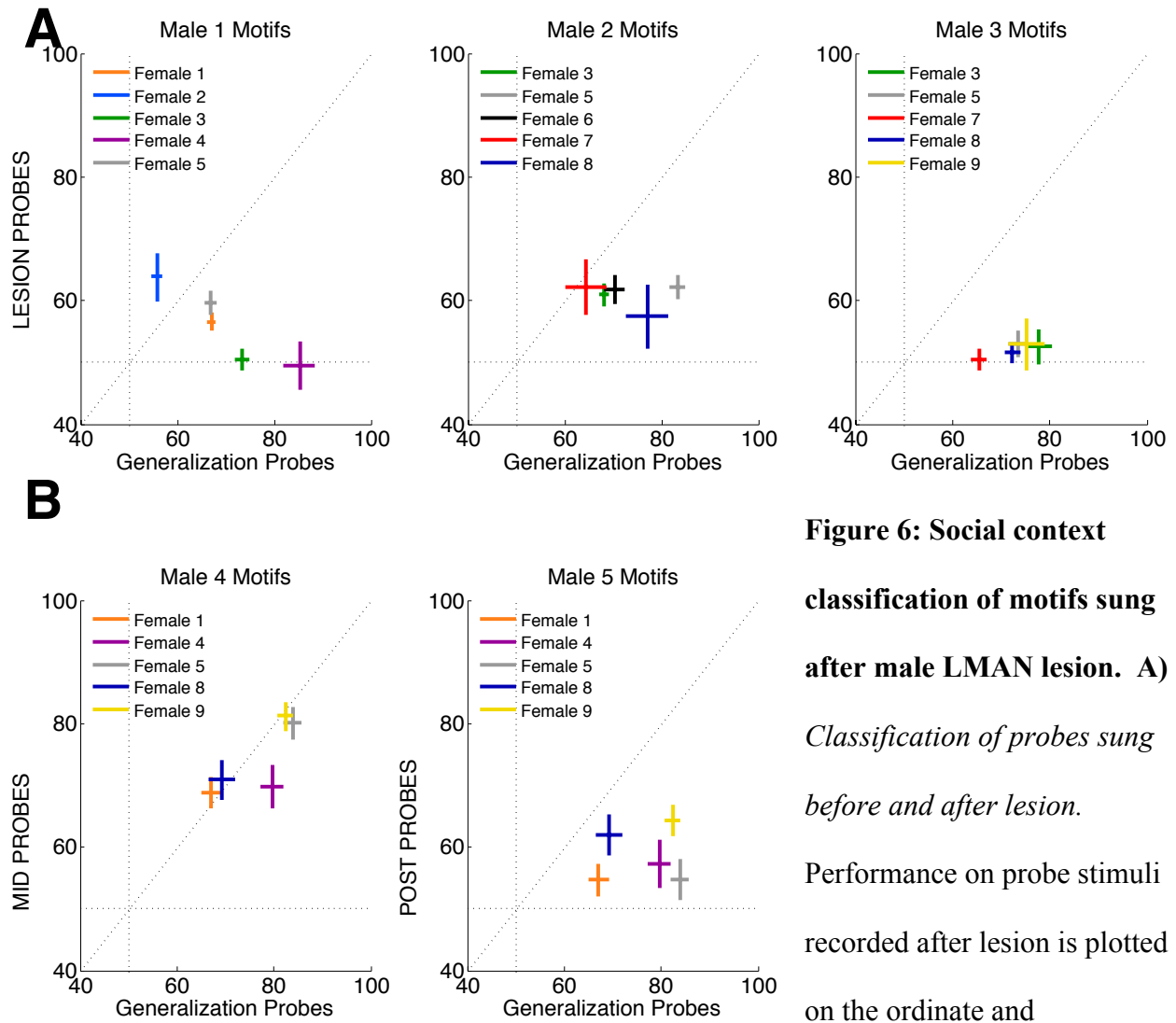


Figure 6: Social context classification of motifs sung after male LMAN lesion. A)

Classification of probes sung before and after lesion.

Performance on probe stimuli recorded after lesion is plotted on the ordinate and

performance on probes recorded before lesion is on the abscissa. Colors correspond to different females. Some females were no longer able to correctly classify stimuli based on social context after Male 1's LMAN lesion (purple and green), while other females performed well on these probes (leftmost plot). All females performed less well on Male 2's post-lesion stimuli than they did on probes from the original set (middle plot), although they still performed well above chance (dotted horizontal line). None of the females were able to classify Male 3's post-lesion

probe stimuli (rightmost plot). **B) Classification of control probes.** To control for changes in recording quality across days, we constructed a stimulus set for Male 5 such that songs from recording days 1, 2, 4, and 5 were used for the training and original probe sets, while a “mid-probe” set was made from day 3 songs. For all females tested, performance on mid-probes was nearly identical to that on the original probes (left plot). Post-lesion songs from Males 1-3 were recorded three to 12 weeks after LMAN lesion. To control for potential changes in recording quality over time, we also made recordings separated by seven weeks without performing any surgical lesions on Male 5. Female performance on these probes was significantly worse than on the original probes, indicating that the passage of time alone may be responsible for much of the lesion effect depicted in *A*.

Discussion

We began with the hypothesis that females were using spectral variability across motifs to classify directed and undirected bouts of song. However, we found that females could reliably use single motifs to identify social context. Performance on probe stimuli was well above chance, indicating that the birds were classifying the stimuli on the basis of a categorical perception; however, the lower overall performance on probes indicated that task performance was enhanced by an element of memorization. Shared patterns of individual misclassification supported the notion that birds were using common cues to classify the motifs. Further confirmation that females were using cues other than spectral variability to classify the stimuli came with the finding that lesions of the male's LMAN, known to eliminate differences in the variability of the FF of syllables across social contexts, did not completely remove the female's ability to classify his motifs. Finally, the females' performance could not be fully explained by simple differences in tempo. These results suggest that female zebra finches can use a variety of acoustic cues to reliably classify the social context in which a male produced his song.

Evolutionary significance

Unlike many other songbird species, in which song is used for both mating and territorial purposes, male song acts primarily as courtship signal for zebra finches. When a male sings to a female, the spectral variability of his motifs subtly decreases across renditions (Hessler and Doupe 1999, Kao et al. 2005). Females prefer bouts of song recorded when the male was singing to another female over bouts recorded as he sang alone (Woolley and Doupe 2008). We began with the hypothesis that females were using variability across renditions as a measure of fitness. This notion fits with the popular idea in evolutionary biology that regularity and symmetry are signs of fitness that animals use during sexual selection (citation). We

hypothesized that female zebra finches would be unable to discriminate single motifs of directed and undirected song, since all of the most obvious differences across social context wouldn't be evident at the level of the single motif. However, females were capable of learning the classification task, indicating that females are sensitive to cues within a single rendition. This suggests that these other cues may also play a role in female preference, although behavioral preference experiments will be necessary to determine whether these cues are important to females.

Potential mechanisms underlying social context differences

Although we have found that females can identify the social context in which a motif was sung, the motif features females use to perform this task remain unclear. Our results indicate that while tempo changes are clearly an important cue for classifying the most males' motifs, this feature cannot fully explain the females' behavior. The answer may come from some combination of the following four features: (1) fast pitch fluctuations within individual syllables, (2) the mistuning of individual harmonics, (3) changes in the relative power of different harmonics within a stack, and (4) changes in timbre across context.

Recent work in our laboratory (Kojima et al. *in preparation*) has shown that not only does the fundamental frequency of individual syllables vary more across renditions when a male sings alone, the fundamental also varies more within each syllable. These rapid fluctuations in pitch also appear to come from a distinct neural mechanism; lesions of the AFP nucleus Area X reduce the rapid frequency fluctuations, but leave across-rendition variability intact. Lesions of the AFP's output nucleus LMAN eliminate social modulation of both types of variability. The differences in rapid frequency fluctuations across social contexts may contribute to the birds'

ability to classify a single motif as directed or undirected, although they cannot explain the females' continued ability to classify many stimuli after male AFP lesion.

In many respects, songbird hearing is similar to, or slightly worse than, that of the human (Dooling 1979, Dooling and Searcy 1985, Klump and Maier 1990, Okanoya and Dooling 1990). However, zebra finches are exceptionally sensitive to mistunings of single harmonics (Lohr and Dooling 1998). Small differences in inharmonicity across social contexts may play a role in differentiating motifs from the two social contexts.

On the other hand, my personal performance on the directed and undirected stimuli designed for the female zebra finches was very similar to the birds' performance. This suggests that the major cues being used by female zebra finches may involve perceptual abilities not unlike those of human listeners. Males stiffen their posture and increase their beak gape when they sing to females. Expansion of the oropharyngeal-esophageal cavity can cause an energy shift towards lower frequencies, while a wide-open beak will emphasize frequencies at 5kHz and above (Ohms et al. 2010). These changes in the relative power of different harmonics are just as easily detectable by human subjects as they are to zebra finches (Lohr and Dooling 1998).

The musical percept called timbre is often defined as everything that cannot be describe purely as a change in spectral or temporal information. Recent work has improved this vague definition of timbre (Elliott et al. 2013) by relating perceptual descriptions to actual acoustic structures. In order to precisely define the relationship between the perception of timbre and spectrotemporal features, Elliott and colleagues multidimensionally scaled a set of perceptual dissimilarity judgments about different sounds and regressed these with the joint spectrotemporal modulation power spectrum of each sound. A similar approach might also be used to better

understand the complex spectrotemporal patterns that may distinguish directed and undirected motifs.

Broader significance

Now that we know that females can detect the subtle cues that distinguish single motifs of directed and undirected song, we can ask whether these same cues affect female preference. Understanding the limits of female perception, as well as the features of song that are most preferred or salient, will enable us to design better electrophysiological experiments. Finally, female zebra finches are a good test audience to figure out precisely how male song differs across social contexts and neurological states. A clearer picture of the differences between practice and performance motor states in the male will give us insight into the mechanics of vocal production and its relation to neural activity in the motor pathway.

Methods

Behavioral task and training

We trained adult female zebra finches ($n = 11$) to perform a classification task in a two-alternative forced choice paradigm. One bird developed motor problems, had difficulty perching, and was removed from the study. We trained birds in custom-built operant cages (Darrell Floyd and Ken McGary, University of California at San Francisco). Custom MatLab software (Kathy Nagel and Helen McLendon) that interfaced with a TDT RX8 board (Tucker Davis Technologies, Alachua, FL) controlled the operant cages. We used small bookshelf speakers to play the stimuli (Bose, Framingham, MA). We calibrated intensity of playback using pure tones and a calibrated microphone (Brüel and Kjaer, Naerum, Denmark).

To begin a trial, the female hopped on a central “song perch” facing a speaker. This hop triggered the playback a single song motif randomly selected from a database of song recorded from one male zebra finch. The female then decided whether the motif was sung in the presence of another female (‘directed’) or when the male was alone (‘undirected’), and reported her choice by hopping on one of two response perches, located to the right (for directed motifs) and to the left (for undirected motifs) of the song perch. If she classified the song correctly, a feeder located under the song perch was raised for 4–5 s, allowing her access to food. If she classified the song incorrectly, no new trials could be initiated for a time-out period of 30–90 s. Females were not allowed to respond until motif playback was complete. If the bird did not respond within 5 s of the end of the motif, the trial was scored as having no response, and the female could initiate a new trial by hopping again on the song perch.

Birds required 10-30 days of training to acquire this task. Training consisted of four stages. In “song mode,” food was freely available. Hopping on the song perch produced a song drawn at random from a separate habituation database, which contained no songs from the individuals used for classification trials. The habituation database included 28 song renditions from each of four zebra finches, for a total of 112 stimuli. In “food mode,” the bird could still hop on the song perch to hear song from the habituation database, but she now had to hop on one of the two response perches to raise the feeder. 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 5 s, to receive a reward. Finally, in “discrimination mode,” the female began classification trials. The habituation song set was replaced with two sets of training songs, and trials were rewarded (with food) or punished (with a 30–90 s 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 moved to the next stage of training when they performed ~200 hops per day (for song mode), or earned ~200 rewards per day (for food and sequence mode). Females generally learned the classification task (performing consistently above chance) within 5-10 days of beginning classification trials.

Females occasionally developed a bias in which they classified most of one individual's songs correctly (~80-90% correct) but performed more poorly (50-70% 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 (60–85%). This procedure was usually successful at correcting large biases in performance, and reward rates were equalized over 1–3 subsequent days. We excluded data from days on which the two perches were rewarded at different rates from further analyses. In addition to this manual correction, the computer automatically adjusted the rate of song presentation to play more songs associated with the non-preferred perch (see Nagel et al. 2010 for detailed methods).

Females were allowed to work continually for the duration of their 14 h 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 by more than 15%. Birds had free access to water throughout the experiment. Water was located on one side of the cage, next to one of the two response perches. After birds reached criterion in their discrimination behavior, the overall reward rate for correct trials was lowered to 70–95% to increase the total number of trials birds performed each day. Whether a bird was rewarded or not on a given trial was determined randomly. Correct trials that did not produce a reward did not produce punishment. All incorrect trials continued to be punished with a time-out. UCSF's Institutional Animal Care and Use Committee approved all experimental procedures.

Classification of novel stimuli

To test the trained females' ability to generalize the learned classification to new stimuli, we inserted probe trials. Probe stimuli were novel motifs from the same male heard in the training set. Probe trials were introduced only after birds reached criterion and were performing stably at the lowered reward rate (75–95%). Probe trials were randomly interspersed between training trials at a low rate (10–20% of trials), were rewarded at a fixed rate equal to the overall reward rate (75–95%) regardless of which response perch the female chose, and were never punished. This ensured that no information about the “correct” category of the probe 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 stimulus list until they elicited a response.

We considered females that performed above chance for both directed and undirected probe types to have ‘generalized.’ To perform significantly below chance, birds would have to systematically classify directed probe songs as undirected songs, and undirected probe songs as directed songs because percent correct was always calculated for directed and undirected songs taken together.

Song stimuli

We trained female birds on one of five training databases, each constructed from recordings of a different male zebra finch. The original song files recorded for previous publications (see Kao et al. 2005, Kao and Brainard 2006). Briefly, each male was recorded over the course of 3-5 days (spread across 3-16 calendar days). Each recorded song bout was separated by at least 2s of silence. Song recordings were made from a stationary speaker located within a housing cage (53 x 53 x 53 cm; Country Cages); males were allowed to move freely

during recording. Females were introduced intermittently, such that many interleaved bouts of directed and undirected song were collected.

The database from Male 1 included 46 song motifs, 23 sung in the presence of a female ('directed') and 23 sung while alone ('undirected'). Males 2 and 3 each had databases consisting of 29 directed and 29 undirected motifs. The database from Male 4 had 25 motifs sung in each social context, and Male 5's database contained 28 motifs for each context. Each motif was taken from the beginning of a song bout, and the two introductory notes that preceded the motif were included as well. Preliminary experiments excluded these notes, but we found that females had difficulty learning these truncated stimuli. Directed and undirected stimulus sets were closely matched in time; for each directed motif we selected, we also selected a motif from the undirected bout immediately preceding the directed bout. We constructed probe stimuli for each bird in an identical manner; we pulled probe stimuli from a random subset of the original recordings.

We carefully screened each stimulus and removed any motifs containing background noise or movement artifacts. To further reduce the possibility of background noise contaminating the behavioral results, we removed the recorded sound between syllables from the databases of males 1, 2, and 3. We preserved the duration of each inter-syllable interval and used a cosine-shaped ramp to prevent harsh onsets and offsets at the beginning and end of syllables.

We normalized the motifs such that the mean intensity was the same for all stimuli (73 dB RMS SPL, calculated over the total duration of the stimulus, before inter-syllable intervals were silenced). All stimuli were filtered between 250 and 8000Hz (Butterworth 2-pole filter). The directed and undirected stimuli from each of the five males thus all had similar distributions

of total duration, RMS volume, and overall power spectrum, but contained subtle differences in syllable structure and timing. Human listeners (lab members) could not easily identify the social context in which each motif was song.

Recordings of male zebra finches after LMAN lesion were also made for use in previously published work (Kao et al. 2005, for detailed methods, see Kao and Brainard 2006). We constructed post-lesion probe stimuli in a manner consistent with the stimuli constructed using pre-lesion song.

Data analysis

Each trial could be scored as correct, incorrect, or no response. We calculated the “percent correct” by dividing the total number of correct responses by the total number of trials to which the bird responded and multiplying by a hundred. We calculated the “percent response” by dividing the total number of responses by the total number of trials and multiplying by a hundred. We defined reaction time as the time from song perch hop to the subsequent response hop. Ninety-five percent 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 training trials (4-9 control trials for every probe) that the binomial estimate of error on control trials is exceedingly small. In this paper we state that a female generalized to a probe stimulus when the lower 95% confidence interval on her percent

correct was greater than 50% (chance); in addition, the term “correct” response here refers to correct identification of the social context in which the motif was sung (‘directed’ or ‘undirected’).

In order to determine how much useful information syllable durations provided for social context classification, we trained a naïve Bayes classifier on the syllable durations from each stimulus in a training set (`naivebayes.fit`, MatLab) and used the resulting model to predict the classification of the fitted set as well as the probe stimulus set from the same male.

Chapter 2: Spectrotemporal feature dependence on naturalistic stimulus structure in songbird auditory forebrain

Abstract

At the highest level of the auditory system in songbirds, neurons in nucleus HVC are exquisitely selective for the bird's own song. We sought to explore how this selectivity emerges in zebra finches by studying two nuclei along the ascending auditory pathway using both artificial acoustic stimuli and natural stimuli composed of elements of birdsong. Neurons in field L, lower in the pathway, project to the caudal mesopallium (CM), which in turn, projects to HVC. By comparing responses to the natural and artificial stimuli, we discovered that responses were more reliable to the natural stimuli, and that receptive fields calculated from the natural stimuli gave a much better prediction of neuronal responses. Surprisingly, this stimulus dependence was similar in the two brain areas, contrary to our expectation that neurons in CM would be more selective for the natural stimuli than neurons in field L. We characterized neuronal responses with a linear filter followed by a nonlinear response function. In both field L and CM, and for both putative excitatory and inhibitory neurons, natural stimuli gave rise to more complex filter shapes than the artificial stimulus. Receptive fields calculated in response to the natural, song stimuli also had excellent predictive value, far surpassing that of the receptive fields calculated from the artificial stimuli. Our results indicate that for many neurons in the songbird auditory forebrain, receptive field structure is highly dependent on stimulus statistics, and that receptive fields constructed in response to different stimulus classes bear surprisingly little information regarding responses to other sounds.

Introduction

For most animals, hearing and recognizing natural vocalizations is critical for survival. Despite its importance, however, we know much less about auditory processing and its organization than we do visual processing. Songbirds are one of the few animals besides humans that learn their vocal behavior in a process dependent on auditory feedback (Doupe and Kuhl 1999). Moreover, they learn to recognize the songs of other individuals and groups throughout life (Brooks and Falls 1975, Miller 1979, Becker 1982, Miller 1982, Nelson 1989, Stripling et al. 2003, Nagel et al. 2011). A hierarchy of brain areas (Figure 1), equivalent to mammalian auditory cortex, is thought to underlie these abilities (London and Clayton 2008), providing a well-defined target for investigation of auditory processing and recognition.

At the top of this auditory hierarchy, in the song nucleus HVC (acronym used as a proper name), are found some of the most complex auditory neurons known, the so-called song-selective neurons. These neurons respond nearly exclusively to the sound of the bird's own song, preferring it not only over other bird's songs but also over even slightly modified versions of bird's own song (Margoliash 1983, 1986, Fortune and Margoliash 1992, Sutter and Margoliash 1994, Lewicki and Arthur 1996, Volman 1996). Such exquisite selectivity likely emerges from a hierarchy of auditory areas afferent to HVC. In the primary auditory forebrain area field L, the first step in the hierarchy beyond the auditory thalamus, studies using stimuli such as tones and white noise have shown that neurons have relatively simple receptive fields and are tonotopically organized as in mammals (Müller and Leppelsack 1985). However, a major advantage of the songbird model system is that it can also be probed with songs, which are natural vocalizations of known behavioral relevance, and synthetic stimuli derived from these songs. Field L neurons respond vigorously to these naturalistic stimuli. A number of studies

have successfully summarized the stimulus features encoded by field L neurons in response to these naturalistic stimuli by calculating spectrotemporal receptive fields for these cells (Theunissen et al. 2000, Sen et al. 2001, Nagel and Doupe 2006, 2008, Woolley et al. 2009). This careful characterization has revealed striking topographical organization and hints at separate pathways for the analysis of spectral and temporal features of sound (Kim and Doupe 2011).

It is critically important to understand how the auditory system progresses from the relatively simple properties of field L neurons towards responses as complex as song selectivity. A number of interconnected regions lie between field L and HVC, including the caudal mesopallium (CM). Previous attempts to characterize the spectral and temporal selectivity of neurons in CM have been less successful than in field L (Theunissen et al. 2000, Sen et al. 2001). Cells in CM are more selective than those in field L and some cells only respond to one, or a few, highly structured song elements (Gentner and Margoliash 2003, Meliza and Margoliash 2012). In addition, simplified, synthetic stimuli generally do not drive consistent or robust neural responses in these cells. Thus, in order to characterize CM, we have assessed neuronal responses to a large battery of conspecific songs. These constraints make obtaining receptive fields using traditional techniques difficult for two main reasons. First, song contains much complex statistical structure, and these stimulus correlations can bias the recovered receptive field. When the average stimulus that precedes a spike is itself correlated with other parts of the stimulus, it becomes difficult to define precisely the features that drive spiking. Although decorrelation methods reduce the effects of these intra-stimulus correlations, they are not always sufficient to recover an accurate filter. Second, the receptive field may provide an incomplete model of CM neuronal responses. Because the STRF is essentially an average stimulus, it can be

described as a single dimension in a high-dimensional stimulus space. Given that cells in CM may respond specifically to complex song features, it would not be surprising if a full description of these neurons' receptive fields required extracting multiple dimensions from the stimulus space.

The method of maximally informative dimensions (MID) is a form of spike-triggered analysis that does not require the experimenter to use simplified or artificial stimuli. By maximizing the mutual information between isolated directions in stimulus space and neural responses, this technique removes the influence of correlations in the stimulus. Previous work has successfully used this technique in the primary auditory and visual cortices to provide a precise quantitative estimate of feature selectivity (Sharpee et al. 2006, Atencio et al. 2008, Atencio et al. 2009). We hypothesized that this analysis would reveal the types of spectrotemporal features present in CM, providing a framework within which to understand intermediate stages of auditory processing.

As a result of both evolutionary history and individual experience, CM contains a population of neurons predisposed to represent conspecific vocalizations. The feature selectivity of these cells may be constantly changing to reflect new auditory and behavioral contexts. Thus, feature selectivity is not an essential or invariant property of neurons (Nelken 2004); rather, it is a measure that may be used and compared across carefully controlled experimental conditions. It gives us a 'snapshot' of a cell's response properties.

Here, we have characterized the feature selectivity of field L and CM neurons in response to the most relevant possible stimulus ensemble, i.e. a wide range of conspecific natural vocalizations. In this way, we have obtained an overview of response properties as this area is processing information of the type selected for during both evolution and individual

development. In addition, we have assessed the feature selectivity of these same neurons in response to a simpler, artificially generated stimulus that contains some of the statistical relationships present in natural stimuli. This provides a solid starting point for further investigation into the role of stimulus statistics, behavioral task, and experience in shaping the response properties of intermediate sensory neurons.

Figure 1: Songbird auditory forebrain diagram.

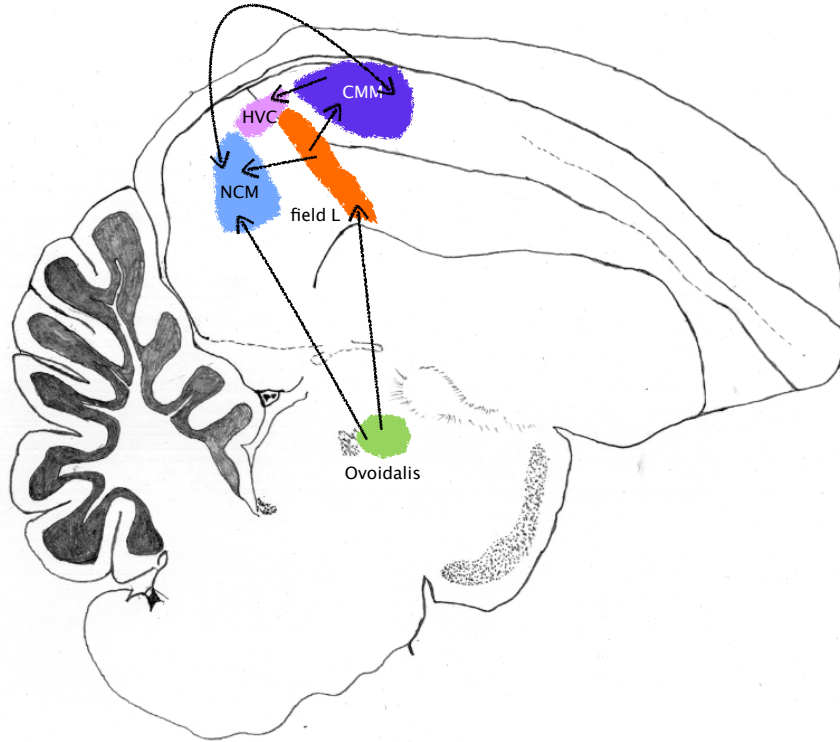


Figure 1: Songbird auditory forebrain diagram. Sagittal view. CM, Caudal mesopallium; NCM, Caudomedial nidopallium; HVC, used as a proper name.

Results

Spectrotemporal receptive fields in field L and CM

To investigate the stimulus dependence of receptive field properties in the auditory forebrain, we recorded neural responses to a natural stimulus composed of zebra finch song elements as well as an artificial stimulus containing spectrotemporal properties similar to those found in song (Figure 2). The song stimulus contained brief recordings from 171 different male zebra finches. Each excerpt consisted of a single rendition of the male's song "motif", a characteristic element (mean duration: $873\text{ms} \pm 389\text{ms}$) that the male zebra finch repeats to form a bout of song. These motifs created a total of 149s of unique song content, and were pseudo-randomly concatenated to form 13min blocks of continuous song stimulation. During recordings, we interleaved these song stimulus blocks with 8-12min blocks of the second stimulus type. This stimulus, called ripple-noise, consisted of 35 independently modulated, logarithmically spaced frequency bands. Each band contained white noise filtered by exponentially distributed amplitude modulations, reflecting the statistics of zebra finch song (Nagel and Doupe 2008, Kim and Doupe 2011). The resulting stimulus resembled the sound of water flowing over rocks.

In order to understand how receptive field properties change across the ascending auditory system, we recorded from single neurons in two regions: field L and CM. Field L receives direct thalamic input and contains neurons that respond broadly to naturalistic stimuli, while the higher order auditory area CM is highly selective for conspecific song. In male zebra finches, neurons from CM project directly to HVC, a nucleus containing neurons that respond only to the bird's own song (Figure 1). We recorded responses to these stimuli from 72 neurons: 33 in field L, 30 in CM, and 9 with uncertain histological localization. Examples of single unit

responses in each brain area are shown in Figure 3. We hypothesized that by using both song and ripple-noise stimuli, as well as a relatively unbiased mutual information-based approach to extract receptive fields, we could uncover meaningful information about the response properties of field L neurons as well as the highly selective cells found in CM.

We calculated receptive fields using both spike-triggered averaging and maximally informative dimensions for all stimulus types and all neurons. Filter shapes for the first maximally informative dimension were quite similar to the decorrelated spike triggered average, validating both approaches. In most cases, a simple linear filter described the neuronal response well. When we simultaneously estimated the first and second maximally informative dimensions, we found that the first dimension still looked similar to a linear estimate and that the second dimension revealed little additional structure.

Figure 2: Stimuli used during recordings.

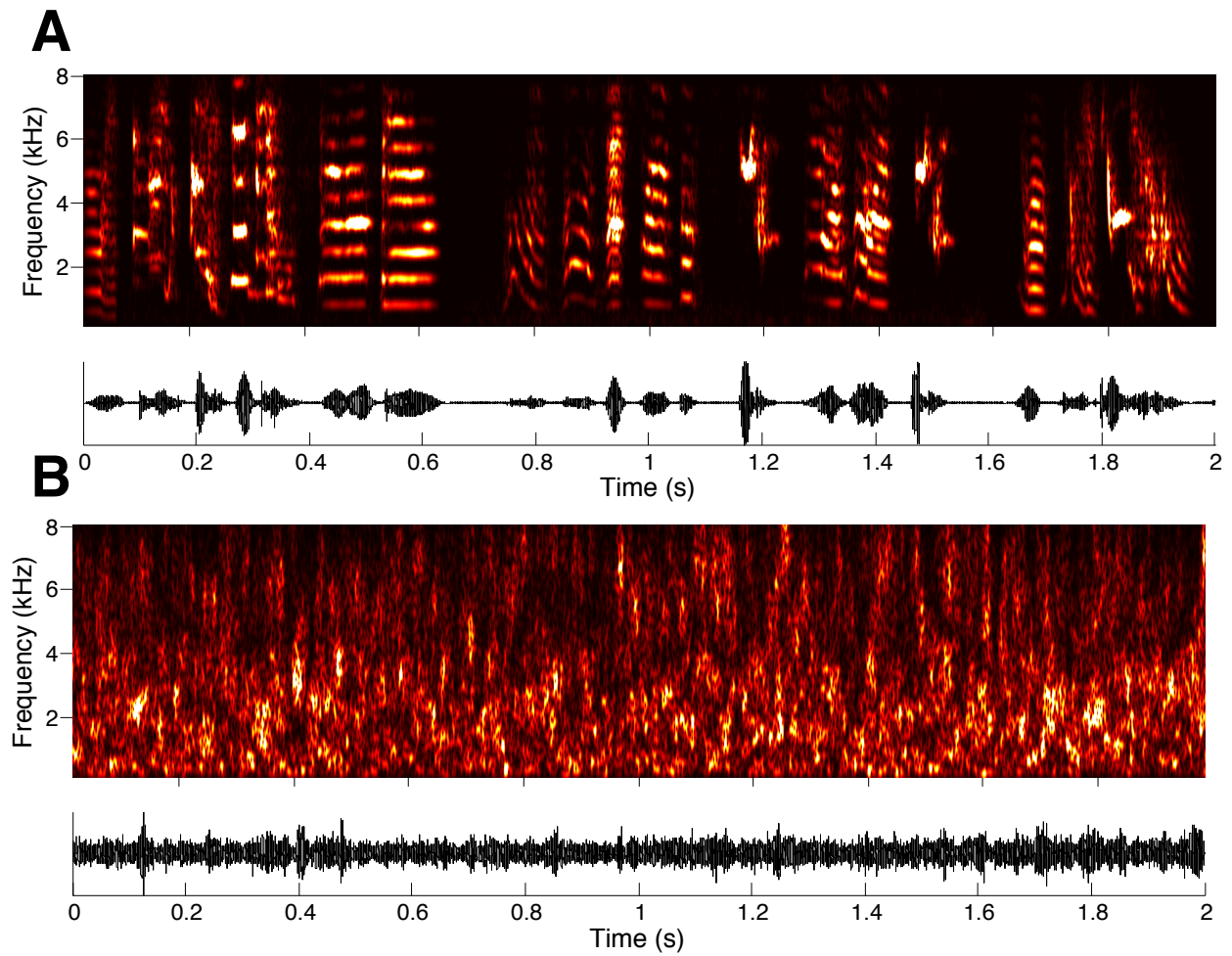


Figure 2: Stimuli used during recordings. **A)** An example two-second segment of *song* stimulus. **B)** A representative two-second segment of computationally generated *ripple-noise* (artificial) stimulus.

Figure 3. Example neural responses.

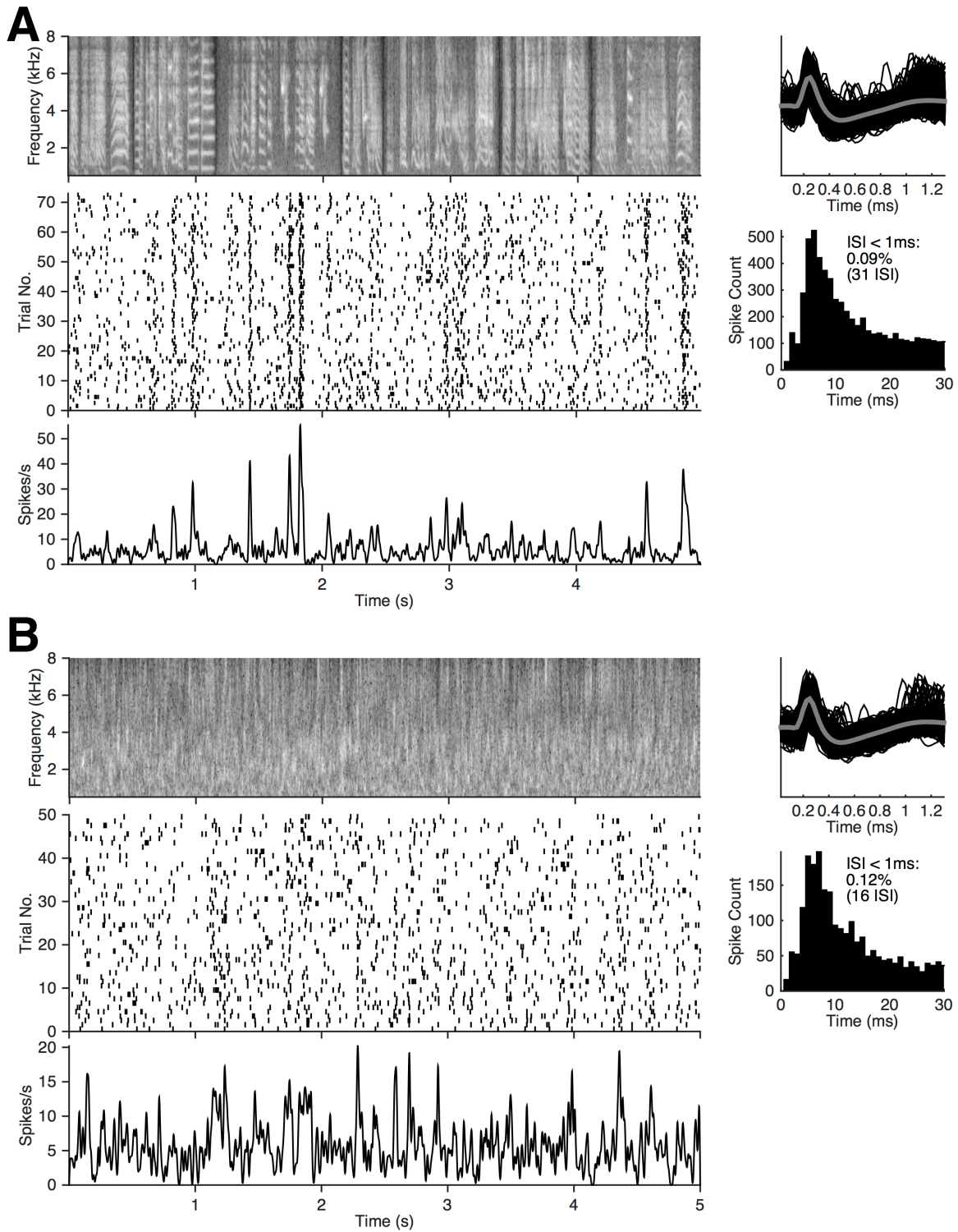


Figure 3: Example neural responses. Example data collection results from song stimulus (A) and ripple-noise stimulus (B).

Filter shapes differ in response to different stimuli

For each neuron, we calculated filters using both song and ripple-noise stimuli. Neurons in field L and CM yielded a variety of filter shapes in response to the two stimulus types (Figure 4). The two leftmost filters on the top row of Figure 4A illustrate the range of filter shapes seen.

First, consider the filter second from the left. In response to the song stimulus, the most prominent feature is a single peak, narrow in both frequency and time, located at 2.5kHz, approximately 20ms before the spike. The blue shading around the red peak indicates that the presence of sound at frequencies just above or below 2.5kHz, especially just prior to, or at the same time as, sound at 2.5kHz, reduces the likelihood that the neuron will respond. The neuron's filter shape looks very similar in response to the ripple stimulus (Figure 4B, top row, second filter from left). For this neuron, a brief isolated pulse of sound at approximately 2.5kHz appears to be the stimulus most likely to evoke a spike, both during song and ripple-noise stimulation.

In contrast to the simple response pattern demonstrated by the neuron described above, the leftmost filter in the top row of Figure 4A reveals a complex set of response contingencies during song stimulation. The acoustic pattern most likely to elicit a spike from this neuron included sound at several separate frequencies, prolonged in time. Broadband sounds would be unlikely to cause a response, as they impinge on the inhibitory bands that interleave the excitatory frequencies.

Although both stimulus types elicited a wide variety of filter shapes, striking differences were seen between filters constructed from the same neurons in response to song and ripple-noise (Figure 4). Filter shapes for the same neuron were often quite different, although the mean firing rates were similar across stimulus types (song: 4.85spikes/s; ripple-motif: 4.19spikes/s). In

response to song, many neurons had filters containing multiple frequency bands extending over many milliseconds. Some, however, had single peaks, which were narrow in time and wide in frequency, wide in time and narrow in frequency, or tight in both dimensions. In contrast, the ripple-noise stimulus elicited far fewer filter shapes. Filters constructed in response to this stimulus only contained single excitatory peaks, although these peaks varied across both dimensions in terms of peak location and width.

Despite the differences in shape across the two stimulus types, the song and ripple-noise filter shapes for a given neuron bore similarities to one another. The frequency of peak response was generally the same for both stimulus types. Since the filter shapes were diverse, and many were quite complex, we used the dot product as a rough measure of the similarity between filters (Figure 6). Across the population, the dot products between each neuron's song and ripple-noise filters were higher than the dot products for shuffled pairs of song and ripple-noise filter shapes.

Since CM neurons are often much more highly selective for conspecific song than neurons in field L, we hypothesized that the filter shapes representing the first MID might be more similar across stimulus types in field L than in CM. However, dot product values between ripple and motif filters varied widely across neurons and were no higher in field L than in CM ($p = 0.7211$, rank-sum test). At least for this limited data set, there is no evidence CM filters differ more than field L filters across stimulus type. This suggests that song and ripple-noise stimuli elicit distinct responses in both brain areas.

Receptive fields derived from song responses have high predictive value

One way to evaluate the quality of a receptive field is to test its “predictive value,” or ability to predict the actual response of the neuron to a repeated stimulus segment not used to calculate the original filter. In order to measure the neuron's firing rate, we interleaved repeated

5s song and ripple-noise stimuli with the non-repeated segments used to calculate the receptive field. Using the 10-100 collected repeats we then calculated the peristimulus time histograms (PSTH) for these segments. To generate the predicted firing rate, we convolved the filter with the repeated stimulus segment and applied a gain function to the output (Brenner et al. 2000, Kim and Doupe 2011). We used the correlation coefficient (CC) between the predicted response and the observed PSTH as a measure of the receptive field's ability to predict novel responses (Figure 5).

The CC values for almost all of the predictions are above chance, and many are excellent. Figure 7 depicts CC values for all 72 neurons; CC values for ripple predictions are on the ordinate, and motif prediction results are on the abscissa. To compare the predictions of the receptive fields derived from the two different kinds of stimuli, we restricted all further analyses to the 43 neurons whose CC values were greater than 0.15 for both stimulus types. As the preponderance of neurons below the line of unity in Figure 7 illustrates, receptive fields constructed from song responses yielded predictions with much higher CC values than those constructed from ripple-noise ($p = 8.36e-06$, sign-rank test). These results indicate that the receptive fields calculated during song describe the neuron's song response patterns more fully than the receptive fields calculated during ripple-noise are able to describe firing patterns during ripple-noise stimulation.

Significant CC values were equally distributed across the two brain areas (motif: $\chi^2 = 1.47$, $p = 0.225$, ripple: $\chi^2 = 0.72$, $p = 0.395$). In addition, the mean CC values of the significant filters were not different across brain areas (motif: $p = 3.67e-01$, ripple: $p = 9.22e-01$, rank-sum tests). These results suggest that the predictive ability of the filters is similar in both field L and CM.

Figure 4. First MID filter shapes.

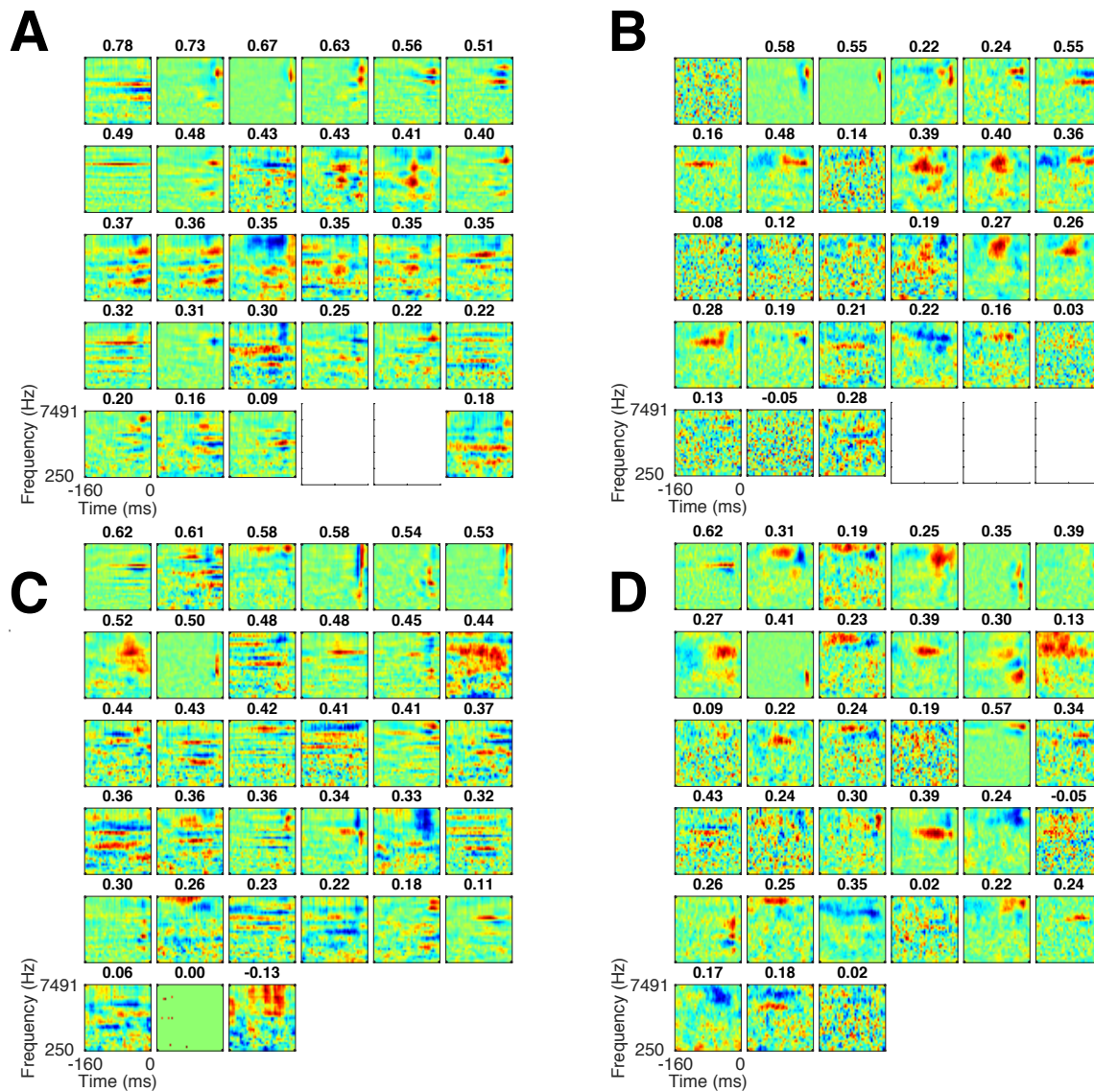
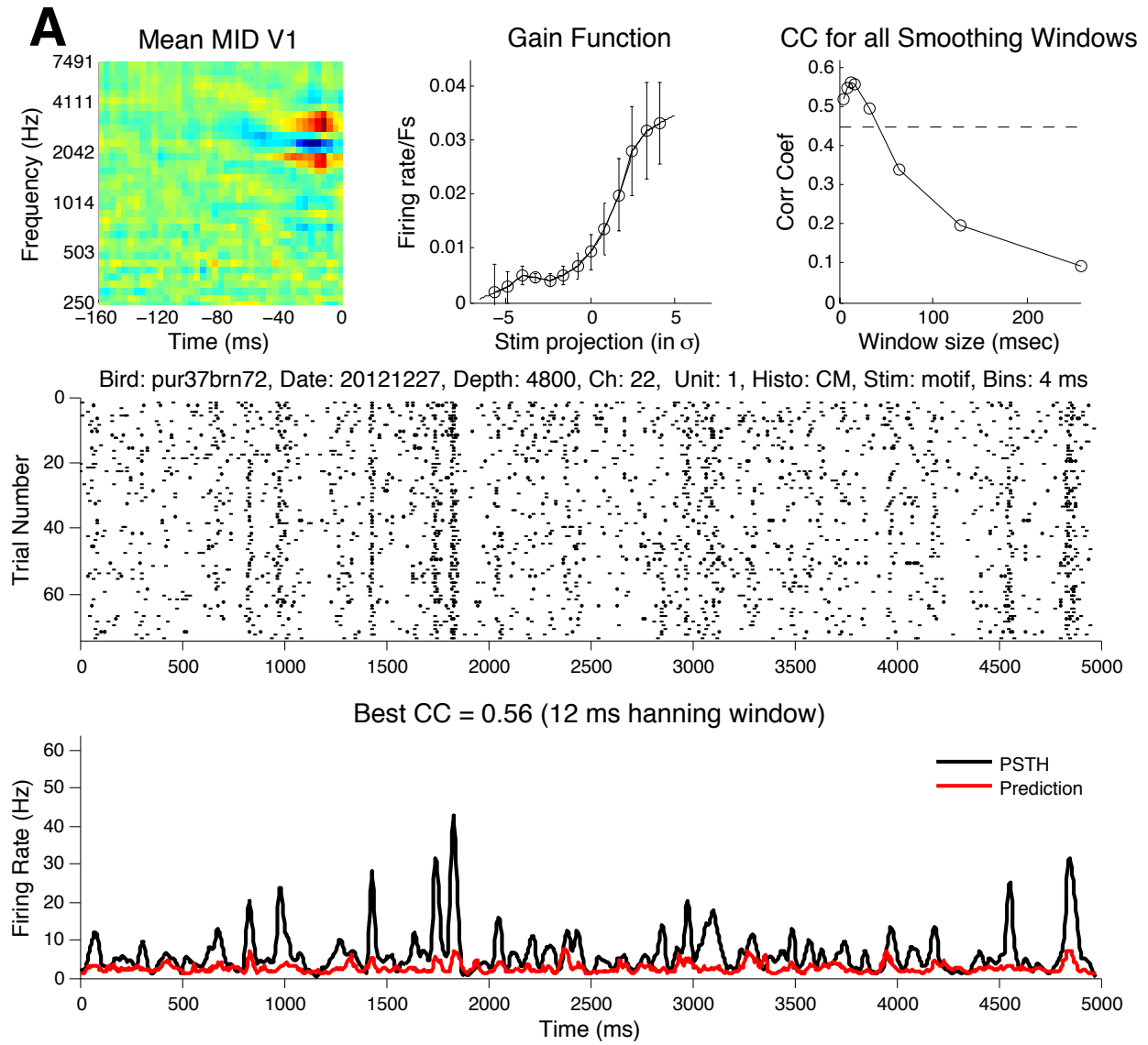


Figure 4: First MID filter shapes. Filters are plotted according to decreasing correlation coefficient (CC) between measured and predicted song responses. Above each filter is its CC value. **A)** Filter shapes for field L neurons constructed from song stimulus responses. **B)** Filter shapes for the same field L neurons constructed from ripple-noise responses. **C)** Filter shapes for neurons in CM constructed from song responses. **D)** Filter shapes for the same CM neurons constructed using ripple-noise responses.

Figure 5. Example predictions.



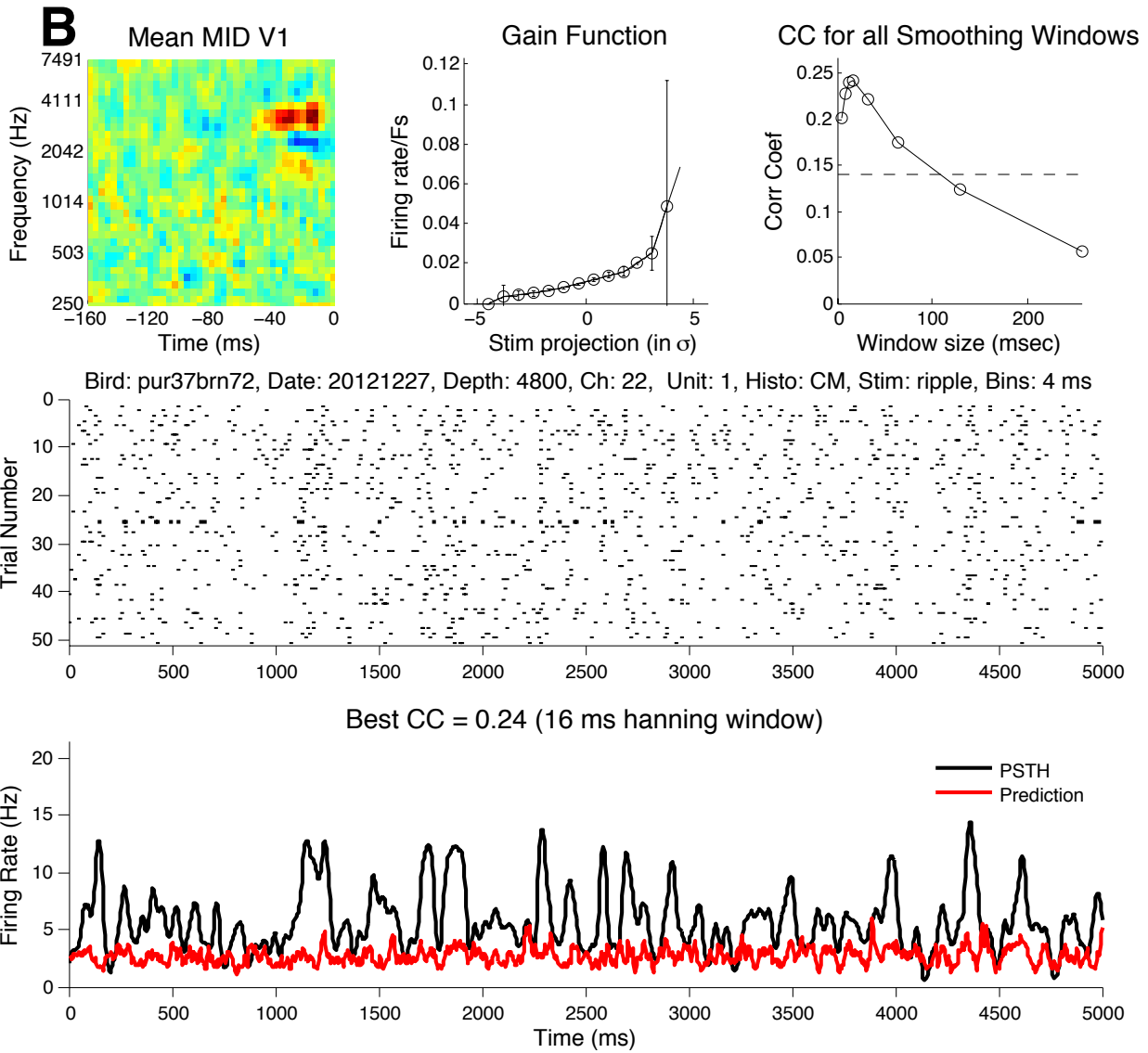


Figure 5: Example predictions. **A)** *Song stimulus response prediction.* We used a receptive field (filter shape, top left; gain function, top middle) calculated using the responses of a single CM neuron to a unrepeated song segment to predict the response of the same neuron to a repeated song segment (raster containing neuron's response times, center panel; PSTH, bottom panel, black line). The predicted response is shown in red. **B)** *Ripple-noise response prediction.* We also predicted and measured the same neuron's response to the artificial ripple-noise stimulus.

Figure 6: Similarity between song and ripple-noise filter shapes.

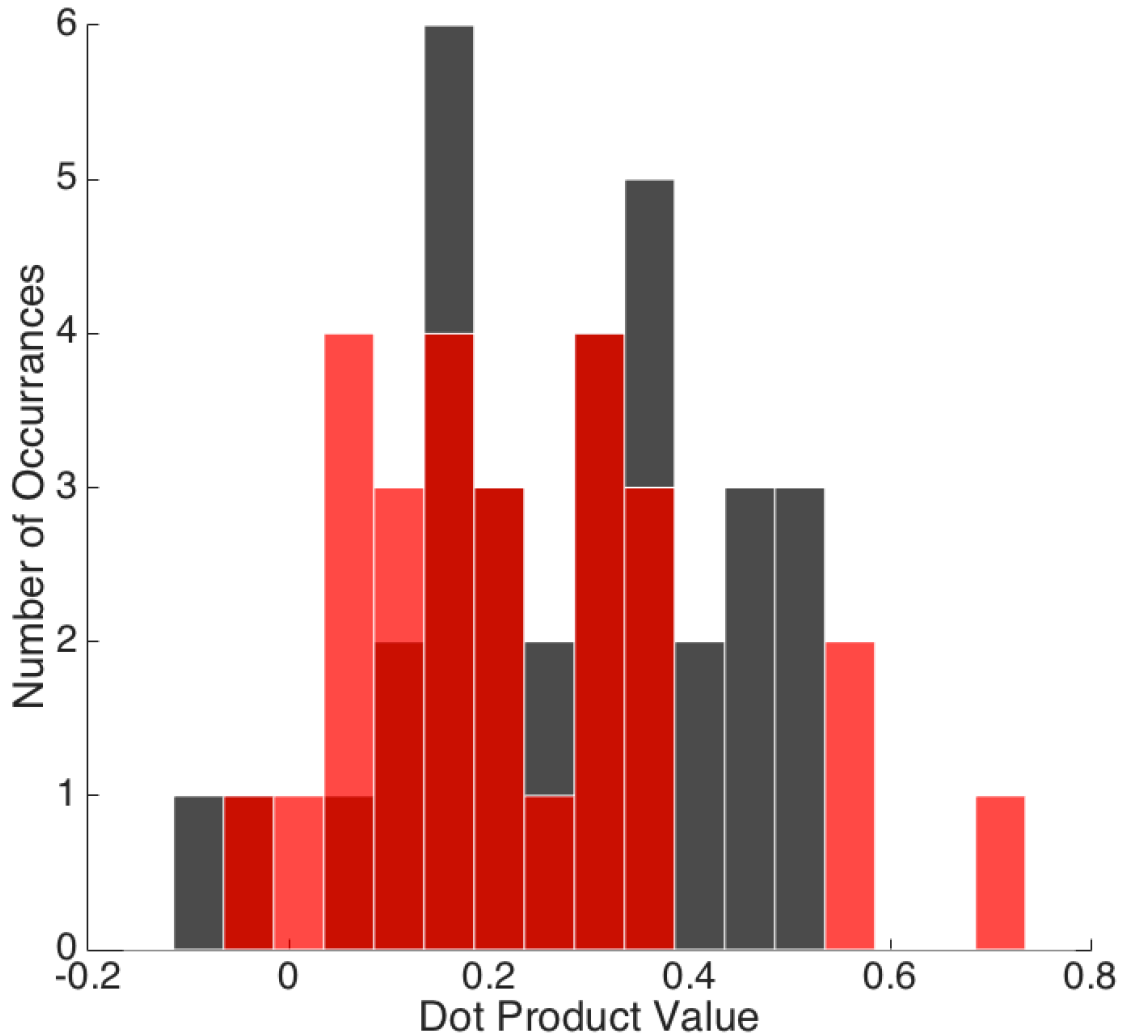


Figure 6: Varying degrees of similarity between song and ripple-noise first MID filter shapes. Histograms containing dot product values for all neurons for which both song and ripple-noise stimulus responses were recorded. Red indicates CM, gray indicates field L.

Receptive fields fail to predict responses to the across stimulus type

Given that song receptive fields predicted song responses better than ripple-noise receptive fields predicted ripple-noise responses, we wondered whether song receptive fields would outperform ripple-noise receptive fields in predicting responses ripple-noise stimuli as well. To test this notion, we examined how well receptive fields calculated in response to natural song stimuli predicted responses to ripple-noise stimuli, and vice versa.

Figure 8 depicts the predictions and cross-predictions for all 43 neurons whose CC values were greater than 0.15 for both stimulus types. The greatest CC values were found for song receptive fields predicting song responses. As reported above, ripple-noise filters predicted responses from their own stimulus class much more poorly. CC values for the two cross predictions, song receptive fields predicting ripple-noise responses and ripple-noise receptive fields predicting song responses, were similar to one another and significantly lower than the within-type predictions (One-way ANOVA, $F = 21.72$, $p < 1 \times 10^{-11}$; post-hoc Bonferroni corrections). Thus, despite their greater predictive value, song receptive fields did not outperform ripple-noise receptive fields when predicting responses to a stimulus class other than the one from which the receptive field was derived. Receptive fields from both stimulus classes performed poorly in this regard, supporting the notion that the two stimulus classes were sufficiently different to explore distinct aspects of the neuron's response properties or to cause adaptation to occur across stimulus classes.

Narrow- and wide-spiking neurons have similar filter shapes

In the mammalian sensory cortex, intracellular recordings have revealed distinct electrophysiological characteristics for excitatory and inhibitory neurons (Connors and Gutnick 1990, Markram et al. 2004). In particular, excitatory neurons typically have longer waveforms

than inhibitory interneurons. Paired intra- and extra-cellular recordings have established a direct correspondence between the durations of waveforms recorded using these two techniques (Henze et al. 2000). The use of extracellularly recorded wavelengths to distinguish putative excitatory and inhibitory neurons has been successfully used in the mammalian visual (Swadlow and Weyand 1987, Gur et al. 1999, Shapley et al. 2003, Mitchell et al. 2007, Nowak et al. 2008), somatosensory (Simons 1978, McCormick et al. 1985, Swadlow and Gusev 2002), and auditory (Atencio and Schreiner 2008) cortices. This correspondence has also been suggested to be true for neurons in the songbird auditory forebrain (Meliza and Margoliash 2012, Jeanne et al. 2013, Schneider and Woolley 2013).

We also observe a bimodal distribution of spike widths in our data. Figure 9A depicts the mean waveform shape for all 72 neurons we recorded. The width of each waveform's trough at half-height proved a convenient measure to illustrate the bimodality of spike shapes observed (Figure 9B). A cut-off value of 0.28ms divides the population into two groups containing 36 neurons each. Narrow- and wide-spiking neurons were similarly distributed across field L and CM ($\chi^2 = 1.8$, $p = 0.18$), regardless of whether we included all 72 sites, or only those with two significant filters (43 neurons). Consistent with earlier reports, the narrow-spiking neurons had firing rates more than double that of the wide-spiking neurons (mean firing rate ratios, narrow/wide: 8.1/2.5 spikes/s during motif; 7.2/2.1 spikes/s during ripple). However, despite their different firing rates, a comparison of the receptive field properties of these two populations did not exhibit any clear differences.

Figure 7: Correlation coefficient between response and predicted response.

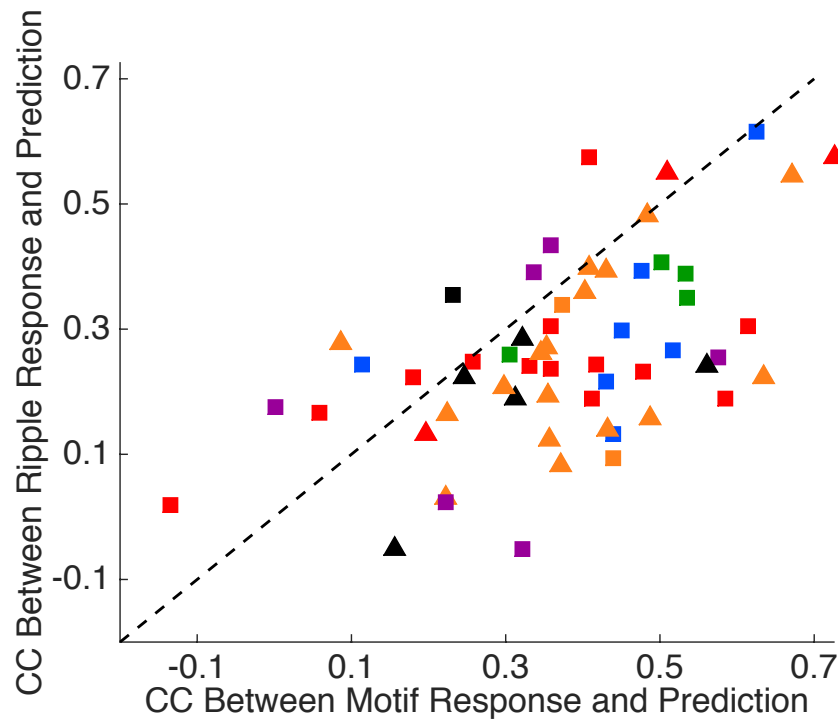


Figure 7: Correlation coefficient between response and predicted response is higher for the song than the ripple-noise stimulus. Correlation coefficient between song response and predicted response plotted against the correlation coefficient between ripple-noise response and predicted response for each neuron. Squares represent field L neurons; triangles, CM neurons. Data from each animal is plotted in a different color. Dotted line indicates the line of unity.

Figure 8: Cross predictions.

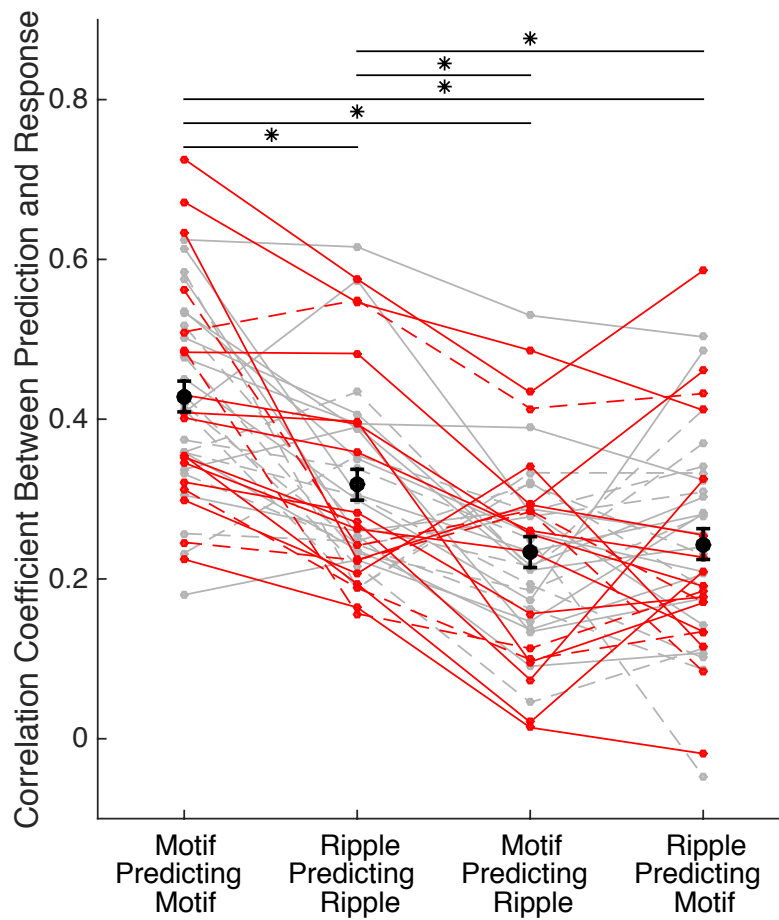


Figure 8: Cross predictions. Correlation coefficient (CC) values between predicted and observed responses for all 43 neurons with significant prediction CC values for both stimulus types are depicted in the two leftmost columns; cross predictions are shown in the right columns. CC values for CM neurons are in red, field L in gray. Lines connect CC values for neurons with narrow waveforms; hashed lines connect CC values for wide waveforms.

Figure 9: Bimodal Distribution of Spike Widths.

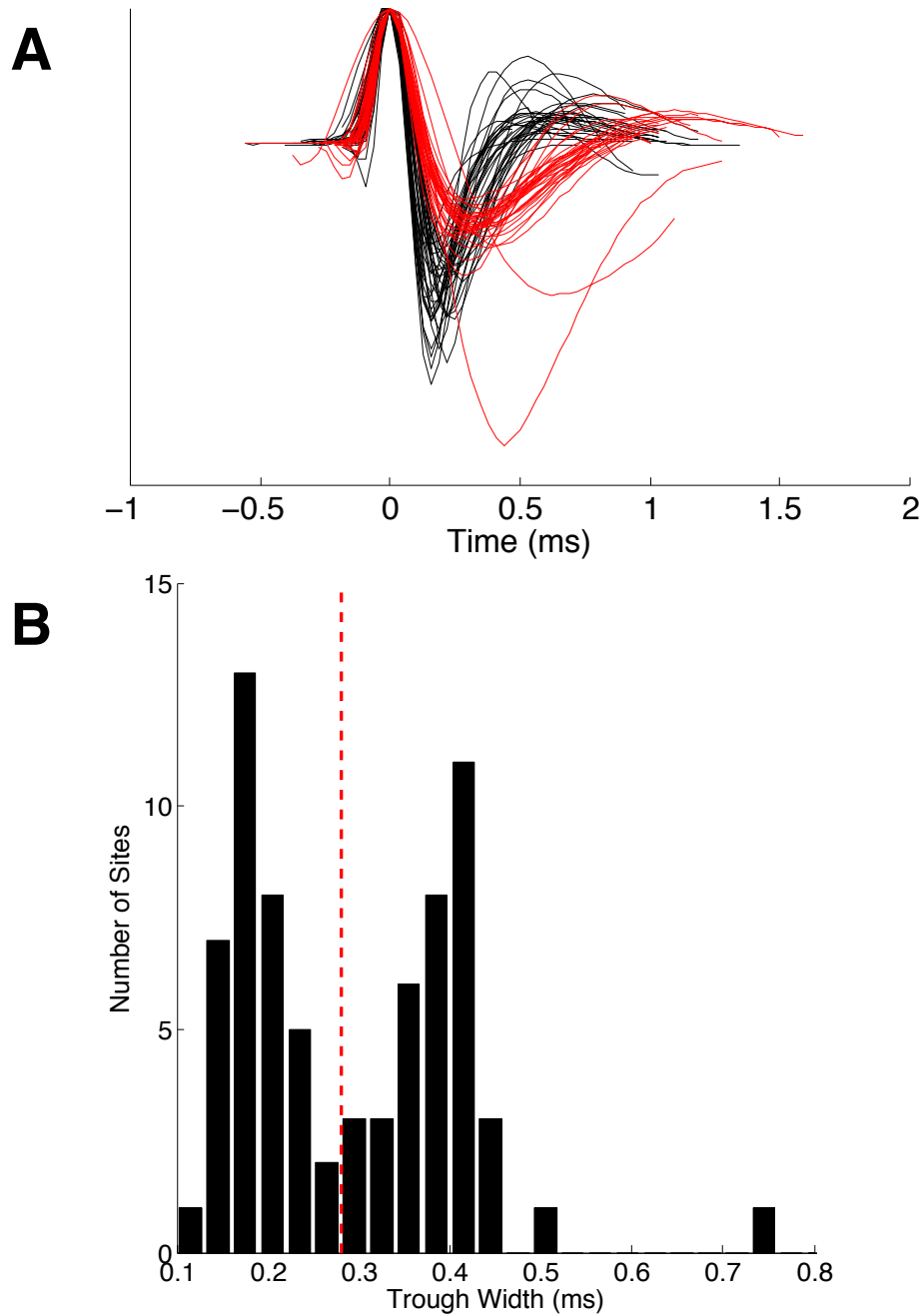


Figure 9: Bimodal Distribution of Spike Widths. A) Mean waveforms for both wide- (red) and narrow- (black) spiking neurons. B) Histogram of trough width at half height for all neurons. Dashed red vertical line at 0.28ms indicates the cutoff between wide- and narrow-spiking neurons.

Discussion

By comparing responses to both natural and artificial stimuli in the zebra finch auditory forebrain, we uncovered a striking context-dependence in receptive field structure and predictive value. Surprisingly, this stimulus dependence was consistent across neuron types and brain areas, contrary to our expectation that neurons in CM would be more sensitive to stimulus type. In both field L and CM, and for both putative excitatory and inhibitory neurons, natural song stimuli elicited more complex filter shapes than an artificial ripple-noise stimulus. Receptive fields calculated in response to the song stimulus also had excellent predictive value, far surpassing that of the ripple-noise receptive fields. Our results indicate that for many neurons in the songbird auditory forebrain, receptive field structure is highly dependent on stimulus statistics, and that receptive fields constructed in response to different stimulus classes bear surprisingly little information regarding responses to other sounds.

Previous attempts to calculate traditional spectrotemporal receptive fields have not provided consistent results in nuclei beyond field L (Theunissen et al. 2000, Sen et al. 2001, Woolley et al. 2009, Meliza et al. 2010). Although many assumed that the highly non-linear nature of these neurons' response properties was responsible for the inadequacy of these receptive fields, we hypothesized that, with the right stimulus, we could elicit more informative responses. Although artificial stimuli often lack enough temporal and harmonic structure to drive consistent or robust neuronal responses, the long periods of silence and strong temporal changes that characterize natural song bouts can drive strong "onset" and "offset" responses that can overwhelm other spectral and temporal response properties. Our song stimulus contained a large, continuous battery of song motifs that "bridged the divide" between natural song bouts and artificial stimuli. Using this stimulus, we were able to calculate significant receptive fields in

both brain areas and found that a single linear filter and gain function seemed to be sufficient to capture most of the information.

In addition to calculating receptive fields in response to our song stimulus, we also recorded responses to an artificial stimulus from the same neurons. Direct comparison of receptive fields calculated in response to these two stimuli demonstrated that neurons with complex receptive field shapes in response to song often have much simpler receptive fields in response to the artificial ripple-noise stimulus. In particular, the song filters had much more complex harmonic structure than the ripple-noise filters.

In addition, for both brain areas, song receptive fields had much greater predictive value than those calculated in response to the ripple-noise stimulus. Given that the artificial stimulus we used was specifically designed to mimic the correlations found in natural sounds, the extent of the divide in predictive value between ripple-noise and song filters was quite surprising. These results are, however, consistent with previous recordings in the avian midbrain, which demonstrated higher predictive values for receptive fields calculated in response to song than for receptive fields calculated in response to an artificial stimulus (Woolley et al. 2006).

Receptive fields calculated using one stimulus type did not predict responses to the other type well. Given the striking differences in filter shapes between song and ripple-noise stimuli, this result was unsurprising. This further supports the notion that neurons are actively adapting their responses, or entering an entirely different portion of a more complex receptive field when they begin responding to a different stimulus type. Recent work in the mammalian visual cortex has shown that adaptation occurs on several timescales, ranging from seconds to minutes, as

neurons switch between natural and artificial visual stimuli (Sharpee et al. 2006). Further analysis and using longer stimulus blocks would be necessary to distinguish these hypotheses.

Consistent with previous reports in both mammalian sensory cortices and songbird auditory forebrain, we observed a bimodal distribution of spike shapes that we attribute to inhibitory interneuron and excitatory neuron classes. Although previous studies in songbirds have reported differing proportions of narrow- and wide-spiking neurons across brain areas (Meliza and Margoliash 2012, Schneider and Woolley 2013), we did not observe any significant differences between field L and CM. Despite the fact that the putative interneurons fired at a rate 2–3 times that of the putative excitatory neurons, we did not see any obvious difference in the shapes or predictive value of the receptive fields calculated for these two neuron types. This also stands in contrast to previous reports, which found that wide-spiking neurons were more selective for specific song elements than narrow-spiking neurons (Meliza and Margoliash 2012, Schneider and Woolley 2013).

Methods

Electrophysiology

We recorded from 12 adult female zebra finches (*Taeniopygia guttata*), all of which were bred and raised in our colony. We performed all procedures in accordance with protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee. We anesthetized each bird with three 25 μ l injections of 20% urethane after depriving her of food and water for one hour. We secured the bird's head using a custom stereotaxic head holder, and then used dental cement to attach a metal post to the skull. Once the cement dried, we used the metal post to fix the bird to a custom stereotaxic frame. Inside the

frame, the anesthetized bird rested on an electric heating pad with continuous temperature control (FHC).

In order to isolate single neurons in both field L and CM, we used a variety of electrode types and configurations, including linear arrays of silicon probes (Neuronexus; 32 channels set 50 μm apart with a 95mm² spot size), silicon probes in tetrode configurations (Neuronexus; 32 channels on 4 prongs set 200 μm apart, each containing 2 tetrodes set 150 μm apart with a 121 μm^2 spot size) and arrays of tungsten electrodes (FHC; 1, 2, 4, or 8 electrodes, 190-240 μm spacing, impedance 3-9M; 75 μm shank diameter). We placed electrodes over CM and field L in a mediolateral orientation, 1.5-1.7mm anterior to the Y-sinus reference point and 0.5-2.2mm from the midline. We opened a small window in the skull, removed the dura, and then lowered the electrodes into the brain. We then applied a silicone elastomer around the opening to prevent drying.

We performed all recordings in a sound-attenuation chamber. A speaker (JBL Scientific Biotechnology Services) located 20cm away from the bird's head played all stimuli at 70dB. We calibrated sound levels using a sound level meter (A-weighting; Brüel and Kjær). We used a Neuralynx headstage (HS-36), Neuralynx amplifiers (Lynx-8), and custom software (Multikrank; written by D. Schlee and modified for multi-channel recording by B. D. Wright) for all recordings.

Spike sorting

We used a commercial spike sorter (Offline Sorter; Plexon) to obtain spike times for each single unit. We used the first two dimensions in principal component space to separate clusters. When defining single units, we used clear visual separation in this space, as well as <0.5% violation of the refractory rate (1ms) as criteria.

Histology

At the end of each experiment, we made electrolytic lesions at several depths on one to two channels. We perfused each bird transcardially with lactated Ringer's solution followed by 3.7% formaldehyde. We used a freezing microtome to cut 40- μ m thick parasagittal sections. We stained every other section with Nissl. Unlike the rest of the auditory forebrain, type 1 cannabinoid receptor antibodies (CB1) strongly stain the thalamorecipient layer of field L (L2; Soderstrom et al. 2004). Thus, in some birds, we stained alternating sections for CB1 to aid us in identifying field L neurons.

Stimulus generation

Song stimulus design. We pulled created our song stimulus from a database of songs recorded in the laboratory over the course of many years. Although all songs were collected from birds as they sang alone in a sound isolation chamber, recording conditions and background noise varied across birds. We extracted a single song excerpt from recordings of each of 171 males. Each excerpt consisted of a single rendition of the male's song "motif", a characteristic element (mean duration: 873ms +/- 389ms) that the male zebra finch repeats to form a bout of song. We hand-selected each song segment to include as little surrounding silence as possible while not clipping the onset or offset of the motif and then smoothed the ends of the signal. These motifs created a total of 149s of unique content.

For receptive field estimation, we concatenated random permutations of the motifs, forming 12.5min blocks of continuous song stimulation. Within each block, we also repeated a single 5s segment every 75s, for a total of 10 repeats. These repeats were later used to test the predictive value of the receptive field calculated using the unrepeated portion of the stimulus (see *Receptive field estimation*). This resulted in an overall duration of 13.33 minutes for each block.

During recordings, we interleaved these song stimulus blocks with blocks of the ripple-noise stimulus (see below) for up to four hours, obtaining a total of 12.5 to 121.2min (mean: 48.5 ± 32.3 min) of unrepeated song stimulus and 10-100 iterations of the repeated song segment (mean: 28.2 ± 26.5 repeats).

Ripple-noise stimulus design. Our ripple-noise stimulus was similar to that used in two previous studies (Nagel and Doupe 2008; Kim and Doupe 2011). Briefly, we created 35 overlapping carrier bands by filtering white noise using a bank of Gaussian filters with logarithmically spaced center frequencies. The frequency of each band was given by $\exp(\log(250) + 0.1*(i-1))$, where i is the band number (1-35). We then multiplied each carrier band by a separate amplitude envelope formed by filtering white noise with an exponential filter (time constant, 20Hz). We summed all the resulting signals to yield our final stimulus. The exponential decay constant of 20Hz for each amplitude envelope mimicked the bias towards low temporal frequency modulations found in natural sounds (Singh and Theunissen 2003), while the overlapping carrier frequency bands gave the final signal a nearly flat power spectrum (Theunissen et al. 2000). The result was a stimulus reminiscent of the sound of running water, with much smoother temporal and spectral modulations than white noise.

For receptive field calculation, we played an unrepeated stimulus segments lasting a total of 8 to 12min. Interleaved within this stimulus was a repeated 5s long segment used to test the predictive value of the receptive field (see *Receptive field estimation*). During recordings, we interleaved these ripple-noise stimulus blocks with blocks of the song stimulus (see above) for up to four hours, obtaining a total of 4.2 to 94.8min (mean: 31.9 ± 20.3 min) of unrepeated ripple-noise stimulus and 10-100 iterations of the repeated ripple-noise segment (mean: 37.1 ± 33.2 repeats).

Receptive field estimation

We used two different techniques to calculate receptive fields: (1) spike-triggered averaging, followed by decorrelation, and (2) maximally informative dimensions.

Reverse correlation. All spike-triggered averaging and decorrelation methods were performed as previously reported (Theunissen et al. 2000, Nagel and Doupe 2008, Kim and Doupe 2011). Briefly, we cross-correlated each row of the stimulus spectrogram with the spike train. Since our stimuli contained correlations in both frequency and time, the resulting spike-triggered average (STA) contained residual correlations due to the structure of the stimulus. We removed these correlations by dividing the STA by the autocorrelation of the stimulus spectrogram.

Maximally informative dimensions. To find the maximally informative dimensions (MID) for each neuron given each stimulus set, we used methods similar to previously published work (Sharpee et al. 2004, Atencio and Schreiner 2008, Meliza et al. 2010, Sharpee et al. 2011). The first MID identifies the direction in stimulus space that best predicts spike production. Consider any single arbitrary direction (v) in stimulus space. For each presented stimulus (s), we can determine how close s is to v (i.e. 'how much' v is in a given s) by finding the projection of s onto v ($s \cdot v$). By projecting all presented s onto v we obtain a distribution of values $P_v(x)$, where x is the value of a single projection. Similarly, we can obtain a distribution of x for all stimuli leading to a spike, $P_v(x|\text{spike})$. We determine how much the direction v contributes to spike generation by comparing these two distributions. If these distributions are very similar, it means that v is just as likely to precede a spike as not, and it doesn't contribute towards eliciting spikes. If, however, the two distributions are separated, then it means the presence of direction v in the

stimulus affects spike generation. We can quantify this separation using a measure of mutual information, also called the Kullback-Leibler divergence:

$$I(v) = \int dx P_v(x | spike) * \log_2 \left(\frac{P_v(x | spike)}{P_v(x)} \right)$$

We used an optimization algorithm (a combination of gradient ascent and simulated annealing methods) to find the direction that maximizes this function. This algorithm turns out to be robust to the choice of starting value when used with naturalistic stimuli (Sharpee et al. 2004).

This allowed us to quantify the relative importance of a single direction in stimulus space. However, the response of many sensory neurons depends on multiple directions in stimulus space. To find the first two MIDs, we maximized with respect to multiple dimensions simultaneously by multiplying the distributions for each individual direction.

To assess the predictive value of extracted MIDs, we first calculated the nonlinear relationship between each receptive field's output and the cell's actual response (following the methods of Brenner et al. 2000, Sharpee et al. 2011). We computed predicted responses to repeated stimulus segments by first convolving the receptive field with the spectrogram of this repeated segment, summing across frequency bands, and then applying the nonlinearity to the output of this convolution. Finally, these predicted responses were compared to the actual responses by calculating the cross correlation between these predictions and the peristimulus time histogram (PSTH) of the response to repeated stimuli.

Discussion

I set out to test the ability of zebra finches to classify songs on the basis of subtle differences across social contexts, and to investigate the emergence of complex auditory responses in the forebrain. First, I demonstrated that birds could learn to classify single motifs according to the social context in which they were sung, consistent with the notion that the subtle spectral and temporal cues within a single motif could be relevant for female preference. I also showed that, given only the duration of each syllable in the training set, a simple machine learning algorithm could classify most males' directed and undirected motifs quite well. However, this model was not able to classify motifs from one of the five males tested, nor was the model's behavior consistent with the classification patterns exhibited by females on individual stimuli within the training and probe sets. These results indicate that spectral and temporal cues beyond those tested by the model influence the birds' behavior. Finally, lesions of a nucleus required for social context-dependent differences in spectral variability caused most males to produce songs whose social context, while not as easily classified, was still detectable to females performing the task. Together, these experiments demonstrate that the birds' ability to detect social context differences is based on many subtle cues and can be robust to changes in the male's motor program.

Second, I found that the response properties of most neurons in field L and CM depend strongly on the statistics of the stimulus. By comparing responses to the natural and artificial stimuli, I discovered that responses were more reliable to the natural stimuli, and that receptive fields calculated from the natural stimuli predicted novel responses better than the receptive fields calculated using artificial stimuli. Surprisingly, this stimulus dependence was similar in the two brain areas, contrary to our expectation that neurons in CM would have more differences

in their responses to the two stimulus types than neurons in field L. In both field L and CM, and for both putative excitatory and inhibitory neurons, natural stimuli gave rise to more complex filter shapes than the artificial stimulus. Receptive fields calculated in response to the natural, song stimuli also had excellent predictive value, far surpassing that of the receptive fields calculated from the artificial stimuli. Our results indicate that for many neurons in the songbird auditory forebrain, receptive field structure is highly dependent on stimulus statistics, and that receptive fields constructed in response to different stimulus classes contain surprisingly little information regarding responses to other sounds.

These findings suggest several avenues for future investigation. What precise spectrotemporal features are relevant for female classification of directed and undirected motifs? Are these the same features females prefer to hear when listening to directed bouts of song? What areas of the brain are responsible for these preferences and classification abilities? How do the properties of single neurons relate to the birds' perceptual abilities and preferences? How do the receptive fields of single neurons relate to their song selectivity?

The operant conditioning experiments chapter one were inspired by the results of Woolley and Doupe (2008), which revealed that females strongly prefer directed over undirected bouts of song. Although song bouts contain many structural differences across social context quite obvious to the human listener, Woolley and Doupe found that the only cue correlated with the strength of female preference was a subtle change in the variability of the fundamental frequency across motifs. Females prefer directed song more strongly when the variability is lower in directed bouts than in undirected bouts. Female zebra finches could be directly sensitive to this measure, i.e. to the spectral variability across motifs in a bout, or they could be attending to the (correlated measure of) spectral variability within the motif itself. While not a

demonstration of preference, my results are consistent with the notion that females are sensitive to the spectral and temporal cues within a single motif. Now that we know that females can detect these subtle changes, the obvious next step is to see if these differences affect female preference. Females are unlikely to demonstrate preference for single motifs of male song taken out of the context of a bout, so I suggest an experiment in which all of the motifs within a bout of song are replaced with a single rendition of a directed motif, preserving the other structural cues within the bout such as the number of repeats and introductory notes. The strength of preference for this artificial bout could be measured with respect to another bout in which all of the motifs have been replaced with a single rendition of the male's undirected song. In this way, one could determine whether females prefer a single rendition of a male's directed motif to his undirected motif.

Another avenue of exploration is to understand the relative influence of spectral and temporal cues on female classification behavior. I have shown that a simple model trained on the duration of syllables alone cannot fully explain the females' classification patterns. An algorithm trained on spectral cues (such as the power spectrum of each syllable), or a combination of temporal and spectral cues might better recapitulate the birds' behavior. One could test the resulting model by specifically altering the spectral and temporal properties of probe stimuli to determine when classification behavior is disrupted. Such results would help elucidate the specific features of the motif most relevant for female classification behavior.

Once we know more about the specific features that affect female classification and preference of social context cues, we can specifically modulate these features during neural recordings. These experiments could tell us at what level in the ascending auditory pathway selectivity for these features emerges. Immediate early gene responses in CM, unlike other areas

of the auditory forebrain, are sensitive to social context (Woolley and Doupe 2008), suggesting this brain area as a potential substrate for detection of these features.

Another approach to understanding the relationship between auditory perception and neural encoding is to examine the relationship between behaviorally relevant stimulus features and observed receptive field structures. The fine structure of the neuronal filter shapes observed in chapter two in response to natural stimuli might be an appropriate substrate for the creation of the perceptual abilities demonstrated in chapter one. The precise spectral bands observed in these filters may serve as excellent detectors of slight changes in either the relative power or tuning of the harmonically spaced frequency bands found in song. Other neurons seem particularly tuned for fast temporal responses, which may be important for detecting the fine temporal changes that occur across social contexts and individuals. One way to quantify this relationship would be to look at the modulation power spectrum for different song motifs and receptive fields (Woolley et al. 2005, 2009). Similarities in the spectra for stimulus and receptive field might aid in understanding the precise stimulus features neurons extract.

Careful comparison of song selectivity and receptive field properties in single neurons may also help us understand the emergence of complex auditory response properties in the songbird forebrain. For some of the neurons presented in chapter two, I have enough song data to measure motif selectivity for each motif presented during the song stimulus. The relative song selectivity of these neurons may help explain the diverse filter shapes and predictive values observed in both field L and CM.

Together, these proposed experiments should help us link the neural representations of sounds to specific behavioral patterns.

Bibliography

- Amin, N., Gill, P. & Theunissen, F.E., 2010. Role of the zebra finch auditory thalamus in generating complex representations for natural sounds. *Journal of neurophysiology*, 104(2), pp.784–798.
- Atencio, C.A. & Schreiner, C.E., 2008. Spectrotemporal processing differences between auditory cortical fast-spiking and regular-spiking neurons. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 28(15), pp.3897–3910.
- Atencio, C.A., Sharpee, T.O. & Schreiner, C.E., 2008. Cooperative nonlinearities in auditory cortical neurons. *Neuron*, 58(6), pp.956–966.
- Atencio, C.A., Sharpee, T.O. & Schreiner, C.E., 2009. Hierarchical computation in the canonical auditory cortical circuit. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51), pp.21894–21899.
- Becker, P., 1982. The coding of species-specific characteristics in bird sounds. In D. Kroodsma, E. Miller, & H. Ouellet, eds. *Acoustic Communication in Birds*. New York, NY: Academic Press, Inc.
- Beecher, M.D., Campbell, S.E. & Stoddard, P.K., 1994. Correlation of song learning and territory establishment strategies in the song sparrow. *Proceedings of the National Academy of Sciences of the United States of America*, 91(4), pp.1450–1454.
- Brenner, N., Bialek, W. & de Ruyter van Steveninck, R., 2000. Adaptive rescaling maximizes information transmission. *Neuron*, 26(3), pp.695–702.
- Brooks, R.J. & Falls, J.B., 1975. Individual recognition by song in white-throated sparrows. III. Song features used in individual recognition. *Canadian Journal of Zoology*, 53(12), pp.1749–1761.

- Connors, B.W. & Gutnick, M.J., 1990. Intrinsic firing patterns of diverse neocortical neurons. *Trends in neurosciences*, 13(3), pp.99–104.
- Cooper, B.G. & Goller, F., 2006. Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *Journal of neurophysiology*, 95(6), pp.3798–3809.
- Depireux, D.A. et al., 2001. Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex. *Journal of neurophysiology*, 85(3), pp.1220–1234.
- Dooling, R.J., 1979. Temporal summation of pure tones in birds. *The Journal of the Acoustical Society of America*, 65(4), pp.1058–1060.
- Dooling, R.J. & Searcy, M.H., 1985. Temporal integration of acoustic signals by the budgerigar (*Melopsittacus undulatus*). *The Journal of the Acoustical Society of America*, 77(5), pp.1917–1920.
- Doupe, A.J. & Kuhl, P.K., 1999. Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience*, 22, pp.567–631.
- Eggermont, J.J., Johannesma, P.M. & Aertsen, A.M., 1983. Reverse-correlation methods in auditory research. *Quarterly reviews of biophysics*, 16(3), pp.341–414.
- Elhilali, M. et al., 2007. Auditory cortical receptive fields: stable entities with plastic abilities. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(39), pp.10372–10382.
- Elliott, T.M., Hamilton, L.S. & Theunissen, F.E., 2013. Acoustic structure of the five perceptual dimensions of timbre in orchestral instrument tones. *The Journal of the Acoustical Society of America*, 133(1), pp.389–404.

- Fortune, E.S. & Margoliash, D., 1992. Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *The Journal of comparative neurology*, 325(3), pp.388–404.
- Fritz, J. et al., 2003. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience*, 6(11), pp.1216–23.
- Fritz, J., Elhilali, M. & Shamma, S., 2005. Active listening: task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex. *Hearing research*, 206(1-2), pp.159–176.
- Fritz, J.B., Elhilali, M. & Shamma, S.A., 2005. Differential dynamic plasticity of A1 receptive fields during multiple spectral tasks. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(33), pp.7623–7635.
- Gentner, T.Q. & Hulse, S.H., 1998. Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal behaviour*, 56(3), pp.579–594.
- Gentner, T.Q. & Margoliash, D., 2003. Neuronal populations and single cells representing learned auditory objects. *Nature*, 424(6949), pp.669–74.
- Gur, M., Beylin, A. & Snodderly, D.M., 1999. Physiological properties of macaque V1 neurons are correlated with extracellular spike amplitude, duration, and polarity. *Journal of neurophysiology*, 82(3), pp.1451–1464.
- Henze, D.A. et al., 2000. Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *Journal of neurophysiology*, 84(1), pp.390–400.
- Hessler, N.A. & Doupe, A.J., 1999. Social context modulates singing-related neural activity in the songbird forebrain. *Nature neuroscience*, 2(3), pp.209–211.

- Jeanne, J.M. et al., 2011. Emergence of learned categorical representations within an auditory forebrain circuit. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(7), pp.2595–2606.
- Jeanne, J.M., Sharpee, T.O. & Gentner, T.Q., 2013. Associative learning enhances population coding by inverting interneuronal correlation patterns. *Neuron*, 78(2), pp.352–63.
- Kao, M.H. & Brainard, M.S., 2006. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of neurophysiology*, 96(3), pp.1441–1455.
- Kao, M.H., Doupe, A.J. & Brainard, M.S., 2005. Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature*, 433(7026), pp.638–643.
- Kim, G. & Doupe, A., 2011. Organized representation of spectrotemporal features in songbird auditory forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(47), pp.16977–16990.
- Kojima, S. et al., *Basal ganglia contribute to the fluctuation in behavioral trajectories and its social modulation*,
- Kowalski, N., Depireux, D.A. & Shamma, S.A., 1996. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *Journal of neurophysiology*, 76(5), pp.3503–3523.
- Kowalski, N., Depireux, D.A. & Shamma, S.A., 1996. Analysis of dynamic spectra in ferret primary auditory cortex. II. Prediction of unit responses to arbitrary dynamic spectra. *Journal of neurophysiology*, 76(5), pp.3524–3534.
- Lewicki, M.S. & Arthur, B.J., 1996. Hierarchical organization of auditory temporal context sensitivity. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 16(21), pp.6987–6998.

- Linden, J.F. et al., 2003. Spectrotemporal structure of receptive fields in areas AI and AAF of mouse auditory cortex. *Journal of neurophysiology*, 90(4), pp.2660–2675.
- Lohr, B. & Dooling, R.J., 1998. Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *Journal of comparative psychology (Washington, D.C. : 1983)*, 112(1), pp.36–47.
- London, S.E. & Clayton, D.F., 2008. Functional identification of sensory mechanisms required for developmental song learning. *Nature neuroscience*, 11(5), pp.579–586.
- Maier, E.H. & Klump, G.M., 1990. Auditory duration discrimination in the European starling (*Sturnus vulgaris*). *The Journal of the Acoustical Society of America*, 88(2), pp.616–621.
- Margoliash, D., 1983. Acoustic parameters underlying song-specific neurons in the white-crowned sparrow. *Journal of Neuroscience*, 3(5), pp.1039–1057.
- Margoliash, D., 1986. Preference for autogenous song by auditory neurons in a song system nucleus of the white-crowned sparrow. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 6(6), pp.1643–1661.
- Markram, H. et al., 2004. Interneurons of the neocortical inhibitory system. *Nature reviews. Neuroscience*, 5(10), pp.793–807.
- Meliza, C.D., Chi, Z. & Margoliash, D., 2010. Representations of conspecific song by starling secondary forebrain auditory neurons: toward a hierarchical framework. *Journal of neurophysiology*, 103(3), pp.1195–1208.
- Meliza, C.D. & Margoliash, D., 2012. Emergence of selectivity and tolerance in the avian auditory cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 32(43), pp.15158–68.

- Mello, C. V, Vicario, D.S. & Clayton, D.F., 1992. Song presentation induces gene expression in the songbird forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, 89(15), pp.6818–6822.
- Miller, D.B., 1979. Long-term recognition of father's song by female zebra finches. *Nature*, 280(5721), pp.389–391.
- Miller, E., 1982. Character and variance shift in the acoustic signals of birds. In D. Kroodsma, E. Miller, & H. Ouellet, eds. *Acoustic Communication in Birds*. New York, NY: Academic Press, Inc.
- Miller, L.M. et al., 2002. Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of neurophysiology*, 87(1), pp.516–527.
- Mitchell, J.F., Sundberg, K.A. & Reynolds, J.H., 2007. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron*, 55(1), pp.131–141.
- Muller, C.M. & Leppelsack, H.J., 1985. Feature extraction and tonotopic organization in the avian auditory forebrain. *Experimental brain research*, 59(3), pp.587–599.
- Nagel, K. et al., 2011. A bird brain's view of auditory processing and perception. *Hearing Research*, 273(1-2), pp.123–133.
- Nagel, K.I. & Doupe, A.J., 2008. Organizing principles of spectro-temporal encoding in the avian primary auditory area field L. *Neuron*, 58(6), pp.938–955.
- Nagel, K.I. & Doupe, A.J., 2006. Temporal processing and adaptation in the songbird auditory forebrain. *Neuron*, 51(6), pp.845–859.
- Nagel, K.I., McLendon, H.M. & Doupe, A.J., 2010. Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of neurophysiology*, 104(3), pp.1426–1437.

Nelken, I., 2004. Processing of complex stimuli and natural scenes in the auditory cortex.

Current opinion in neurobiology, 14(4), pp.474–480.

Nelson, D.A., 1989. Song frequency as a cue for recognition of species and individuals in the

field sparrow (*Spizella pusilla*). *Journal of comparative psychology (Washington, D.C. :*

1983), 103(2), pp.171–176.

Nowak, L.G., Sanchez-Vives, M. V & McCormick, D.A., 2008. Lack of orientation and

direction selectivity in a subgroup of fast-spiking inhibitory interneurons: cellular and

synaptic mechanisms and comparison with other electrophysiological cell types. *Cerebral*

cortex (New York, N.Y. : 1991), 18(5), pp.1058–1078.

Ohms, V.R. et al., 2010. Vocal tract articulation in zebra finches. *PloS one*, 5(7), p.e11923.

Okanoya, K. & Dooling, R.J., 1990. Temporal integration in zebra finches (*Poephila guttata*).

The Journal of the Acoustical Society of America, 87(6), pp.2782–2784.

Olveczky, B.P., Andalman, A.S. & Fee, M.S., 2005. Vocal experimentation in the juvenile

songbird requires a basal ganglia circuit. *PLoS biology*, 3(5), p.e153.

Scheich, H. et al., 1979. Functional organization of some auditory nuclei in the guinea fowl

demonstrated by the 2-deoxyglucose technique. *Cell and tissue research*, 204(1), pp.17–

27.

Schneider, D.M. & Woolley, S.M.N., 2013. Sparse and background-invariant coding of

vocalizations in auditory scenes. *Neuron*, 79(1), pp.141–152.

Sen, K., Theunissen, F.E. & Doupe, A.J., 2001. Feature analysis of natural sounds in the

songbird auditory forebrain. *Journal of Neurophysiology*, 86(3), pp.1445–1458.

- Shapley, R., Hawken, M. & Ringach, D.L., 2003. Dynamics of orientation selectivity in the primary visual cortex and the importance of cortical inhibition. *Neuron*, 38(5), pp.689–699.
- Sharpee, T., Rust, N.C. & Bialek, W., 2004. Analyzing neural responses to natural signals: maximally informative dimensions. *Neural computation*, 16(2), pp.223–250.
- Sharpee, T.O. et al., 2006. Adaptive filtering enhances information transmission in visual cortex. *Nature*, 439(7079), pp.936–942.
- Sharpee, T.O., Nagel, K.I. & Doupe, A.J., 2011. Two-dimensional adaptation in the auditory forebrain. *Journal of neurophysiology*, 106(4), pp.1841–1861.
- Singh, N.C. & Theunissen, F.E., 2003. Modulation spectra of natural sounds and ethological theories of auditory processing. *The Journal of the Acoustical Society of America*, 114(6 Pt 1), pp.3394–3411.
- Soderstrom, K. et al., 2004. Endocannabinoids link feeding state and auditory perception-related gene expression. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 24(44), pp.10013–10021.
- Sossinka, R. & Böhner, J., 1980. Song Types in the Zebra Finch *Poephila guttata castanotis*. *Zeitschrift für Tierpsychologie*, 53(2), pp.123–132.
- Stripling, R. et al., 2003. Rapidly learned song-discrimination without behavioral reinforcement in adult male zebra finches (*Taeniopygia guttata*). *Neurobiology of Learning and Memory*, 79(1), pp.41–50.
- Stripling, R., Kruse, A.A. & Clayton, D.F., 2001. Development of song responses in the zebra finch caudomedial neostriatum: role of genomic and electrophysiological activities. *Journal of neurobiology*, 48(3), pp.163–180.

- Suthers, R.A., Goller, F. & Pytte, C., 1999. The neuromuscular control of birdsong. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 354(1385), pp.927–939.
- Sutter, M.L. & Margoliash, D., 1994. Global synchronous response to autogenous song in zebra finch HVC. *Journal of neurophysiology*, 72(5), pp.2105–2123.
- Swadlow, H.A. & Gusev, A.G., 2002. Receptive-field construction in cortical inhibitory interneurons. *Nature neuroscience*, 5(5), pp.403–404.
- Swadlow, H.A. & Weyand, T.G., 1987. Corticogeniculate neurons, corticotectal neurons, and suspected interneurons in visual cortex of awake rabbits: receptive-field properties, axonal properties, and effects of EEG arousal. *Journal of neurophysiology*, 57(4), pp.977–1001.
- Terpstra, N.J. et al., 2006. Localized brain activation specific to auditory memory in a female songbird. *The Journal of comparative neurology*, 494(5), pp.784–791.
- Theunissen, E. & Doupe, A.J., 1998. Temporal and Spectral Sensitivity of Complex Auditory Neurons in the Nucleus HVC of Male Zebra Finches. , 18(10), pp.3786–3802.
- Theunissen, F.E., Sen, K. & Doupe, A.J., 2000. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 20(6), pp.2315–2331.
- Vates, G.E. et al., 1996. Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches (*Taenopygia guttata*). *Journal of Comparative Neurology*, 366(4), pp.613–642.

- Volman, S.F., 1996. Quantitative assessment of song-selectivity in the zebra finch “high vocal center”. *Journal of comparative physiology. A, Sensory, neural, and behavioral physiology*, 178(6), pp.849–862.
- Wang, Y., Brzozowska-Prechtel, A. & Karten, H.J., 2010. Laminar and columnar auditory cortex in avian brain. *Proceedings of the National Academy of Sciences of the United States of America*, 107(28), pp.12676–12681.
- Woolley, S.C. & Doupe, A.J., 2008. Social context–induced song variation affects female behavior and gene expression. *PLoS biology*, 6(3), p.e62.
- Woolley, S.M.N. et al., 2009. Functional groups in the avian auditory system. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(9), pp.2780–2793.
- Woolley, S.M.N. et al., 2005. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature neuroscience*, 8(10), pp.1371–1379.
- Woolley, S.M.N., Gill, P.R. & Theunissen, F.E., 2006. Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(9), pp.2499–2512.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts 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, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature  Date January 12, 2015