UNIVERSITY OF CALIFORNIA, SAN DIEGO

Representational plasticity during task performance and learning in the avian auditory cortex

A dissertation submitted in partial satisfaction of the requirements for the degree

Doctor of Philosophy

in

Neurosciences

by

Daniel Philip Knudsen

Committee in Charge:

Professor Timothy Gentner, Chair
Professor Takaki Komiyama
Professor William Kristan
Professor Pamela Reinagel
Professor Larry Squire

2013

The Dissertation of Daniel Philip Knudsen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2013

DEDICATION

For Dad, Mom, Calli, and Kara.

Thank you for giving me everything I needed.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

This work would not have been possible without the support, friendship, encouragement, and enthusiasm of all the members of the Gentner lab. Thank you Jason, Jamie, Emily, Andrei, Micah, Becca, Justin, Krista, and Jordan. You have each helped and encouraged me in too many ways to list. You are a very special group of individuals, and I look forward to witnessing the amazing things each of you will (continue to) do. Many thanks are due also to Tim Gentner. Besides giving me a place to do my research, you encouraged me to try new things and most of all to trust myself in executing and interpreting my science. You are patient, always supportive, and an insightful mentor. Thank you for the effort and energy you put into my training.

Besides the immediate members of my lab, I'd like to thank my fellow graduate students, professors, and friends from the UCSD Neurosciences Graduate Program. What an amazing group of people! It has been an incredible honor to be affiliated with the best neuroscientists in the world, and to learn, play, and live alongside you all. Thanks also to Telefunk for giving me a creative (and cathartic) outlet – I encourage all future graduate students to join a band that will let them literally bang on drums after a frustrating day in lab.

This document would not exist if it were not for the experiences I was afforded by my undergraduate institution, Northeastern University in Boston, Massachusetts. I especially want to acknowledge the special mentorship I received and still enjoy from Drs. Jim Stellar, Joe Ayers, and Don O'Malley. You each shared so much of

yourselves and your experience with me; thank you. Special thanks are also in order for my good Northeastern friends Gordo, Mariko, Tim, and Kelly, who gave and still give me their support, friendship, and advice.

Mom, Dad, and Calli, Thank you for giving me everything I ever needed to succeed, and especially for encouraging me to ask questions and always look for the "big boy" answer. Thank you for your patience, love, encouragement, and pride. I am lucky to be part of such a loving and supportive family. As if one amazing family wasn't enough, I'm fortunate to have married in to another. Thank you Tim, Rhonda, Kelsey, Tyler, and Kylie; you have always made me feel loved and supported.

Kara, there are no words that suffice to express the depths of my gratitude for your support, sacrifice, and love. You are wonderful.

Chapter 1, in full, is a reprint of the material as it appears in Brain and Language, 2010. Knudsen, Daniel P.; Gentner, Timothy Q., Elsevier, 2010. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in Learning and Motivation, 2010. Knudsen, Daniel P.; Thompson, Jason V.; Gentner, Timothy Q., Elsevier, 2010. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in the Journal of Neurophysiology, 2013. Knudsen, Daniel P.; Gentner, Timothy Q., The American Physiological Society, 2013. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Knudsen, Daniel P; Gentner, Timothy Q. The dissertation author was the primary investigator and author of this material.

VITA

2006 Bachelor of Science in Behavioral Neuroscience *summa cum laude*, minor Physics, University Honors Program Distinction, Honors Project Distinction
Northeastern University, Boston, MA

2013 Doctor of Philosophy in Neurosciences
University of California, San Diego

PEER-REVIEWED RESEARCH ARTICLES

Knudsen D.P., Gentner T.Q. (2013) Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. *Journal of neurophysiology,* 109: 1690–703. doi:10.1152/jn.00461.2012.

Knudsen, D.P., Thompson, J.V., & Gentner, T.Q. (2010). Distributed recognition of natural songs by European starlings. *Learning and motivation*, *41*(4), 287-306. doi:10.1016/j.lmot.2010.08.005

Ayers, J., Rulkov, N., Knudsen, D., Kim, Y.B., Volkovskii, A., & Selverston, A. (2010). Controlling underwater robots with electronic nervous systems. *Applied bionics and biomechanics*, *7*(1), 57-67. Taylor & Francis. doi:10.1080/11762320903244843

Hildebrand, D.G.C.*, Knudsen, D.P.*, Hesse, G.W., & Stellar, J.R. (2009). A flexible system for hands-free intracranial microinjection. *Journal of neuroscience methods*, *185*(1), 62-5. doi:10.1016/j.jneumeth.2009.09.011(*equal contribution)

Knudsen, D.P., Arsenault, J., Hill, S., Omalley, D., & Jose, J. (2006). Locomotor network modeling based on identified zebrafish neurons. *Neurocomputing*, *69*(10-12), 1169-1174. doi:10.1016/j.neucom.2005.12.068

REVIEWS

Knudsen, D.P., & Gentner, T.Q. (2010). Mechanisms of song perception in oscine birds. *Brain and language*, 115(1), 59-68. doi:10.1016/j.bandl.2009.09.008

CONFERENCE PROCEEDINGS

Ayers, J., Rulkov, N., Taub, E., Knudsen, D., Kim, Y.B., Selverston, A., & Uswatte, G. (2006). Electronic Neurons: From Biomimetic Robots to Blast Neurorehabilitation. *Proceedings of the 7th international conference on technology and the mine problem* (p. 10). Monterey, CA. doi:10.1.1.75.9255

ABSTRACT OF THE DISSERTATION

Representational plasticity during task performance and learning
in the avian auditory cortex

by

Daniel Philip Knudsen

Doctor of Philosophy in Neurosciences

University of California, San Diego, 2013

Professor Timothy Gentner, Chair

The brain's representation of the auditory world is not static, but changes based on an animal's history and current goals. We explored experience-dependent changes in both behavioral and neural representations of behaviorally relevant auditory stimuli in the caudal mesopallium, a secondary auditory forebrain region, in European

starlings. To accomplish this, we first designed and built a system that facilitated simultaneous neural and behavioral recording, allowing—for the first time—neural responses to be recorded while birds performed auditory-mediated operant tasks. We found that when birds were engaged in an auditory task, neurons had more information in their stimulus-driven firing rates about the task-relevant stimulus classes than when birds were not engaged in the task. We also trained birds to quickly learn novel stimulus classifications and showed that neural responses change over the course of learning. For most neurons, stimulus-driven neural responses decreased with learning, and they did so most strongly for the newly learned stimuli. We suggest that these effects may be due in part to stimulus-specific adaptation, and its modulation by behavioral state. We also observed the formation of task-dependent firing rate modulation with learning. These results highlight the plasticity of the avian auditory system, and further our understanding of the way that nervous systems adapt to the changing environment and behavioral goals of an organism.

INTRODUCTION

A fundamental challenge in neuroscience is to understand the way that nervous systems build representations of the external world and then use those representations to interact with that world. We know that different brains do not encode the same representation of the external environment, and that the manner in which the world is perceived depends strongly on an organism's goals at timescales that range from evolutionary to immediate. On evolutionary timescales, the nervous system of a particular species represents the information present in the environment in a manner that is adapted to its ecological niche, leading to changes in the structure and function of both sensory organs and motor effectors and of the neural systems that mediate their interaction. For example, pit vipers have specialized organs that sense the heat given off by their prey in the infrared, while primates have elaborately developed visual systems that allow us to navigate our environment using shorter wavelengths of light. In the lifetime of a single organism, experience with the environment during both early development and in adult life can shape sensory representations. A newborn is able to detect phonetic contrasts present in all languages equally well, but between 6-12 months of age, this ability is greatly reduced to the set of sounds prevalent in the

language she hears most frequently (Kuhl, 2004). In addition to developmental changes, long-term alterations due to training can be observed as tuning shifts in tonotopic maps and receptive fields in adult mammalian primary auditory cortex (Dahmen & King, 2007; Weinberger, 2007), and human musicians show enhancements in auditory encoding when compared to non-musicians (Musacchia, Strait, & Kraus, 2008). Even short-term modifications to sensory representation can be accomplished by altering an animal's immediate goals. This is perhaps best exemplified in exquisite studies of selective attention in primate visual cortex, where the single-neuron representation of identical stimuli can be altered on a trial-by-trial basis depending on which features or locations in space are attended (Reynolds & Chelazzi, 2004). Sensory representations are strongly dependent on an animal's experience and behavioral goals, and the nervous systems that have evolved to represent the sensory world have also evolved to represent the fact that as this world changes, an organism's relationship with the world must change along with it.

More than two centuries of detailed behavioral ecology shows that songbirds interact with the auditory world in a sophisticated manner that depends on their individual life histories. Much of this interaction comes in relation to songbirds' rather unique facility for vocal learning (a behavior shared with only a few other taxa of birds and mammals, including humans). In addition to receiving auditory information from the world around them, songbirds actively engage in producing complex songs that they learn from other members of their species. The production and perception of these songs mediates complex social interactions such as mating and individual

recognition that directly affect birds' fitness. A male bird's ability to accurately perceive, to remember, and to copy his tutor's song, and a female's ability to perceive the quality of a male's song and compare it to the songs of other males are therefore both under strong selection pressure. These auditory-mediated behaviors depend on a bird's previous auditory experience, and it is therefore no surprise that the songbird auditory system shows evidence of experience-dependent plasticity (Chew, Vicario, & Nottebohm, 1996; Gentner & Margoliash, 2003; Jeanne, Thompson, Sharpee, & Gentner, 2011; Mello, Vicario, & Clayton, 1992; Thompson & Gentner, 2010). Sophisticated auditory cognition, coupled with the evidence of behaviorally dependent plasticity of representation in the auditory system makes songbirds an attractive model for studying the influence of behavior and experience on auditory processing.

The following chapters are an account of my graduate work, performed in the laboratory of Dr. Timothy Gentner at the University of California, San Diego, where I have sought to characterize the manner in which the avian auditory system adapts to changes in its relationship to its auditory environment. Working with other members of the lab and using the European starling (*Sturnus vulgaris*), a species of songbird, as a model species, we performed experiments to characterize the behavioral and neural representation of conspecific song, and the ways in which these representations changed as a function of behavioral demands and training history.

I begin in Chapter 1 by discussing avian auditory perception and cognition, and the usefulness of avian vocal communication as a model for understanding the mechanisms underlying human speech, a necessary precursor to spoken language. In

vocal communication—and especially in human language—the notion of a brain dynamically interacting with the environment arguably reaches a pinnacle, as vocal communicators use the output of their nervous systems to directly influence the neural state of their communication target.

Chapter 2 describes a set of behavioral studies performed to assess the manner in which auditory stimuli are behaviorally encoded in the songbird brain. The results of this paper set limits on the ability of birds to encode subsets of a stimulus and associate them with a particular behavioral outcome leading to reward. Birds appear to build distributed stimulus representations, and can perform a simple binary classification task with small segments of training stimuli, but they will use extra information to perform the task if it is available.

Previous studies of neurons in the auditory system of songbirds have been limited to anesthetized or awake-restrained recording, and lacked any controlled behavioral manipulations during neural recording. Chapter 3 describes the development of a system that allowed us to record from single neurons in the auditory cortex of starlings while they performed an auditory classification task. We used this system to study the manner in which engaging in an auditory task alters stimulus representation in a region of the avian auditory cortex that responds to complex vocalizations, the caudal mesopallium. We find that when the bird engages in an auditory-mediated task, neurons in this region alter their responses to stimuli, and for a large subpopulation of these neurons, engagement in the task improves the ability of a neuron to encode the stimulus differences that are important for successful

classification. That is, when a bird is engaged in an auditory task, there is more information about that task encoded by these neurons than when the bird is not engaged in this auditory recognition task.

Chapter 4 discusses use of the behavioral electrophysiology system described in Chapter 3 to monitor changes in neural coding that take place across learning. We trained birds to quickly recognize sets of novel stimuli and monitored neurons in the medial portion of the caudal mesopallium while they learned to recognize these novel stimuli. We found that the stimulus representations were not static, but changed over the course of learning, generally decreasing their responsiveness to presented stimuli, especially for newly learned auditory stimuli. We also report that some neurons appear to develop the types of task-engagement dependent differences described in Chapter 3 over the course of learning.

Following these chapters is a discussion of the ways in which these studies enhance our understanding of auditory cognition, and potential future directions that build from the results obtained here.

REFERENCES

Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of individual birds. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 1950–1955.

Dahmen, J. C., & King, A. J. (2007). Learning to hear: plasticity of auditory cortical processing. *Current opinion in neurobiology*, 17(4), 456–64. doi:10.1016/j.conb.2007.07.004

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature*, 424(6949), 669–74. doi:10.1038/nature01731

Jeanne, J. M., Thompson, J. V., Sharpee, T. O., & Gentner, T. Q. (2011). Emergence of learned categorical representations within an auditory forebrain circuit. *The Journal of Neuroscience*, 31(7), 2595–606. doi:10.1523/JNEUROSCI.3930-10. 2011

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. Nature reviews. Neuroscience, 5(11), 831–43. doi:10.1038/nrn1533

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), 6818–22.

Musacchia, G., Strait, D., & Kraus, N. (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing research*, *241*(1-2), 34–42. doi:10.1016/j.heares.2008.04.013

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Reviews of Neuroscience,* 27, 611–47. doi:10.1146/annurev.neuro.26.041002.131039

Thompson, J. V, & Gentner, T. Q. (2010). Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *Journal of neurophysiology*, 103(4), 1785–97. doi:10.1152/jn.00885.2009

Weinberger, N. M. (2007). Associative representational plasticity in the auditory cortex: a synthesis of two disciplines. *Learning & memory (Cold Spring Harbor, N.Y.)*, 14(1-2), 1–16. doi:10.1101/lm.421807

CHAPTER 1

MECHANISMS OF SONG PERCEPTION IN OSCINE BIRDS

KNUDSEN, D. P. AND GENTNER, T. Q.

AS IT APPEARS IN

BRAIN AND LANGUAGE

2010

VOLUME 115 ISSUE 1, PAGES 59-68

DIGITAL OBJECT IDENTIFIER (DOI): 10.1016/J.BANDL.2009.09.008

Mechanisms of song perception in oscine birds

Daniel P. Knudsen [a], Timothy Q. Gentner [b,*]

[a] Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, United States
[b] Department of Psychology, University of California, San Diego, La Jolla, CA 92093, United States

ARTICLE INFO

Article history:
Accepted 23 September 2009
Available online 14 May 2010

Keywords:
Psychoacoustics
Auditory
Vocal communication
Language
Songbird
Evolution
Neuroethology
Speech

ABSTRACT

Songbirds share a number of parallels with humans that make them an attractive model system for studying the behavioral and neurobiological mechanisms that underlie the learning and processing of vocal communication signals. Here we review the perceptual and cognitive mechanisms of audition in birds, and emphasize the behavioral and neural basis of song recognition. Where appropriate, we point out a number of intersections with human vocal communication behavior that suggest common mechanisms amenable to further study, and limitations of birdsong as a model for human language.

1. Introduction

Like other communication signals, one adaptive function of birdsong is to influence the behavior of others, usually conspecific individuals (Kroodsma & Miller, 1996). Communication signals achieve this function by transmitting information between the sender of the signal and the receivers. The success of this transmission rests on predictability. When a sender produces a specific signal, in this case a song, it does so under the expectation that the signal will elicit a predictable (i.e. intended) behavior in the receiver. Without the predictable correspondence between production and perception, signals would lose their functionality. Thus, the presence of a functional signal implies a reliable correspondence between production and perception mechanisms shaped and maintained by selection pressures. This correspondence confers a special status to communication signals. Like other natural stimuli, communication signals are often physically complex. Unlike most complex natural stimuli, however, many of the physical dimensions along which communication signals vary can be directly tied to adaptive behaviors. Research in oscine birds has capitalized on this idea to study the mechanisms underlying the perception and cognition of complex natural stimuli (song) in the context of natural behaviors.

What follows is an account of our current understanding about the ways in which the songbird auditory system interprets a continuous stream of acoustic information as a collection of behaviorally relevant communication signals and uses this information to affect behavior (Fig. 1). Parallels will be drawn to human vocal communication, and we will present a case for the use of songbirds as a model of certain aspects of human language processing. We first provide a broad overview of perceptual psychophysics in songbirds so that one can appreciate the strong perceptual similarities between birds and humans. We then review the behavioral and neurophysiological work on conspecific song perception, focusing on individual vocal recognition mechanisms in European starlings, a species of songbird.

2. Perceptual psychoacoustics

It is helpful for any description of a complex system to begin with a characterization of general abilities along fundamental dimensions. Common dimensions along which acoustic signals are deconstructed are frequency, amplitude, and time, and it is instructive to understand sensory processing at this level. The goal of such research is to inform our understanding of how more 'atomic' descriptions of sounds give rise to the perception of more complex behaviorally relevant features. Such studies provide an important context for studies of more complex signals, and establish the range over which any perceptual ability can operate. The following section provides a brief description of avian psychoacoustics studies supporting the notion that humans and birds experience a similar acoustic world. More thorough reviews are available (Dooling, 1982, 1992, 2000; Fay, 1988).

* Corresponding author. Address: UCSD, Department of Psychology, 9500 Gilman Dr. MC0109, La Jolla, CA 92093-0109, United States.
E-mail address: tgentner@ucsd.edu (T.Q. Gentner).

0093-934X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2009.09.008

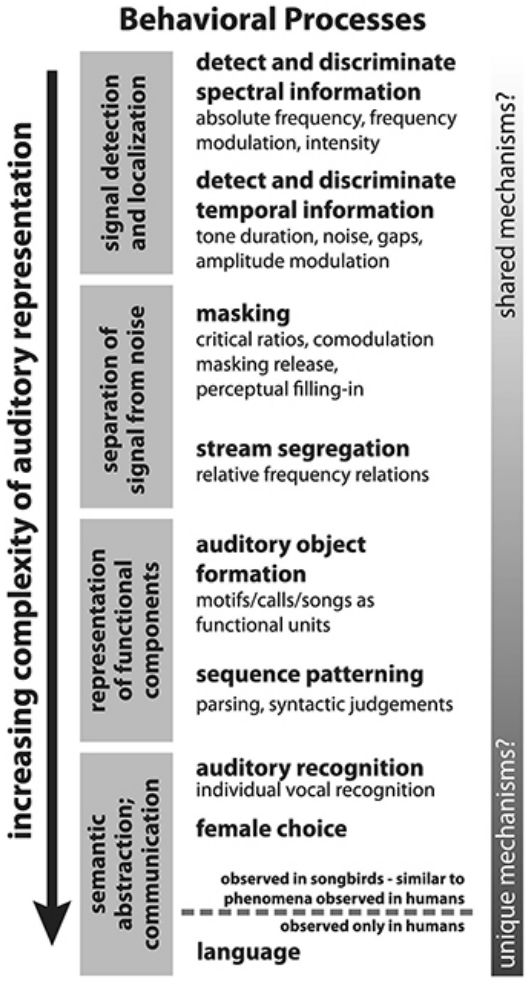


Fig. 1. Functional auditory behaviors in songbirds. The gray boxes represent the functions that the auditory system must perform to mediate behavior. Text to the right of these functions gives examples of phenomena observed in songbirds (see text). Although this review focuses on songbird auditory processing, many of the functions of the auditory system are likely conserved across a range of vertebrates, including humans. As behavioral complexity increases, so does the likelihood that particular mechanisms are unique to different species. The similarities and the differences between species yield powerful comparative hypotheses about the behavioral and neural mechanisms for auditory perception and cognition in vertebrates.

2.1. Spectral sensitivity

Audibility curves, which describe the loudness (sound pressure level, SPL) required for detectability as a function of frequency, have the same basic shape in many songbird species (Dooling, Okanoya, Downing, & Hulse, 1986; Okanoya & Dooling, 1987). The threshold sensitivity for pure tones in most songbird species is best around 2–5 kHz, increases gradually as the frequency of the tone becomes lower, and increases quite sharply as the frequency of the tone rises. The typical high-frequency cutoff for songbirds is 8–11 kHz. Although audibility thresholds are lower in humans than in birds, the overall shape of the audibility curve is similar.

In addition to hearing single tones, the auditory system must also discriminate between different tones. Understanding the minimal detectable differences between tones can point to the types of frequency modulations within a signal that are available for a species to use in vocal communication. In general, birds are quite sensitive to changes in frequency, and can discriminate a change in frequency as small as 1% (Dooling, 1982), while humans have even lower detection thresholds across the range of audible frequencies. Birds' sensitivity to frequency changes also depends on the type of frequency modulation and range of carrier frequencies (Langemann & Klump, 1992), results again similar to observations made in human studies (Demany & Semal, 1989; Fastl, 1978). To detect a change in intensity (loudness) between two successive tones, birds require a difference of about 3 dB, humans about 1 dB (Dooling, 1982). While humans have quantitatively lower detection thresholds for frequency and loudness discrimination, the similarity of findings in humans and birds point to auditory systems that are qualitatively similar in the range of psychophysically observable spectral sensitivities.

2.2. Temporal sensitivity

Most acoustic signals unfold over time, and processing in the temporal domain is therefore particularly important. Temporal processing abilities of songbirds have been studied in a variety of ways. A simple measure is the detectability of a sound as its temporal duration increases. In general, the longer the tone is played for, the lower the SPL needed for detection. Consistent with findings from a host of other animals including humans (see Brown & Maloney, 1986), birds' thresholds for hearing a pure tone improve as the duration is increased from a few milliseconds to 200–300 ms (Dooling, 1980). Another common measure of temporal acuity, known as the gap detection threshold, measures the minimum temporal interval that can be detected between two sounds. Several studies from birds show gap detection thresholds ranging from 2 to 3 ms, which is similar to thresholds found in humans (see Klump & Maier, 1989). This suggests that intervals in natural vocalizations less than 2–3 ms may not be perceived. Duration discrimination measures, which describe an organism's ability to determine whether one sound has a longer duration than another, are also similar between birds and humans (Maier & Klump, 1990).

2.3. Masking

In psychoacoustics, the critical ratio describes the ability of an organism to perceive a tone in a noisy background. It is defined as the SPL of a target tone needed for detection divided by the SPL of the background masking noise. Critical ratios are a function of the frequency of the target tone, and in accordance with previous studies in humans and in other mammals, critical ratios in most songbirds increase at about 3 dB per octave (Dooling et al., 1986; Langemann, Klump, & Dooling, 1995; Okanoya & Dooling, 1987). Most songbirds' critical ratio curves show a similar shape to those of humans and other mammals, though humans show lowered threshold levels on the order of a few decibels (Okanoya & Dooling, 1987).

Comodulation masking release (CMR) is a slightly more complex masking phenomenon that has been described in both birds and humans. CMR occurs when sounds that are modulated together across time serve to release each other from masking by overlapping noise. CMR is measured as the effective decrease in threshold SPL afforded by the comodulation. CMR has been proposed as mechanism for auditory stream segregation (discussed below), as sounds that are produced from the same source take the same path to the listener and are therefore modulated

identically. Starlings show robust CMR, and in many ways are similar to humans in terms of amount of masking release and the frequency and bandwidth dependencies of the release (Klump & Langemann, 1995).

3. Auditory cognition in birds

Within the broad framework of acoustic processes laid out by the results of psychoacoustic studies, birds (and any other organism that uses acoustic communication) face a number of specific challenges. The auditory system must first detect a potentially relevant signal against backgrounds of environmental noise or other potentially relevant signals. Signals must be processed to determine their relevance in specific behavioral contexts, and mapped onto appropriate behaviors. These behaviors are often mediated by learning, memory and attentional mechanisms, implying a suite of cognitive processes that act on complex stimulus representations such as stream segregation, discrimination, recognition, categorization, temporal pattern detection, and decision processes. The acoustic diversity of birdsong makes the system attractive for studies of the behavioral and neurobiological bases for many of these cognitive processes.

3.1. Stream segregation

One challenge to the auditory system is the separation of incoming auditory information into behaviorally relevant groups, or streams. This is referred to as stream segregation, and under most natural conditions equates to separation of streams based on originating source. In order to effectively use the acoustic information in one's environment, it is helpful to group acoustic information arising from the same source, and to treat this information as separate from information arising from other sources. In human language, for example, this ability allows one to follow the vocalizations of a speaker across time in a noisy environment.

Stream segregation is one of the main tasks of any perceptual system engaged in vocal communication, and it is therefore no surprise that songbirds are capable of separating one auditory stream from another in a variety of conditions. Like humans (Bregman, 1990), starlings can hear competing streams emerge from a sequence of alternating tones when the frequency difference between putative "streams" is large (MacDougall-Shackleton, Hulse, Gentner, & White, 1998). Likewise, starlings continue to recognize long segments of conspecific songs when noise is added in the form of other starling songs, the songs of other species, or an entire 'dawn chorus' (Hulse, MacDougall-Shackleton, & Wisniewski, 1997). Canaries can maintain accurate recognition of conspecific song in the face of very loud broadband white noise (Appeltants, Gentner, Hulse, Balthazart, & Ball, 2005). The neural correlates of the so-called cocktail party effect, that is the extraction of a meaningful communication signal from noise, have been investigated in starlings and zebra finches (Bee & Klump, 2004; Narayan et al., 2007).

3.2. Song discrimination and recognition

Many species of bird are adept at discriminating songs of their own species as well as songs from other species. Within this context, birds are generally more adept at discriminating among different conspecific songs than among heterospecific songs (Cynx & Nottebohm, 1992; Dooling, Brown, Klump, & Okanoya, 1992; Sinnott, Sachs, & Hienz, 1980). Fieldwork in sparrows is consistent with these laboratory findings, and suggests that different species may rely, at least for species recognition, on those acoustic features that most reliably distinguish their own songs from those of other species in the same sound environment (Nelson & Marler, 1990). For example, frequency appears to be the most important cue for species recognition among field sparrows (Nelson, 1989). Recognition systems appear to be tuned to the particular acoustic space of one's own species' vocalizations (Nelson, 1988), but also maintain more general capacities to distinguish among many different types of song. These biases for the discrimination of species-specific vocalizations, and hence mechanisms for vocal recognition, likely result from evolutionary or ontogenic changes in the (presumably) central processing mechanisms that underlie higher-level auditory perception. This is consistent with the more general assumption that the perceptual and cognitive processes underlying individual vocal recognition, a widespread phenomenon among songbirds (reviewed by Stoddard (1996)), operate on complex (multi-dimensional) acoustic signals. Within the same species many acoustic features may be relevant for classification of song components (Sturdy, Phillmore, & Weisman, 2000). Recent laboratory studies of European starlings have addressed these questions directly by determining more precisely the form of the acoustic signal controlling recognition in this species.

3.2.1. Individual vocal recognition in starlings

Male starlings present their songs in long episodes of continuous singing referred to as bouts. Song bouts, in turn, are composed of much smaller acoustic units referred to as motifs (Adret-Hausberger & Jenkins, 1988; Eens, Pinxten, & Verheyen, 1991), which in turn are composed of still smaller units called notes. Notes can be broadly classified by the presence of continuous energy in their spectrographic representations, and although several notes may occur in a given motif, their pattern is usually highly stereotyped between successive renditions of the same motif. One can thus consider starling song as a sequence of motifs, where each motif is an acoustically complex event. The number of unique motifs that a male starling can sing (i.e. his repertoire size) can be quite large, and consequently different song bouts from the same male are not necessarily composed of the same set of motifs. This broad acoustical variation in their song provides several potential cues that starlings might use when learning to recognize the songs of an individual conspecific and while maintaining that recognition over time. One straightforward recognition mechanism is the association of specific motifs with specific singers. Although some sharing of motifs does occur among captive males (Hausberger, 1997; Hausberger & Cousillas, 1995), the motif repertoires of different males living in the wild are generally unique (Adret-Hausberger & Jenkins, 1988; Eens, Pinxten, & Verheyen, 1989; Eens et al., 1991). Thus, learning which males sing which motifs can provide discriminative cues for song recognition. Similar strategies may underlie individual song recognition behaviors observed in other species where individual possess elaborate vocal repertoires such as song sparrows (Beecher, Campbell, & Burt, 1994; Stoddard, Beecher, Horning, & Campbell, 1991), but other strategies are available as well (Brooks & Falls, 1975a, 1975b; Nelson, 1989).

Data from operant studies in starlings indicate that song recognition is based at the level of the motif. Starlings trained using operant conditioning to recognize sets of songs from conspecific individuals can readily generalize correct recognition to novel songs from the same singers (Gentner & Hulse, 1998, 2000b). However, recognition falls to chance when these novel song bouts have no motifs in common with the training songs (Gentner & Hulse, 2000b). This failure to generalize correct recognition to songs composed of novel motifs, or to single novel motifs, is consistent with the notion that starlings learn to recognize the songs of individual conspecific singers by attending to information contained at (or below) the level of the motif. That is, they appear to associate distinct sets of motifs (or variant motif features) with individual singers.

If starlings learn to recognize individuals by the sets of unique motifs that they sing, then once learned, it should be possible to control recognition systematically by varying the proportions of motifs in a given bout that come from two "vocally familiar" males. That is, recognition behavior ought to follow the proportional distribution of motifs from two vocally familiar males rather than the presence or absence of single diagnostic motifs from either male. The behavioral data confirm this prediction. When starlings are compelled to classify conspecific songs, they do so by memorizing large numbers of unique song components (i.e. motifs) and organizing subsets of these motifs into separate classes (Gentner & Hulse, 2000b). Consistent with this idea, recent results demonstrate that motifs form perceptually salient auditory objects embedded in a hierarchy of acoustic patterns (Gentner, 2008). When the sub-motif acoustic structure is permuted in varying ways, recognition of the songs is significantly impaired. These effects are only seen for familiar motifs, however, suggesting that starlings learn the sub-motif structure of motifs and then perceive them as holistic auditory objects (Gentner, 2008).

Additional support for the notion that motifs form holistic auditory objects comes from recent results in starlings that mirror perceptual restoration studies in humans (Seeba & Klump, 2009). In the classic demonstration of perceptual restoration (Warren, 1970), units of speech (phonemes) within a presented sentence are perceived as present by human observers despite being replaced by noise (e.g. a human cough). The restoration effect only occurs when the phoneme is replaced by noise; when silence is inserted instead of noise, phonemes are correctly perceived as missing. This shows that with appropriate contextual cues, observers perceive missing acoustic information at the level of a whole phoneme, suggesting that they have a memory for phonemes as holistic auditory units. Similarly, starlings trained to detect small changes in motif similarities report significantly greater similarity between intact motifs and those with sections replaced by noise when compared to motifs with sections replaced by silence. Importantly, these results hold only for familiar motifs, suggesting that memory for motifs as learned objects informs their perception at later times (Seeba & Klump, 2009). These results provide another line of evidence for considering motifs as acoustic units and suggest that the representation of these motifs emerges via a combination of feed forward auditory input and memory for the motif as an object.

As a cognitive recognition strategy, classifying songs according to their component (motif) structure represents a straightforward method of dealing with these complex acoustic signals. Because individual starlings tend to possess unique motif repertoires, disjoint sets of motifs will generally correspond to individual identity. Therefore, attending to the motif structure captures a significant portion of the individual variation in the signal, albeit at the expense of a heavy demand on memory capacity. From a human perspective, this might seem a suboptimal strategy for individual vocal recognition, because in humans the total set of vocal signals that can be produced by a single individual is theoretically infinite. Instead, individual talker recognition in humans relies on voice characteristics such as timbre, glottal pulsation frequency, and spectral contours imparted by laryngeal morphology (Bricker & Pruzansky, 1976) that are largely independent of the linguistic content (Remez, Fellowes, & Rubin, 1997). Although recent results indicate that some voice characteristics are present in starling song (Gentner, 2008), these birds appear to recognize individuals by preferentially using information at longer temporal scales rather than the short (fast) timescale features that carry voice characteristics. This observation is consistent with the conjecture that the incorporation of individual identity information into the phonetics of the signal may be a plesiomorphic (evolutionarily older) state for vocal communication. As the size of the lexicon and sharing of signal elements between individuals increased, the association of unique vocabularies with different individuals would become increasingly inefficient, forcing the differentiation of vocal identity from other semantic channels.

3.3. Syntactic pattern sensitivity

Part of what makes human language unique is its incorporation of long timescale patterning rules (grammars) that constrain the meaning of specific utterances, e.g. telling us precisely who did what to whom. Animal vocalizations lack the combinatorial grammatical complexity of human language. Yet many animal communication signals have a rich temporal structure, and understanding how the temporal patterns in such signals are used and perceived is of broad interest. Such studies can serve to emphasize important similarities and differences in underlying mechanisms for processing acoustically complex vocal signals.

The prior sections on individual vocal recognition in starlings establish the functional importance of motif-level song organization. Under normal conditions, however, single motifs are almost never produced in isolation but typically occur as part of long and elaborate song bouts where 25–30 different motifs may be strung together in close succession. In the following sections we discuss the role of temporal motif patterning in starling song perception.

3.3.1. Motif sequences

Sensitivity to the temporal patterning of motifs could arise in its simplest form based on the explicit sequences in which familiar motifs are produced. As an example, a bird might learn to recognize motifs A–D (where letters denote different motifs) *and* the sequence ABCD. If this is the case, then presenting the same motifs in a different order may affect song recognition. Consistent with this, starlings trained to recognize naturally patterned songs have a much harder time recognizing novel songs from familiar singers when motifs are strung together in a random order, than when motifs follow each singer's natural motif ordering (Gentner & Hulse, 1998), indicating that starlings do attend to motif sequencing.

3.3.2. Motif patterns

Additional studies of temporal pattern perception in starlings move beyond simple transition probabilities between adjacent motifs to ask if starlings can acquire abstract rules that describe the patterning of familiar motifs (Gentner, Fenn, Margoliash, & Nusbaum, 2006). That is, instead of learning the explicit sequence of motifs ABCD as above, the study asked if starlings could learn the pattern *ABCD*, where the letters now denote *sets* of motifs that can occur at each position. Rules that describe sequences of patterned strings have a rich history in the theory of formal grammars, and form the basis for theories of syntax in human language. Formal grammars can be classified hierarchically according to the complexity of the patterns they can produce or recognize (Hopcroft & Ullman, 1979). Finite-state grammars are among the most limited type of formal grammar, and have been thought to describe all animal communication systems (Hauser, Chomsky, & Fitch, 2002; but see Suzuki, Buck, & Tyack, 2006). Human languages minimally require a grammar more complex than finite-state, called a context-free grammar, in part to support the hierarchical embedding common in many syntactic structures (Chomsky, 1957; Hopcroft & Ullman, 1979).

Starlings can learn to classify patterned strings of song motifs generated by both a finite-state grammar and a context-free grammar (Gentner et al., 2006). In this case, strings generated by the context-free grammar followed the form A^2B^2, while the strings generated by the finite-state grammar followed the form $(AB)^2$, where A and B refer to sets of acoustically distinct starling song

motifs known as "rattles" and "warbles". It is important to note that the context-free grammar learned in this study, while more computationally complex than the finite-state grammar, is not sufficient to fully capture the hierarchical structure in human languages. Specifically, it lacks any long-range dependencies between specific elements that, in for example some English sentences, help define the pairings between subjects and verbs, and thereby enable understanding of who did what to whom. Therefore, one should not conclude from this study that songbirds have the capacity to master "human" grammar. Nonetheless, it is clear that they can extract patterning rules from strings of vocal signals, and that those rules can have a remarkable level of complexity. Whether or not birds and humans rely on similar neuronal mechanisms to extract temporal pattern rules is unknown. In any case, the basic ability to attend to high-level patterns within strings of vocal signals, including at least some patterns defined by formal grammars, does not appear to be a recent or unique adaptation.

Although there has been some discussion of the role of syntactic rules in structuring bird song in chickadees and wrens (Clucas, Freeberg, & Lucas, 2004; Hailman & Ficken, 1986; Holland, Dabelsteen, & Paris, 2000), the existence of complex pattering rules in songbird vocalizations has not received adequate attention. Currently, there is no strong evidence to support the notion that songbirds (or any non-human species) use syntax to vary the *semantic* content of vocal signals in the combinatorial manner. By itself, the ability to process simple syntactic structures may be of little functional significance or adaptive value, and may represent a necessary but insufficient precondition to the use of unbounded signal sets observed in humans. These sorts of primitive pattern recognition abilities might need to be paired with more sophisticated gestural systems as a permissive step to subsequent evolution of human language processing capacities in their modern forms.

4. The neural coding in the songbird auditory system

The behavioral data yield several hypotheses regarding the neural mechanisms underlying the recognition of natural (i.e. complex) acoustic signals. First, the functionality of auditory objects in recognition behavior implies an explicit and discrete representation for vocal signals in the central nervous system. That is, the response functions of single neurons or of populations of neurons in appropriate auditory regions should reflect the segmentation of song at the level of the motif (in starlings) or other behaviorally relevant temporal scales. Second, because vocal recognition behavior requires the *learned* association between songs and singers, the neural representations of motifs should reflect their explicit behavioral relevance. That is, there should be a bias for representations of familiar compared to unfamiliar motifs. Third, the representational mechanisms and capacity (i.e. memory) of the system should permit the acquisition of very large numbers of acoustically complex, natural objects (motifs). The nature of object representation, representational plasticity, and memory capacity are central questions for any researcher interested in the neural coding of natural stimuli, including human speech and language. Below, we consider these questions in the context of the neural representation of starling song.

4.1. The avian auditory system

Auditory information enters the avian brain via sensory transduction in the inner ear, where vibrations on the basilar membrane of the cochlea are transduced into the action potentials propagated to the cochlear nuclei in the brainstem (Fig. 2). The cochlear nuclei are characterized by strong tonotopy, and the neurons within these nuclei have a tendency to fire tonic bursts that align with amplitude peaks in the sound pressure waveform of the stimulus (Sachs & Sinnott, 1978). Afferents from the brainstem converge in the midbrain structure mesencephalicus lateralis dorsalis (MLd). MLd is anatomically analogous to the mammalian inferior colliculus (IC), and in many respects it is also functionally analogous. Like the mammalian IC, MLd neurons show a wide variety of tuning characteristics, are not particularly sensitive to level differences, and have temporally precise responses that preserve the timing information in the acoustic signal (Woolley & Casseday, 2004). They also respond to white noise and conspecific song in a way that suggests selectivity for particular spectrotemporal features, and the vast majority of these cells follow sinusoidal amplitude

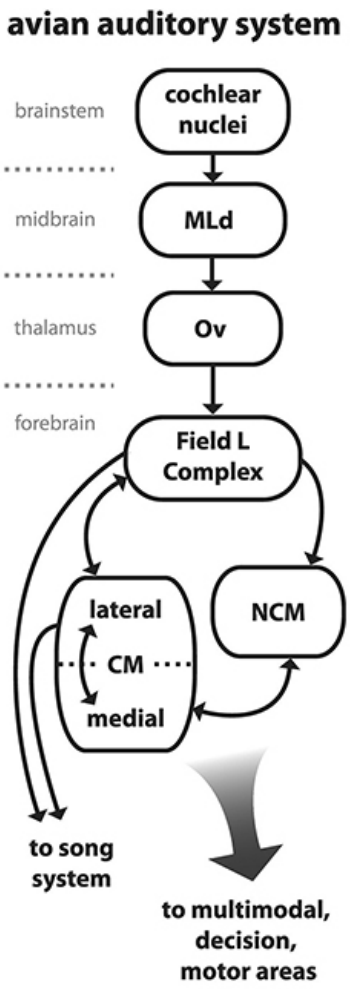


Fig. 2. Schematic of the major auditory processing regions in the avian brain. Auditory information projects from the ear to the brainstem cochlear nuclei. In general, more complex acoustic information is represented in more central structures, i.e. those further from the cochlea. Apart from those to the classic song-control system, projections from the primary auditory pathways to areas responsible for multimodal integration and non-singing behaviors are uncharacterized, but presumably exist. See the text for more details. Thinner arrows indicate direct neural projections. Abbreviations: CM: caudal mesopallium, MLd: mesencephalicus lateralis dorsalis, NCM: caudomedial nidopallium, Ov: nucleus ovoidalis.

modulations – suggesting again that these cells encode the temporal features of an acoustic signal with high fidelity (Woolley, Fremouw, Hsu, & Theunissen, 2005).

MLd sends afferents into the avian auditory thalamus, the Nucleus Ovoidalis (Ov). Ov shows a strong tonotopic arrangement, and its neurons have high spontaneous firing rates, and respond to stimuli at their preferred frequency with a tonic increase in firing rate (Bigalke-Kunz, Rübsamen, & Dörrscheidt, 1987). There is also evidence of lateral inhibition in this region, as many cells show suppressive sidebands in their frequency tuning curves (Bigalke-Kunz et al., 1987). Thalamic efferents project from Ov mainly into the Field L complex (shortened here to Field L), the avian analog of the mammalian primary auditory cortex (Karten, 1968, 1969). Specifics of cytoarchitecture and neural connectivity allow further subdivision of the Field L complex into regions L1, L2a, L2b, L3 and L (here 'L' refers to a subregion of the larger complex Fortune & Margoliash, 1992; Saini & Leppelsack, 1977, 1981). L2a is the primary thalamorecipient zone, but L1 and L3 also receive weak thalamic input (Vates, Broome, Mello, & Nottebohm, 1996). Field L shows strong tonotopy and robust responses to both natural song and pure tones (Capsius & Leppelsack, 1996, 1999; Haüsler, 1996; Leppelsack & Vogt, 1976; Sen, Theunissen, & Doupe, 2001).

From Field L, forebrain processing proceeds along a number of parallel and interconnected pathways. Field L itself projects to the caudomedial nidopallium (NCM), which surrounds Field L anatomically, and to the lateral portion of the caudal mesopallium (CLM) (Vates et al., 1996). The medial portion of the caudal mesopallium (CMM) shares reciprocal projections with both CLM and NCM (Vates et al., 1996). Neurons throughout the songbird auditory telencephalon show complex patterns of tonotopic organization (Capsius & Leppelsack, 1996; Haüsler, 1996; Leppelsack & Schwartzkopff, 1972; Rübsamen & Dörrscheidt, 1986), and varying selectivity to species-specific vocalizations (Bonke, Bonke, & Scheich, 1979; Leppelsack & Vogt, 1976; Müller & Scheich, 1985). Neurons in Field L2a/b are less selective for species-specific vocalizations than those in L1 and L3 (Theunissen & Doupe, 1998; Theunissen, Sen, & Doupe, 2000; Theunissen et al., 2004), which in turn are less selective than those in NCM and CM (Grace, Amin, Singh, & Theunissen, 2003; Müller & Scheich, 1985). Increasing song selectivity across these regions is coincident with an increase in the non-linear components of the neural responses (Sen et al., 2001), and higher regions in the pathway (i.e. NCM and CMM) are involved in the representation of complex acoustic features (e.g. Bonke et al., 1979; Gentner, Hulse, & Ball, 2004; Gentner & Margoliash, 2003; Leppelsack & Vogt, 1976; Müller & Scheich, 1985).

Additional support for the role of NCM and CM in the processing of conspecific song comes from studies of stimulus driven expression of the immediate-early-gene (IEG) ZENK, a commonly used marker of song-evoked auditory activity (Mello, Velho, & Pinaud, 2004a) and song-induced experience-dependent plasticity (Jarvis, Mello, & Nottebohm, 1995; Jones et al., 2001; Mello, Nottebohm, & Clayton, 1995; Mello, Velho, & Pinaud, 2004b; Mello, Vicario, & Clayton, 1992; Ribeiro, Cecchi, Magnasco, & Mello, 1998). In starlings, the IEG response in NCM appears tied to stimulus novelty, whereas IEG activity in CMM appears to correlate with the ongoing recognition of familiar songs (Gentner et al., 2004). Together, these observations are consistent with a functional hierarchical organization of the songbird auditory system that is tuned throughout to conspecific song (Hsu, Woolley, Fremouw, & Theunissen, 2004; Woolley, Gill, Fremouw, & Theunissen, 2009; Woolley et al., 2005).

In addition to connections between regions within the auditory system, anatomical (Vates et al., 1996) and functional (Bauer et al., 2008; Shaevitz & Theunissen, 2007) connections have been observed between auditory processing areas and the song production system, providing a path for feedback during the acquisition and maintenance of a bird's own song. While the full range of efferent targets for the auditory telencephalon is unknown, it is likely that brain regions outside the classic "song system" are involved in the subsequent processing of song representations, including areas responsible for integrating multimodal information that naturally accompanies hearing song, as well as regions necessary for carrying out decision processes and motor programs necessary for executing song-mediated behaviors, e.g. copulation displays.

An increasing amount is known about the functionality and gross anatomical connectivity of the structures described above, but there is still little knowledge of the underlying functional micro-circuitry. Early EM and light microscopy studies of local circuitry within Field L indicate that neurons are arranged in small clusters (Saini & Leppelsack, 1977), and it is tempting to hypothesize that these clusters may form functional units. Perhaps related to this, large numbers of GABAergic cells are distributed throughout Field L, NCM, and CMM. In NCM and CMM many of these cells also show song-inducible expression of ZENK (Pinaud et al., 2004), suggesting that inhibitory mechanisms play an important functional role in the representation of learned acoustic material (see the following for further discussions of inhibitory processing in the avian auditory forebrain Capsius & Leppelsack, 1996; Pinaud & Terleph, 2008; Pinaud et al., 2008). Much work remains to be done to test these hypotheses. In general, elucidating the structure and activity of local circuitry will be crucial for understanding how song representation differs in distinct brain regions, the transformations that occur between these regions, and the role of these different representations in a bird's behavior. This will be an area of considerable research effort over the next decade.

The avian forebrain lacks the gross anatomical structure of mammalian cortex. Nonetheless, birds and mammals use their auditory systems to solve many similar behavioral problems. Structural and functional similarities present in the auditory regions of both classes of animals at the cellular (Saini & Leppelsack, 1981) and systems levels (Wild, Karten, & Frost, 1993), (Sockman, Gentner, & Ball, 2002), and across the whole of the forebrain (Jarvis et al., 2005) may reflect these common behavioral goals, Observations suggesting a link between avian studies and the rich body of research on the mammalian cortex highlight the need for comparative studies of auditory system function. Comparing anatomical and functional similarities between birds and mammals and across species within these classes will lend insight into which processes are likely to be fundamental to the representation of the auditory world by a neural substrate, and which may be idiosyncratic to a particular organism or group of organisms.

4.2. Experience-dependent representation of song

The acoustic environment of a songbird is not likely to be constant over time, or between individuals. Territories change hands and the available range of mates fluctuates. In order to effectively interpret the acoustic features in song, the auditory system needs to structure its representation in such a way that allows for dynamic behavioral goals of the organism. A combination of gene expression and electrophysiological studies has helped to characterize several regions within the auditory telencephalon where representations of learned song are thought to be stored.

Many neurons in NCM and CM show a rapid and selective upregulation of the ZENK (or its protein product) in response to the presentation of conspecific songs (Mello et al., 1992), in manner sensitive to the acoustics of particular song syllables (reviewed in Pinaud & Terleph, 2008; Ribeiro et al., 1998; Terleph, Mello, & Vicario, 2006). Interestingly, these IEG responses (Mello et al., 1995), as well as NCM multiunit recordings (Chew, Vicario, & Nottebohm, 1996), habituate to the repeated presentation of the same conspecific song. Similar habituation in the ZENK response in NCM also

appears closely tied to the context under which song is presented (Kruse, Stripling, & Clayton, 2004). In addition, IEG responses are elevated during specific components of vocal recognition (Gentner et al., 2004), female choice (Gentner, Hulse, Duffy, & Ball, 2001), and other song-based learned behaviors (Jarvis et al., 1995) including vocal acquisition (Phan, Pytte, & Vicario, 2006). While these habituation processes are likely to be important for understanding how the system processes conspecific song (e.g. Dong & Clayton, 2009), non-associative forms of learning are insufficient to capture the complexity of most song-based natural behaviors where reinforcement and stimulus associations are vital.

An organism's sensory experiences evoke long-lasting changes that bias cortical circuits towards representations of behaviorally relevant signals (Gilbert, Sigman, & Crist, 2001). Experience-dependent plasticity in neural responses underlies many forms of associative learning and has been observed across a variety of vertebrate sensory systems and brain regions (e.g. Bakin & Weinberger, 1990; Kay & Laurent, 1999; Kilgard, 2003). Experience dependant plasticity has also been observed in starlings that have been trained, using methods similar to those described above, to recognize sets of conspecific songs. In birds trained this way, single neurons in CMM respond much more robustly (i.e. emit many more action potentials) to songs the birds have learned to recognize than to unfamiliar conspecific songs. The strong response bias for familiar songs is consistent in animals trained with either a two-alternative choice procedure, where all of the training songs are paired with positive reinforcement, or go/no-go procedure, where only half the training songs are paired with positive reinforcement. In the latter case, however, songs associated with positive reinforcement elicited significantly stronger responses from CMM neurons than those associated with no reinforcement (Gentner & Margoliash, 2003). Thus, plasticity in CMM neuronal responses cannot be driven simply by song exposure. Other processes that mark the relevance of each song independent of any particular signal acoustics, such as attention, motivation and/or specific reinforcement, are likely involved.

CMM neurons do not respond at high rates to the entire ensemble of familiar songs. Instead, many individual neurons are selective for only a single song, or for only a small number of motifs within a single song (Gentner & Margoliash, 2003). Thus the same features that drive individual vocal recognition of starling song at the behavioral level (i.e. motifs) also elicit selective responses from single neurons. The mechanisms whereby this selectivity arises over the course of learning are not clear. It may be that the receptive fields of individual neurons are modified to match the spectrotemporal acoustics of behaviorally relevant signals. Rapid shifts in auditory receptive fields driven by behavioral goals have been observed in other systems (Fritz, Elhilali, & Shamma, 2007; Fritz, Shamma, Elhilali, & Klein, 2003). Alternatively, neurons that best match the acoustics on relevant songs may be somehow selected from an otherwise static population, and/or the responses neurons that don't match the relevant acoustics suppressed. These possibilities, along with the larger question of how to best model the receptive fields of these high-level auditory neurons, remain the topic of future research (e.g. Gill, Woolley, Fremouw, & Theunissen, 2008). Nonetheless, recent work consistent with the ideas above suggests that there may be a variety of forms of experience-dependent plasticity throughout the songbird auditory forebrain (Caporello & Gentner, 2008; Jeanne, Sharpee, & Gentner, 2008; Thompson & Gentner, 2008). Spike trains in Field L already carry sufficient information to discriminate between conspecific songs (Narayan, Graña, & Sen, 2006), suggesting that representation of the relevant acoustics may not be the primary role for processing within these regions. Rather, the system may be primarily concerned with the extraction of task-relevant features and their mapping onto appropriate responses.

The observations of experience-dependent plasticity in the songbird auditory system suggest a system that is simultaneously constrained in its immediate representational capacities by each animal's history, yet is tremendously adaptive in its ability to acquire a broad (currently undetermined) range of complex representations. It is tempting to speculate that similar experience-dependent, hierarchical processes may give rise to the phonetic, phonemic, syllabic and word-level representations that must be correlated with language experience in humans. At the same time, one must be careful to temper these kinds of speculations with an appropriate appreciation for the natural history of the organism. While the song recognition system expresses tremendous plasticity, it is reasonable to ask whether all of the links between auditory features and the behaviorally relevant phenomena they represent must be learned during an individual bird's lifetime. The existence of, for example, species-specific warning calls and the preference for conspecific song suggest that the meaning associated with some signals may not require any sort of learning. By extension, such processes may entail substantially different mechanisms at both neural and behavioral levels.

In many songbirds, the females' decision of who to mate with (termed female choice) is driven by the acoustics of competing males' songs. (Catchpole & Slater, 1995; Searcy, 1992). These behaviors provide an interesting context within which to explore the relationship between learned and innate signal processing mechanisms, as they are strongly influenced by both kinds of information. Untrained female starlings show strong preferences for the long songs of older males (Gentner & Hulse, 2000a), and these preferences are reflected in ZENK protein expression levels in NCM (Gentner et al., 2001; see also Maney, MacDougall-Shackleton, MacDougall-Shackleton, Ball, & Hahn, 2003). Yet, the magnitude of these preferences can be modified with relatively little experience (Sockman, Gentner, & Ball, 2005; Sockman et al., 2002). Lesions to CM, the same region involved in song recognition in starlings, lead female zebra finches to produce copulation solicitations to both conspecific and heterospecific song (Macdougall-Shackleton et al., 1998). Similar effects are reported in female canaries with lesions to the song nucleus HVC (Brenowitz, 1991), suggesting that brain regions involved in song production may also function in song perception. Future work should continue to target the neural mechanisms that support this natural decision behavior.

5. Conclusions

Although no two species are identical, this review describes a number of ways in which the auditory processing capacities of songbirds are similar to those of other vertebrates, and in particular to those of humans. Perhaps one reason why birdsong has captivated humans for so many centuries is that we "hear the world" in very similar ways. Indeed, the fact that birds and humans share a similar perceptual representation of the acoustic world holds the promise that higher-level auditory features relevant to more complex behaviors in birds may inhabit a similar acoustic space to those that are relevant for vocal communication in humans. In turn, there may be common neural and behavioral mechanisms adaptive for processing signals within this shared acoustic space. After human language, birdsong is arguably the most acoustically complex and diverse communication signal known. Although it is interesting in it own right, the parallels to human perception make it an excellent model for investigations for higher-level auditory processes, and enable a full range of neurobiological and behavioral techniques that are difficult to use in humans.

Although it is likely that the bird's brain can help us understand important strategies for solving auditory perception tasks that are mutual between humans and birds, birdsong is not language.

In particular, birdsong lacks the rich combinatorial structure and semantics of human vocal signals, and may be more analogous to speech than language (Doupe & Kuhl, 1999). By some theories, much of what makes human language unique is modality independent (e.g. Hauser & Bever, 2008), and so the greatest differences between humans and other animals may lie in the mapping between the outputs of speech processing brain areas and higher-order structures. The mechanisms for extracting, parsing, and representing patterned strings of vocal signals may be conserved across a wide range of species. Within this context, there remain a number of substantial challenges for the songbird auditory system that are germane to our general understanding of systems neuroscience. How are the complete receptive fields for high-level auditory (or other sensory modality) neurons best modeled? The best current models perform poorly beyond the primary thalamorecipient zones. What is the micro-circuitry of Field L, NCM, CM? The organization of the avian auditory telencephalon may be substantially different from that of mammalian cortex, yet very similar function is achieved. How? What are the sources – behavioral, anatomical and physiological – for top-down modulation through attention, reinforcement, memory, etc. within the system? Finally, what are the roles of judgment and decision in the auditory telencephalon and its efferent targets? Further studies into the behavioral and neural basis of the birdsong auditory system promise to be instrumental in answering these questions.

References

Adret-Hausberger, M., & Jenkins, P. F. (1988). Complex organization of the warbling song in starlings. *Behaviour, 107*, 138–156.

Appeltants, D., Gentner, T. Q., Hulse, S. H., Balthazart, J., & Ball, G. F. (2005). The effect of auditory distractors on song discrimination in male canaries (*Serinus canaria*). *Behavioural Processes, 69*(3), 331–341.

Bakin, J. S., & Weinberger, N. M. (1990). Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Research, 536*(1–2), 271–286.

Bauer, E. E., Coleman, M. J., Roberts, T. F., Roy, A., Prather, J. F., & Mooney, R. (2008). A synaptic basis for auditory-vocal integration in the songbird. *Journal of Neuroscience, 28*(6), 1509–1522.

Bee, M., & Klump, G. M. (2004). Primitive auditory stream segregation: A neurophysiological study in the songbird forebrain. *Journal of Neurophysiology, 92*(2), 1088–1104.

Beecher, M., Campbell, S. E., & Burt, J. (1994). Song perception in the song sparrow: Birds classify by song type but not by singer. *Animal Behaviour, 47*, 1343–1351.

Bigalke-Kunz, B., Rübsamen, R., & Dörrscheidt, G. J. (1987). Tonotopic organization and functional characterization of the auditory thalamus in a songbird, the European starling. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 161*(2), 255–265.

Bonke, B. A., Bonke, D., & Scheich, H. (1979). Connectivity of the auditory forebrain nuclei in the guinea fowl (*Numida meleagris*). *Cell and Tissue Research, 200*(1), 101–121.

Bregman, A. S. (1990). *The auditory scene auditory science analysis: The perceptual organization of sound*. Cambridge: MIT Press.

Brenowitz, E. A. (1991). Altered perception of species-specific song by female birds after lesions of a forebrain nucleus. *Science, 251*(4991), 303–305.

Bricker, P. D., & Pruzansky, S. (1976). Speaker recognition. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295–326). New York: Academic Press.

Brooks, R. J., & Falls, J. B. (1975a). Individual recognition by song in white-throated sparrows. I. Discrimination of songs of neighbors and strangers. *Canadian Journal of Zoology, 53*, 879–888.

Brooks, R. J., & Falls, J. B. (1975b). Individual recognition by song in white-throated sparrows. III. Song features used in individual recognition. *Canadian Journal of Zoology, 53*, 1749–1761.

Brown, C. H., & Maloney, C. G. (1986). Temporal integration in two species of old world monkeys: Blue monkeys (*Cercopithecus mitis*) and grey-cheeked mangabeys (*Cercocebus albigena*). *The Journal of the Acoustical Society of America, 79*(4), 1058–1064.

Caporello, E., Gentner, T. Q. (2008). *Song recognition learning drives experience-dependent changes in the avian telencephalic auditory area field L*. Paper presented at the society for neuroscience.

Capsius, B., & Leppelsack, H. J. (1996). Influence of urethane anesthesia on neural processing in the auditory cortex analogue of a songbird. *Hearing Research, 96*(1–2), 59–70.

Capsius, B., & Leppelsack, H. (1999). Response patterns and their relationship to frequency analysis in auditory forebrain centers of a songbird. *Hearing Research, 136*(1–2), 91–99.

Catchpole, C. K., & Slater, P. B. J. (1995). *Bird song*. Cambridge: Cambridge University Press.

Chew, S. J., Vicario, D. S., & Nottebohm, F. (1996). A large-capacity memory system that recognizes the calls and songs of conspecifics. *Proceedings of the National Academy of Science, 93*, 1950–1955.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Clucas, B. A., Freeberg, T. M., & Lucas, J. R. (2004). Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*). *Behavioral Ecology and Sociobiology, 57*, 187–196.

Cynx, J., & Nottebohm, F. (1992). Testosterone facilitates some conspecific song discriminations in castrated zebra finches (*Taeniopygia guttata*). *Proceedings of the National Academy of Sciences of the United States of America, 89*(4), 1376–1378.

Demany, L., & Semal, C. (1989). Detection thresholds for sinusoidal frequency modulation. *Journal of the Acoustical Society of America, 85*(3), 1295–1301.

Dong, S., & Clayton, D. F. (2009). Habituation in songbirds. *Neurobiology of Learning and Memory, 92*(2), 183–188.

Dooling, R. J. (1980). Behavior and psychophysics of hearing in birds. In A. N. Popper & R. R. Fay (Eds.), *Comparative studies of hearing in vertebrates*. New York: Springer-Verlag.

Dooling, R. J. (1982). *Auditory perception in birds. Acoustic communication in birds* (Vol. 1). Academic Press.

Dooling, R. J. (1992). Hearing in birds. In D. Webster, R. Fay, & A. N. Popper (Eds.), *The Evolutionary Biology of Hearing* (pp. 545–559). New York: Springer-Verlag.

Dooling, R. J. (2000). Hearing in birds and reptiles. In R. J. Dooling, R. R. Fay, & A. N. Popper (Eds.), *Comparative hearing: Birds and reptiles* (pp. 308–359). New York: Springer-Verlag.

Dooling, R. J., Brown, S. D., Klump, G. M., & Okanoya, K. (1992). Auditory perception of conspecific and heterospecific vocalizations in birds: Evidence for special processes. *Journal of Comparative Psychology, 106*(1), 20–28.

Dooling, R. J., Okanoya, K., Downing, J., & Hulse, S. H. (1986). Hearing in the starling (*Sturnus vulgaris*): Absolute thresholds and critical ratios. *Bulletin of the Psychonomic Society, 24*(6), 462–464.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*, 567–631.

Eens, M., Pinxten, M., & Verheyen, R. F. (1989). Temporal and sequential organization of song bouts in the European starling. *Ardea, 77*, 75–86.

Eens, M., Pinxten, M., & Verheyen, R. F. (1991). Male song as a cue for mate choice in the European starling *Sturnus vulgaris*. *Behaviour, 116*(3–4), 210–238.

Fastl, H. (1978). Frequency discrimination for pulsed versus modulated tones. *Journal of the Acoustical Society of America, 63*(1), 275–277.

Fay, R. R. (1988). *Hearing in vertebrates: A psychophysics databook*. Winnetka: Hill – Fay Associates.

Fortune, E. S., & Margoliash, D. (1992). Cytoarchitectonic organization and morphology of cells of the field L complex in male zebra finches (*Taenopygia guttata*). *Journal of Comparative Neurology, 325*(3), 388–404.

Fritz, J. B., Elhilali, M., & Shamma, S. A. (2007). Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of Neurophysiology, 98*(4), 2337–2346.

Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience, 6*(11), 1216–1223.

Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. *Journal of the Acoustical Society of America, 124*(2), 1350–1359.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*(7088), 1204–1207.

Gentner, T. Q., & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour, 56*(3), 579–594.

Gentner, T. Q., & Hulse, S. H. (2000a). Female European starling preference and choice for variation in conspecific male song. *Animal Behaviour, 59*(2), 443–458.

Gentner, T. Q., & Hulse, S. H. (2000b). Perceptual classification based on the component structure of song in European starlings. *Journal of the Acoustical Society of America, 107*(6), 3369–3381.

Gentner, T. Q., Hulse, S. H., & Ball, G. F. (2004). Functional differences in forebrain auditory regions during learned vocal recognition in songbirds. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 190*(12), 1001–1010.

Gentner, T. Q., Hulse, S. H., Duffy, D., & Ball, G. F. (2001). Response biases in auditory forebrain regions of female songbirds following exposure to sexually relevant variation in male song. *Journal of Neurobiology, 46*(1), 48–58.

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature, 424*(6949), 669–674.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron, 31*(5), 681–697.

Gill, P., Woolley, S. M. N., Fremouw, T. E., & Theunissen, F. E. (2008). What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. *Journal of Neurophysiology, 99*(6), 2809–2820.

Grace, J. A., Amin, N., Singh, N. C., & Theunissen, F. E. (2003). Selectivity for conspecific song in the zebra finch auditory forebrain. *Journal of Neurophysiology, 89*(1), 472–487.

Hailman, J. P., & Ficken, M. S. (1986). Combinational animal communication with computable syntax: Chick-a-dee calling qualifies as 'language' by structural linguistics. *Animal Behaviour, 34*, 1899–1901.

Hausberger, M. (1997). Social influences on song acquisition and sharing in the European starling (*Sturnus vulgaris*). In C. Snowden & M. Hausberger (Eds.),

Social influences on vocal development (pp. 128–156). Cambridge: Cambridge University Press.
Hausberger, M., & Cousillas, H. (1995). Categorization in birdsong: From behavioural to neuronal responses. *Behavioural Processes, 35*(1–3), 83–91.
Hauser, M. D., & Bever, T. (2008). Behavior: A biolinguistic agenda. *Science, 322*(5904), 1057–1059.
Haüsler, U. (Ed.). (1996). *Measurement of short-time spatial activity patterns during auditory stimulation in the starling*. New York, NY: Plenum Press.
Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569–1579.
Holland, J., Dabelsteen, T., & Paris, A. L. (2000). Coding in the song of the wren: Importance of rhythmicity, syntax and element structure. *Animal Behaviour, 60*(4), 463–470.
Hopcroft, J., & Ullman, J. (1979). *Introduction to automata theory, languages, and computation*. Reading, MA: Addison-Wesley.
Hsu, A., Woolley, S. M., Fremouw, T. E., & Theunissen, F. E. (2004). Modulation power and phase spectrum of natural sounds enhance neural encoding performed by single auditory neurons. *Journal of Neuroscience, 24*(41), 9201–9211.
Hulse, S. H., MacDougall-Shackleton, S. A., & Wisniewski, A. B. (1997). Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology, 111*(1), 3–13.
Jarvis, E. D., Gunturkun, O., Bruce, L., Csillag, A., Karten, H., Kuenzel, W., et al. (2005). Avian brains and a new understanding of vertebrate brain evolution. *Nature Reviews Neuroscience, 6*(2), 151–159.
Jarvis, E. D., Mello, C. V., & Nottebohm, F. (1995). Associative learning and stimulus novelty influence the song-induced expression of an immediate early gene in the canary forebrain. *Learning and Memory, 2*, 62–80.
Jeanne, J. M., Sharpee, T. O., & Gentner, T. Q. (2008). *Song recognition learning drives experience-dependent representations in the auditory forebrain region CLM*. Paper presented at the society for neuroscience.
Jones, M. W., Errington, M. L., French, P. J., Fine, A., Bliss, T. V., Garel, S., et al. (2001). A requirement for the immediate early gene Zif268 in the expression of late LTP and long-term memories. *Nature Neuroscience, 4*(3), 289–296.
Karten, H. J. (1968). The ascending auditory pathway in the pigeon (*Columba liva*). II. Telencephalic projections of the nucleus ovoidalis thalami. *Brain Research, 11*, 134–153.
Karten, H. J. (1969). The organization of the avian telencephalon and some speculations on the phylogeny of the amniote telencephalon. In J. M. Petras & C. R. Noback (Eds.), *Comparative and evolutionary aspects of the vertebrate central nervous system* (Vol. 167, pp. 164–179). New York: The New York Academy of Science.
Kay, L. M., & Laurent, G. (1999). Odor- and context-dependent modulation of mitral cell activity in behaving rats. *Nature Neuroscience, 2*(11), 1003–1009.
Kilgard, M. (2003). Cholinergic modulation of skill learning and plasticity. *Neuron, 38*(5), 678–680.
Klump, G. M., & Langemann, U. (1995). Comodulation masking release in a songbird. *Hearing Research, 87*(1–2), 157–164.
Klump, G. M., & Maier, E. H. (1989). Gap detection in the starling (*Sturnus vulgaris*). *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 164*(4), 531–538.
Kroodsma, D. E., & Miller, E. H. (Eds.). (1996). *Ecology and evolution of acoustic communication in birds*. Ithaca: Comstock/Cornell.
Kruse, A. A., Stripling, R., & Clayton, D. F. (2004). Context-specific habituation of the ZENK gene response to song in adult zebra finches. *Neurobiology of Learning and Memory, 82*(2), 99–108.
Langemann, U., & Klump, G. M. (1992). Frequency discrimination in the European starling (*Sturnus vulgaris*): A comparison of different measures. *Hearing Research, 63*(1–2), 43–51.
Langemann, U., Klump, G. M., & Dooling, R. J. (1995). Critical bands and critical-ratio bandwidth in the European starling. *Hearing Research, 84*(1–2), 167–176.
Leppelsack, H. J., & Vogt, M. (1976). Responses of auditory neurons in the forebrain of a songbird to stimulation with species-specific sounds. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 107*(3), 263–274.
Leppelsack, H.-J., & Schwartzkopff, J. (1972). Properties of acoustic neurons in the caudal neostriatum of birds. *Journal of Comparative Physiology, 80*, 137–140.
MacDougall-Shackleton, S. A., Hulse, S. H., Gentner, T. Q., & White, W. (1998). Auditory scene analysis by European starlings (*Sturnus vulgaris*): Perceptual segregation of tone sequences. *Journal of the Acoustical Society of America, 103*(6), 3581–3587.
Maier, E. H., & Klump, G. M. (1990). Auditory duration discrimination in the European starling (*Sturnus vulgaris*). *Journal of the Acoustical Society of America, 88*(2), 616–621.
Maney, D. L., MacDougall-Shackleton, E. A., MacDougall-Shackleton, S. A., Ball, G. F., & Hahn, T. P. (2003). Immediate early gene response to hearing song correlates with receptive behavior and depends on dialect in a female songbird. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 189*(9), 667–674.
Mello, C., Nottebohm, F., & Clayton, D. (1995). Repeated exposure to one song leads to a rapid and persistent decline in an immediate early gene's response to that song in zebra finch telencephalon. *Journal of Neuroscience, 15*, 6919–6925.
Mello, C. V., Velho, T. A., & Pinaud, R. (2004a). Song-induced gene expression: A window on song auditory processing and perception. *Annals of the New York Academy of Sciences, 1016*, 263–281.
Mello, C. V., Velho, T. A. F., & Pinaud, R. (2004b). Song-induced gene expression: A window on song auditory processing and perception. *ANYAS, 1016*, 263–281.
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), 6818–6822.
Müller, S. C., & Scheich, H. (1985). Functional organization of the avian auditory field L. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 156*(1), 1–12.
Narayan, R., Best, V., Ozmeral, E., McClaine, E., Dent, M., Shinn-Cunningham, B., et al. (2007). Cortical interference effects in the cocktail party problem. *Nature Neuroscience, 10*(12), 1601–1607.
Narayan, R., Graña, G., & Sen, K. (2006). Distinct time scales in cortical discrimination of natural sounds in songbirds. *Journal of Neurophysiology, 96*(1), 252–258.
Nelson, D. A. (1988). Feature weighting in species song recognition by the field sparrow (*Spizella pusilla*). *Behaviour, 106*(1–2), 158–181.
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, 103*(2), 171–176.
Nelson, D. A., & Marler, P. (1990). The perception of birdsong and an ecological concept of signal space. In W. C. Stebbins & M. A. Berkely (Eds.), *Comparative perception: Complex signals* (Vol. 2, pp. 443–477). New York: John Wiley and Sons.
Okanoya, K., & Dooling, R. J. (1987). Hearing in passerine and psittacine birds: A comparative study of absolute and masked auditory thresholds. *Journal of Comparative Psychology, 101*(1), 7–15.
Phan, M. L., Pytte, C. L., & Vicario, D. S. (2006). Early auditory experience generates long-lasting memories that may subserve vocal learning in songbirds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(4), 1088–1093.
Pinaud, R., & Terleph, T. A. (2008). A songbird forebrain area potentially involved in auditory discrimination and memory formation. *Journal of Biosciences, 33*(1), 145–155.
Pinaud, R., Terleph, T. A., Tremere, L. A., Phan, M. L., Dagostin, A. A., Leão, R. M., et al. (2008). Inhibitory network interactions shape the auditory processing of natural communication signals in the songbird auditory forebrain. *Journal of Neurophysiology, 100*(1), 441–455.
Pinaud, R., Velho, T. A. F., Jeong, J. K., Tremere, L. A., Leão, R. M., von Gersdorff, H., et al. (2004). GABAergic neurons participate in the brain's response to birdsong auditory stimulation. *European Journal of Neuroscience, 20*(5), 1318–1330.
Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology – Human Perception and Performance, 23*(3), 651–666.
Ribeiro, S., Cecchi, G. A., Magnasco, M. O., & Mello, C. V. (1998). Toward a song code: Evidence for a syllabic representation in the canary brain. *Neuron, 21*(2), 359–371.
Rübsamen, R., & Dörrscheidt, G. J. (1986). Tonotopic organization of the auditory forebrain in a songbird, the European starling. *Journal of Comparative Physiology, 158*, 639–646.
Sachs, M. B., & Sinnott, J. M. (1978). Responses to tones of single cells in nucleus magnocellularis and nucleus angularis of the redwing blackbird (*Agelaius phoeniceus*). *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 126*(4), 347–361.
Saini, K., & Leppelsack, H. (1977). Neuronal arrangement in the auditory field L of the neostriatum of the starling. *Cell and Tissue Research, 176*(3), 309–316.
Saini, K. D., & Leppelsack, H. J. (1981). Cell types of the auditory caudomedial neostriatum of the starling. *Journal of Comparative Neurology, 198*, 209–230.
Searcy, W. A. (1992). Song repertoire and mate choice in birds. *American Zoologist, 32*, 71–80.
Seeba, F., & Klump, G. M. (2009). Stimulus familiarity affects perceptual restoration in the European starling (*Sturnus vulgaris*). *PLoS One, 4*(6), e5974.
Sen, K., Theunissen, F. E., & Doupe, A. J. (2001). Feature analysis of natural sounds in the songbird auditory forebrain. *Journal of Neurophysiology, 86*(3), 1445–1458.
Shaevitz, S. S., & Theunissen, F. E. (2007). Functional connectivity between auditory areas field L and CLM and song system nucleus HVC in anesthetized zebra finches. *Journal of Neurophysiology, 98*(5), 2747–2764.
Sinnott, J. M., Sachs, M. B., & Hienz, R. D. (1980). Aspects of frequency discrimination in passerine birds and pigeons. *Journal of Comparative and Physiological Psychology, 94*(3), 401–415.
Sockman, K. W., Gentner, T. Q., & Ball, G. F. (2002). Recent experience modulates forebrain gene-expression in response to mate-choice cues in European starlings. *Proceedings of the Royal Society of London Series B – Biological Sciences, 269*(1508), 2479–2485.
Sockman, K. W., Gentner, T. Q., & Ball, G. F. (2005). Complementary neural systems for the experience-dependent integration of mate-choice cues in European starlings. *Journal of Neurobiology, 62*(1), 72–81.
Stoddard, P. K. (1996). Vocal recognition of neighbors by territorial passerines. In D. E. Kroodsma & E. H. Miller (Eds.), *Ecology and evolution of acoustic communication in birds* (pp. 356–374). Ithaca: Comstock/Cornell.
Stoddard, P. K., Beecher, M. D., Horning, C. L., & Campbell, S. E. (1991). Recognition of individual neighbors by song in the song sparrow, a species with song repertoires. *Behavioral Ecology and Sociobiology, 29*, 211–215.
Sturdy, C. B., Phillmore, L. S., & Weisman, R. G. (2000). Call-note discriminations in black-capped chickadees (*Poecile atricapillus*). *Journal of Comparative Psychology, 114*(4), 357–364.
Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. *JASA, 119*(3), 1849–1866.

Terleph, T. A., Mello, C. V., & Vicario, D. S. (2006). Auditory topography and temporal response dynamics of canary caudal telencephalon. *Journal of Neurobiology, 66*(3), 281–292.

Theunissen, F. E., Amin, N., Shaevitz, S. S., Woolley, S. M., Fremouw, T., & Hauber, M. E. (2004). Song selectivity in the song system and in the auditory forebrain. *Annals of the New York Academy of Sciences, 1016*, 222–245.

Theunissen, F. E., & Doupe, A. J. (1998). Temporal and spectral sensitivity of complex auditory neurons in the nucleus HVc of male zebra finches. *Journal of Neuroscience, 18*(10), 3786–3802.

Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral–temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *Journal of Neuroscience, 20*(6), 2315–2331.

Thompson, J. V., & Gentner, T. Q. (2008). *Song recognition learning elicits differential suppression of neural responses to learned songs in NCM*. Paper presented at the society for neuroscience.

Vates, G. E., Broome, B. M., Mello, C. V., & Nottebohm, F. (1996). Auditory pathways of caudal telencephalon and their relation to the song system of adult male zebra finches. *Journal of Comparative Neurology, 366*(4), 613–642.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392–393.

Wild, J. M., Karten, H. J., & Frost, B. J. (1993). Connections of the auditory forebrain in the pigeon (*Columba livia*). *Journal of Comparative Neurology, 337*(1), 32–62.

Woolley, S. M. N., & Casseday, J. H. (2004). Response properties of single neurons in the zebra finch auditory midbrain: Response patterns, frequency coding, intensity coding, and spike latencies. *Journal of Neurophysiology, 91*(1), 136–151.

Woolley, S. M. N., Fremouw, T. E., Hsu, A., & Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nature Neuroscience, 8*(10), 1371–1379.

Woolley, S. M., Gill, P. R., Fremouw, T., & Theunissen, F. E. (2009). Functional groups in the avian auditory system. *Journal of Neuroscience, 29*(9), 2780–2793.

Chapter 1, in full, is a reprint of the material as it appears in Brain and Language, 2010. Knudsen, Daniel P.; Gentner, Timothy Q., Elsevier, 2010. The dissertation author was the primary investigator and author of this paper.

CHAPTER 2

DISTRIBUTED RECOGNITION OF NATURAL SONGS BY EUROPEAN STARLINGS

KNUDSEN, D. P., THOMPSON, J. V., AND GENTNER, T. Q.

AS IT APPEARS IN

LEARNING AND MOTIVATION

2010

VOLUME 41 ISSUE 4, PAGES 287-306

DIGITAL OBJECT IDENTIFIER (DOI): 10.1016/J.LMOT.2010.08.005

Distributed recognition of natural songs by European starlings

Daniel Knudsen[a], Jason V. Thompson[a], Timothy Q. Gentner[a,b,*]

[a] *Neuroscience Graduate Program, University of California, San Diego, CA, United States*
[b] *Department of Psychology, University of California, San Diego, CA, United States*

ARTICLE INFO

Keywords:
Auditory
Auditory perception
Vocal communication
Songbird
Psychoacoustics
Neuroethology

ABSTRACT

Individual vocal recognition behaviors in songbirds provide an excellent framework for the investigation of comparative psychological and neurobiological mechanisms that support the perception and cognition of complex acoustic communication signals. To this end, the complex songs of European starlings have been studied extensively. Yet, several basic parameters of starling individual vocal recognition have not been assessed. Here we investigate the temporal extent of song information acquired by starlings during vocal recognition learning. We trained two groups of starlings using standard operant conditioning techniques to recognize several songs from two conspecific male singers. In the first experiment we tested their ability to maintain accurate recognition when presented with (1) random sequences of 1–12 motifs (stereotyped song components) drawn from the training songs, and (2) 0.1–12-s excerpts of continuous song drawn from the training songs. We found that song recognition improved monotonically as more vocal material is provided. In the second experiment, we systematically substituted continuous, varying length regions of white noise for portions of the training songs and again examined recognition accuracy. Recognition remained above chance levels for all noise substitutions tested (up to 91% of the training stimulus) although all but the smallest substitutions led to some decrement in song recognition. Overall, above chance recognition could be obtained with surprisingly few motifs, short excerpts of song, and in the absence

* Corresponding author at: University of California, San Diego, Department of Psychology, 9500 Gilman Dr. MC0109, La Jolla, CA 92093, United states.
E-mail address: tgentner@ucsd.edu (T.Q. Gentner).

0023-9690/$ – see front matter
doi:10.1016/j.lmot.2010.08.005

of large portions of the training songs. These results suggest that starlings acquire a representation of song during individual vocal recognition learning that is robust to perturbations and distributed broadly over large portions of these complex acoustic sequences.

Vocal recognition is a widespread phenomenon among songbirds (reviewed by Stoddard, 1996) and serves two broad goals: species recognition, and recognition of individual conspecifics. Myriad acoustic features may guide vocal recognition, but it follows that selection pressures will emphasize the features that most reliably distinguish one's own species' songs from those of other sympatric species (see Nelson & Marler, 1990), and those that differentiate individual conspecifics. Among field sparrows, for example, frequency appears to be the most important cue for species recognition (Nelson, 1989). Likewise, laboratory studies demonstrate that songbirds are more adept at discriminations involving conspecific songs as compared to discriminations involving heterospecific songs (Cynx & Nottebohm, 1992; Dooling, Brown, Klump, & Okanoya, 1992; Heinz, Sinnott, & Sachs, 1980). Recognition systems appear to be tuned to the particular acoustic space of one's own species' vocalizations (Nelson, 1988) but also maintain more general capacities to distinguish among many different types of song. Within a species, particularly those with acoustically diverse songs, many acoustic features may be relevant for classification of song components (Sturdy, Phillmore, & Weisman, 2000). Recent laboratory studies of European starlings have addressed these questions directly by determining more precisely the form of the acoustic signal controlling individual recognition.

Male starlings present their songs in long episodes of continuous singing referred to as bouts. Song bouts, in turn, are composed of much smaller acoustic units referred to as motifs (Adret-Hausberger & Jenkins, 1988; Eens, Pinxten, & Verheyen, 1991), which in turn are composed of still smaller units called notes. Notes can be broadly classified by the presence of continuous energy in their spectrographic representations, and although several notes may occur in a given motif, their pattern is usually highly stereotyped between successive renditions of the same motif (a sample of song is shown in Fig. 1). One can thus consider starling song as a sequence of motifs, where each motif is an acoustically complex event. The number of unique motifs that a male starling can sing (i.e. his repertoire size) can be large, and different song bouts from the same male do not necessarily comprise the same set of motifs. The broad acoustic variation in their song provides several potential cues that starlings might use when learning to recognize the songs of an individual conspecific and while maintaining that recognition over time. One straightforward recognition mechanism is the association of specific motifs with specific singers.

Classifying songs according to their component (motif) structure represents an efficient strategy for representing these complex acoustic signals (see Fig. 1). Because individual starlings tend to possess unique motif repertoires (Adret-Hausberger & Jenkins, 1988; Eens, Pinxten, & Verheyen, 1989; Eens et al., 1991), disjoint sets of motifs will generally correspond to individual identity. Attending to the motif structure, therefore, captures a significant portion of the individual variation in the signal, albeit at the expense of requiring a large memory capacity. From a human perspective, this might seem a suboptimal strategy for individual vocal recognition, because the total set of vocal signals that can be produced by a single individual is theoretically infinite. Individual talker recognition in humans relies on voice characteristics such as timbre, glottal pulsation frequency, and spectral contours imparted by laryngeal morphology (Bricker & Pruzansky, 1976) that are largely independent of the linguistic content (Remez, Fellowes, & Rubin, 1997). The incorporation of individual identity information into the phonetics of the vocal signal, as appears to be the case for starlings, may be a plesiomorphic (evolutionarily older) state for human vocal communication. As the size of the lexicon and sharing of signal elements increases between individuals, however, the exclusive use of this 'object-dependent' strategy for individual vocal recognition becomes increasingly inefficient. It is likely that starlings, along with other organisms that rely on acoustically complex vocal signals to mediate adaptive behaviors, posses multiple mechanisms for extracting behaviorally relevant information from communication signals.

Although recognition of the complex songs of European starlings has been studied extensively, several basic parameters of starling individual vocal recognition have not been assessed. Recent results

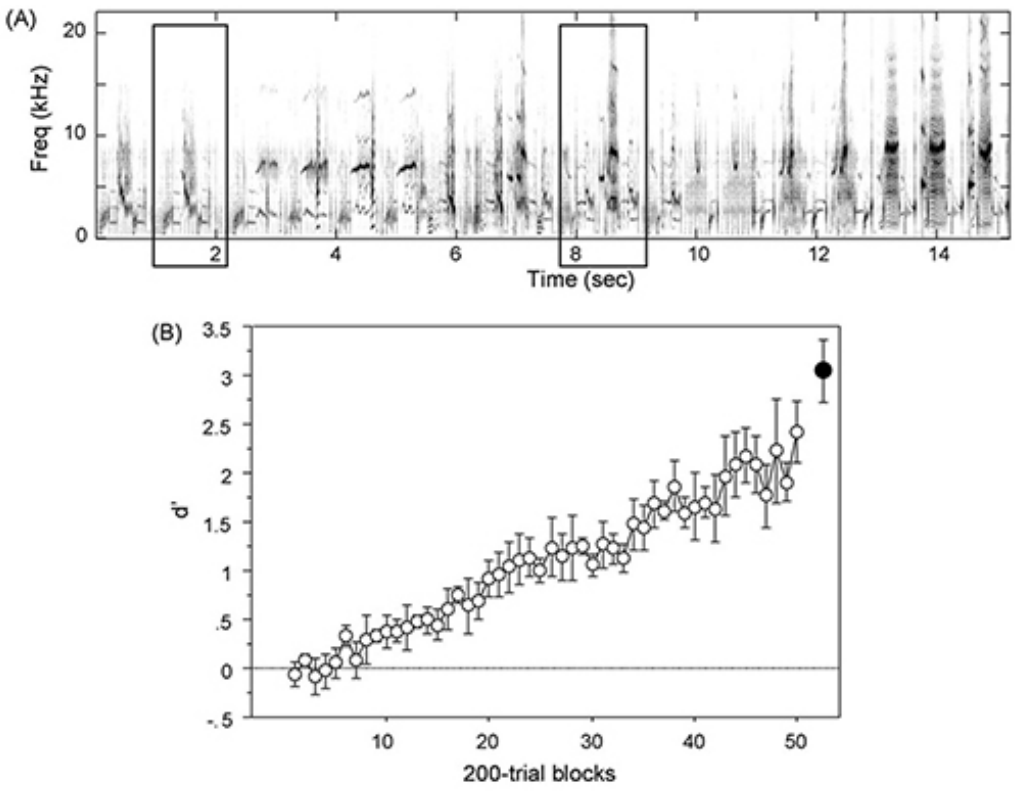


Fig. 1. Baseline stimuli and acquisition for Experiment 1. (A) Example spectrogram of one baseline training song. Note the diverse acoustic structure of the song. Example motifs are outlined by the boxes just before 2 and 8 s. (B) Mean (±sem) d' (open circles) for acquisition of the baseline song recognition task plotted over the first 50 200-trial blocks. The single point (open square) on the right shows the recognition accuracy averaged over the last five 200-trial blocks prior to probe testing.

indicate that some voice characteristics are present in starling song, and that perceptual knowledge spans multiple hierarchical levels of song organization (Gentner, 2008). Here, we investigate the temporal extent of song-knowledge that starlings acquire during canonical vocal recognition tasks.

General methods

Subjects

Ten adult European starlings, *Sturnus vulgaris*, served as subjects in this study. All subjects were wild caught in southern California in May 2006. All had full adult plumage at the time of capture, and thus were at least one-year old. From the time of capture until their use in this study, all subjects were housed in large mixed-sex, conspecific aviaries with *ad libitum* access to food and water. The photoperiod in the aviary and the testing chambers followed the seasonal variation in local sunrise and sunset times. No significant sex differences have been observed in previous studies of individual vocal recognition (Gentner & Hulse, 2000), and the sex of subjects in this study was not controlled.

Apparatus

Starlings learned to classify the training stimuli using a custom-built operant apparatus, housed in a 61 cm × 81 cm × 56 cm ID sound attenuation chamber (Acoustic Systems). Inside the chamber, a subject was held in a weld-wire cage (41 cm × 41 cm × 35 cm) that permitted access to a 30 cm × 30 cm operant panel mounted on one wall. The operant panel contained three circular response ports spaced

6 cm center-to-center, aligned in a row with the center of each port roughly 14 cm off the floor of the cage and with the whole row centered on the width of the panel. Each response port was a PVC housed opening in the panel fitted with an IR receiver and transmitter that detected when the bird broke the plane of the response port with its beak. This 'poke-hole' design allows starlings to probe the apparatus with their beak, in a manner akin to their natural appetitive foraging behavior. Independently controlled light-emitting diodes (LEDs) could illuminate each response port from the rear. Directly below the center port, in the section of cage floor immediately adjacent to the panel, a fourth PVC lined opening provided access to food. A remotely controlled hopper, positioned behind the panel, moved the food into and out of the subject's reach beneath the opening. Acoustic stimuli were delivered through a small full-range audio speaker mounted roughly 30 cm behind the panel and out of the subject's view. The SPL inside all chambers was calibrated to the same standard broadband signal. Custom software monitored the subject's responses and controlled the LEDs, food hoppers, chamber light and auditory stimulus presentation according to procedural contingencies.

Stimuli

Song recording

Recordings of eight male European starlings were used to generate all the stimuli for this experiment. The procedures for obtaining high-quality song recordings from male starlings have been detailed elsewhere (Gentner & Hulse, 1998). Briefly, a minimum of 0.5 h of song was recorded from each male while housed individually in a large sound-attenuating chamber. During recording, males had visual and auditory access to a female starling (the same female was used to induce song from all the males). The multiple songs of each bird were parsed into roughly 10 to 15-s exemplars of continuous singing taken from the beginning, middle, or end of a typically much longer song bout. Human observers labeled the motifs in each exemplar. These same stimuli have been used to explore the role of motif familiarity in the recognition of individual songs in several studies (Gentner & Hulse, 1998, 2000; Gentner, Hulse, Bentley, & Ball, 2000; Gentner & Margoliash, 2003). None of the males whose songs were used to generate the stimuli for the present study served as a subject in the operant testing described here.

Procedure

Shaping

Subjects learned to work the apparatus through a series of successive shaping procedures. Upon initially entering the operant chamber, the subject was allowed unrestricted access to the food hopper on a variable interval schedule, and then taught through auto-shaping to peck at a flashing LED in the center port to gain access to the food hopper. Once the subject pecked reliably at the center port to obtain food, pecks to the center port caused the center LED to cease flashing and the LED in either the left or right response port to begin flashing (side selected at random with equal probability). The subject was then required to peck at the port with the flashing LED to raise the food hopper. Once behavior to obtain food was again reliable (typically 100 trials), the center LED ceased flashing, while the requirement to peck at the center port to trigger the left or right LED to flash remained in effect. Shortly thereafter, pecks to the center port initiated the presentation of a song stimulus, and the trial proceeded as described below. In all cases, initial shaping occurred within 1–2 days, and was followed immediately by the start of song recognition training.

Song recognition training

Subjects learned to classify the two sets of song exemplars using a two-alternative choice (2AC) procedure. Subjects initiated a trial by pecking the center response button to trigger immediately the presentation of a single song exemplar. The specific baseline training stimuli are described in the methods for each experiment. Following completion of the song exemplar, the subject was required to peck either the left or the right response port within 2 s. Half of the song exemplars were associated with the left response port and the other half with the right port. The training songs were divided according to individual identity, such that a single individual sang the songs in each set. Thus, the

baseline training corresponds to basic individual vocal recognition (Gentner & Hulse, 1998, 2000). Correct responses (i.e. pecking the left/right port after hearing a song associated with that port) were rewarded with access to the food hopper for 3 s. Incorrect responses (i.e. pecking the left/right port after hearing a song associated with the opposite port) were punished by extinguishing the house light for 2–10 s and withholding food. Responses prior to completion of the stimulus were ignored. The trial ended when either the food hopper retracted following a correct response, or the house light re-illuminated following an incorrect response. The inter-trial interval was 2 s. Incorrect responses, and any trial in which the subject failed to peck either the left or the right key within 2 s of stimulus completion initiated a correction-trial sequence, during which the initiating stimulus was repeated on all subsequent trials until the animal responded correctly. Correction-trial data are not included in the analysis. The song exemplar presented on any given trial was selected randomly (with replacement) from the pool of all stimuli the animal was learning to recognize. Water was always available. Subjects were on a closed economy during training, with daily sessions lasting from sunrise until sunset. Food intake was monitored daily to insure wellbeing.

Test procedure

During later sessions, we presented test stimuli to gauge the basis for recognition abilities that each subject acquired during baseline training. Prior to initiation of the first test session, the rate of food reinforcement for correct responses and "punishment" (dimmed house lights) for incorrect responses was reduced from 100% where it had been during baseline training. After performance stabilized at reduced reinforcement rates, typically within one or two sessions, we began presenting test stimuli on a subset of the trials. The specific test stimuli, their presentation rates, and the reinforcement rates for probe sessions are given in the methods for each experiment. The test stimulus for a given trial was selected randomly from the set of all possible test stimuli for that subject. We reinforced responses to test stimuli non-differentially regardless of accuracy, such that each response to a test stimulus had an equal, non-zero chance of eliciting a food reward, the same chance of eliciting punishment (timeout without food), and some comparable (though not necessarily equal) chance of eliciting no operant consequence. Because reinforcement of the test stimuli was random and non-differential with respect to response outcome, subjects had no opportunity to learn to associate a given test stimulus with a left or right response. Thus, the correct classification of test stimuli can be taken as strong evidence for vocal recognition rather than rote learning of specific training exemplars (Hulse, 1995). If there is no recognition, classification accuracy will be at chance.

Analysis

We used *d*-prime (d') to estimate the sensitivity for classification of baseline training song stimuli, and the various test stimuli as given by the equation:

$$d' = z(H) - z(F),$$

where H gives the proportion of correct responses to one stimulus class (hit rate), F gives the proportion of incorrect responses to the opposing stimulus (false alarm rate), and $z()$ denotes the z-score of those random variables. The measure d' is convenient because it eliminates any biases in the response rates (e.g. due to guessing) that may vary across individuals and within individuals over time. To gauge the effect of various song manipulations during the test sessions, we compared d' values for different stimulus classes using repeated measures ANOVA, and where appropriate used post hoc analyses to quantify the significance of specific differences between mean d' measures. We take $d' = 0$ to be chance performance.

Confidence interval simulations

When considering the data for a single subject (Figs. 2, 4 and 8), we used simulations to generate confidence intervals around either chance or a given bird's baseline performance over a number of trials. For confidence intervals around chance, we simulated 1000 blocks of trials in which the number of trials and overall response rates in a simulated block matched the real data (in 200-trial blocks for

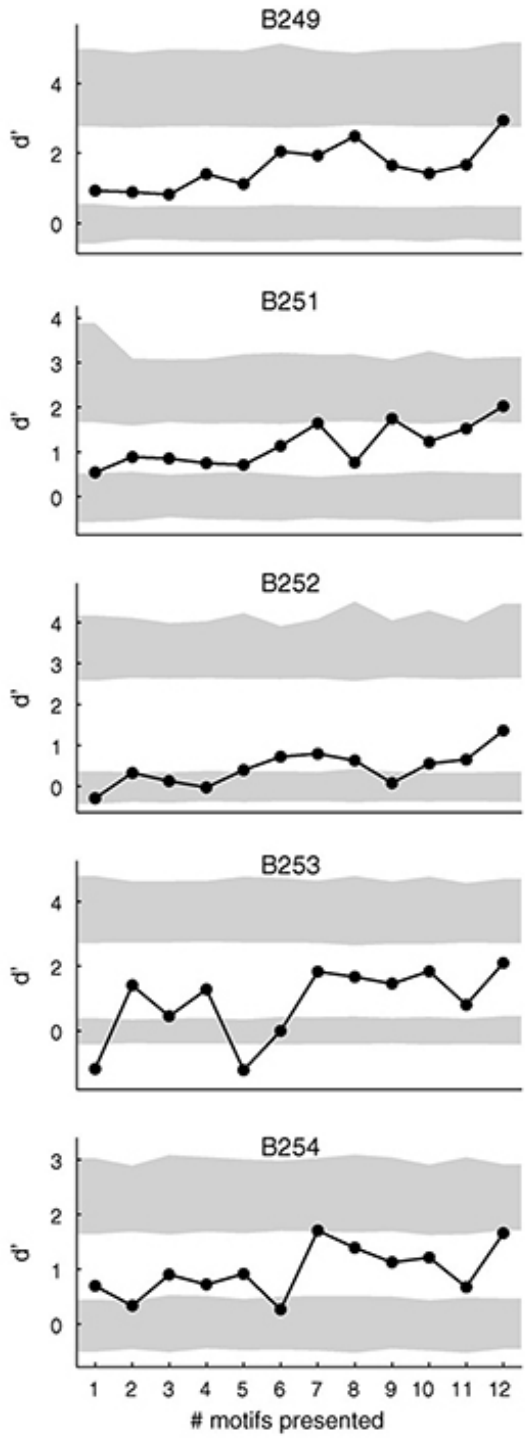


Fig. 2. Individual data for motif-probes. Each panel shows the behavior of an individual subject on motif-probe test trials. Circles indicate performance (d') as a function of the number of motifs presented. The upper region of gray shading in each figure is the 95% confidence interval around baseline performance, while the lower shaded region at the bottom of each figure describes the 95% confidence interval around chance. Individual subject identities are at the top of each panel.

training, or in blocks with the same number of trials as the experimental block of interest), but for which the response selection (e.g. left or right) was randomized. For confidence intervals around a bird's baseline performance, we again created 1000 simulated blocks of trials. However, this time we used a bootstrap technique in which each simulated block was created by randomly selecting, from among all the baseline trials in a session, a number of baseline trials equal to the number of test trials in the experimental block of interest. To calculate the confidence interval, we calculated d' for each of the 1000 simulated blocks. From this distribution of d' values expected by either chance or the bird's baseline performance, we took the 95th percentile as the critical statistical bound.

Comparisons across subjects

The absolute range of d' varies with trial number and response rate. To compensate for this in statistical analyses involving comparisons between subjects, we verified all reported effects using a normalized d' in which we express the value for a given class of test stimuli as a proportion of baseline performance: norm-d' = (d' test)/(d' baseline). Analyses conducted on normalized d' values yielded results qualitatively identical to those reported for the raw d' measures. For ease of interpretation, and unless otherwise noted, we show the raw d' values and report the corresponding statistics.

Experiment 1

Starlings readily learn to classify the songs of individual conspecific singers at very high levels of accuracy (Gentner & Hulse, 1998, 2000), and this ability is tied closely to the motif structure of songs (e.g. Gentner, 2008). However, the large numbers of motifs within each song and their broad acoustic variability are likely to yield many redundant cues to aid song classification. In Experiment 1, we trained birds to recognize songs from two singers, then measured classification accuracy for subsections of the training songs by presenting (1) strings of 1–12 randomly ordered motifs from the training songs or (2) continuous 0.1–12 s excerpts of a training song.

Methods

Subjects

Five starlings participated in Experiment 1. At the start of training, all birds were naïve to the operant apparatus and the stimuli used in the experiment.

Baseline training stimuli

Each stimulus set consisted of eight song exemplars drawn from the library of song bouts sampled from a single bird, and eight exemplars drawn from the songs of another bird (16 exemplars total). The singer of each set of song exemplars and the assignment of those exemplars to either the left or right response port was counterbalanced across test subjects. Each exemplar was 15 ± 0.5 s of continuous song taken from either the beginning, middle, or end of a song bout, as described in the general methods. Many of the exemplars sampled from the beginning of a song bout included whistles, along with other 'warble' motifs (i.e. 'variable' motifs, rattles, and high-frequency motifs; for motif nomenclature, see Adret-Hausberger & Jenkins, 1988; Eens et al., 1991). Those sampled at later time points in a bout comprised only 'warble' motifs. Previous data indicate that recognition is easily learned with this length of a song exemplar (Gentner, 2008) and is unaffected by the relative position within a longer song bout from which the exemplar is sampled and/or the broader motif classes it may or may not contain (Gentner & Hulse, 1998).

Testing procedure and stimuli

After subjects had reached a stable, accurate level of performance on the baseline song classification task, we lowered the rate of reinforcement and punishment from 100% to 60%. Thus, on 40% of the trials there was no consequence for an operant response. We presented two types of test stimuli in separate sessions: (1) a random assortment of 1–12 motifs drawn from the songs used during baseline training, referred to as a "motif-probe", and (2) 0.1–12.0 s excerpt of continuous song drawn from one of the baseline training songs, referred to as an "excerpt-probe".

To construct the motif-probe stimulus on each trial, we chose randomly, with replacement and equal probability, from the set of all motifs that made up the eight songs in one of the two baseline training classes for a given subject. Motifs from the two baseline training classes were never combined in a single test exemplar. The same motif could appear multiple times within a given test sequence, but never in two or more consecutive positions (i.e. there were no consecutive motif repeats). Preliminary, unpublished data suggest that neither motif repetition nor motif selection from a common training song affects recognition under these conditions. The motif-probe test stimuli were presented on 10% of trials, chosen at random. Subjects were presented an average of 146 ± 4 (range: 99–197) motif-probe stimuli with a given number of motifs (1–12). We presented a maximum of 12 motifs because that was the smallest number contained in any of the baseline training stimuli.

To construct the excerpt-probe stimulus on a given trial we randomly chose a continuous portion of song from a randomly selected baseline training song such that the excerpt had a given duration and offset from the start of the original training exemplar. The duration and offset of each excerpt-probe was drawn from a uniform distribution of values ranging between 0.100–12.000 s and 0–12.000 s, respectively. The magnitude of the offset chosen on any given trial necessarily constrains the maximum duration of the stimulus that can be excerpted from the song. As the baseline training stimuli were all roughly 15-s long, the values used here allow us to examine recognition of excerpts from 0.1 to 3 s long taken at all offset values between 1 and 12 s. For excerpts longer than 3 s, the maximum offset we could examine decreased proportionally. The exact value chosen for the duration and offset on any given trial varied at millisecond resolution. Because the excerpt algorithm was agnostic to natural motif boundaries, we added a 10-ms linear ramp to the start and end of the excised waveform to eliminate any potential onset or offset transients induced by the signal manipulation. Subjects were presented an average of 642 ± 69 (range: 57–1906) excerpt-probe stimuli in each of 12 classes (defined by their duration, see below). Shorter duration excerpts could appear with a greater range of possible offsets, and so appeared more frequently than longer duration excerpts.

Analysis

We computed d' for all relevant classes of the stimuli. For the motif-probes, we binned stimuli into classes according to the number of motifs in each test stimulus. For the excerpt-probes, we binned the stimuli into 12 1-s bins: 0.000–1.000 s, 1.001–2.000 s, 2.001–3.000 s, etc., and computed the corresponding d' for stimuli in each bin. For all probe stimuli in a given bin, the 'hit rate' was taken to be the proportion of correct responses to test stimuli derived from one baseline singer, and the 'false alarm rate' is the proportion of incorrect responses to test stimuli derived from the opposing singer. The explicit choice of the singer tied to 'hit' or 'false alarm' proportions does not affect the value of d'.

Results

All subjects learned to recognize the 16 baseline training songs very quickly and ultimately reached a high degree of accuracy in classifying the songs of the two singers. Fig. 1B shows the mean acquisition curve for the five birds exposed to the baseline training in Experiment 1. Recognition performance improved significantly over the course of learning ($F_{(4, 49)} = 13.8$, $p < 0.0001$, rmANOVA, main effect of training), and performance, averaged across subjects, was consistently above chance levels by the ninth 200-trial block (Fig. 1B). On average subjects required 3699 ± 369 trials (range: 2714–4905) to reach a stable performance criterion where d' was greater than 1.0 for three consecutive 200-trial blocks. The mean ($\pm$sem) d' immediately prior to the start of the first test session was 3.05 ($\pm$0.32), demonstrating very accurate recognition of the songs in each class.

Once subjects had reached a stable asymptotic level of performance on the baseline training songs, we lowered the rate of reinforcement and shortly thereafter began the test sessions with the motif-probe songs (methods). Each motif-probe stimulus was a random sequence of 1–12 motifs drawn from the baseline training songs with the constraint that motifs from different singers never combined into the same motif-probe stimulus. Thus, the individual identity of the original singer was preserved in the large set of motif-probe stimuli.

In general, the more motifs that appeared in each test stimulus, the easier it was for each subject to recognize the correct singer (Fig. 2). Although there was considerable variability across subjects,

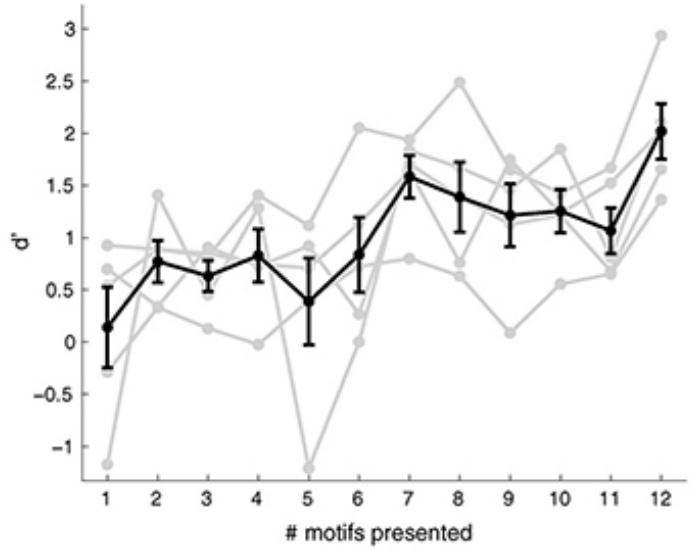


Fig. 3. Mean performance on motif-probes. Mean d' for individuals (light gray lines) and averaged across subjects (black line, ±sem) as a function of the number of motifs in a test song.

increasing the number of motifs in the test stimulus from 1 to 12 had a significant impact on average recognition accuracy ($F_{(4, 11)} = 5.39$, $p < 0.0001$, rmANOVA with all probe classes, Fig. 3). With only a single motif, mean performance ($d' = 0.14 \pm 0.39$) was not above chance ($t = 0.36$, $df = 4$, NS, t-test). With as few as two randomly chosen motifs, mean recognition performance was still modest ($d' = 0.77 \pm 0.20$), but nonetheless significantly above chance ($t = 3.82$, $df = 4$, $p = 0.018$). For 12-motif sequences, recognition performance, average across subjects, was very high ($d' = 2.02 \pm 0.27$, Fig. 3) and was significantly above chance ($t = 7.59$, $df = 4$, $p = 0.0016$). Even when the probe stimulus contained 12 motifs, however, mean recognition was still significantly lower than that for the baseline songs presented during the same sessions (d' for baseline stimuli $= 2.86 \pm 0.26$; $t = 2.98$, $df = 4$, $p = 0.04$ paired t-test), and the mean normalized d' (a ratio of probe:baseline performance) was significantly less than 1 (ratio $= 0.69 \pm 0.08$; $t = 3.765$, $df = 4$, $p = 0.019$).

Although the increase in performance with more motifs was roughly monotonic, it is not linear. The mean response curve (Fig. 3) shows abrupt steps between 1- and 2-motif sequences, another between 6 and 7-motif sequences, where average performance moves well-above chance, and a third large shift between performance on 11- and 12-motif sequences. Examining the data more closely confirms these intuitions. Among the motif-probes containing 2–6 motifs, there is no significant effect of increasing the number of motifs in a given test stimulus ($F_{(4, 4)} = 0.50$, NS). Likewise, among motif-probes containing 7–11 motifs, there is no significant effect of increasing the number of motifs in a given exemplar ($F_{(4, 4)} = 1.42$, NS). Thus, most of the improvement in average performance across the range of motif-probes comes from the increases between 1–2, 6–7, and 11–12 motifs. We note, however, that there is significant variability across subjects (Fig. 2).

After responding to a sufficient number of motif-probe stimuli from each of the 12 classes, we returned subjects to the baseline training songs only. We confirmed their stable performance on the baseline songs over several sessions, and then began the test sessions with the excerpt-probe songs (see methods). Recall, that each excerpt-probe stimulus was a random, 0.1–12.0 s continuous subsection of song drawn from the baseline training songs. We measured performance as the ability to maintain accurate recognition of the individual identity of the singer from whose song the excerpt was drawn. Due to a programming error one subject produced no useable data.

As with the motif-probes, the longer the duration of the song excerpt in each probe stimulus, the easier it was for subjects to recognize the correct singer (Fig. 4). To analyze the results of these probe sessions, we collapsed the different length stimuli into 12 1-s bins, ranging from 0.01–1.00 to

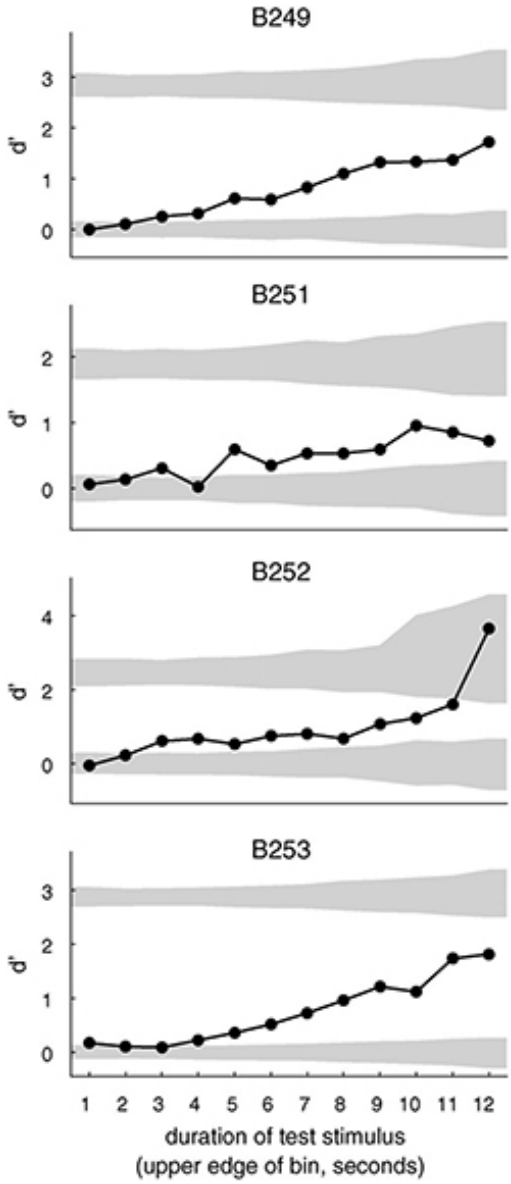


Fig. 4. Individual data for excerpt-probes. Each panel shows the behavior of an individual subject on excerpt-probe test trials. Circles indicate performance (d') as a function of the duration of the excerpt-probe, with confidence intervals and labels as in Fig. 2. Values on the x-axis are the upper edge of the 1-s test stimulus bin, e.g. '3' represents test stimuli 2.001–3.000 s in duration.

11.01–12.00 s (methods) and computed the d' value (for each bird) for all the stimuli that fell into each bin. Initially, we ignored any potential effects due to differences in the offset of a given excerpt from the start of the source song by collapsing across all excerpts of a given length. Increasing the duration of the test stimulus from 1 to 12 s had a significant impact on recognition accuracy ($F_{(3,11)} = 12.07$, $p < 0.0001$, rmANOVA with all probe classes; Fig. 5). As shown in Fig. 5, the overall change in performance with increasing excerpt duration was monotonic and smooth. When presented exemplars 0.1–1.0 s long, recognition accuracy ($d' = 0.04 \pm 0.06$) was not significantly above chance ($t = 0.76$, $df = 3$, NS). Lengthening the probe stimuli slightly, to between 1 and 2 s, produced a rise in performance ($d' = 0.13 \pm 0.02$) to a level that was still modest, but significantly greater than chance ($t = 3.68$, $df = 3$,

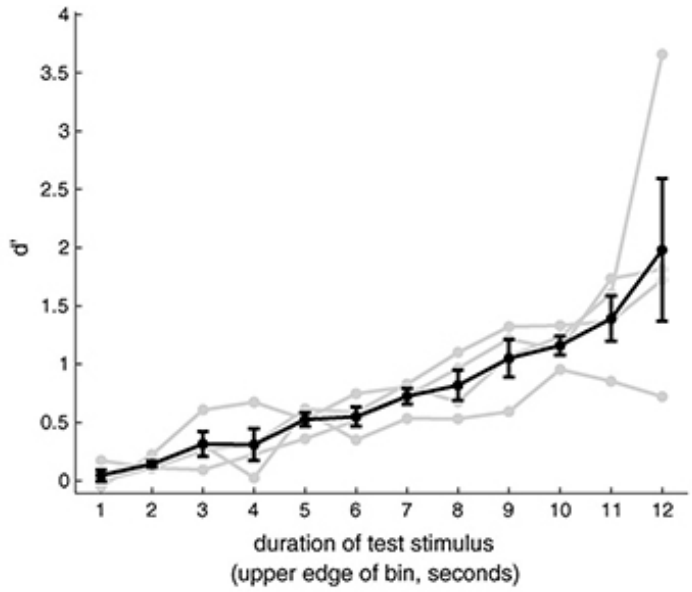


Fig. 5. Mean performance for excerpt-probes. Mean d' for individuals (light gray lines) and averaged across subjects (black line, ±sem) as a function of the duration of the test stimulus. The *x*-axis is labeled as in Fig. 4.

$p = 0.034$). For excerpt-probe stimuli 11–12 s long, mean performance ($d' = 1.42 \pm 0.37$) was significantly above chance ($t = 3.79$, $df = 3$, $p = 0.032$), but still significantly below baseline performance (d' for baseline = 2.61 ± 0.23; $t = 4.711$, $df = 3$, p = 0.018, paired t-test) and the mean normalized d' was significantly less than 1 (ratio = 0.54 ± 0.13; $t = 3.660$, $df = 3$, $p = 0.035$).

In addition to examining the effect of stimulus duration on recognition performance, the excerpt-probe stimuli also provide a means for examining how different regions of the training stimuli may contribute, on average, to recognition performance. For example it could be that either the initial or terminal portions of songs hold particular relevance in guiding performance. To test for such effects, we examined performance on 1-, 2- and 3-s excerpt-probe stimuli as function of each probe's offset relative to start of the source training song. As probe duration increases, the range of possible offsets

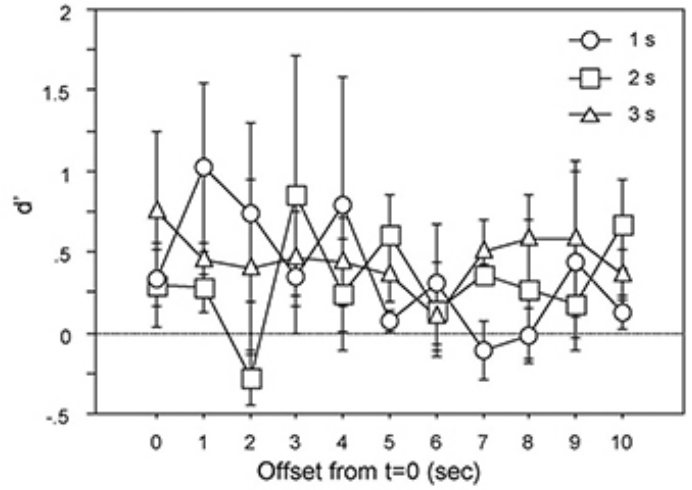


Fig. 6. Stimulus offset. Mean (±sem) d', across subjects, plotted as a function of test song offset (in seconds) relative to the start of the stimulus. Data for excerpt-probe stimuli of 1, 2 and 3 s are plotted separately.

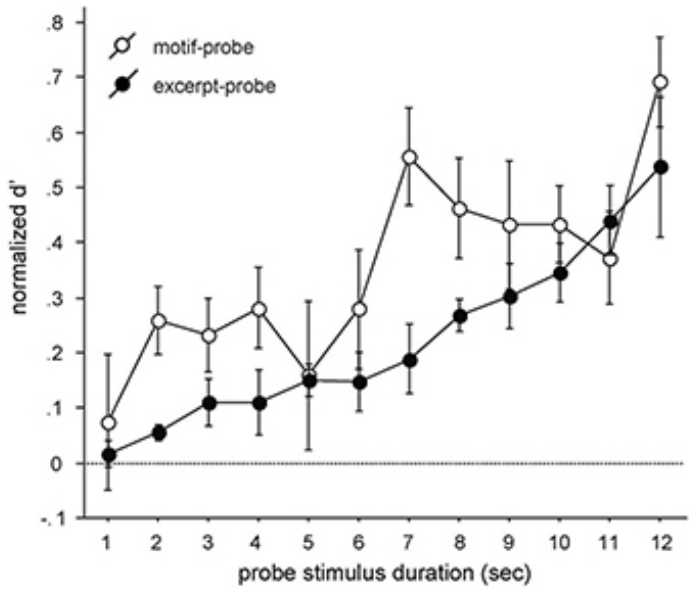


Fig. 7. Comparison performance for Experiment 1. Mean (±sem) normalized *d'* for motif-probe (open circles) and excerpt-probe (filled circles) stimuli plotted as a function of the approximate duration of each test stimulus (see text).

decreases. Thus, we restrict our analysis to these three durations because responses to them can be observed over a wide range of offsets (0–10 s). We observed no effect of relative offset for any of the three excerpt-probe durations ($F_{(3, 10)} = 0.77$, $F_{(3, 10)} = 0.27$, $F_{(3, 10)} = 0.91$, for 1, 2, and 3-s excerpt-probes, respectively, $p > 0.5$ all cases). Fig. 6 shows the similarity in performance across all of these probe conditions, and in the mean performance across subjects.

It is also helpful to consider performance in the motif- and excerpt-probe conditions together. Rigorous comparisons between specific stimuli in these conditions is hampered somewhat by differences in how the two types of probe stimuli are binned for analysis. On average, however, motifs are roughly 1 s long and so direct comparisons across excerpt-probe time-bins and numbers of motifs is reasonable (Fig. 7). Comparing between the two probe conditions in this way, we observe roughly similar patterns of response for similar duration test stimuli. Overall, subjects' tended to perform slightly better on the motif-probe stimuli than excerpt-probes of comparable length, but the advantage was not statistically significant ($F_{(33, 11)} = 1.32$, NS, interaction of probe length with condition).

Conclusions

The results of Experiment 1 confirm the intuition that as more vocal material is added to the test songs, in both conditions, recognition performance improves. With very little vocal material, either 1 motif or less than 1 s of song, performance is at chance levels. With 12 motifs or 12 s of song, performance is markedly better and well-above chance. On average, performance over the range of intermediate stimuli, from small to large amounts of song material, showed a monotonic increase in performance correlated with the amount of song in each test stimulus. The transition from short to long excerpts was much smoother, however, than the transition from few to many motifs. Although not statistically significant, it was somewhat surprising to observe that performance on the motif-probe stimuli tended to slightly lead performance on the excerpt-probe stimuli. Given that the excerpt-probes preserve the relative temporal relationships between motifs that subjects experience during baseline training, one might predict performance on these stimuli to lead that on the motif-probes where all these relative temporal relationships are abolished. The poorer performance may reflect the fact that excerpts could violate the natural start and stop points for single motifs, thereby rendering the first and last motifs in each excerpt more difficult to recognize.

Most subjects were generally above chance even for very small sets of motifs (2–6 motifs) but only marginally so. The inter-subject variability observed in the pattern of responding to the motif-probes may reflect different solution strategies and/or a reliance on different sized sets of motif for recognition. Thus, for any given bird, estimating the minimum number of motifs necessary for accurate recognition is difficult. We can say, however, that above chance recognition is possible in principle, and on average, with as few as two randomly selected motifs. The most reliable gains in recognition (across subjects) do not come until more than seven motifs are strung together. It is not clear what significant change in information is achieved when more than seven motifs are selected. Presumably this threshold is related to the overall size of (i.e. number of motifs in) the training sets, and the corresponding subset of motifs that subjects use to solve the task. Whatever the underlying cause, the non-linear nature of results from motif-probes contradicts a strategy where subjects learn only a single "diagnostic" motif (or sub-motif acoustic feature). Such a strategy would predict that as the number of motifs in each probe sequence increased, the likelihood of choosing the diagnostic motif would increase linearly. Similar arguments apply to larger sets of diagnostic motifs, although the probability of selecting all such motifs by chance decreases exponentially as the number of putatively diagnostic motifs increases. Based on these results, we conclude that the random selection of motifs from training songs is a reasonable proxy, but nonetheless incomplete dimension, along which to model the features that control individual song recognition.

Consistent with the foregoing ideas, there appear to be no strong effects tied to the position of song material within the source stimulus. Excerpt-probe stimuli drawn from the start or middle of a training song were just as recognizable as those drawn from the terminal portions. We note, however, that variance was high in the offset-probe tests, and that this may obscure more subtle effects tied to the position of exemplars within a stimulus. Likewise, individual birds may possess idiosyncratic strategies tied to specific training exemplars that are washed out in the kinds of averaging required to address the role of offset in the present case.

Overall, the results support the conclusion that starlings extract numerous cues to individual vocal recognition from each training song, that these cues are distributed over temporally broad regions of each song, and that adding more cues improves recognition in a graded fashion. No single feature/motif or set of features/motifs can be seen as both necessary and sufficient for individual vocal recognition in starlings.

Experiment 2

In Experiment 2, we trained birds to classify conspecific song and then tested their classification accuracy on test stimuli that had portions of their acoustic material replaced by continuous regions of white noise. This is a complementary manipulation to that of Experiment 1, and we were therefore able to assess different aspects of the birds' classification of conspecific song based on degraded versions of training stimuli. Whereas Experiment 1 presented test material that could start and stop at any region of the training stimulus, and could be out of order with respect to the training stimuli in the case of the motif-probe stimuli, the test stimuli in Experiment 2 maintained the relative and absolute timing information present in the training stimuli while still limiting the amount of acoustic material available to the bird.

Methods

Subjects

Five starlings participated in Experiment 2. At the start of training, all birds were naïve to the operant apparatus and the stimuli used in the experiment.

Baseline training stimuli

For each bird, two roughly 10-s segments of starling song from one singer were assigned to the left response port and two segments of song from a different singer were assigned to the right response port. The specific songs used during training and their left or right key assignments were varied across subjects. In total, eight different stimuli were used across the five birds tested (duration, mean ± sd:

9.860 ± 0.657 s]. We calculated d' over 200-trial blocks to assess performance. Birds were maintained on the baseline training task until completing at least three consecutive 200-trial blocks with d' significantly above chance (see general methods).

Testing procedure and stimuli

Once each bird reached behavioral criterion on the baseline training, we decreased the rate of reinforcement to 40%, so that on 60% of the trials, there was no consequence for an operant response. We then began introducing test stimuli. During testing, each trial initiated by the bird had a 50% chance of being a baseline trial and a 50% chance of being a test trial, here referred to as a "noise-probe" stimulus. Baseline trial stimuli were the same as the training stimuli, and had the same behavioral contingencies. For noise-probe stimuli, a baseline stimulus was chosen and then a random offset was selected uniformly from the beginning of the stimulus at a 1-ms resolution (range: 1–9858 ms, mean 3163 ms, sd 2230 ms). Starting at this offset, a random duration was then selected uniformly from the remaining portion of the stimulus, again at a 1-ms resolution (range: 2–9586 ms, mean 3112 ms, sd 2232 ms). The region of song beginning at the selected offset and lasting for the selected duration was then replaced with broadband white noise. Note that the range of the possible durations are constrained by the particular offset selected for a given trial, so that noise-probe stimuli that had later offsets necessarily had fewer possible duration values that could be selected. The mean amplitude of the white noise was the same in each noise-probe so that the noise itself carried no information about the identity of the acoustic information it was replacing. Noise-probe stimuli were reinforced non-differentially, meaning that the operant response of the bird had an equal chance of eliciting a reward or a punishment (see general methods).

For analysis, we binned the noise-probe stimuli into 9 1-s bins: 0.000–1.000 s, 1.001–2.000 s, 2.001–3.000 s, etc., and computed the corresponding d' for stimuli in each bin as described in Experiment 1.

Results

Subjects learned to accurately recognize and classify the four baseline training songs ultimately reaching a high level of performance. As in Experiment 1, recognition performance improved significantly over the course of learning ($F_{(4, 43)} = 3.699$, $p < 0.0001$, rmANOVA, main effect of training). The mean (±sem) d' immediately prior to the start of the first test session was 2.62 (±0.45), demonstrating accurate recognition of the songs from each singer.

Responses to the noise-probe stimuli showed a pattern consistent with the results of Experiment 1. Recognition performance for all birds became degraded as the duration of white noise in each noise-probe stimulus increased (Fig. 8). To assess this change quantitatively, we collapsed test stimuli into nine bins based on the duration of noise in each (methods), and computed the d' value for all of the test stimuli in each bin (Fig. 9). Increasing the duration of the white noise in each probe stimulus (thereby decreasing the amount of song) significantly affected the birds' classification performance across these nine bins ($F_{(4, 8)} = 14.935$, $p < 0.0001$, rmANOVA main effect of noise duration).

Across all birds, mean performance on noise-probes was indistinguishable from baseline when noise was substituted for 0–1 s of song (d' for noise-probes = 2.20 ± 0.31; d' for baseline = 2.20 ± 0.26; $t = 0.047$, $df = 4$, $p = 0.517$, paired t-test). For all the other noise-probe stimuli, with more than 1 s of noise substituted for songs, recognition was significantly below baseline performance ($p < 0.05$, $df = 4$, all cases, paired t-test), and decreased monotonically with increasing noise durations (Fig. 9).

Although performance fell below baseline values when more than 1-s of noise was substituted for song, recognition accuracy remained above chance for all durations of noise examined. Even when presented with test stimuli that included 8–9 s of noise, the most severe stimulus degradation tested, birds were able to perform the classification task significantly above chance levels ($d' = 0.783 \pm 0.26$; $t = 3.075$, $df = 4$, $p = 0.019$). Thus, birds could still recognize the correct singer of a song when presented with as little as 9–11% of a baseline stimulus, on average.

The results are consistent with those of Experiment 1 where increasing the total amount of song available to the bird improved recognition. To compare the results of Experiments 1 and 2 directly, we classified excerpt-probe and noise-probe stimuli as the proportion of song remaining (relative to the

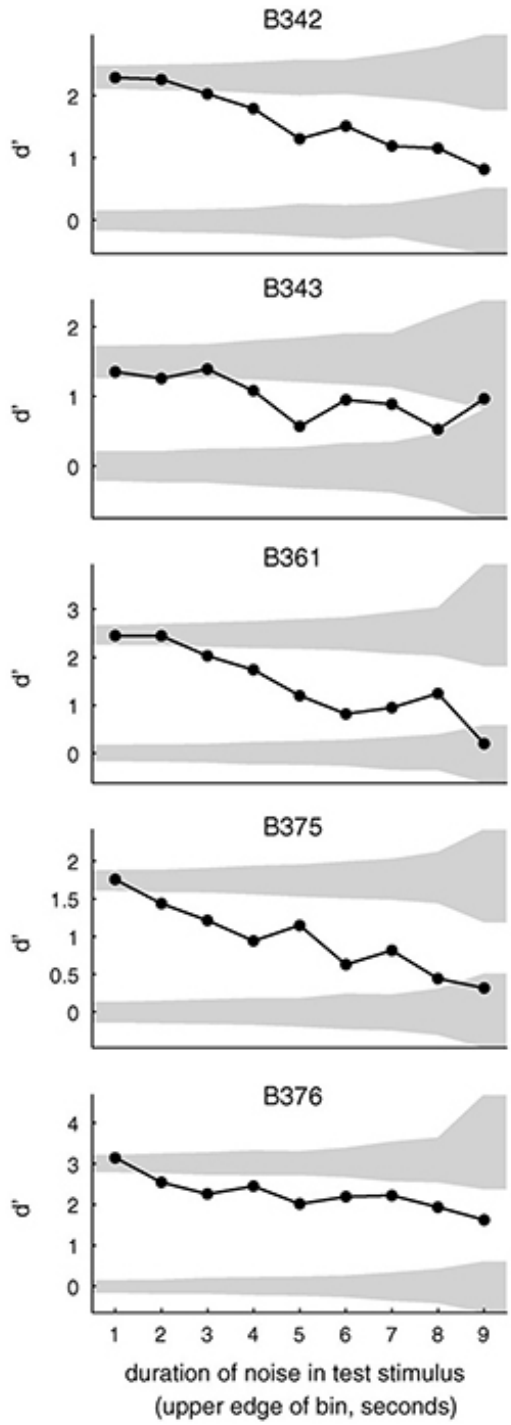


Fig. 8. Individual data for noise-probes. Each panel shows the behavior of an individual subject on noise-probe test trials. Circles indicate performance (d') as a function of the duration of the noise in each probe stimulus, with confidence intervals and labels as in Fig. 2. Values on the x-axis are the upper edge of the 1-s test stimulus bin (e.g. '3' represents test stimuli with noise 2.001–3.000 s in duration).

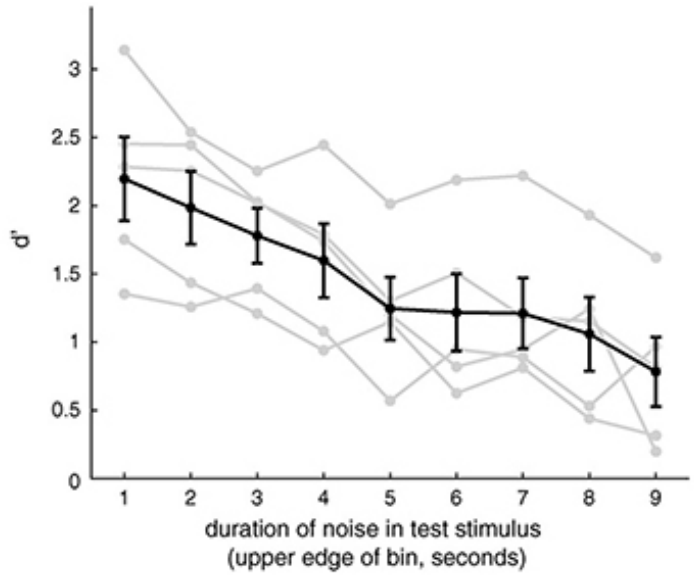


Fig. 9. Mean performance for noise-probes. Mean d' as in Fig. 3, with x-axis labeled as in Fig. 8.

original) in a given test stimulus, binned the probes into 10 equal-proportion bins, and calculated a d' value for each bin. Because of the way the stimuli were constructed, we had insufficient data to estimate performance for bins with a proportion of 0.8–1.0 for the excerpt-probes and 0.9–1.0 for noise-probes. The results of this comparison are shown in Fig. 10. Both conditions showed similar improvement in recognition over comparable ranges of probe stimuli, but recognition of the noise-probe stimuli was clearly better than that of the excerpt-probes. To assess the rate at which behavior improved as a function of available training song, we fit lines to both datasets. The results of the linear regression for

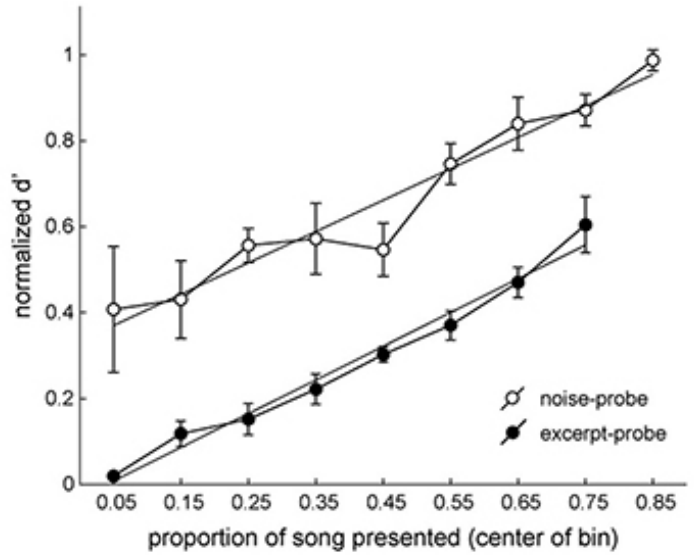


Fig. 10. Comparison performance on excerpt-probe and noise-probe test stimuli. Mean (±sem) normalized d' for excerpt-probe (solid line, filled circles) and noise-probe (dashed line, open circles) stimuli plotted against the proportion of baseline song remaining in the test stimulus. Thin lines show the linear regression for data in each condition. Excerpt-probe: $y = 0.79 \times x - 0.03$, $R^2 = 0.98$; noise-probe: $y = 0.73 \times x + 0.33$, $R^2 = 0.94$.

the excerpt-probe stimuli (*Y*-intercept = −0.031, slope = 0.785, R^2 = 0.980) and the noise-probe stimuli (*Y*-intercept = 0.333, slope = 0.730, R^2 = 0.944), produced a very similar slope, suggesting that the gain in performance seen with increasing baseline stimulus material in Experiment 1 was similar to the gain seen in Experiment 2. The difference in the *Y*-intercept values indicates that recognition of the excerpt-probe stimuli in Experiment 1 was, overall, more difficult than recognition of the probe stimuli in Experiment 2.

Conclusions

The results of Experiment 2 demonstrate that substituting noise for continuous regions of a familiar song can systematically impair song recognition. Recognition performance on noise-probe stimuli is indistinguishable from baseline levels until more than 1 s of the song is masked, after which performance decreases monotonically with increasing noise duration. Importantly, however, accurate classification performance was maintained at all noise durations tested. The ability of birds to correctly recognize the individual singer based on as little as 9–11% of a learned song segment supports the conclusion that the representation of a given conspecific song is distributed across a large portion of the acoustic material available to the bird, and not simply in a few salient auditory features.

In isolation, the results of Experiment 2 could be explained by a generalization decrement. That is, substituting noise for increasingly larger portions of a training song may simply render each probe stimulus more dissimilar to its noise-free version. This dissimilarity, rather than impaired recognition, may account for the observed change in performance. We take up this question in the general discussion and argue that recognition provides a more parsimonious explanation consistent with all of the present results and those from previous studies of song perception in starlings.

Although not directly assessed in the same subjects, the normalized performance values were higher on average in Experiment 2 than in Experiment 1, even when accounting for the level of performance on the training stimuli. This difference may be tied to the availability of absolute timing information present in the noise-probe stimuli, but absent in the motif-probe and excerpt-probe stimuli. It is also possible that differences in performance across the two experiments reflect disparities in the training sets used for each condition. There was much more material for each bird to recognize in Experiment 1 than in Experiment 2 (Experiment 1: 16 songs per subject, mean duration 14.97 s; Experiment 2: 4 songs per subject, mean duration 9.86 s). Dissociating these potential effects requires additional experimentation.

General discussion

The results of Experiments 1 and 2 provide converging evidence in support of the conclusion that knowledge of the songs acquired during individual vocal recognition is distributed across significant portions of conspecific song. Recognition is not restricted to a single motif, or a single region of song. Furthermore, recognition is redundant in that no single acoustic feature of song is necessary, and multiple features (or motifs or regions) are sufficient to enable accurate recognition.

Observations from the present studies provide much greater detail in our understanding of the features that guide song recognition. Across both probe conditions in Experiment 1, performance was consistently below baseline recognition. Even when provided with very large amounts of material, either 12 motifs in the first probe condition or 12 s of song in the second, performance never reached parity with the baseline accuracy. Subjects always performed better on the baseline songs than on the probe stimuli. In both probe conditions, the maximum number of motifs or test stimulus duration never reached the average number of motif/duration for the baseline songs. Baseline training songs had, on average, about 14 motifs and lasted about 15 s. Therefore, we cannot rule out the possibility that adding more motifs (or increasing the duration) of the test stimuli might have brought performance in line with that seen for the baseline. One would expect this to be the case for increasing the duration of the excerpt-probe stimuli. However, some deficit in recognition for motif-probe stimuli, even for long runs of randomly sequenced motifs, is consistent with previous work showing that temporal patterning of motifs in the training songs can affect recognition (Gentner, 2008; Gentner & Hulse, 1998).

Multiple lines of converging evidence support the conclusion that motifs function as an important natural perceptual unit for song recognition in starlings. Synthetic songs that follow the natural motif sequence patterns of each individual are recognized without discernable impairment compared to natural songs (Gentner & Hulse, 1998). Recognition performance of starlings tested with chimeric songs composed of differing proportions of familiar motifs is predicted by the ratio of motifs from each singer (Gentner & Hulse, 2000). Abolishing the sub-motif structure in learned songs leads to significant deficits in song recognition, and sensitivity to motif-level organization emerges without explicit reinforcement (Gentner, 2008). Moreover, familiar motifs (presented in isolation) elicit robust extracellular responses from single auditory neurons in the starling forebrain (Gentner & Margoliash, 2003; Jeanne, Sharpee, & Gentner, 2008; Thompson & Gentner, 2008). Finally, the vocal production of single motifs is highly stereotyped across multiple iterations by the same singer, the natural repetition structure of song operates at the motif level, and where song sharing does occur, it is motifs that are shared. To this long list we can add the results of the present study showing that on average, presenting a randomly chosen, non-sequential, pair of familiar motifs is sufficient for recognition at above chance levels.

In all cases tested, full recognition performance is only maintained when motif structure is intact, but motifs are not the only temporal scale for song perception. Previous results suggest that starlings are also sensitive to features at temporal scales below the motif level (Gentner, 2008) and at temporal scales that span multiple motifs. Starlings can more easily recognize strings of familiar motifs that follow familiar motif transitions than those that have random transitions between motifs (Gentner & Hulse, 1998). Starlings can also learn to recognize patterns defined over disjoint sets of motifs, e.g. the patterns $a_i b_k a_j b_l$ and $a_i b_k b_l a_j$, where $a_{i,j}$ are elements of one set of familiar motifs and $b_{k,l}$ are elements of another set (Gentner, Fenn, Margoliash, & Nusbaum, 2006). Together, the data indicate that different tasks can focus attention on the song in differing ways. The fact that songs carry salient information at temporal scales longer and shorter than that of the motif should not be surprising. The same is true in humans. Human listeners trained with speech synthesizers to recognize either words or sentences show no generalization across context (Greenspan, Nusbaum, & Pisoni, 1988). Likewise, humans trained under specific conditions to extract relevant speech cues from sentences of natural speech generalize identification only to words within sentences, not to isolated words (Nygaard & Pisoni, 1998).

Our discussion of starlings' behavior in the kind of operant tasks presented here turns on the notion that birds learn to recognize sets of songs (or sets of motifs composing songs) associated with each response. This follows from the well-documented ability of individual vocal recognition in all songbird species studied (Stoddard, 1996). It is important, however, to consider simpler explanations that may fit the pattern of data. The results from Experiment 2, in particular, can be interpreted in terms of a generalization decrement, whereby performance falls as probe stimuli become increasingly dissimilar to the training stimuli when more noise is added. Several lines of evidence contradict the generalization decrement hypothesis. Following nearly identical baseline training, starlings can continue to classify novel songs very accurately, i.e. without any detectable response decrement, so long as the novel songs contain at least some number of familiar motifs. This is true even when the novel songs represent substantial permutations of the training songs, either because familiar motifs from multiple trainings songs are combined to make the test songs (Gentner & Hulse, 1998, 2000), or because substantial portions of the motifs in the test songs are completely unfamiliar (Gentner & Hulse, 1998; Gentner et al., 2000). A similar maintenance of recognition in the face of training song permutation can be seen in the results of Experiment 1, where probe songs are composed of motifs drawn (randomly) from multiple training songs and presented in a random order. In this case, similarity to any single training song is low, yet accuracy steadily improves as more motifs are added to the probe songs. Given that the baseline training for Experiments 1 and 2 was very similar, we contend that recognition, not stimulus generalization, is the most parsimonious description of the behavior during all the probe sessions described here (and in our earlier studies).

One interesting conclusion based on the present results is that part of the way singer identity is represented in songs is tied to the relative and absolute timing of motifs (or other acoustic features) within the song. The improvement in recognition for Experiment 2 over the excerpt-probes from Experiment 1 is consistent with this idea. Results demonstrating that star-

lings are sensitive to both relative and absolute motif position cues (unpublished) support this hypothesis.

Having a recognition system sensitive to distributed information, such as that implied by the current results would be adaptive in several ways. Under natural conditions, a singer's vocalizations may be degraded in multiple ways before reaching the receiver. The kind of recognition system described here would be resilient to masking by environmental sounds or perhaps even other singers, and to song bout interruptions on the part of the singer and partial reception by the receiver. The general finding that starlings need only a relatively small portion of acoustic material, taken from any number of points in a bout, to correctly recognize singers points to an auditory system that is adapted to these sorts of environmental challenges.

The finding that recognition increases (or decreases) gradually with the addition (removal) of vocal material in songs means that, at least under these kinds of training conditions, behavior cannot be considered as a dichotomous (either/or) dependent measure of song, but rather a more probabilistic judgment of singer identity. This may reflect the fact that vocal identity itself, as coded in the acoustics of the song, is more probabilistic than discrete, or that the underlying decision mechanisms that ultimately transform information coded in song to behavior are themselves probabilistic. Answering these kinds of questions is not a simple endeavor. It requires precise control over singer identity, song acoustics, recognition behavior, and access to dependent measures at multiple points in the neural processing chain. The latter requirement points to one of the great strengths of the songbird system in general, and of vocal recognition in starlings, more specifically, as this species provides one of the few organisms in which all these measure are feasible. The work described here sets the stage for further manipulations that will directly control recognition behavior under more invasive physiological procedures, such as awake, chronic electrophysiology (Knudsen & Gentner, 2009).

References

Adret-Hausberger, M., & Jenkins, P. F. (1988). Complex organization of the warbling song in starlings. *Behaviour, 107*, 138–156.

Bricker, P. D., & Pruzansky, S. (1976). Speaker recognition. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 295–326). New York: Academic Press.

Cynx, J., & Nottebohm, F. (1992). Testosterone facilitates some conspecific song discriminations in castrated zebra finches (*Taeniopygia guttata*). *Proclamations of the National Academy of Sciences USA, 89*, 1376–1378.

Dooling, R., Brown, S., Klump, G., & Okanoya, K. (1992). Auditory perception of conspecific and heterospecific vocalizations in birds: Evidence for special processes. *Journal of Comparative Psychology, 106*, 20–28.

Eens, M., Pinxten, M., & Verheyen, R. F. (1989). Temporal and sequential organization of song bouts in the European starling. *Ardea, 77*, 75–86.

Eens, M., Pinxten, R., & Verheyen, R. F. (1991). Organization of song in the European starling – Species-specificity and individual-differences. *Belgian Journal of Zoology, 121*, 257–278.

Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. *Journal of The Acoustical Society of America, 124*, 1350–1359.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204–1207.

Gentner, T. Q., & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour, 56*, 579–594.

Gentner, T. Q., & Hulse, S. H. (2000). Perceptual classification based on the component structure of song in European starlings. *Journal of the Acoustical Society of America, 107*, 3369–3381.

Gentner, T. Q., Hulse, S. H., Bentley, G. E., & Ball, G. F. (2000). Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *Journal of Neurobiology, 42*, 117–133.

Gentner, T. Q., & Margoliash, D. (2003). Neuronal populations and single cells representing learned auditory objects. *Nature, 424*, 669–674.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 421–433.

Heinz, R. D., Sinnott, J. M., & Sachs, M. B. (1980). Auditory intensity discrimination in blackbirds and pigeons. *Journal of Comparative and Physiological Psychology, 94*, 993–1002.

Hulse, S. H. (1995). The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals. In G. M. Klump, R. J. Dooling, R. R. Fay, & W. C. Stebbins (Eds.), *Methods in comparative psychoacoustics* (pp. 319–330). Basel, Switzerland: Birkhäuser Verlag.

Jeanne, J. M., Sharpee, T. O., & Gentner, T. Q. (2008). Song recognition learning drives experience-dependant responses in the auditory forebrain region CLM. *Society of Neuroscience Abstracts*, 34.

Knudsen, D., & Gentner, T. Q. (2009). Behavioral state modulation of responses to conspecific song in European starling auditory forebrain. *Society of Neuroscience Abstracts*, 35.

Nelson, D. A. (1988). Feature weighting in species song recognition by the field sparrow (*Spizella pusilla*). *Behaviour, 106*, 158–181.

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, 103*, 171–176.
Nelson, D. A., & Marler, P. (1990). The perception of birdsong and an ecological concept of signal space. In W. C. Stebbins, & M. A. Berkely (Eds.), *Comaprative perception: Complex signals* (pp. 443–477). New York: John Wiley and Sons.
Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics, 60*, 355–376.
Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 651–666.
Stoddard, P. K. (1996). Vocal recognition of neighbors by territorial passerines. In D. E. Kroodsma, & E. H. Miller (Eds.), *Ecology and evolution of acoustic communication in birds* (pp. 356–374). Ithaca: Comstock/Cornell.
Sturdy, C. B., Phillmore, L. S., & Weisman, R. G. (2000). Call-note discriminations in black-capped chickadees (*Poecile atricapillus*). *Journal of Comparative Psychology, 114*, 357–364.
Thompson, J. V., & Gentner, T. Q. (2008). Song recognition learning elicits differential suppression of neural responses to learned sons in NCM. Paper presented at the Society for Neuroscience Abstracts.

Chapter 2, in full, is a reprint of the material as it appears in Learning and Motivation, 2010. Knudsen, Daniel P.; Thompson, Jason V.; Gentner, Timothy Q., Elsevier, 2010. The dissertation author was the primary investigator and author of this paper.

CHAPTER 3

ACTIVE RECOGNITION ENHANCES THE REPRESENTATION OF BEHAVIORALLY RELEVANT INFORMATION IN SINGLE AUDITORY FOREBRAIN NEURONS

KNUDSEN, D. P., AND GENTNER, T. Q.

AS IT APPEARS IN

JOURNAL OF NEUROPHYSIOLOGY

2013

VOLUME 109 ISSUE 7, PAGES 1690-1703

DIGITAL OBJECT IDENTIFIER (DOI): 10.1152/JN.00461.2012

Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons

Daniel P. Knudsen[1] and Timothy Q. Gentner[1,2,3]

[1]*Neurosciences Graduate Program, University of California San Diego, La Jolla, California;* [2]*Department of Psychology, University of California San Diego, La Jolla, California; and* [3]*Kavli Institute for Brain and Mind, La Jolla, California*

Submitted 1 June 2012; accepted in final form 3 January 2013

Knudsen DP, Gentner TQ. Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. *J Neurophysiol* 109: 1690–1703, 2013. First published January 9, 2013; doi:10.1152/jn.00461.2012.—Sensory systems are dynamic. They must process a wide range of natural signals that facilitate adaptive behaviors in a manner that depends on an organism's constantly changing goals. A full understanding of the sensory physiology that underlies adaptive natural behaviors must therefore account for the activity of sensory systems in light of these behavioral goals. Here we present a novel technique that combines in vivo electrophysiological recording from awake, freely moving songbirds with operant conditioning techniques that allow control over birds' recognition of conspecific song, a widespread natural behavior in songbirds. We show that engaging in a vocal recognition task alters the response properties of neurons in the caudal mesopallium (CM), an avian analog of mammalian auditory cortex, in European starlings. Compared with awake, passive listening, active engagement of subjects in an auditory recognition task results in neurons responding to fewer song stimuli and a decrease in the trial-to-trial variability in their driven firing rates. Mean firing rates also change during active recognition, but not uniformly. Relative to nonengaged listening, active recognition causes increases in the driven firing rates in some neurons, decreases in other neurons, and stimulus-specific changes in other neurons. These changes lead to both an increase in stimulus selectivity and an increase in the information conveyed by the neurons about the animals' behavioral task. This study demonstrates the behavioral dependence of neural responses in the avian auditory forebrain and introduces the starling as a model for real-time monitoring of task-related neural processing of complex auditory objects.

behavioral electrophysiology; auditory cortex; birdsong; task engagement; behavioral state

SENSORY SIGNALS MUST GUIDE many different behaviors (e.g., feeding, predator avoidance, mating), and the mappings from neural sensory systems to behaviors are not fixed. As an organism interacts with the environment, both the sensory signals and the animal's behavioral goals for these signals change over time. For example, the salient color and smell of a tasty berry that an animal has previously eaten when hungry may be ignored when the animal is sated. These dynamic components make the challenge of understanding sensory encoding more difficult, as any complete functional description must account for changes in the stimulus representation in relation to each animal's goals. Indeed, goal-dependent modulation of sensory representations can be powerful and broad, ranging from presynaptic inhibition on sensory afferents only during certain behaviors in invertebrates (Gaudry and Kristan 2009) to attentional modulation in visual and auditory cortices (Hubel et al. 1959; Mesgarani and Chang 2012; Moran and Desimone 1985). It is likely that the goals of an organism are continuously shaping the sensory system transformations to highlight relevant stimulus features and attenuate distracting elements.

Studying behavioral goal-dependent modulation of sensory encoding requires tight control over an organism's behavior. One simple method for studying behavioral modulation of sensory representations is to observe neural activity when an animal is engaged in a task that requires the use of sensory information and then compare this activity to that observed when the animal is not actively engaged in the task. Any changes in the neural representation of the stimuli under these two conditions must be attributed to the animal's engagement in the task. In the auditory system, the effects of task engagement on the activity of neurons have been studied in a variety of species, including monkeys (Hocherman et al. 1976; Miller et al. 1972; Niwa et al. 2012), cats (Lee and Middlebrooks 2011), ferrets (Fritz et al. 2003), and rats (Otazu et al. 2009). Changes due to task engagement range from general increases and decreases in spontaneous and driven neural activity to more task-specific effects such as the suppression of responses to distracters (Otazu et al. 2009) or the increase in sensitivity to task-related target features (Fritz et al. 2003; Lee and Middlebrooks 2011; Niwa et al. 2012; see Sutter and Shamma 2011 for a detailed review). These studies provide important insights into the ways that task engagement can alter auditory representations but are limited by the use of artificial stimuli in organisms performing simple laboratory tasks. Because adaptive behaviors can powerfully modulate an animal's goals, understanding will likely benefit from the use of natural stimuli in realistic contexts.

The songbird auditory system provides an excellent model for studying the neural substrate of ethologically relevant behavior involving complex natural stimuli (Doupe and Kuhl 1999; Gentner and Ball 2005; Knudsen and Gentner 2010; Marler 2004; Pinaud and Terleph 2008; Theunissen and Shaevitz 2006). Birds participate in a variety of acoustically mediated natural behaviors such as territory defense, mate attraction (Catchpole and Slater 1995), and recognition of song type (Beecher et al. 1994) and conspecific individual identity (Gentner and Hulse 2000). These behaviors depend critically on differences in the acoustic features in other birds' songs, and recent studies suggest that the avian auditory system is poised to preferentially represent those acoustic features that are particularly relevant to an individual bird's experience. Moving from the songbird auditory periphery into more central

Address for reprint requests and other correspondence: T. Q. Gentner, Psychology Dept., Univ. of California San Diego, 9500 Gilman Dr. #0109, La Jolla, CA 92093-0109 (e-mail: tgentner@ucsd.edu).

auditory regions, neural receptive fields become increasingly complex; linear response models describe less and less of the variance in the neurons' song-evoked, time-varying firing rates (Woolley et al. 2005). That is, higher-order auditory regions are less likely to be driven by simple stimuli like pure tones or noise bursts, requiring more complex acoustic stimuli such as conspecific song to drive their responses. Many neurons in these regions show selectivity among different segments of song, which may arise as a function of the combination of feedforward inputs representing less complex features (Meliza et al. 2010). In addition, many of these secondary auditory regions show strong experience-dependent effects, displaying increased (Gentner and Margoliash 2003) or decreased (Thompson and Gentner 2010) responsiveness for songs that birds have learned to classify, and increased information about learned stimulus identity and task-relevant classification (Jeanne et al. 2011). These studies suggest that powerful changes can be effected in the representation of acoustic stimuli based upon a bird's behavioral experience but have focused on changes observed at one time point, after learning and during neurophysiological recordings in anesthetized animals. These are therefore strong candidate regions for displaying behavioral state-dependent modulation. The goal of the present research was to determine the nature of these modulations, if they exist.

Here we present a novel technique for studying extracellular neurophysiology in European starlings (*Sturnus vulgaris*) while birds perform behaviors mediated by natural stimuli. We trained birds in operant tasks where they learned to recognize a number of conspecific starling songs and then recorded extracellular action potentials from single units in the caudomedial mesopallium (CM), an avian analog of mammalian auditory cortex (Butler et al. 2011) that undergoes strong experience-dependent learning effects (Gentner and Margoliash 2003; Jeanne et al. 2011). We compared responses while birds actively engaged in the auditory recognition task to responses to the same stimuli while the birds were not engaged in the task and found that the behavioral state of the animal modulates the activity of the majority of neurons recorded in CM. Some neurons show systematic increases in their driven firing rate, others show systematic decreases, and yet others display stimulus-specific changes in firing rates. At the population level, task engagement causes neurons to be excited by fewer stimuli, leading to a corresponding increase in stimulus selectivity. Additionally, task engagement leads to a decrease in trial-to-trial firing rate variability at the population level. Although the changes due to task engagement are heterogeneous, we find that these changes allow the output from single neurons to better discriminate between relevant behavioral classes when birds are engaged in the behavioral task than when they are not. From these results, we conclude that engagement in an auditory recognition task alters the neural representation of auditory stimuli in the songbird auditory forebrain, and that these changes occur in such a way as to better transmit information regarding the task that birds are performing.

METHODS

All experiments were performed in accordance with a protocol (no. S05383) approved by the Institutional Animal Care and Use Committee of the University of California San Diego and followed the American Physiological Society "Guiding Principles for the Care and Use of Vertebrate Animals in Research and Training."

Subjects

Eight (7 male and 1 female) European starlings served as subjects for this study. Both male and female starlings readily acquire laboratory operant behaviors involving conspecific vocal recognition (Gentner and Hulse 1998; Gentner et al. 2000), and previous studies of neural activity in medial CM show experience-dependent plasticity in both sexes (Gentner and Margoliash 2003). Subjects were wild-caught in southern California and had adult plumage at the time of capture. Prior to training and testing, birds were housed with conspecifics in large flight aviaries with ad libitum access to food and water. Light-dark cycles in the aviaries were matched to the naturally varying photoperiod. Prior to experiments, birds were naive to all stimuli and to the operant apparatus used in this study.

Operant Apparatus and Shaping

At the start of training, birds were removed from the aviary and acclimated to individual operant chambers. Detailed specifications of the custom-built operant chambers have been described elsewhere (Gentner 2008). Briefly, birds lived in weld-wire cages mounted inside sound-attenuation chambers (Acoustic Systems; Eckel Industries). One wall of the cage contained a metal panel with three response ports into which birds pecked their beaks to trigger various outcomes (Fig. 1*A*). Directly below the response ports was a feeding station at which food could be presented based on the reward contingencies of a given task. A speaker mounted behind the panel was used for auditory stimulus presentation. Above the cage, a recessed broad-spectrum compact fluorescent light bulb provided naturalistic illumination of the chamber ("house light"). Birds earned all food through

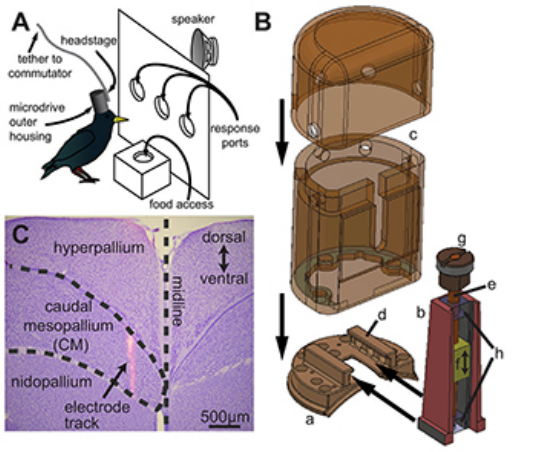


Fig. 1. Novel electrode microdrive for recording extracellular action potentials in awake behaving birds. *A*: schematic of a bird performing the operant task while implanted with the microdrive, depicting the operant and recording apparatus. *B*: the novel electrode microdrive (*b*), base plate (*a*), and protective cap (*c*). See text for a detailed description of the base plate tracks (*d*), threaded rod (*e*), plastic electrode-holding shuttle (*f*), operating knob (*g*), and restraining blocks (*h*). *C*: composite image of a Nissl-stained coronal section (purple cell bodies) through a starling brain at ~2,400 μm rostral of the bifurcation of the Y sinus and a fluorescent (pink-orange) image of a DiI-coated electrode track. The arrow points to the track of an electrode that passed through the caudal mesopallium (CM).

successful completion of the operant tasks described below. Access to water was unrestricted. After quickly learning to eat from the feeding station, birds underwent an auto-shaping procedure that used visual cues (LEDs mounted at the back of the response ports) to teach them how to use the apparatus.

Stimuli

All auditory stimuli were segments of conspecific song recorded from six male starlings at 44.1 kHz and 16 bit (for a more detailed description of the recording parameters, see Gentner 2008). Stimuli were normalized to 65 dB mean SPL, and the first and last 20 ms of the stimulus were linearly ramped to zero to avoid onset and offset artifacts. Starling song can be decomposed into repeated groups of spectrotemporal features called motifs that are natural behaviorally relevant subunits of song (Gentner 2008; Gentner and Hulse 2000; Seeba and Klump 2009). For four of the birds in this study training stimuli comprised either 6 (3 birds) or 12 (1 bird) single motifs (range: 0.38–1.13 s), and for the remaining four birds stimuli were 4 (3 birds) or 8 (1 bird) sections of song made up of 11–15 individual motifs ("long songs," range: 9.08–10.16 s). During some neural recording sessions, a subset of birds (4/8) were presented with additional stimuli (similarly long songs or single motifs) that were not from the set of that bird's training stimuli. These additional stimuli are not discussed further; all analyses are restricted to a given bird's training stimuli to isolate the effects of task engagement on well-learned auditory-mediated behaviors. No bird heard its own song, nor was it familiar with any of the songs used in training or testing before the start of the experiments described here.

Training

Six birds performed a simple two-alternative choice (2AC) recognition task (2 with the long songs and 4 with single motifs). In the 2AC task, birds initiated trials by pecking their beak into the center response port. This elicited the playback of a training stimulus from the speaker, after the end of which the birds had 2 s to make a response by pecking into either the left or right response port. Half of a bird's training stimuli were rewarded with access to food for pecking the left response port, and the other half were rewarded after pecking of the right response port. If the bird made an inappropriate response (e.g., pecking left when it should peck right), the lights in the operant apparatus were turned off for 5 s and the bird was restricted from initiating another trial during this time. If the bird made no response in the left or the right port during the 2-s response window, the trial was considered a "no-response" trial and was not included in analysis. Two additional birds performed a go/nogo (GNG) recognition task using the long songs, where they were trained to respond to half of the songs with a peck in the center response port and withhold responding to the second half of the songs. Correct responses (pecks to "go" songs) were rewarded with access to food, and incorrect responses (pecks to "nogo" songs) were punished by restricting access to trial initiation and briefly turning off the house lights (5 s). Withheld responses (not pecking to a "go" song) were never explicitly reinforced. In both the 2AC and GNG tasks, birds learned all reward-pairing contingencies through trial and error. Incorrect responses were always punished. Correct responses were rewarded at 100% during early training but at lower rates during later training sessions (typically 40–60%) to keep response rates high by maintaining motivation and delaying satiation.

Acclimation to Recording Apparatus

Once the birds reached stable and asymptotic performance on training stimuli, they were transferred to a modified operant apparatus that allowed simultaneous behavioral testing and extracellular recording of action potentials. This recording apparatus was identical in design to the training apparatus, except for modifications that allowed for electrical isolation of the animals and ensured high-quality, low-noise electrophysiological recording. Modifications included coating the wire cage in plastic, a plastic response panel, and insulating the feeding station with a layer of neoprene rubber. In addition to this, the house light was changed from an alternating current CFL to a 24-V DC-powered incandescent bulb, and a green LED ("cue light") was installed above the center response port to signify trial availability and to act as a secondary reinforcer. The cue light remained on whenever the bird was allowed to initiate a trial and turned off as soon as a trial was initiated. Upon the successful completion of a trial, the cue light blinked five times in 0.5 s with a 50% duty cycle to indicate a correct response, even if food was not presented on that trial. A further modification consisted of the installation of a 32-channel motorized commutator (Plexon, Dallas, TX) through a hole in the ceiling of the soundproof chamber. A small hole in the top of the wire cage admitted a multiwire tether that connected the bird to the commutator. This setup allowed the free movement of a tethered bird inside the cage without fear of tangling (Fig. 1*A*).

The behavioral control of the recording apparatus was managed by custom-written Spike2 [Cambridge Electronic Design (CED)] scripts in combination with a CED Power 1401 input-output device that handled digital-to-analog conversion for stimulus playback and managed digital inputs and outputs for control of the response ports and reward/punishment apparatus. In addition, the software controlled the analog-to-digital conversion performed by the 1401 and maintained the temporal registration of the neural data with the behavioral data.

Neurophysiology

Electrode microdrive. Figure 1*B* depicts the electrode microdrive assembly developed in conjunction with the Scripps Institute of Oceanography Machine Shop (**http://sioms.ucsd.edu/**) to allow the recording of extracellular action potentials from awake, behaving starlings. The drive assembly consists of three main components: the base plate (*a*), the microdrive (*b*), and the outer housing (*c*) (Fig. 1*B*). The base plate attaches to the bird's skull with adhesive and dental acrylic and provides an anchor point for the rest of the assembly. The bottom flanges on the microdrive are fitted into the tracks in the base plate (Fig. 1*B*, *d*), and it is held in place by a set screw threaded through one of the holes in the base plate tracks. Finally, the bottom portion of the outer housing is lowered over the microdrive and screwed into the base plate, while the top portion of the outer housing is screwed directly into the lower outer housing.

Base plate. The base plate is made of titanium, and its bottom aspect is machined to conform roughly to the curve of a starling skull. The flange around the perimeter of the base plate and the larger vertically oriented holes provide attachment points for the dental acrylic used to affix the drive to the skull. The smaller vertically oriented holes allow for the passage of wires (e.g., from reference or ground electrodes) from the implant site to the electrode connector. The large rectangular cutout in the base plate allows for access to the implant site by the electrode, visual inspection of the electrode insertion site by the researcher after implantation, and minor curettage and cleaning prior to later electrode penetrations. The horizontally oriented tracks (Fig. 1*B*, *d*) provide an attachment location for the microdrive as mentioned above.

Microdrive. The microdrive slides into the tracks in the base plate (Fig. 1*B*, *d*) and is held in place with a set screw. This allows for adjustment along one horizontal axis, even after the base plate is cemented in place, permitting multiple electrode penetrations in the same subject. Within the titanium microdrive body is a vertically oriented threaded rod (Fig. 1*B*, *e*). Threaded onto this rod and fit tightly into the microdrive housing is a plastic shuttle (*f*) that moves up and down relative to the rest of the drive assembly when the rod is turned with the operating knob at the top (*g*) (Fig. 1*B*). The rod is threaded with 91 threads per inch, and one full rotation of the rod

through 360° leads to a vertical displacement of the shuttle of 279.12 μm. Keeping the rod in place on the bottom is a restraining block containing a closed cylinder in which the rod is allowed to rotate freely, and at the top restraining block there is channel in which a collar firmly attached to the threaded rod is allowed to rotate freely (Fig. 1*B*, *h*). A rubber washer (not depicted) is fitted into the space between the control knob and the top of the upper restraining block to keep tension on the rod in order to prevent it from falling out of the bottom restraining block. By attaching a recording electrode or electrode array to the shuttle with adhesive, and fixing the microdrive in place over the region of interest, the operating knob may be turned to raise and lower the electrode along a vertical track.

Electrode array. The recording electrodes used in this study were 16-channel electrode arrays (NeuroNexus Technologies) equipped with the F16 connector package (currently deprecated and superseded by the H-series connector package). This design consists of 1 or 2 silicone shanks that enter the brain (3–5 mm × 80 μm × 15 μm) and contain 16 iridium contact sites (arranged in either a 1 × 16 linear array or an array of four tetrodes; 121-μm^2, 312-μm^2, or 413-μm^2 site area) attached to a flexible cable (21 mm) that then attaches to a 20-channel connector. The microdrive shuttle is sized to accept both the 16-channel electrode arrays used in this study and 32-channel arrays available from NeuroNexus to increase channel counts in future studies. When the entire microdrive assembly is in place, this connector is attached to a 16-channel tethered headstage (HST/16V-G20, Plexon). The headstage is permanently attached by adhesive to the upper portion of the outer housing and provides 20× gain. The flexible cable on the electrode array allows the recording shanks to be raised and lowered on the shuttle without needing to move the connector/headstage assembly. Before implantation, the electrode array shanks are coated with a small amount of fluorescent dye (DiI), to aid electrode track localization during postmortem histology (DiCarlo et al. 1996).

Outer housing. The outer housing (Fig. 1*B*, *c*) is machined from polyetheretherketone (PEEK), a strong, lightweight, and biocompatible plastic, and provides protection for the rest of the drive assembly from the movements of the bird and protection for the implant site from any external factors or foreign bodies. The outer housing consists of two separate pieces: the bottom section screws directly into the base plate, and the upper section screws into this bottom section. Once the outer housing is screwed into place on the base plate, a small window into the internals of the drive assembly remains open at the point where the rectangular cutout in the base plate is open. This is sealed with a removable silicone gel (Kwik-Cast, World Precision Instruments), so that the entirety of the microdrive assembly and the craniotomy are sealed from the external environment.

While the above describes the present iteration of the microdrive assembly, the eight birds included in this study were implanted either with drives that were identical to the one depicted in Fig. 1*B* or with earlier versions that shared main characteristics but differed in some small respects, such as the inability to remove the outer housing after implantation and thus make multiple penetrations in the same animal.

Electrode microdrive implantation surgery. Once birds achieved high accuracy on their behavioral task, they underwent surgery to implant the microdrive and allow for the recording of single units from their auditory forebrain. Birds were anesthetized with isoflurane, and an incision was made in the scalp to expose the skull. A small opening in the top layer of the skull was made to allow visualization of the bifurcation of the Y sinus. This location was used to determine the proper stereotaxic coordinates for targeting the CM (Fig. 1*C*) 2,500 μm rostral and 500 μm lateral (all birds had implants to the left hemisphere). A small craniotomy was placed dorsal to the target location, and the dura was resected to expose the brain's surface over the target region. A layer of silicone gel (3-4680, Dow Corning, Midland, MI; Jackson and Muthuswamy 2008) was applied to act as an artificial dura, which sealed the entire durotomy and lower portion of the craniotomy. Two smaller craniotomies were made roughly 3 mm bilaterally from this main craniotomy and similarly sealed with artificial dura. These allowed the implantation of a reference electrode (custom 0.003-in.-diameter PtIr wire, etched to a point and glass coated to an impedance of ~1.5 MΩ, 2–4 mm in length) and a ground electrode (custom 0.003-in.-diameter PtIr wire, etched to a point and uninsulated, 2–4 mm in length) These secondary electrodes were patched to the connector on the main electrode array via a fine PVC-insulated wire (Pacific Wire and Cable, Santa Ana, CA) soldered in place prior to implantation. The microdrive, fitted with an electrode array, was then screwed into the base plate, and the entire assembly was lowered into place over the skull so that the vertical trajectory of the electrode shank lined up with the durotomy dorsal to the target of interest. The base plate was then attached to the skull with adhesive and dental acrylic. Once the dental acrylic hardened, the electrode array was slowly lowered into the brain, by turning the control knob on the microdrive, to a depth of ~500–800 μm, dorsal to the CM target area. The outer housing of the microdrive assembly was screwed into place, the electrode connector was mated to the headstage attached to the outer housing, and the bird was allowed to recover with free access to food and water until the start of recording.

Recording procedure. Recording sessions started by attaching the top of the headstage to the bottom of the motorized commutator via the tether described above. The output of the commutator was attached to a multichannel amplifier (either a Plexon PBX2 preamplifier or a model 3600 16-channel microelectrode amplifier, A-M Systems, Sequim, WA). The amplifier provided between 2,000× and 10,000× gain and band-pass filtered the signal (low cutoff: 300–500 Hz; high cutoff: 5–8 kHz). The output of the amplifier was sent to a CED Power 1401 analog-to-digital converter that could provide an additional amplification of the signals by up to 10× and digitized each channel at 19–25 kHz with the custom-written Spike2 scripts.

Once the bird was tethered to the recording apparatus, we followed the recording procedure diagrammed in Fig. 2*A*. A library of stimuli, including the training stimuli, was presented in randomized order to the bird while the operant apparatus was inactive and the bird was quiescent. The electrode array was slowly advanced by turning the control knob on the microdrive until the signal from a single unit could be isolated from background noise. Once a single unit was isolated on one or more channels, the testing sessions began. Testing always began with a non-task-engaged block ("nonengaged"), where the operant apparatus remained inactive and the training stimuli were pseudorandomly presented to the bird (mean 42.3 trials per block). The intertrial interval was drawn randomly on each trial from a uniform distribution between 1 and 5 s. Either the response ports were physically blocked or the LED cue light was turned off to indicate that no trial could be initiated. We used a video camera to monitor each bird's behavioral state continuously and ensure that it did not attempt to initiate or respond to trials during the nonengaged condition. Additionally, we recorded all pecking responses during the nonengaged condition and excluded individual trials in the nonengaged block during which a bird pecked into any response port during stimulus presentation. In practice, such trials were rare, as the lack of reinforcement quickly extinguished pecking during the nonengaged condition. After the nonengaged block completed, the apparatus was turned on, and the bird began a task-engaged ("engaged") block of trials. Trials were allowed to continue while the neuron's isolation was stable, and while the bird continued to perform trials. Because the bird initiated all the trials during the engaged block, the intertrial intervals were variable and the total number of trials per block (mean 30.4) was generally lower than in the nonengaged condition. After the completion of an engaged block, another nonengaged block of trials was run. If the neuron was still separable from background noise and the bird was still motivated, another engaged block was run after this, followed by a final nonengaged block. For the analyses reported here, unless otherwise noted, we grouped the trials from the multiple nonengaged or multiple engaged blocks together.

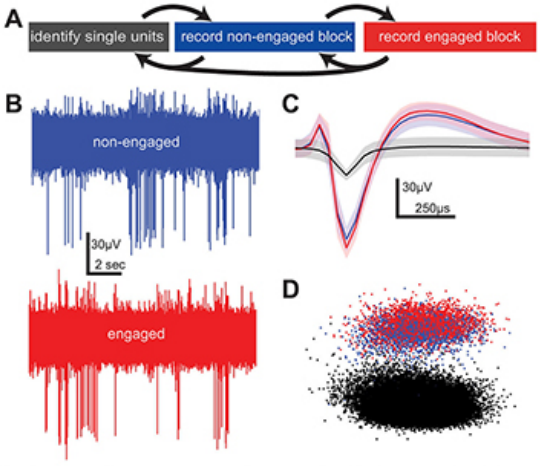


Fig. 2. Recording across non-task-engaged and task-engaged conditions. *A*: diagram of the experimental procedure on a typical recording day. The electrode was advanced ventrally in the brain by adjusting the operating knob on the microdrive in order to isolate single units (gray box). A nonengaged recording block was then performed to record a number of responses to stimuli when the bird was not engaged in the auditory recognition task (blue box). The bird was then allowed to initiate trials and respond to auditory stimuli, and responses were recorded during this task-engaged period (red box). After this task-engaged block, at least 1 more nonengaged block was run before advancing the electrode and searching for more single units. *B*: example raw waveform recordings (blue, nonengaged; red, engaged) from which waveforms were extracted. *C*: mean nonengaged (blue) and engaged (red) waveforms of the same sorted neuron and the mean noise waveform (black) for the same period. Shading denotes ±1 SD. *D*: principal component analysis shows that the waveforms of a sorted single unit cluster together in space during both the nonengaged (blue) and engaged (red) recording conditions and both separate from the cluster of noise waveforms (black).

Data Analysis

Histology. At the conclusion of each experiment, the electrode array was retracted and the microdrive was removed from the base plate. The bird was deeply anesthetized with Nembutal (150 mg/kg) and then perfused with heparinized saline followed by 10% formalin. After cryoprotection in a solution of 30% sucrose in phosphate-buffered saline, brains were frozen, sectioned on a freezing microtome, and mounted on glass slides. Fluorescence images of each section were taken to visualize the DiI fluorescence along the electrode track(s). We then stained the sections for Nissl (which destroys the DiI fluorescence signal) and took another series of images. The Nissl and fluorescent images were then aligned, to visualize the electrode tracks in the context of established cytoarchitectonic boundaries between auditory forebrain regions (e.g., Fig. 1*C*). Single-unit recording locations were registered to the histologically verified electrode position, and only those neurons determined to be within the boundaries of CM were analyzed for this study.

Spike sorting. Putative action potentials were detected and sorted from the raw voltage waveforms with the built-in spike sorting features in Spike2. Some noise artifacts due to birds' movements were observable in the raw waveforms despite attempts to electrically isolate the bird and ground the microdrive. These artifacts generally contained power in low-frequency bands relative to putative action potential shapes, and so could be greatly reduced by high-pass filtering at 300 Hz. Noise artifacts are also common to many channels, and so subtracting one or more averaged reference channels from the data channel of interest (Ludwig et al. 2009) also improved the signal greatly. Putative action potentials were sorted on the basis of spike shape similarity by a combination of template matching and clustering in principal component space. Only well-isolated single units are included in these analyses (see Fig. 2, *B–D*). For each neuron, <0.01% of the interspike intervals violated a 1-ms refractory period.

Data inclusion criteria. Only neurons located in CM and that responded to at least one stimulus with a firing rate significantly different from the spontaneous rate in either the engaged or nonengaged condition were included in our analyses. Birds were trained and tested on a number of stimuli, but only responses to those training stimuli that were presented at least five times in both the engaged and nonengaged conditions are analyzed here, and only those neurons with at least two such stimuli are included. To facilitate analyses across subjects, neural data were analyzed at the scale of the motif. For neurons from birds trained with long songs (n = 24) there were on average 51.5 [standard deviation (SD) 3.2] motifs per neuron that reached the data inclusion criteria, and for neurons from birds trained with individual motifs (n = 41) the mean number of motifs was 3.4 (SD 2.0).

Spontaneous firing rate calculation. Spontaneous firing rates were calculated during the intertrial intervals. Two estimates were taken for each stimulus presentation trial: one before the stimulus [from the start of recording until 250 ms before the stimulus start (mean duration 1.15 s)] and one after the end of the stimulus and trial reward/punishment phase (each estimate was 1 s in duration). Reported average spontaneous firing rates included both estimates from all trials in a given task engagement condition. Differences in spontaneous firing rates for a given neuron were determined by comparing the distribution of firing rates observed during the nonengaged condition to the distribution observed during the engaged condition with the Mann-Whitney U-test (α = 0.05). P values obtained from each neuron were corrected for multiple comparisons across neurons with the false discovery rate procedure (Benjamini 2001). To determine whether neurons in our population were more likely to show increases or decreases in spontaneous firing rate with task engagement, we compared the number of observed decreases and increases to the 95% confidence interval given by the binomial distribution that assumed an equal number of increases and decreases among the neurons showing a significant difference.

Determining significant responses to stimuli. To determine whether a neuron's firing rate was significantly modulated by the presentation of a given motif, we compared the distribution of firing rates evoked by the motif in a given condition to the distribution of spontaneous firing rates in the same condition with the Mann-Whitney U-test (α = 0.05). P values were corrected for multiple comparisons of stimuli within a neuron with the false discovery rate procedure. Responses to motifs that were significantly above the spontaneous rate were termed "excitatory," and responses to motifs that were significantly lower than the spontaneous rate were termed "suppressive." Population comparisons between proportions of driven motifs across conditions were made with the Wilcoxon signed-rank test (α = 0.05). To determine whether single neurons were driven by a different number of motifs in the nonengaged and engaged conditions, we compared the proportion of motifs that evoked responses in the engaged condition to the upper and lower bounds of the 95% confidence interval (computed from the binomial) around the proportion of motifs that evoked responses in the nonengaged condition.

Firing rate changes due to task engagement. We compared changes in firing rates evoked by each motif presented to each neuron in the engaged and nonengaged conditions, using the Mann-Whitney U-test (α = 0.05, corrected for the false discovery rate). Because for most cases the nonengaged rate was computed by averaging rates across nonengaged blocks that preceded and followed the engaged block, any nonspecific effect of block order, such as a slow increase or decrease in rate, would work against our observation of a task-engagement effect. To rule out nonspecific effects of ordering more directly, we also compared the spike rates evoked in each neuron by each motif between two nonengaged blocks separated by an engaged block. We

treated these three blocks as three independent variables in a one-way ANOVA, looked for a significant main effect across blocks ($P < 0.05$), and then used post hoc comparisons [Tukey's honestly significant difference (HSD)] to compare firing rates in each block. As evidence of a nonspecific effect of block ordering, we looked for instances where either *1*) rates differed significantly between the two nonengaged blocks, and both nonengaged rates differed significantly but in opposite directions from the engaged rate, or *2*) rates differed significantly between the two nonengaged blocks, and only one differed significantly from the engaged rate. The same control analysis was done for the smaller subset of neurons where we recorded two engaged blocks separated by an intervening nonengaged block. Both sets of control analyses revealed little support for ordering effects.

Selectivity. Stimulus selectivity describes a neuron's tendency to elicit different numbers of action potentials in response to the presentation of different stimuli and is useful for describing the response characteristics of a neuron with respect to nonparametric stimuli. To investigate stimulus selectivity in our data set, we used a variant of the activity fraction as described by Vinje and Gallant (2000; see also Rolls and Tovee 1995), given by

$$\text{Selectivity} = \left\{1 - \left(\sum r_i/n\right)^2 / \sum\left[r_i^2/n\right]\right\} / \left(1 - 1/n\right)$$

where r_i is the neuron's mean response to the ith stimulus (motif) and n is the number of stimuli presented to the neuron. This measure varies between 0 and 1; neurons with selectivity values near 0 will generally have similar responses to many stimuli, while neurons with selectivity values near 1 respond to fewer stimuli (but at least 1). We calculated the selectivity for each neuron in each behavioral condition and compared the engaged and nonengaged conditions across neurons with a paired t-test.

Variability. One important measure of neural responsiveness is the reliability with which a neuron responds to repeated presentations of the same stimulus. We used the coefficient of variation (CV) to measure this trial-to-trial variability. The CV is equal to the SD of the firing rate across individual trials divided by the mean firing rate across those trials. Normalization by the mean allows comparisons across motifs that evoke different mean firing rates. The CV was calculated for each motif presented to a neuron, and a separate value was calculated for each behavioral condition. Motifs that elicited no spikes in any trial within a condition were not included in the analyses, and only neurons with CV calculated for at least four motifs in both conditions were analyzed, to allow statistical testing. To determine whether or not individual neurons showed increases or decreases in CV with behavioral state, we compared the distribution of each neuron's CV values over all motifs during the nonengaged condition to the distribution of CVs observed during the engaged condition, using a Wilcoxon signed-rank test. Additionally, a mean CV was calculated for each neuron in each condition by averaging the CVs to individual stimuli to obtain a mean nonengaged and a mean engaged CV for each neuron. The distribution of mean nonengaged CVs was then compared to the distribution of mean engaged CVs with a paired t-test. We additionally analyzed the population-level CV for each condition by calculating firing rates using submotif stimulus-response bins that were smaller than those given by motif boundaries (10, 20, 50, 100, 200 ms). Results for these population-level analyses were the same as when we used stimulus responses based on motif boundaries.

Receiver operating characteristic. Two receiver operating characteristic (ROC) curves were plotted for each neuron, one for each condition. For a given condition, we divided all of the trials presented to the neuron into *class 1* ("go" stimuli for birds trained on GNG tasks or "left" stimuli for birds trained on 2AC tasks) or *class 2* ("nogo" stimuli for birds trained on GNG tasks or "right" stimuli for birds trained on 2AC tasks). The firing rate distribution for the class with the higher mean was selected as the "target" class, since we only care about discriminability and not the directionality of the discrimination. We then stepped through all observed firing rate values, treating each one as a separate threshold, and plotted the proportion of false positive firing rate values (proportion of firing rates observed from the nontarget class that were above the current threshold) on the x-axis against the proportion of true positive rates (proportion of firing rates observed from the target class that were above the current threshold). We then calculated the area under this curve (AUC) as our dependent measure. The AUC represents the probability that one can correctly classify a pair of observed firing rates evoked by a *class 1* and a *class 2* motif. To determine whether a neuron's AUC in a given condition was different from that expected by chance, AUCs were calculated on distributions with shuffled trial-class relationships. That is, we randomly assigned each observed firing rate to either *class 1* or *class 2*, regardless of the class of the stimulus that actually elicited it, and calculated the AUC for these distributions. This was repeated 1,000 times in order to build a null distribution. If the AUC calculated from the distributions with the proper trial-class relationships fell above the 95th percentile of this null distribution, it was considered a significant AUC.

RESULTS

The analyses below include 65 well-isolated single units recorded from eight adult starlings. To these neurons, we presented an average of 3.6 (SD 1.6) training stimuli an average of 42.3 (SD 30.4) times each in the nonengaged condition and 30.4 (SD 27.0) times each in the engaged condition. Average behavioral performance classifying the training stimuli for the 2,000 trials preceding microdrive implant was 91% correct for birds trained with the GNG procedure and 85% for those trained with the 2AC procedure. All subjects maintained high levels of performance during neural recording sessions; the mean behavioral performance during the trials included in the neural analyses below was 87% correct for birds performing GNG tasks and 90% correct for birds performing 2AC tasks, suggesting that they tolerated the implants well and were still able to behave at a high level of performance. We confirmed that birds were not attempting to engage in the task during the nonengaged condition with video monitoring and by recording all pecks into response ports (see METHODS). We detected pecks during nonengaged trials in only 17/65 (26%) of recorded neurons, and for only 3 of these neurons were pecks recorded in >5% of nonengaged trials. Trials in which pecks were detected are excluded from the analyses below, but including these trials did not change any of the results.

Task Engagement Changes Average Firing Rates in Some CM Neurons

Median spontaneous firing rates measured during intertrial intervals did not differ between the nonengaged [1.43 spikes per second (sp/s), quartiles 0.72, 3.30] and engaged (1.48 sp/s, quartiles 0.69, 3.43) conditions at the population level (Wilcoxon signed-rank test, $P = 0.20$). However, at the level of individual neurons, 30/65 (46%) neurons showed a significant increase in their spontaneous firing rate in the engaged condition compared with the nonengaged condition, and 7 (11%) showed a significant decrease (Mann-Whitney U-test, $P < 0.05$ corrected for false discovery rate). These increases and decreases are not distributed evenly [probability of observing 7 decreases < 0.001, binomial distribution B(37,0.5)], suggesting that while there is a broad range of spontaneous firing rate differences induced by task engagement, CM neurons are more

likely to increase, rather than decrease, their spontaneous firing rate with task engagement.

Neurons also tended to increase their stimulus-driven firing rates when birds were engaged in a task (median firing rate 2.42 sp/s, quartiles 0.70, 6.41) compared with the nonengaged trials (median firing rate 1.99 sp/s, quartiles 0.81, 6.18), though the difference in median stimulus-driven firing rates averaged across all the stimuli presented to all neurons did not meet our criterion for significance at the population level (Wilcoxon signed-rank test, $P = 0.056$). At the individual neuron level, 6/65 (9%) neurons showed a significant increase in driven firing rate averaged across all stimuli in the engaged condition and 6/65 (9%, see Fig. 4*B* for an example) showed a significant decrease (Wilcoxon signed-rank test, $P < 0.05$ corrected for false discovery rate). Although the population-level effect for the differences in driven firing rates seems stronger than the population-level effect for changes in spontaneous firing rates, fewer individual neurons show firing rate differences in driven firing rates than show differences in spontaneous firing rates. This is likely due to the fact that the distribution of driven firing rates is broader than the distribution of spontaneous firing rates (note in both conditions the similar lower quartiles between the driven and spontaneous firing rates and the larger upper quartiles for the driven firing rates). The increased variability in the driven firing rate distributions compared with the spontaneous distributions makes it less likely to detect small differences in the broader driven firing rate distributions than in the narrower spontaneous firing rate distributions. Taken together, these results show that CM neurons can show both generalized increases and decreases in spontaneous and stimulus-driven firing rates based on task engagement.

CM Neurons Respond to Fewer Stimuli During Task Engagement

As in anesthetized starlings (Gentner and Margoliash 2003; Jeanne et al. 2011; Meliza et al. 2010), CM neurons in awake birds responded to auditory stimulation with both increases and decreases in their firing rates relative to baseline levels. The responses of a typical neuron to two training stimuli are shown in Fig. 3*A*. To examine these responses in greater detail, we divided the training stimuli into their constituent motifs and determined for each neuron which motifs evoked a significant change in firing rate relative to spontaneous rate (Fig. 3*A*). Responses were designated as either excitatory for stimuli that evoked a response above the neuron's spontaneous rate or suppressive for stimuli that lowered the firing rate below the spontaneous rate. This was repeated for both the nonengaged and engaged task conditions.

We quantified these effects by calculating the proportion of motifs driving significant excitatory or suppressive responses in each condition for each neuron. To control for poor estimates of proportionality that come from discretization due to small numbers of stimuli, we restricted this analysis to only those neurons presented with at least four motifs ($n = 38$). The median proportion of motifs that evoked firing rates significantly different (either excitatory or suppressive) from the spontaneous rate was 0.60 (quartiles 0.43, 0.96) in the nonengaged condition and 0.50 (quartiles 0.22, 0.86) in the engaged condition. This difference is significant (Wilcoxon signed-rank test, $P = 0.003$) and remains significant if we include all neurons regardless of the number of stimuli presented (Wilcoxon signed-rank test, $P = 0.02$). Figure 3*B* depicts these proportions graphically for all neurons in both nonengaged (Fig. 3*B*, *top*) and engaged (Fig. 3*B*, *bottom*) conditions.

By subtracting the proportion of significantly driven motifs in the nonengaged condition from that observed in the engaged condition, we calculated differences in the proportion of driven motifs due to task engagement. The distribution of these differences across the population (median −0.07, quartiles −0.26, 0.00; see Fig. 3*C*, *top*) is significantly less than zero (Wilcoxon signed-rank test, $P = 0.004$), indicating that CM neurons respond to fewer motifs when birds are engaged in an auditory recognition task. When we asked whether individual neurons showed differences in the proportion of motifs that elicited a response in the two conditions, we found that 17/38 (45%) of neurons showed such a difference, with 15/17 (88%) of these neurons responding to fewer motifs in the engaged condition and 2/17 (12%) responding to more motifs in the engaged condition (binomial test per neuron, $\alpha = 0.05$), suggesting that it was more likely that a given CM neuron responded to fewer motifs while a bird was engaged in a task than it did when the bird was not engaged.

CM Neurons Are Excited by Fewer Stimuli During Task Engagement

The observed reduction in the number of motifs that drive significant responses in CM neurons during task engagement may come from a decrease in the number of motifs that evoke excitatory responses or an increase in the number of motifs that suppress responding. To answer this question, we separately compared the proportions of motifs driving excitatory and suppressive responses in the engaged and nonengaged conditions (again restricting our analysis to neurons with at least 4 motifs presented in each behavioral condition, $n = 38$). The difference between the proportion of motifs that excited the neuron in the engaged condition and the nonengaged condition was significantly different from chance (median difference = −0.02, quartiles −0.13, 0.00; Wilcoxon signed-rank test, $P = 0.008$; Fig. 3*C*, *middle*). Thus CM neurons are excited by fewer stimuli in the engaged condition than in the nonengaged condition. The difference between the proportion of motifs that suppressed the neuron in the engaged and nonengaged conditions was not different from chance (median difference = 0.00, quartiles −0.06, 0.00; Wilcoxon signed-rank test, $P = 0.8$; Fig. 3*C*, *bottom*). Thus, across the population of neurons, the decrease in the number of motifs that evoke CM responses during task engagement comes from a decrease in net excitation rather than an increase in suppression.

Task Engagement Changes Motif-Evoked Firing Rates in Most CM Neurons

Although only a minority of CM neurons show a uniform change in firing rate for all stimuli, averaging over all of the motifs presented to a neuron can mask significant variability in the response to different stimuli. Many individual neurons show differences in the driven firing rate in response to one or more motifs. Figure 4*A*, *left*, shows a plot of one neuron's average responses to individual motifs in both nonengaged and engaged conditions for six training stimuli. For most motifs the neuron responds similarly in the two conditions, but for *motif*

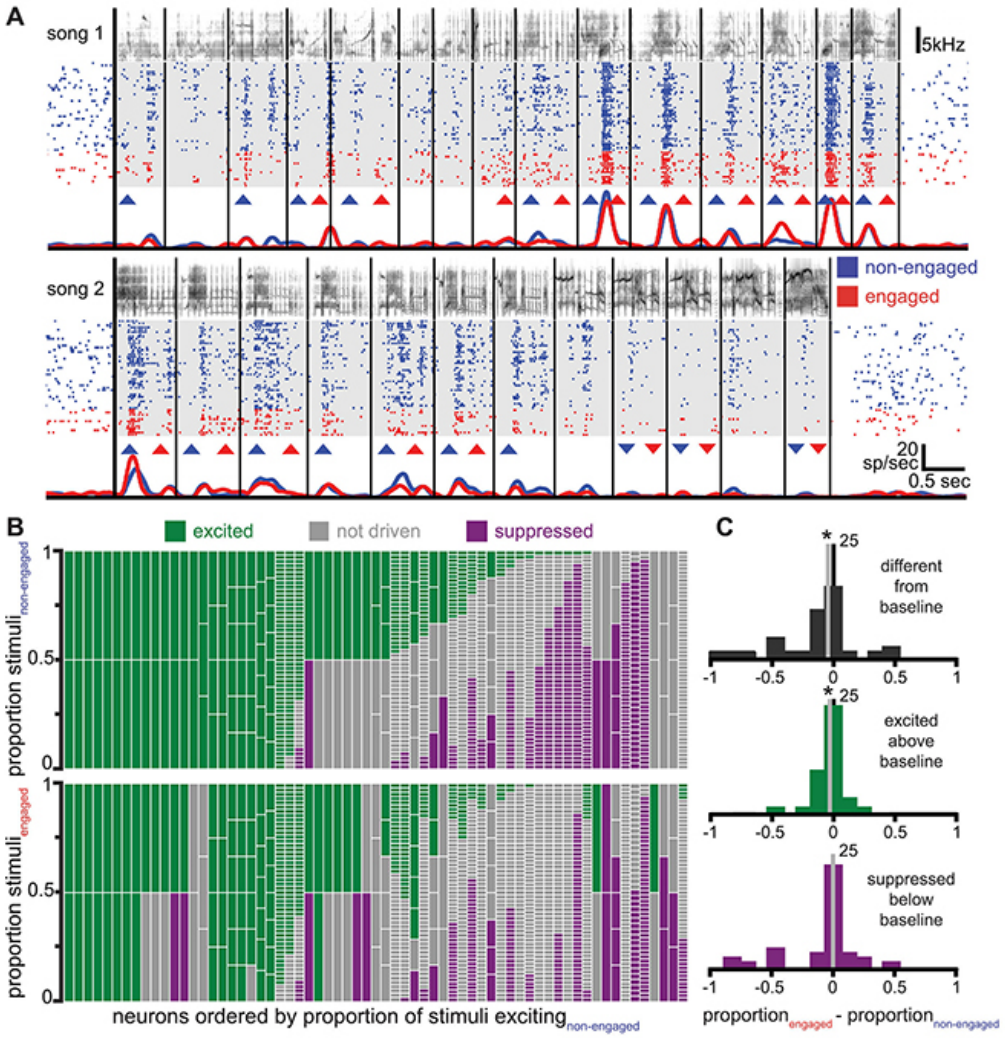


Fig. 3. Neurons are driven by fewer acoustic stimuli when birds are engaged in an auditory recognition task. *A*: responses of a neuron to 2 different example stimuli. For each stimulus there is depicted the spectrogram of the song (*top*), raster plots of the occurrence of action potentials on each trial (*middle*; each row is a trial and each colored dot represents the occurrence of an action potential; *song 1*: 56 trials, *song 2*: 52 trials), and the time-varying firing rate for trials in each condition (*bottom*). Vertical black lines depict motif boundaries in the stimulus, and upward- and downward-pointing triangles denote motifs for which the firing rate was significantly higher or lower than baseline levels, respectively, in either of the 2 conditions (nonengaged and engaged). *B*: each vertical column depicts the stimuli presented to a neuron, with colors representing the proportion of these stimuli that were excited above baseline levels (green), suppressed below baseline levels (purple), or not significantly different from baseline levels (gray), in both the nonengaged (*top*) and engaged (*bottom*) conditions. Horizontal white lines in each column separate individual stimuli. The neurons are sorted by *1*) the proportion of stimuli exciting the neuron during the nonengaged condition, *2*) the proportion of stimuli suppressing the neuron during the nonengaged condition, and *3*) the number of stimuli presented to each neuron. *C*: each of the 3 histograms represents the frequency of observing a given difference between the engaged and nonengaged conditions in the proportion of stimuli that were significantly different from spontaneous values (*top*), excited above baseline (*middle*), and suppressed below baseline (*bottom*). Values on *left* indicate that there was a higher proportion of stimuli in the nonengaged condition, and values on *right* indicate that there was a higher proportion in the engaged condition. The top 2 distributions have medians that are significantly different from zero (*).

4 the firing rate is significantly higher when the bird is engaged in the task and for *motif 5* the neuron's firing rate is significantly lower during task engagement. Figure 4*A*, *right*, shows a comparison of the firing rates in each condition, and a histogram of these differences at top right shows that despite the changes in responses to individual stimuli there is no systematic increase or decrease in firing rate due to task engagement (Wilcoxon signed-rank test, $P = 0.44$). Figure 4*B* depicts another neuron, which shows a distribution of differences in the driven firing rate between the two conditions that

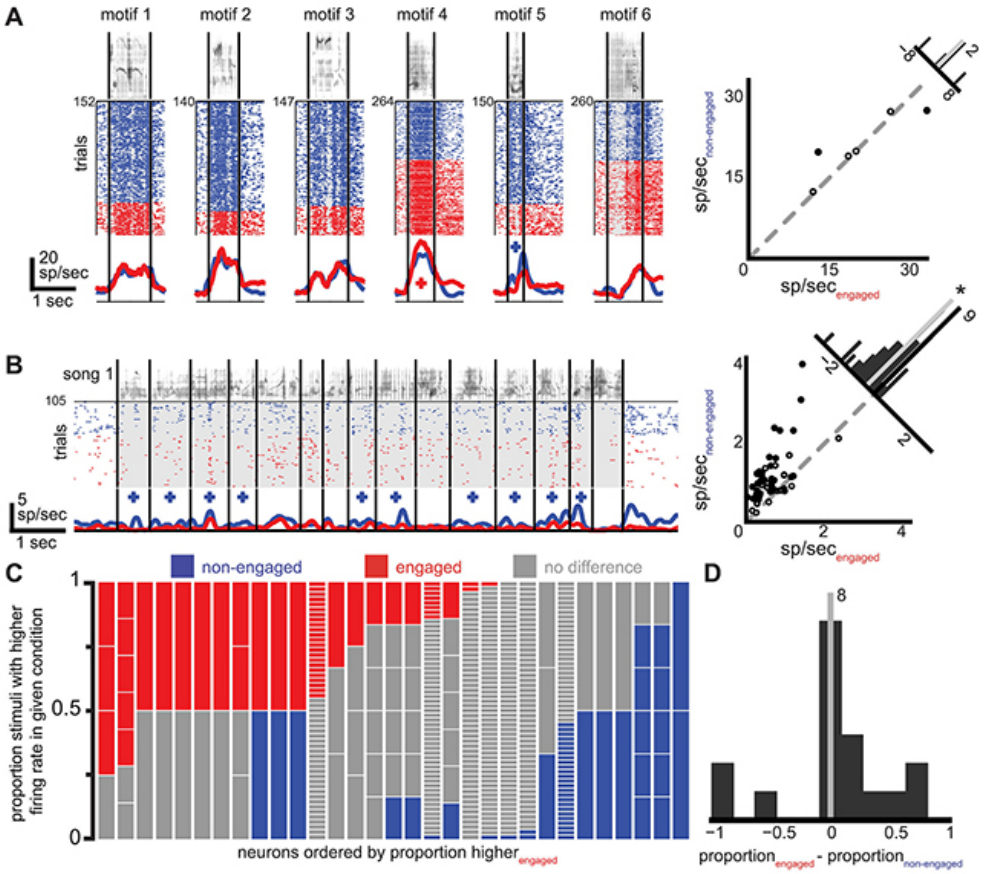


Fig. 4. Neurons respond differently when birds are engaged in an auditory recognition task than when the birds are not task engaged. *A*, *left*: responses of an example neuron to training stimuli as in Fig. 3*A*, except that instead of triangles plus signs now denote motifs for which the firing rate is significantly different between the conditions (blue plus sign, nonengaged higher; red plus sign, engaged higher). All 6 motifs that made up the bird's training set are depicted. *Right*: circles represent the firing rate of all stimuli presented to this neuron in both the nonengaged (*y*-axis) and engaged (*x*-axis) conditions. Filled circles denote individual stimuli that are significantly different between the 2 conditions. The histogram at *top right* displays the distribution of the difference in firing rate between the engaged and nonengaged conditions for all stimuli (gray line, median). *B*: a different example neuron from a bird trained with long songs shows an overall decrease in firing rate during the engaged condition compared with the nonengaged condition. The median (gray line) of firing rate differences depicted in the histogram is significantly below zero (*). *C*: similar to Fig. 3*B*, for neurons that show any significant differences between the engaged and nonengaged conditions each vertical column depicts the stimuli presented to a neuron, with colors representing the proportion of these stimuli that elicited significantly different firing rates between the 2 conditions (red, engaged higher; blue, nonengaged higher; gray, not significantly different). Neurons are sorted by the proportion of stimuli eliciting a higher firing rate during the engaged condition. *D*: for stimuli that show differences in firing rate in the 2 conditions, these differences are equally distributed between increases and decreases with task engagement (the distribution is not shifted from zero).

is significantly shifted from zero (Wilcoxon signed-rank test, $P < 0.001$), indicating that this is one of the six neurons for which the mean driven firing rate shows a more general decrease with task engagement. Even for this neuron, however, not every motif evoked a different firing rate in the two conditions. Thus firing rate changes across the two conditions are not uniform for all motifs.

To characterize this response heterogeneity more fully, we asked how many neurons showed a difference in firing rate between the engaged and nonengaged conditions for any single motif. We found that 31 of 65 (48%) neurons showed significantly different firing rates between the two conditions in at least one individual motif. Ten neurons showed only decreases in the response to single motifs with task engagement, 13 showed only increases, and 8 showed both increases and decreases. These neurons are represented in Fig. 4*C*, where the proportion of increases (engaged firing rate higher) and decreases (nonengaged firing rate higher) due to task engagement can be seen for each neuron. For neurons that had differences in at least one motif and that were presented with at least four motifs ($n = 17$), the median proportion of motifs that showed modulation in the firing rate between the two conditions was 0.33 (quartiles 0.13, 0.55). For each of these neurons, we subtracted the proportion of stimuli that were higher in the

engaged condition from the proportion that were higher in the nonengaged condition and compared these values across the population of neurons (Fig. 4*D*). This difference was not significant across the population (Wilcoxon signed-rank test, $P = 0.55$), indicating that there was not a consistent tendency to have significant increases in one or the other behavioral condition. In all, 39 of 65 (60%) neurons showed a difference in firing rate between the engaged and nonengaged conditions, either at the level of a single motif (31/65, 48%) or at the level of a general facilitation or suppression across all stimuli (12/65, 19%).

To confirm that the observed changes in stimulus-evoked spike rates were not attributable to a time-varying parameter other than behavioral state, we compared firing rates across the nonengaged recording blocks that preceded and followed each engaged block (see METHODS). In 28 of the 31 neurons in which the firing rate for at least one motif differed between engaged and nonengaged conditions, the mean firing for those motifs did not differ significantly between the pre- and postengaged blocks (ANOVA, Tukey's HSD post hoc, $P > 0.05$ all cases). Thus the observed change in spike rate is tied directly to the change in task engagement. In the remaining three neurons, we found a small number of motifs (3 of 23, 1 of 5, and 3 of 3, respectively) where the firing rates differed significantly between the pre- and postengaged blocks and moved in opposite directions relative to the intervening engaged block. Thus, for these seven motifs, we could not rule out the possibility that temporal ordering effects might explain the observed change in spike rate across the task engagement conditions. None of the 12 neurons showing a general facilitation or suppression of firing rates with task engagement was affected by the ordering of behavioral state. Thus most spike rates were stable across multiple, nonsequential behavioral conditions, and among the very few neurons that did show instabilities between conditions they were not uniform across all stimuli. The relative timing or ordering of recording sessions cannot account for the observed changes in firing rates.

Task Engagement Increases Stimulus Selectivity

Stimulus selectivity values ranged from 0.02 to 0.80 in the nonengaged condition (mean 0.36, SD 0.25), suggesting a wide range of representational densities; some neurons responded similarly to most stimuli, while others responded to only a few stimuli. Across the population, neurons showed a modest but significant increase in selectivity during task engagement (range 0.02–0.89, mean 0.39, SD 0.26) relative to the nonengaged condition (paired *t*-test, $P = 0.04$). Figure 5*A* shows these selectivity values for all neurons with greater than four stimuli ($n = 38$) and depicts a distribution of the differences between conditions in the histogram at top right. The tendency for neurons to respond to fewer stimuli in the engaged condition as reported above, and to respond to those fewer stimuli with higher firing rates, leads to a general increase in selectivity at the population level and a modest sparsening of the stimulus representation.

Task Engagement Decreases Response Variability

We measured trial-to-trial response variability, using the CV (see METHODS). We found that in 6/37 neurons (16%) the CV was lower in the engaged condition, while only 1/37 neurons

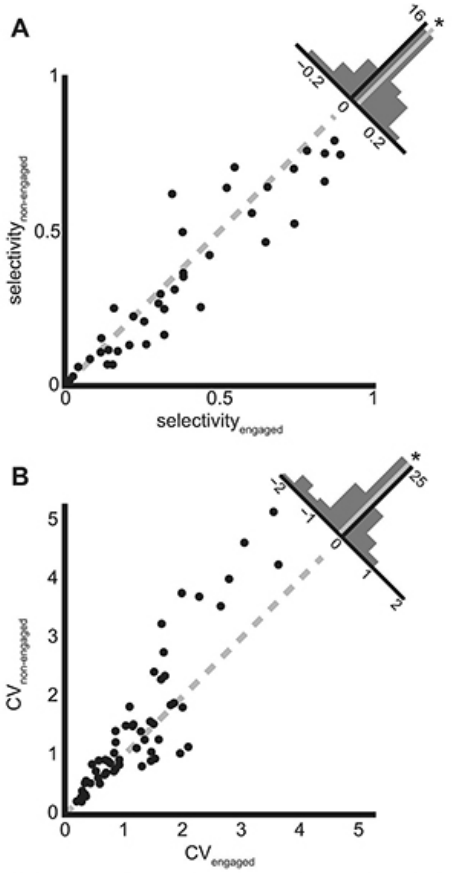


Fig. 5. Firing rate variability decreases and stimulus selectivity increases during task engagement. *A*: scatterplot of the selectivity of a neuron to all presented stimuli in both the nonengaged (*y*-axis) and engaged (*x*-axis) behavioral conditions. The histogram at *top right* shows the difference in the selectivity between the engaged and nonengaged conditions. The median (gray line) is significantly shifted to the right (*); neurons show higher stimulus selectivity in the engaged condition. *B*: scatterplot of the mean of the coefficient of variation (CV) of trial-to-trial firing rate across all stimuli presented to a neuron in the nonengaged (*y*-axis) and engaged (*x*-axis) conditions. The histogram in *top right* corner depicts the distribution of the differences in CV between the engaged and nonengaged conditions. The median (gray line) is significantly shifted to the left (*); the trial-to-trial firing rate variability decreases when a bird engages in an auditory recognition task.

(3%), showed a higher CV in the engaged condition (Wilcoxon signed-rank test, $P < 0.05$ corrected for false discovery rate). When we averaged the CV values for all motifs presented to a given neuron and compared this across the engaged and nonengaged conditions for all neurons (Fig. 5*B*), we found that the CV was significantly higher in the nonengaged condition (mean 1.46, SD 1.18) than in the engaged condition (mean 1.22, SD 0.84; paired *t*-test, $P = 0.001$). Analyzing responses of neurons to submotif stimulus time bins instead of responses based on motif boundaries yielded similar results (see METHODS). Thus actively engaging in a task reduces the trial-to trial

variability in the spike rates of CM neurons, in principle permitting a reliable representation of the stimulus.

Task Engagement Improves Stimulus Class Discrimination

Up to this point, our results describe a population of neurons that modify their activity in response to task engagement by responding to fewer stimuli and responding to those stimuli more strongly and with increased reliability. There are likely many top-down factors contributing to these phenomena that can be attributed to the animal's behavioral goals. In this task, birds need to discriminate between two classes of stimuli: either they must discriminate between stimuli that require a response (go stimuli) and those that do not (nogo stimuli) or they must discriminate between stimuli associated with a left response and stimuli associated with a right response. If the top-down mechanisms that birds employ when engaged in the task are affecting CM neurons, we hypothesized that they might be doing so in a way that increases the discriminability of the neural representation of the stimulus classes upon which the animal is performing the task. To test this hypothesis, we calculated the ROC curve for each behavioral condition in each neuron (Fig. 6, *A* and *B*). The area under the ROC curve (AUC) describes how well an ideal observer can discriminate between the two stimulus classes given the neuron's firing rate. In our present formulation, it represents the probability that an observer is able to use a neuron's distribution of firing rates to correctly categorize a pair of responses when one comes from a *class 1* motif and the other from a *class 2* motif. If neurons have a greater AUC in the engaged condition than they do in the nonengaged condition, it would demonstrate that there was more discriminative power in the neurons' responses when a bird was engaged in a task.

When we compared the AUC values during nonengaged recording epochs to engaged epochs, we found that CM neurons tended to have higher AUCs during task engagement (median 0.59, quartiles 0.54, 0.69) than during passive listening (median 0.58, quartiles 0.53, 0.66), though this difference was not statistically significant (Wilcoxon signed-rank test, $P = 0.082$). We then asked how many neurons could classify *class 1* and *class 2* stimuli significantly better than chance by comparing the measured AUC to a null distribution of ROCs created from the same data but with the associations between firing rate and stimulus class randomized. By this measure, 31/65 (48%) neurons had AUCs significantly above chance values in both the nonengaged and engaged conditions—that is, an observer could use the neuron's firing rate distributions to correctly determine motif classes at better than chance performance. For these neurons, we found the AUC values during engaged epochs (median 0.67, quartiles 0.59, 0.84) to be significantly higher than those calculated from nonengaged recording epochs (Fig. 6*C*; median 0.61, quartiles 0.55, 0.72; Wilcoxon signed-rank test, $P = 0.002$). From these analyses, we conclude that neurons that can discriminate between *class 1* and *class 2* stimuli can do so better when the animal is engaged in an auditory recognition task. In other words, engaging in auditory recognition increases the amount of information about the animal's task that is represented by these CM neurons.

DISCUSSION

Telencephalic sensory systems provide representations of the external sensory world that can be modulated by the real-time behavioral demands of an organism. This report describes techniques that allowed us to determine the effects of task engagement on neurons in the starling auditory system by recording single-unit activity in the CM in awake, behaving European starlings while they performed auditory recognition tasks using complex natural stimuli. We found that engaging in a task that requires using the auditory information present in complex natural stimuli modulates the responses of the majority of the neurons in starling CM. Of the 65 single units from which recordings were made, 39 (60%) showed significant

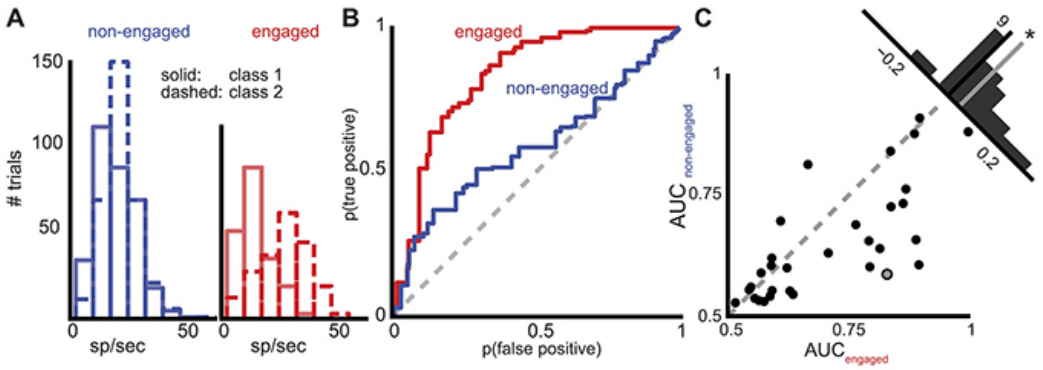


Fig. 6. CM neurons have a greater ability to discriminate between behavioral classes when birds are engaged in a task. *A*: distribution of firing rates for an example neuron in both the nonengaged (*left*) and engaged (*right*) conditions. Solid lines indicate the distribution of firing rates to *class 1* stimuli, and dashed lines indicate the distribution of firing rates for *class 2* stimuli. By eye, one can observe a greater separation in the distributions in the engaged condition. This difference is better described by the receiver operating characteristic curve (ROC) in *B*, where the false positive rate is plotted against the true positive rate at different thresholds in each condition. The area under the curve (AUC) is larger for the red curve, indicating that there is more discriminability between *class 1* and *class 2* firing rates for this neuron in the engaged condition than in the nonengaged condition. *C*: for all neurons that showed significant discrimination ability in both task engagement conditions, the AUC is plotted in each condition. The distribution of differences (*top right*, $AUC_{engaged} - AUC_{nonengaged}$) is significantly greater than zero(*), meaning that neurons show greater discrimination of task-relevant classes when they are engaged in a task than when they are passively listening. The gray-centered point indicates the example neuron from *A* and *B*.

differences in the stimulus-driven firing rate due to the behavioral state of the animal. We also observed a reduction in the proportion of stimuli that excite neurons above baseline, accompanied by an increase in selectivity and a decrease in response variability. In addition, for those neurons that reliably encode the behavioral class of (i.e., correct response to) the training stimuli, behavioral engagement increased the information about the stimulus classification that these neurons carried. Changing the behavioral goal of the animal alters the neural representation of sensory stimuli in CM neurons in real time.

Although studies of auditory-evoked neural activity in awake starlings have been performed for over three decades (Kirsch et al. 1980), most have made recordings in passively listening birds from the primary thalamorecipient region field L (but see Meliza et al. 2010 for recordings from awake, passively listening, starlings in medial CM) and were aimed at mapping response properties (Capsius and Leppelsack 1996, 1999; Cousillas et al. 2005) and neural correlates of perceptual phenomena such as categorization of natural sounds (Hausberger et al. 2000) and stream segregation (Bee et al. 2010; Bee and Klump 2004, 2005; Itatani and Klump 2009, 2011). As far as we are aware, this is the first report of recordings from auditory-responsive neurons made in awake songbirds engaged in a controlled auditory-mediated behavioral task. These awake-behaving recordings were enabled by the microdrive described here. Combined with commercially available microelectrode arrays, this drive provides a robust recording system that allows for the isolation of high-quality single units in awake, behaving animals. A motorized version of the microdrive is under development and will allow remote positioning of the electrode array and limit the need to physically handle the animal. These recording techniques will require less experimenter intervention and ultimately allow a broader range of physiological experimentation in the context of increasingly complex behaviors.

At least two previous studies have attempted to make comparisons of responses in the starling auditory forebrain across behavioral states. Recording from neurons in field L, Capsius and Leppelsack (1996) reported a decrease in the spontaneous rate of multineuron and single-neuron activity in the auditory forebrain of anesthetized and awake starlings. Meliza et al. (2010) did not see this general decrease in spontaneous firing rate with anesthesia in medial CM, although they did see differences in spike timing precision and facilitative interactions in awake compared with anesthetized birds. In agreement with the latter study, we observed no uniform change of the population average spontaneous firing rate between behavioral states. Although one should be careful of conflating changes in anesthesia with changes in natural behavior states, these reported differences may reflect differential regulation of spontaneous firing rate in field L and CM. One advantage of our data set is that we track changes in the same neurons across behavioral state, whereas the other studies recorded from different neurons in the two behavioral conditions. The fact that we observe one subpopulation that increases spontaneous firing rate and another that decreases spontaneous rate indicates that behavioral state modulation, at least in CM, may be different for different classes of neurons. Determining whether these two subpopulations correspond to physiologically different neuronal subtypes will require additional study.

In the Meliza et al. (2010) study, responses of CM neurons to motifs in awake-restrained and anesthetized starlings are described as being additive and suppressive combinations of responses to motif subcomponents called "notes." The authors report that there are more facilitative interactions between combinations of these notes when birds are awake than when they are anesthetized and suggest that this should lead to increased stimulus selectivity. While they observe no change in selectivity across behavioral state using the same selectivity measure we used here, they suggest that the broad range of selectivity values measured and the relatively low sample size might not allow them to observe these selectivity effects. We too observed a broad range of selectivity values across neurons in our data set, but because we measure selectivity in the same neurons under two behavioral conditions the between-neuron variability can be removed from the analysis. Doing this reveals a somewhat modest, but consistent, increase in selectivity that coincides with task engagement.

The observed selectivity differences are due to differential modulation of responses to different motifs in the two behavioral conditions. Although we found no uniform difference in spontaneous or driven responses across all neurons due to task engagement, some individual neurons displayed systematic increases (6/65 neurons, 9%) or decreases (6/65 neurons, 9%) in driven firing rate as a result of task engagement. Moreover, about half (31/65, 48%) of the neurons in CM showed a firing rate difference for at least one motif between behavioral conditions. These differences were equally distributed between increases and decreases, demonstrating that modulatory mechanisms can act selectively on certain acoustic features while leaving others unchanged. This agrees with previous conclusions (Meliza et al. 2010) that CM neurons receive a wide variety of inputs tuned to complex motif subcomponents like notes and their complex receptive fields are built up by suppressive and facilitative combinations of these inputs. Although our knowledge of CM receptive fields is impoverished, the fact that response changes with task engagement are often motif specific suggests that individual inputs (or sets of inputs) to CM neurons may be independently altered in the short term by behavioral tasks and/or goals. Moreover, because the discriminability of responses evoked by different motifs improves during task engagement, but only in the subset of neurons that discriminate between behavioral classes to begin with, the modulatory mechanisms invoked by task engagement also appear able to target specific subsets of neurons.

The proportion of stimuli that neurons responded to decreased with task engagement, and this change was due to a reduction in the ability for stimuli to drive neurons above their baseline firing rates. This effect might be caused by a reduction in feedforward excitatory drive, a general increase in inhibitory drive that affects excitatory responses to a greater degree than suppressive responses, or an enhancement of the suppressive interactions between features that drive inputs onto CM neurons. Enhanced inhibition or suppression may explain, at least in part, the observed reduction in trial-to-trial firing rate variability. However, a global increase in inhibitory drive during the engaged condition is not supported by changes in spontaneous firing rate or by the trend toward higher driven firing rates, which should both be uniformly lower in such a scenario. A general caveat to these interpretations is that the effects of suppression are hard to measure in the extracellular responses

of neurons that have low firing rates, and so may be underrepresented here. To adequately determine the relative contributions of inhibitory and excitatory drive on tuning properties and the effects of behavioral engagement, it will be necessary to develop intracellular recording techniques that allow the monitoring of subthreshold currents in behaving animals.

Short-term modulation of auditory cortical responses across behavioral states has been observed in other systems, but its explicit functions remain poorly understood. In ferrets and cats, receptive fields of auditory cortical neurons shift with task engagement toward stimulus features that are diagnostic for the task at hand (Fritz et al. 2003, 2007; Lee and Middlebrooks 2011). In rats, responses of auditory cortical neurons to distracting, broadband clicks that precede tones are suppressed when rats are actively detecting the tones (Otazu et al. 2009). The present results extend this prior work to include songbirds and complex natural stimuli and tie these behavioral-state-dependent changes directly to the discriminability of the stimulus-driven neural responses. We speculate that the changes in neural responsiveness observed here are the product of short-term modulatory processes such as attention and ultimately contribute to the enhanced representations of learned acoustic material observed throughout in CM at longer timescales (Gentner and Margoliash 2003; Jeanne et al. 2011). Activation of neuromodulatory centers associated with arousal (Froemke et al. 2007; Kilgard and Merzenich 1998; McLin et al. 2002; Metherate and Weinberger 1990) or reward (Bao et al. 2001) have been implicated in experience-dependent plasticity of auditory cortices, and receptive field changes in single neurons that result from stimulation of the nucleus basalis coincident with auditory signal presentation can last for multiple hours (Froemke et al. 2007). Short-term modulation of natural stimulus responses having been demonstrated, it will be important for future studies to directly control selective attention to specific acoustic features of the same stimulus, and to track changes in the response of single neurons (and populations) over the course of recognition learning.

In addition to the modulatory roles of internal processes such as attention and learning, the overt structure of the task is likely to have significant effects on neuronal encoding as well. We observed general effects of task engagement across different behavioral tasks (2AC, GNG) and stimuli (long songs, single motifs), but it would not be surprising if differences in training protocols (even those employed in the present study) did not lead to differential task-related modulation. The reward contingencies differ dramatically between the 2AC and GNG procedures, and this may have a strong effect on the motivation of the animals during task engagement. Although the present experimental design lacked sufficient statistical power to test the independent contribution that different training might make to the observed task-engagement effects, we note that different training contingencies have been linked to CM neuron response differences in anesthetized birds (Gentner and Margoliash 2003). Thus it seems reasonable to predict that similar differences could be observed in in awake birds given the appropriate design in future studies. Importantly, because the training conditions and stimulus sets were held constant within each neuron, differences between training and/or stimuli cannot explain the reported effects of task engagement within any single neuron.

The results of the present study highlight the importance of precise control over each animal's behavior during awake physiological recording. While it is clear that moving away from anesthetized preparations is a necessary step for systems neurophysiology, the benefits of making this transition are not likely to be realized by simply removing anesthesia. As soon as one wakes up their experimental subject they must contend with a large number of previously fixed and unknown internal variables. Our results demonstrate that simple engagement in an auditory recognition task can produce significant changes in how neurons encode natural acoustic stimuli, but we note that this manipulation, while effective, leaves most of the ethologically relevant internal variables only loosely controlled at best. Although interesting in their own right, showing how the auditory system can modify its representations in real time, the results are also somewhat daunting. They remind us that providing a full account of auditory cortex will require tasks that bring stimulus type, reward contingencies, and working memory loads, attention, and learning under exquisite behavioral control.

ACKNOWLEDGMENTS

The authors acknowledge David Malmberg for his helpful advice and technical expertise in designing and improving the novel microdrive, as well as the members of the Gentner Lab and two anonymous reviewers for helpful comments on the manuscript.

GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-008358.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: D.P.K. and T.Q.G. conception and design of research; D.P.K. performed experiments; D.P.K. analyzed data; D.P.K. and T.Q.G. interpreted results of experiments; D.P.K. prepared figures; D.P.K. drafted manuscript; D.P.K. and T.Q.G. edited and revised manuscript; D.P.K. and T.Q.G. approved final version of manuscript.

REFERENCES

Bao S, Chan VT, Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412: 79–83, 2001.

Bee MA, Klump GM. Primitive auditory stream segregation: a neurophysiological study in the songbird forebrain. *J Neurophysiol* 92: 1088–1104, 2004.

Bee MA, Klump GM. Auditory stream segregation in the songbird forebrain: effects of time intervals on responses to interleaved tone sequences. *Brain Behav Evol* 66: 197–214, 2005.

Bee MA, Micheyl C, Oxenham AJ, Klump GM. Neural adaptation to tone sequences in the songbird forebrain: patterns, determinants, and relation to the build-up of auditory streaming. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 196: 543–557, 2010.

Beecher MD, Campbell SE, Burt JM. Song perception in the song sparrow: birds classify by song type but not by singer. *Anim Behav* 47: 1343–1352, 1994.

Benjamini Y. The control of the false discovery rate in multiple testing under dependency. *Ann Stat* 29: 1165–1188, 2001.

Butler AB, Reiner A, Karten HJ. Evolution of the amniote pallium and the origins of mammalian neocortex. *Ann NY Acad Sci* 1225: 14–27, 2011.

Capsius B, Leppelsack HJ. Influence of urethane anesthesia on neural processing in the auditory cortex analogue of a songbird. *Hear Res* 96: 59–70, 1996.

Capsius B, Leppelsack HJ. Response patterns and their relationship to frequency analysis in auditory forebrain centers of a songbird. *Hear Res* 136: 91–99, 1999.
Catchpole CK, Slater PJ. *Bird Song: Biological Themes and Variations*. Cambridge, UK: Cambridge Univ. Press, 1995.
Cousillas H, Leppelsack HJ, Leppelsack E, Richard JP, Mathelier M, Hausberger M. Functional organization of the forebrain auditory centres of the European starling: a study based on natural sounds. *Hear Res* 207: 10–21, 2005.
DiCarlo JJ, Lane JW, Hsiao SS, Johnson KO. Marking microelectrode penetrations with fluorescent dyes. *J Neurosci Methods* 64: 75–81, 1996.
Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. *Annu Rev Neurosci* 22: 567–631, 1999.
Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat Neurosci* 6: 1216–1223, 2003.
Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *J Neurophysiol* 98: 2337–2346, 2007.
Froemke RC, Merzenich MM, Schreiner CE. A synaptic memory trace for cortical receptive field plasticity. *Nature* 450: 425–429, 2007.
Gaudry Q, Kristan WB. Behavioral choice by presynaptic inhibition of tactile sensory terminals. *Nat Neurosci* 12: 1450–1457, 2009.
Gentner T, Ball G. A neuroethological perspective on the perception of vocal communication signals. In: *The Handbook of Speech Perception*, edited by Pisoni DB, Remez RE. Oxford, UK: Blackwell, 2005, p. 653–675.
Gentner T, Hulse S. Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Anim Behav* 56: 579–594, 1998.
Gentner TQ, Hulse SH, Bentley GE, Ball GF. Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *J Neurobiol* 42: 117–133, 2000.
Gentner TQ, Hulse SH. Perceptual classification based on the component structure of song in European starlings. *J Acoust Soc Am* 107: 3369–3381, 2000.
Gentner TQ, Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669–674, 2003.
Gentner TQ. Temporal scales of auditory objects underlying birdsong vocal recognition. *J Acoust Soc Am* 124: 1350–1359, 2008.
Hausberger M, Richard J, Leppelsack E, Leppelsack HJ. Neuronal bases of categorization in starling song. *Behav Brain Res* 114: 89–95, 2000.
Hocherman S, Benson DA, Goldstein MH, Heffner HE, Hienz RD. Evoked unit activity in auditory cortex of monkeys performing a selective attention task. *Brain Res* 117: 51–68, 1976.
Hubel DH, Henson CO, Rupert A, Galambos R. "Attention" units in the auditory cortex. *Science* 129: 1279–1280, 1959.
Itatani N, Klump GM. Auditory streaming of amplitude-modulated sounds in the songbird forebrain. *J Neurophysiol* 101: 3212–3225, 2009.
Itatani N, Klump GM. Neural correlates of auditory streaming of harmonic complex sounds with different phase relations in the songbird forebrain. *J Neurophysiol* 105: 188–199, 2011.
Jackson N, Muthuswamy J. Artificial dural sealant that allows multiple penetrations of implantable brain probes. *J Neurosci Methods* 171: 147–152, 2008.
Jeanne JM, Thompson JV, Sharpee TO, Gentner TQ. Emergence of learned categorical representations within an auditory forebrain circuit. *J Neurosci* 31: 2595–2606, 2011.
Kilgard MP, Merzenich MM. Cortical map reorganization enabled by nucleus basalis activity. *Science* 279: 1714–1718, 1998.
Kirsch M, Coles RB, Leppelsack HJ. Unit recordings from a new auditory area in the frontal neostriatum of the awake starling (*Sturnus vulgaris*). *Exp Brain Res* 38: 375–380, 1980.
Knudsen DP, Gentner TQ. Mechanisms of song perception in oscine birds. *Brain Lang* 115: 59–68, 2010.
Lee CC, Middlebrooks JC. Auditory cortex spatial sensitivity sharpens during task performance. *Nat Neurosci* 14: 108–114, 2011.
Ludwig KA, Miriani RM, Langhals NB, Joseph MD, Anderson DJ, Kipke DR. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *J Neurophysiol* 101: 1679–1689, 2009.
Marler P. Bird calls: their potential for behavioral neurobiology. *Ann NY Acad Sci* 1016: 31–44, 2004.
McLin DE, Miasnikov AA, Weinberger NM. Induction of behavioral associative memory by stimulation of the nucleus basalis. *Proc Natl Acad Sci USA* 99: 4002–4007, 2002.
Meliza CD, Chi Z, Margoliash D. Representations of conspecific song by starling secondary forebrain auditory neurons: toward a hierarchical framework. *J Neurophysiol* 103: 1195–1208, 2010.
Mesgarani N, Chang EF. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485: 233–236, 2012.
Metherate R, Weinberger NM. Cholinergic modulation of responses to single tones produces tone-specific receptive field alterations in cat auditory cortex. *Synapse* 6: 133–145, 1990.
Miller JM, Sutton D, Pfingst B, Ryan A, Beaton R, Gourevitch G. Single cell activity in the auditory cortex of rhesus monkeys: behavioral dependency. *Science* 177: 449–451, 1972.
Moran J, Desimone R. Selective attention gates visual processing in the extrastriate cortex. *Science* 229: 782–784, 1985.
Niwa M, Johnson JS, O'Connor KN, Sutter ML. Active engagement improves primary auditory cortical neurons' ability to discriminate temporal modulation. *J Neurosci* 32: 9323–9334, 2012.
Otazu GH, Tai LH, Yang Y, Zador AM. Engaging in an auditory task suppresses responses in auditory cortex. *Nat Neurosci* 12: 646–654, 2009.
Pinaud R, Terleph TA. A songbird forebrain area potentially involved in auditory discrimination and memory formation. *J Biosci* 33: 145–155, 2008.
Rolls ET, Tovee MJ. Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *J Neurophysiol* 73: 713–726, 1995.
Seeba F, Klump GM. Stimulus familiarity affects perceptual restoration in the European starling (*Sturnus vulgaris*). *PloS One* 4: e5974, 2009.
Sutter ML, Shamma SA. The relationship of auditory cortical activity to perception and behavior. In: *The Auditory Cortex*, edited by Schreiner CE, Winer JA. New York: Springer, 2011, p. 626–628.
Theunissen FE, Shaevitz SS. Auditory processing of vocal sounds in birds. *Curr Opin Neurobiol* 16: 400–407, 2006.
Thompson JV, Gentner TQ. Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *J Neurophysiol* 103: 1785–1797, 2010.
Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science* 287: 1273–1276, 2000.
Woolley SM, Fremouw TE, Hsu A, Theunissen FE. Tuning for spectrotemporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat Neurosci* 8: 1371–1379, 2005.

Chapter 3, in full, is a reprint of the material as it appears in the Journal of Neurophysiology, 2013. Knudsen, Daniel P.; Gentner, Timothy Q., The American Physiological Society, 2013. The dissertation author was the primary investigator and author of this paper.

CHAPTER 4

LEARNING TO CLASSIFY NOVEL STIMULI ALTERS NEURAL RESPONSES IN THE AVIAN AUDITORY CORTEX

KNUDSEN D. P. AND GENTNER T.Q.

INTRODUCTION

The manner in which vertebrate sensory systems represent the external world changes on multiple timescales across the life of an organism. Early sensory input adapts individuals' visual and auditory systems quickly to the gross statistics of the external world that an organism experiences before critical windows shut in early development, while adult sensory systems can be tuned toward features important to adult animals' experiences with behaviorally relevant stimuli. Short-term modulatory changes in the tuning and gain of individual neurons assist with local stimulus normalization and highlight important features relevant to motivational and attentional processes. Stimulus representations are not static in vertebrate sensory systems.

The songbird auditory system is particularly well suited to studying changes in auditory encoding that depend on an animal's experience. Songbirds are notable for their complex vocally mediated social interactions. These social interactions include mate selection and bonding, territory disputes with con- and heterospecifics, and even individual vocal recognition, which requires forming an association between a particular individual and that individual's song (Catchpole and Slater, 1995; Gentner and Hulse, 1998). Each of these behaviors necessitates relating a particular auditory signal with an appropriate behavioral response, which can vary based on the type of interaction, a bird's history, and its current behavioral state. The full range of relationships between acoustic input and appropriate behavior cannot be known until

the world is experienced by an animal, and it must therefore learn the appropriate associations over the course of its lifetime.

Correspondingly, many regions of the avian auditory system show evidence of long-term experience-dependent alterations in their representation of auditory stimuli. After European starlings are brought into a laboratory setting and trained to classify sets of naturalistic stimuli in an operant task, population responses in many auditory brain regions show differences in their neural response as a function of training history. Neurons in both the lateral and medial caudal mesopallium (CLM and CMM, respectively), secondary auditory forebrain regions, tend to show stronger responses for learned songs and contain more information in their patterns of activity about learned stimuli than about novel stimuli (Gentner and Margoliash, 2003; Jeanne et al., 2011). Additionally, CLM seems to preferentially encode stimulus familiarity, while CMM appears to preferentially encode the likelihood that a given stimulus will lead to reward (Jeanne et al., 2011), suggesting that long-term changes are not limited to familiarity, but may represent other features of acoustic stimuli, such as reward valence. A neighboring and interconnected region, the caudomedial nidopallium (NCM), appears to preferentially encode novel stimuli over learned stimuli (Thompson and Gentner, 2010), and responses of neurons in this region are heavily influenced by local inhibitory mechanisms that are themselves modulated by stimulus familiarity (Thompson et al., 2012).

Experience-dependent changes in neural activity are also well documented in the primary auditory cortices (A1) of mammals (for reviews, see: Dahmen and King,

2007; Weinberger, 2007). Both the long-term tonotopic arrangement across neurons and the receptive fields of single neurons in A1 tend to shift according to training with acoustic stimuli (Edeline and Weinberger, 1993; Recanzone et al., 1993). Furthermore, these training-induced effects can be replicated by pairing auditory stimuli with stimulation of the dopaminergic ventral tegmental area (Bao et al., 2001) or the cholinergic nucleus basalis (Metherate and Weinberger, 1990; Kilgard and Merzenich, 1998; McLin et al., 2002; Froemke et al., 2007), modulatory nuclei involved in mediating reward assessment and arousal/attention, respectively. Accordingly, when reward and arousal are modulated in awake animals performing operant tasks with auditory stimuli, task-related changes in auditory coding are observed in single neurons in the A1 of rats (Otazu et al., 2009), cats (Lee and Middlebrooks, 2011), and ferrets (Fritz et al., 2003, 2007), and more recently in the CMM of starlings (Knudsen and Gentner, 2013). In each instance, the behavioral modulation serves to highlight the features the animal needs to use while engaged in a task, or to decrease the representation of distracters.

It is likely that the long-term remodeling of receptive fields and population responses are brought about or even directly composed of these short-term task-dependent changes that take place as an animal learns stimulus-response-reward associations during training, but such a connection remains undescribed. To assess these relationships, and to observe the genesis of the experience dependent plasticity in relation to behaviorally relevant stimuli in the avian auditory system, we performed

experiments to monitor the changes in neural activity as starlings learned to classify novel auditory stimuli.

We trained birds to quickly classify novel acoustic material while recording from single neurons in CMM. We found evidence that neurons with different degrees of training history were represented differently across the population; stimuli with which a bird was very familiar tended to produce a weaker neural response than those with which a bird had less training. Additionally, by recording before, during, and after the learning period, we saw a decrease in the post-learning neural response for most neurons when compared to the pre-learning neural response. This decrease was larger for stimuli a bird had just learned and smaller for the stimuli with which the bird had extensive experience. The reduction in firing rate appears to be due, at least in part, to an adaptation of the neurons' responses to repeated presentations of a stimulus, and we found evidence that the rates of adaptation were different for different stimuli. Furthermore, the rate of adaptation to some stimuli varied as a function of whether the animal was actively engaged in learning the novel stimulus classifications, leading to the possibility that stimulus-specific adaptation may be an important mechanism for modifying neural responses in CMM, and that it may be under active top-down control.

We also observed task-dependent changes in firing rate similar to those recently described in neurons from the same auditory region in starlings (Knudsen and Gentner, 2013), and found evidence that some neurons appeared to develop these task engagement effects as a bird learned a novel set of stimulus classifications. That is, the

response to a given stimulus was sometimes seen to increase or decrease over the course of learning, and then to return to its pre-learning state once the animal was no longer actively engaged in the classification task.

These results demonstrate that CMM neurons modify their responses to stimuli as they learn to classify novel acoustic material, and do so differently for stimuli with different behavioral relevancies. We suggest a model of learning in which the responses of single neurons are altered in relation to their ability to encode relevant stimulus features, leading to a general suppression of most, irrelevant, neural activity and the active retention of a few discriminative features, but note that there is much to be done to test this general hypothesis. These initial forays into studying the physiological basis of auditory cognition and experience dependent plasticity in birds that are awake and engaged in a naturalistic behavior have done much to focus the scope of future studies and to suggest interesting mechanisms that may be involved in shaping the auditory system to better represent an animal's world.

METHODS

Subjects and Operant Shaping

Detailed subject housing, basic operant training, and neurophysiological recording methods are described in greater detail in (Knudsen and Gentner, 2013). Subjects were 6 European starlings (4 male, 2 female) wild-caught in Southern California between 2009 and 2011. Prior to experimentation, subjects were housed in

large mixed-sex aviaries with ad libitum access to food and water, and a naturalistic light-dark cycle that tracked sunrise and sunset times for San Diego, California.

At the start of experimentation, birds were moved to individual testing chambers that each consisted of a weld-wire cage suspended inside a sound-attenuating chamber. Chambers were illuminated by a broad-spectrum compact fluorescent light bulb ("house light") mounted above the wire cage, and a continuously running fan provided ventilation for each chamber. One wall of the wire cage contained an operant panel consisting of three response ports arranged horizontally into which the birds were trained to peck their beaks to indicate operant responses. Directly above the center response port was a green signal light emitting diode (LED) that was used to indicate trial availability and also served as a secondary reinforcer. Below the response ports was an opening into which a food reward could be presented by a solenoid-driven food hopper. Directly behind the operant panel was a single speaker through which we presented all auditory stimuli. Birds earned all their food by performing operant tasks, and had *ad-libitum* access to water throughout the entirety of training and testing.

Stimuli

Starlings sing their songs in bouts that can be described as a string of varying and repeated subunits of spectrotemporal structure, here termed "motifs" (Eens et al., 1989; Gentner, 2008). A bout generally lasts on the order of 30 to 90 seconds, with individual motifs having much shorter durations. An individual male starling sings a

great variety of motifs, and these generally do not overlap with the motifs sung by neighboring males (Eens et al., 1989). Stimuli in the current study consist of a library of 295 distinct single starling motifs that were excised from song bouts of 11 male starlings recorded in sound-attenuating chambers at 44.1Khz. Small periods of silence that occur between individual motifs were used to excise motifs from song bouts. Motifs were then high-pass filtered at 100Hz (4^{th} order Butterworth), down sampled to 40Khz (in order to be played back appropriately by our data acquisition system) scaled to 65dB mean, and ramped at both the beginning and end with 20msec linear ramps to avoid onset or offset clicks. Motifs were then grouped, per bird, into categories defined by similar spectrotemporal features and one representative motif exemplar was selected from each motif category from each bird, yielding 295 spectrotemporally distinct motifs with a mean duration of 0.83 seconds (standard deviation (SD): 0.24 seconds). None of the subjects' songs contributed to the motif library, and all subjects were naive to the stimuli in the motif library prior to experimentation.

Initial Training

Birds were acclimated to the operant apparatus and shaped by an automated procedure to peck into the response ports to obtain a food reward as described previously (Gentner 2008). Once birds were reliably pecking into the response ports, motifs were introduced as conditioned auditory stimuli in a two-alternative-choice (2AC) paradigm. For each bird, three motifs were picked at random from the motif library and were assigned to the association with the left response port, and three

different motifs were selected at random to be associated with the right response port. These six motifs make up the bird's initial training set and are known as the "over-trained" motifs in the learning task described below. All trials were self-initiated by a bird pecking its beak into the center response port, at which time one of the six training motifs would play from the speaker. The motif played on a given trial was chosen from the uniform distribution across all six motifs. At the conclusion of the motif playback, if a bird pecked its beak into the response port associated with the motif (for example, if it pecked left after a "left" motif), the green signal LED would blink rapidly for 0.5 seconds as a secondary reinforcer and the bird would obtain a brief access to the food reward. If it made an incorrect response (pecking the right port to a "left" motif, the house light would extinguish for 5-10 seconds as a mild punishment (timeout). Birds were given two seconds after the end of the motif playback to make a response into either the right or left port. If no such response was made, the trial was marked as a "no-response" trial, and neither reward nor punishment was delivered. At the end of the trial consequence (reward, punishment, or response window), there was an inter-trial interval of two seconds, during which time a new trial could not be initiated. The possibility of initiating a trial was marked by the steady illumination of the green signal LED. Birds lived in their training chambers and to ensure high levels of motivation they were only allowed to initiate trials in restricted time windows throughout the lights-on period of each day. When the lights turned on in the morning, trials were available (as denoted by the green signal LED) for some duration chosen at random (typically 30-60 minutes), after which there was a

subsequent period (duration again randomly chosen, typically 30-60 minutes) where the bird could not initiate trials. This cycle of trial availability and unavailability assured high levels of responding during trial availability and allowed us control over he animals' engagement in the task, especially during electrophysiological recording.

Multiple Training Set Task

Once a bird achieved a high level of performance on the initial over-trained motifs, we began introducing novel sets of training motifs (Figure 1). A novel training set consisted of an equal number of over-trained motifs chosen randomly without replacement from the six over-trained motifs (either one "left" and one "right" over-trained motif or two "left" and two "right" over-trained motifs) and two untrained motifs picked at random from the remainder of the motif library, and now termed the "newly-trained" motifs. Trials proceeded as before, except that the newly-trained motifs were four times as likely to be presented upon trial initiation as the over-trained motifs. For one of the newly-trained motifs, the bird was rewarded for pecks to the left response port and for the other motif, the bird was rewarded for pecks to the right response port. The inclusion of the over-trained motifs allowed us to monitor that birds were still performing the baseline categorization task as they gradually learned to correctly categorize the newly-trained motifs.

At the end of each trial, we determined the number of times each motif in the set had been correctly classified in its previous 20 presentations. If this number was greater than 15 (corresponding to ~97% confidence interval around chance) for all

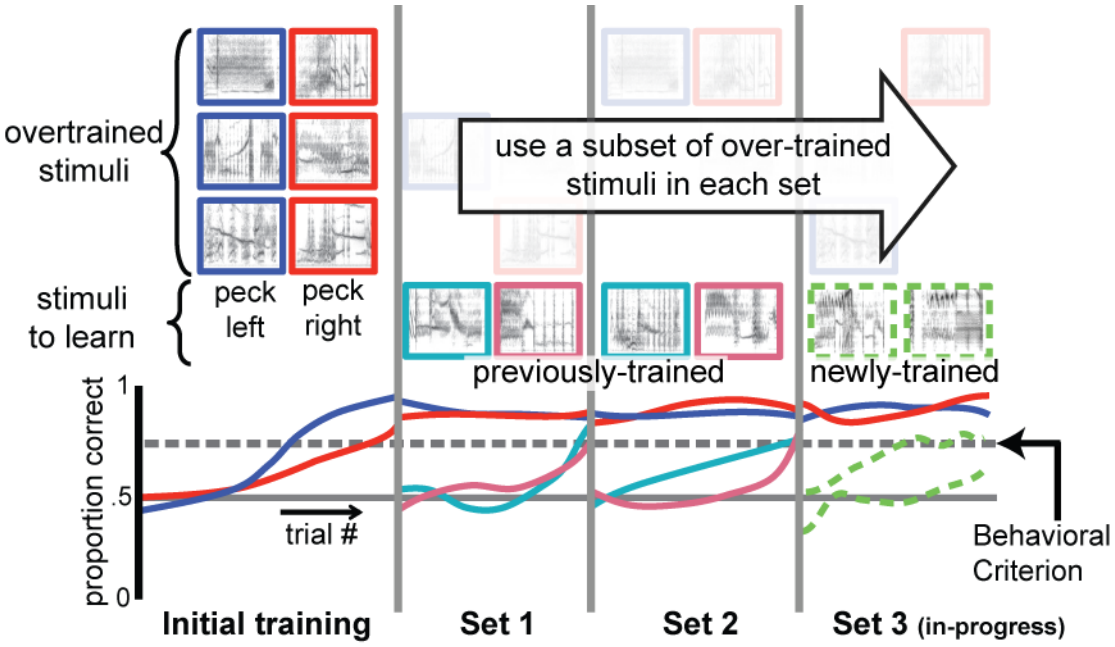


Figure 1: Starlings were trained to quickly classify novel auditory objects. Birds were trained on an initial training set of three left (blue, boxes outline spectrograms of training motifs) and three right (red) training motifs. Once a bird became proficient at classifying these initial motifs, a learning set was chosen that consisted of two of these initial motifs chosen at random (one left and one right; since two initial training motifs were included in every learning set and were seen many times, they were termed "over-trained"), and two untrained motifs (also one left and one right, termed "newly-learned" motifs, green dashes in set 3). Performance improved as the bird learned to correctly classify these new motifs until the bird's behavior reached the behavioral criterion for above-chance performance. At this point, the motifs in the current learning set were considered "previously-trained" (magenta and teal; sets 1 and 2), and were not used in future learning sets. A new set was chosen, and the bird continued to learn novel motifs in this fashion.

motifs, we considered the set to be “learned”. If the bird finished the lights-on period in a given day in the middle of learning a set, the check for 15 of 20 correct was restarted the next morning for each motif. Once the newly-trained motifs were learned, they were classified as “previously-trained” motifs for all future sets, and were not included in any other set. A new set was subsequently chosen, the bird learned to correctly classify the novel motifs while maintaining strong classification performance on the over-trained motifs, and the process continued this way, with birds serially learning to recognize and correctly classify pairs of novel motifs.

Birds generally took fewer trials to learn later sets than initial sets (Figure 2, and results), as they acquired the larger learning set task along with correctly classifying individual motifs in each set. At this point, birds were moved to a different operant chamber that was fitted for electrophysiological recording (Knudsen Gentner 2013 for details). After a brief acclimation period where they continued to perform the learning set task, birds were implanted with a chronic microdrive and neural recording began.

Electrophysiology

Behavioral electrophysiology proceeded as described in (Knudsen Gentner, 2013). Briefly, birds were anesthetized with isoflurane and a custom microdrive containing a 16-channel neuronexus microelectrode array was attached to their skull over a craniotomy that allowed access to the auditory cortex. The dura was resected and replaced by a silicone gel that acted as an artificial dura (Jackson and

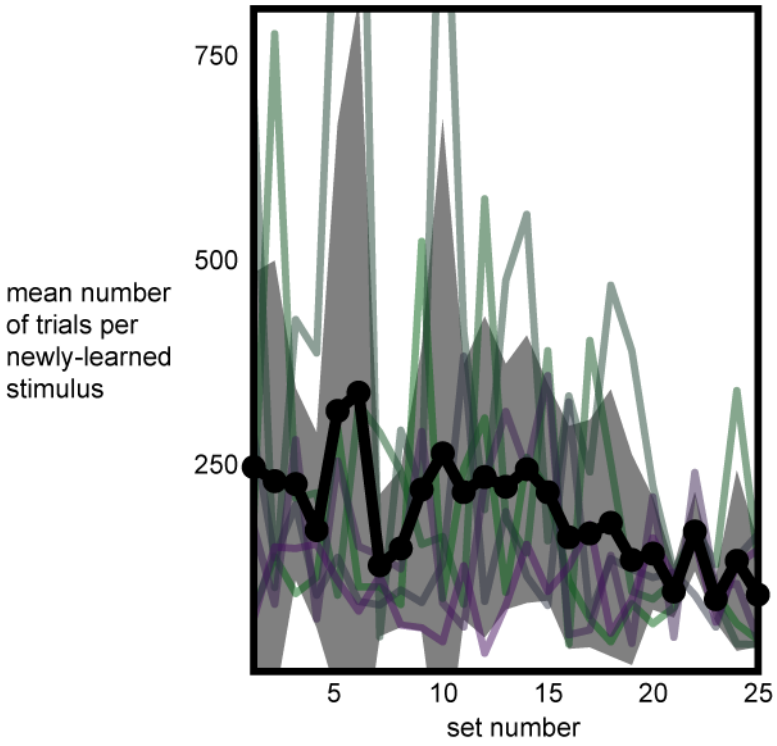


Figure 2: Birds required fewer trials to learn novel acoustic stimuli as training progressed. Each colored line depicts the mean number of trials of a newly-learned motif performed by a single bird over its first 25 learning sets. The thick black line is the mean of these values across the six birds from which neural data was collected in this study. There is a significant negative correlation (Pearson's $r = -0.23$, $p = 0.005$) between the number of trials it took to learn the novel motifs and the serial number of the training set, indicating that birds learned newly-trained motifs with fewer trials in later sets.

Muthuswamy, 2008). The electrode array was coated with fluorescent dye (DiI) to facilitate histological recovery of the track (DiCarlo et al., 1996), and then positioned to target the medial portion of the caudal mesopallium (CMM, target coordinates: 2500um rostral and 500um lateral to the bifurcation of the Y sinus; all implants targeted the left hemisphere). The electrode array was advanced to a depth of approximately 1000-1300um so that the electrode pads were situated within the boundaries of CMM.

The electrode was attached to a headstage (HST/16V-G20, Plexon) mounted to the microdrive implant that provided 20x gain and this was in turn attached, via a tether, to a motorized commutator installed in the roof of the soundproof chamber (Plexon, Dallas, TX). The signal was passed to a 16-channel amplifier where the signal was amplified 2,000x-10,000x (model 3600 16-channel microelectrode amplifier, A-M Systems, Sequim, WA), and bandpass filtered between 500Hz-5kHz. The output of this amplifier was fed to a Power 1401 analog to digital converter (Cambridge Electronic Designs (CED)), where the signal was digitized at 25-32kHz and saved via CED Spike2 software. This device also allowed us to control the behavioral apparatus via custom-written Spike2 scripts, and to maintain the proper temporal alignment of neural and behavioral data.

Recording Procedure

On a given recording day (Figure 3), experiments proceeded by first advancing the electrode array in small steps (~35um) until one or more single units could be

isolated from the background noise. A panel of motifs was then presented while the operant apparatus was inactive (the green signal LED was not illuminated) and the bird listened passively as a number of repeats for each motif was presented (generally 10-40) to determine the neurons' responses to a variety of motifs (Figure 3, Non-Engaged-Pre-Learning condition). The panel of motifs consisted of all six of the bird's over-trained motifs, some number of previously-trained motifs (usually three "left" and three "right"), and some number of novel motifs from the motif library, which the bird had never heard in an operant context before (usually six), and one five-second-long "silent" stimulus. Stimuli were presented in pseudo-random fashion, with a 1-5 second inter-trial interval.

After this, a training set was created by selecting two of the over-trained (one 'left' and one 'right' motif) and two of the novel motifs at random to be the newly-trained motifs (one was assigned to the "left" response, and the other to the "right" response). The apparatus was activated, and the bird performed trials as during training while neural responses were recorded (Figure 3, Engaged-Pre-Learning condition). Recording proceeded until the bird learned the set, or the single unit neural activity was no longer separable from background noise. The behavioral criterion for having learned a set was the same as during the pre-physiological recording (15 of the previous 20 trials correct for each motif in the set). After the bird learned the new set (Figure 3, Engaged-Post-Learning), it was allowed to perform more trials at a high level of performance so that we could collect post-learning neural responses. The operant apparatus was then again deactivated, and the responses to multiple repeats of

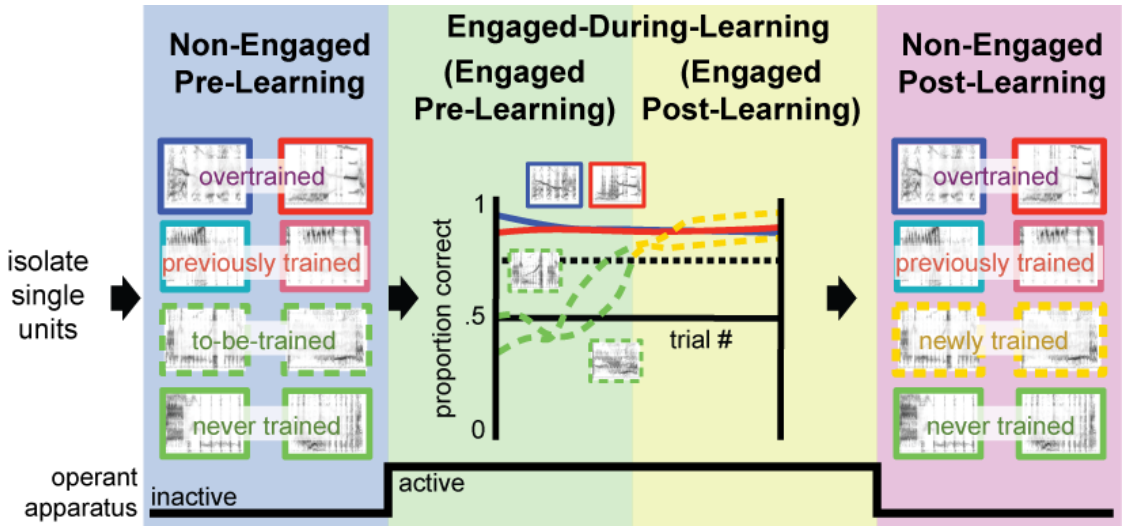


Figure 3: Procedure for neural recording while birds performed the learning task. Time proceeds from left to right. Experiments began by isolating one or more auditory single units. Then, a panel of motifs was presented to the bird while the operant apparatus was disabled and the bird listened in quiet wakefulness during the Non-Engaged-Pre-Learning recording condition. The apparatus was then activated and two of the over-trained and two of the untrained stimuli were presented in the Engaged-During-Learning context. The bird was initially unfamiliar with the novel motifs in the Engaged-Pre-Learning condition, but soon reached the behavioral criterion for above-chance performance and continued performing the task in the Engaged-Post-Learning condition. The apparatus was then again deactivated and the same motifs that were presented in the Non-Engaged-Pre-Learning condition were again presented, this time in the Non-Engaged-Post-Learning condition to assess the effects of learning on the response of the simultaneously recorded neurons.

the entire panel of motifs were again recorded while the bird passively listened (Figure 3, Non-Engaged-Post-Learning). If the single unit was still separable, we attempted to run through the entire procedure again with a new panel of motifs and a new training set. Each instance of recording the activity of a single unit before, during, and after a bird successfully learned a set is referred to as a "learning instance" in the remainder of this document.

Data Analysis

Histology

At the conclusion of recording from each subject, the bird was deeply anesthetized with nembutal and perfused with heparinized saline and then 10% formalin in order to fix brain tissue. The brain was removed from the skull, cryoprotected in a 30% sucrose solution, and then 50um coronal sections were made on a freezing microtome. Slices were mounted on slides, allowed to dry, and then photographed at 250x in both bright field and fluorescent illumination to observe the fluorescent dye marking the electrode tracks. Sections were then Nissl-stained, and photographed again. The fluorescent and Nissl stained images were aligned and tracks were reconstructed to ensure that all single unit recordings were made in CMM. Single unit recordings were excluded from analysis if they did not fall within the histological boundaries of CMM.

Spike sorting and data organization

Single unit activity isolation from background noise was determined on the basis of shape-based sorting using the built-in spike sorting features in Spike2, following the procedures outlined in (Knudsen and Gentner, 2013). Only well-isolated single units were included in the analyses described below. Once spikes were sorted, spike times and all related behavioral and neural data were stored in a custom-designed postgreSQL database that was accessed by custom-written MATLAB scripts, which was used for all analysis unless otherwise noted.

Firing rates and significant responses

Spontaneous firing rates are reported in spikes/second (Hz) and were calculated for each trial from the end of the trial consequence until the end of recording for that trial, or until two seconds before the following trial, whichever came first.

Driven firing rates were calculated for each trial by counting the number of spikes that occurred between the start and end of the motif presented on that trial and dividing by the duration of the motif and are reported as spikes/second (Hz). Individual trial firing rates were averaged across all presentations of a given motif within a given recording condition (e.g., Non-Engaged-Pre-Learning) to get a neuron's mean firing rate for each motif in each recording condition.

For each motif presented to a neuron in a given recording condition, we determined whether the neuron had a significant response to this motif in this

condition by comparing the distributions of firing rates across all presentations of the motif to the distribution of spontaneous firing rates observed during the condition with the Mann-Whitney U-test ($\alpha = 0.05$).

Motif class differences before learning

To determine whether there were differences in the driven response to different classes of motifs based on a bird's long-term training history, we compared the mean firing rates elicited by classes of motifs with different training histories in the Non-Engaged-Pre-Learning recording condition. We first calculated the mean firing rate to each motif and then transformed these values to Z-scores to allow for comparisons across neurons. We averaged these Z-scores across motifs with similar training histories, leading to the mean normalized response to each of the following three motif classes: over-trained, previously-trained, and untrained. We then performed a one-way ANOVA across neurons (normalized firing rate by motif class, $\alpha = 0.05$). In order to focus on the long-term changes, if a neuron was recorded across multiple learning instances in one recording session we only included data from the first learning set for each neuron. Single-neuron analyses were performed similarly, except that the firing rates were not first transformed to Z-scores, and analyses were done on the distributions of firing rates to individual trials grouped by motif class, instead of mean responses to each motif. Post-hoc comparisons between groups in the single-neuron analyses were made with Tukey's honestly significant differences test (Tukey's HSD).

Population-level changes in firing rate across learning

To determine whether learning led to changes in neurons' firing rates in response to motifs with different training histories, we compared the distribution of firing rates recorded in the Non-Engaged-Pre-Learning condition with those in the Non-Engaged-Post-Learning condition. For each learning instance, we calculated a mean firing rate for each motif in both the Non-Engaged-Pre-Learning and Non-Engaged-Post-Learning conditions. We then calculated the Z-scores for these estimates to allow for comparisons across neurons, and then subtracted the normalized pre-learning value from the normalized post-learning value for each motif. We then averaged these within-motif differences into groups of motifs with the same training history yielding four estimates of the difference between the post- and pre-learning normalized firing rate per neuron: one each for the over-trained, previously-trained, newly-trained, and untrained motif groups. These values were compared across learning instances with a one-way ANOVA (normalized response difference X motif class, $\alpha = 0.05$). Post-hoc comparisons were made with Tukey's HSD. We tested whether or not each class of motifs had a change significantly different from zero with a *t-test* ($\alpha = 0.05$) of deviation from the standard normal distribution.

Individual changes in firing rate across learning

In addition to looking at population-level effects in differences across learning, we analyzed changes across individual learning instances. For each motif in each learning instance we compared the distribution of firing rates recorded during the

Non-Engaged-Pre-Learning condition to those recorded during the Non-Engaged-Post-Learning condition with the Mann-Whitney U-test ($\alpha = 0.05$). If the distributions were significantly different, we subtracted the median Pre-Learning firing rate from the median Post-Learning firing rate to label the difference as either a significant increase or a significant decrease in the neuron's response to that motif.

Task-engagement dependent changes in firing rate

To assess task-engagement dependent changes in firing rate, and whether or not they were stable across learning, we compared responses to each motif recorded across each learning instance with a one-way ANOVA on the distribution of firing rates recorded in each of the different behavioral conditions (Non-Engaged-Pre-Learning, Engaged-Pre-Learning, Engaged-Post-Learning, Non-Engaged-Post-Learning). We necessarily restricted the motifs included in the model to those presented in the Engaged-During-Learning behavioral condition, and to those that showed significantly driven firing rates in both the Non-Engaged-Pre-Learning and the Engaged-During-Learning condition. If the ANOVA showed a main effect of behavioral condition, we performed a post-hoc multiple comparison test (Tukey's HSD, $\alpha = 0.05$) to determine across which behavioral conditions the firing rate of the neuron differed. Similar results were obtained by assessing differences with the nonparametric Kruskal-Wallis test instead of using an ANOVA.

To decide whether a neuron developed or lost a particular task engagement effect in response to a particular motif, we first required that the firing rate measured

in the Non-Engaged-Pre-Learning condition was not significantly different from the distribution recorded in the Non-Engaged-Post-Learning condition; that the response is the same in both Non-Engaged epochs is a necessary prerequisite for having a task-engagement-dependent change.

If the Engaged-Pre-Learning firing rate was significantly different from the Non-Engaged-Pre-Learning firing rate, and the direction of the difference between the mean Engaged-Pre-Learning firing rate to the mean Non-Engaged-Post-Learning firing rate was the same as the direction of the difference between the Engaged-Pre-Learning to the Non-Engaged-Pre-Learning (for example, if the firing rate increased with task engagement, this would require it to go back down when the bird was again in the non-engaged condition), we determined that there was already an effect of task engagement on the firing rate in response to the given motif. If the firing rate in the Engaged-Post-Learning condition was significantly different than the firing rate in the Engaged-Pre-Learning condition, and in the same direction as the Non-Engaged conditions relative to the Engaged-Pre-Learning firing rate, then we determined the neuron to have lost a significant task engagement effect across the learning condition.

If the Engaged-Pre-Learning firing rate was not significantly different from the Non-Engaged-Pre-Learning firing rate, we determined that there was no pre-existing effect of task engagement on the response of the neuron to the given motif, and further asked whether or not such an effect might have developed over the course of learning. If the Engaged-Post-Learning firing rate was significantly different from both the Non-Engaged-Pre-Learning and the Engaged-Pre-Learning firing rate, and the mean

difference was in the same direction relative to these distributions and to the Non-Engaged-Post-Learning firing rate, we determined that the neuron had gained an effect of task engagement over the course of the learning set.

Stimulus adaptation

We assessed the effects of repeated presentations of a motif in the Non-Engaged-Pre-Learning condition by using the statistical package JMP (version 10) to perform a least-squares indicator variable linear regression analysis for each neuron. We modeled firing rate as a function of motif presentation number and included motif identity as a categorical indicator variable, allowing us to assess the main effects of motif presentation number and motif identity, as well as the interaction between these two factors. The analysis fits a line for each motif to the relationship between firing rate and presentation number, and allows us to determine main effects of motif identity, presentation number, and the interaction between the two. Neurons were considered to show significant adaptation to a particular motif if the 95% confidence interval around the slope parameter for the linear fit to the motif did not include 0.

We performed a similar analysis to determine whether the rate of adaptation was different across the Non-Engaged-Pre-Learning and Engaged-During-Learning conditions by again using JMP to perform a least-squares indicator variable linear regression analysis on each motif in each learning instance. We modeled firing rate as a function of motif presentation number and included behavioral condition as a categorical indicator variable. We necessarily restricted the motifs included in the

model to those presented in the Engaged-During-Learning behavioral condition, and to those that showed significantly driven firing rates in both the Non-Engaged-Pre-Learning and the Engaged-During-Learning condition.

RESULTS

Birds learned to quickly classify novel motifs

Birds rapidly acquired the overall training set task. Over the first 25 sets across all birds, there was a significant negative correlation between the set number and the number of presentations of the newly-trained motifs required to learn the set (Pearson's $r = -0.23$, $p = 0.005$; Figure 2). We recorded a total of 80 well-isolated single units in CMM in response to the Non-Engaged-Pre-Learning panel of motifs. Of these 80 neurons, we were able to maintain isolation over the course of birds learning newly-trained motifs for 44 neurons. Five of these neurons were held through more than one learning instance, leading to a total of 54 successful learning instances across 44 neurons.

Neurons respond differently to motifs with different training histories.

Birds had different levels of experience with different motifs; over-trained motifs were represented in every set a bird learned, previously-trained motifs were learned at one point, but only ever in one set, and untrained motifs were never presented in an operant context. In order to determine whether there were any long-term changes in the population response to different classes of motifs in the learning

set task, we characterized the neural responses in CM with respect to the difference in motif training history. The over-trained motifs had lower normalized driven responses (mean Z-scored response, see methods, -0.10 (SD 0.46)) than the previously-trained (mean 0.06 (SD 0.4)) or untrained (mean 0.06 (SD 0.46)) motifs (Figure 4). Though there was considerable variance, there was a main effect of motif history on firing rate across the three groups (one-way ANOVA, $F_{(2,237)} = 3.20$, $p = 0.04$).

These population-level measures averaged over much neuron-to-neuron variability. When we looked for differences between firing rates elicited by different motif categories within a given neuron, we found that 15/80 neurons showed a significant main effect of motif class on firing rate (one-way ANOVA for each neuron, $p < 0.05$). Post-hoc analyses on these 15 neurons allowed us to determine which groups elicited significantly different firing rates. Over-trained motifs elicited higher firing rates than either the previously-trained or untrained motifs in only 4/15 neurons, while having significantly lower firing rates in 10/15 neurons. Previously-trained motifs had higher driven rates in 4/15 neurons, and lower rates in 5/15 neurons; and untrained motifs showed higher rates in 8/15 neurons, and lower rates in 1/5 neurons. While there is great heterogeneity, there is a trend for training to lead to lower driven firing rates at both the population and single-neuron level.

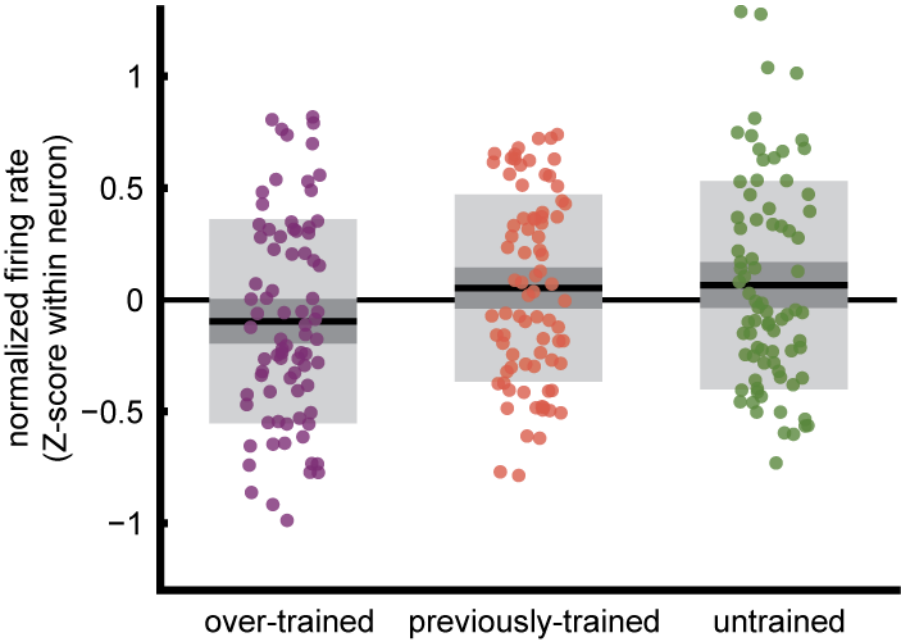


Figure 4: Training history affects the population response to motifs. The figure depicts the normalized response rate of 80 neurons recorded during the Non-Engaged-Pre-Learning behavioral condition, grouped by motif class within neuron. Each dot represents the mean normalized response of a single neuron in the motif classes based on a bird's experience (over-trained, previously-trained, untrained). A one-way ANOVA shows a significant main effect of motif class (see text; light grey shading: one SD; dark grey shading: SEM; horizontal black line: mean).

Firing rates decrease after learning; the largest decrease is observed for the newly-trained motifs

To determine whether or not learning altered driven neural responses across the learning instance, and whether these differences depended on the training history of the motif, we compared driven firing rates measured before and after learning by subtracting the firing rates measured during the Non-Engaged-Pre-Learning period from those measured during the Non-Engaged-Post-Learning period. A positive difference suggests that neurons increased their driven responses to motifs after learning, while a negative value indicates that responses were suppressed relative to pre-learning rates. By normalizing within-neuron and grouping responses from motifs with similar training histories, we were able to compare these differences across the set of learning instances (Figure 5). Driven firing rates tended to decrease for all classes of motifs over the course of learning, but only the previously-trained and newly-trained motifs showed a within-class decrease in firing rate that was significantly lower than zero (*t-test*; $p = 0.045$, previously-trained; $p = 0.0001$, newly-trained), indicating a significant change between the Non-Engaged-Pre-Learning and Non-Engaged-Post-Learning responses to these motif classes. The change in firing rate was significantly modified by the motif training history (one-way ANOVA, main effect of motif training history, $F_{(3,210)} = 2.78$, $p = 0.04$). Post-hoc comparisons showed that the newly-trained motifs showed a significantly greater decrease in firing rate across learning than did the over-trained motifs (Figure 5, asterisk).

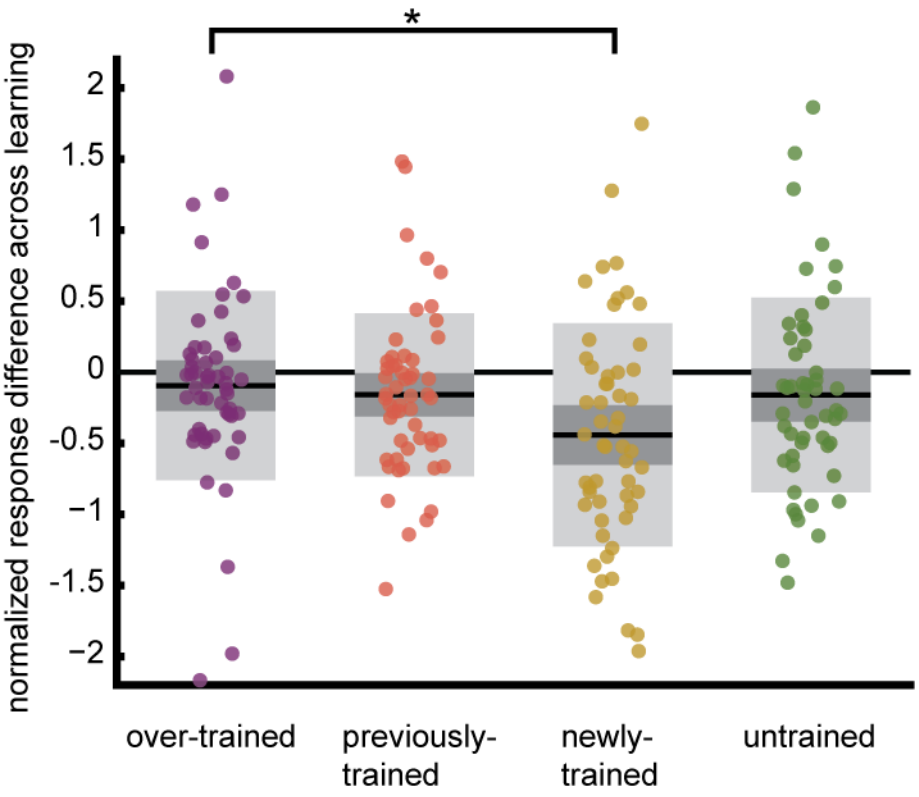


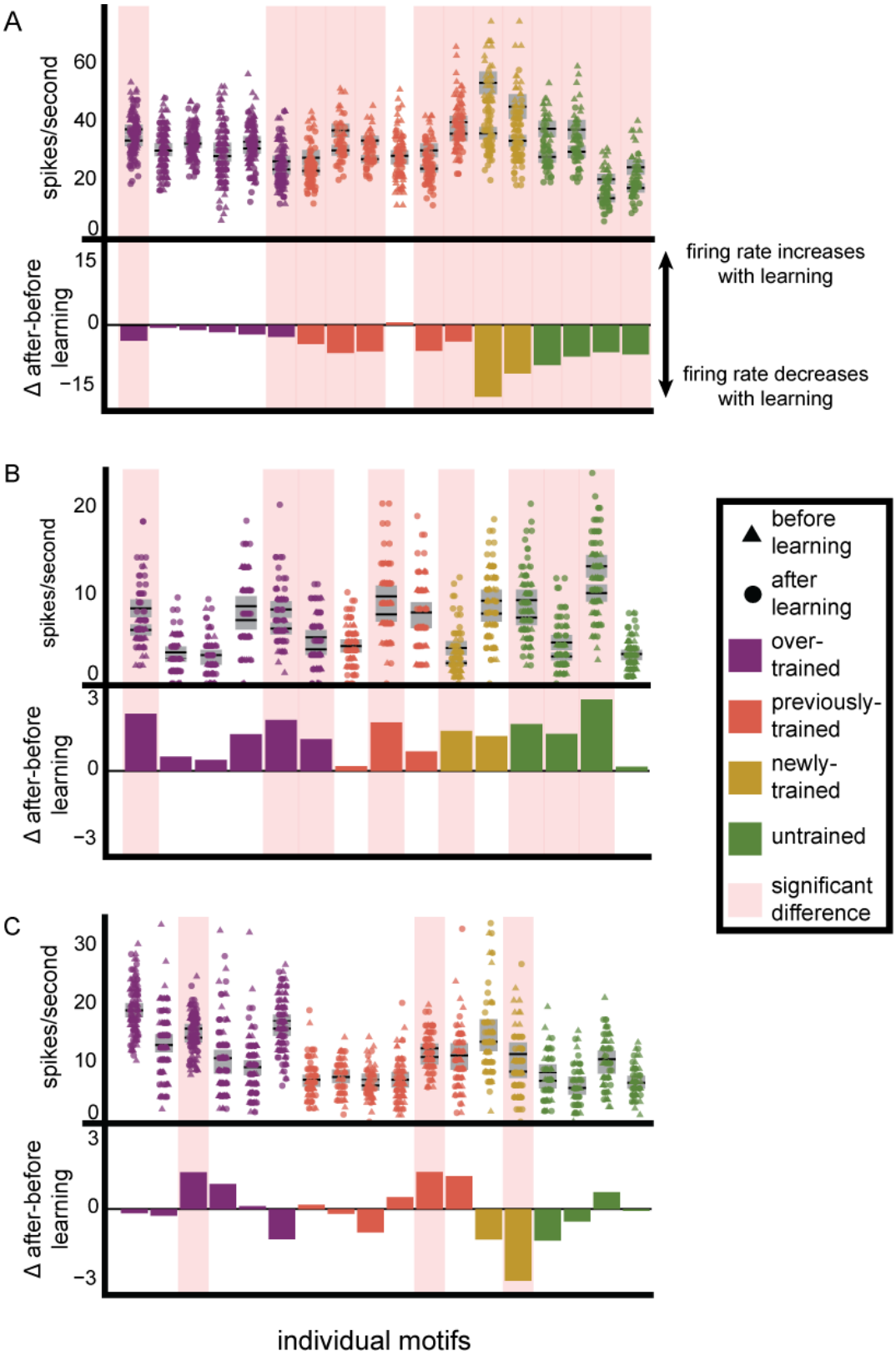
Figure 5: Neurons show decreased responsiveness to motifs after learning, and show a greater decrease for newly-trained motifs. The figure shows the normalized response rate difference across learning for 54 learning instances in 44 neurons, grouped by motif class within neuron. Each dot represents the mean normalized response difference between the Non-Engaged-Post-Learning and the Non-Engaged-Pre-Learning condition for a single neuron in a given learning instance. Responses are grouped by the motif classes based on a bird's experience (over-trained, previously-trained, newly-trained, untrained). A one-way ANOVA shows a significant main effect of motif class, and a post-hoc comparison shows that the newly-trained motifs show a significantly larger decrease across learning than do the over-trained motifs (black line with asterisk: significant difference, see main text. Light grey shading: one SD; dark grey shading: SEM; horizontal black line: mean).

Individual neurons display heterogeneous changes to neural responses across learning

While the population-level effect on the change in firing rate between the Non-Engaged-Pre-Learning and Non-Engaged-Post-Learning suggested a population that decreased its driven response to all motifs, there was a great deal of variability in the direction of change for individual motifs recorded in a given learning instance. In 51/54 learning instances, we found that there was at least one motif that had a significantly different driven firing rate between the Non-Engaged-Pre-Learning and Non-Engaged-Post-Learning behavioral conditions. In 27 (53%) of these instances, all significant changes were decreases, mirroring the population effect of seeing a general decrease in firing rate with learning (e.g., Figure 6A). However, we observed 9 (18%) instances in which all of the significant differences were increases in firing rate with learning (e.g., Figure 6B), and 15 (29%) in which there were both significant increases and significant decreases observed for different motifs (e.g., Figure 6C). This suggests that though the population level effect was to decrease driven firing rates over the course of learning, neurons could show both increases and decreases to individual motifs, suggesting a more complex modification of response rates due to learning than the simple decrease suggested by the population-level analyses.

In addition to changes observed between the Non-Engaged-Pre-Learning and Non-Engaged-Post-Learning conditions, many neurons showed interesting patterns of alterations to their time varying firing rate in response to a given motif while the

Figure 6: Neurons show heterogeneous changes in response to motifs across learning. (A) depicts a learning instance for which all significant changes across learning were decreases. The top graph depicts the distribution of firing rates (spikes/second) across trials both before learning during the Non-Engaged-Pre-Learning condition (triangles) and after learning during the Non-Engaged-Post-Learning condition (circles). Each distribution is summarized by a black line (mean) and gray box (SEM). Each column represents a different motif, colored by the class to which they belong. Light red shading in the background of a motif column denotes that the distribution of firing rates recorded before learning is significantly different than the distribution recorded after learning (Mann-Whitney U test, $p < 0.05$). The lower half of panel a shows the difference in the mean firing rate (Non-Engaged-Post-Learning minus Non-Engaged-Pre-Learning). For this learning instance, all motifs that show a significant difference cause the neuron to respond less after learning. (B) and (C) are arranged similarly, but depict learning instances for which all significant differences are increases (B), or for which there are both significant increases and significant decreases (C).

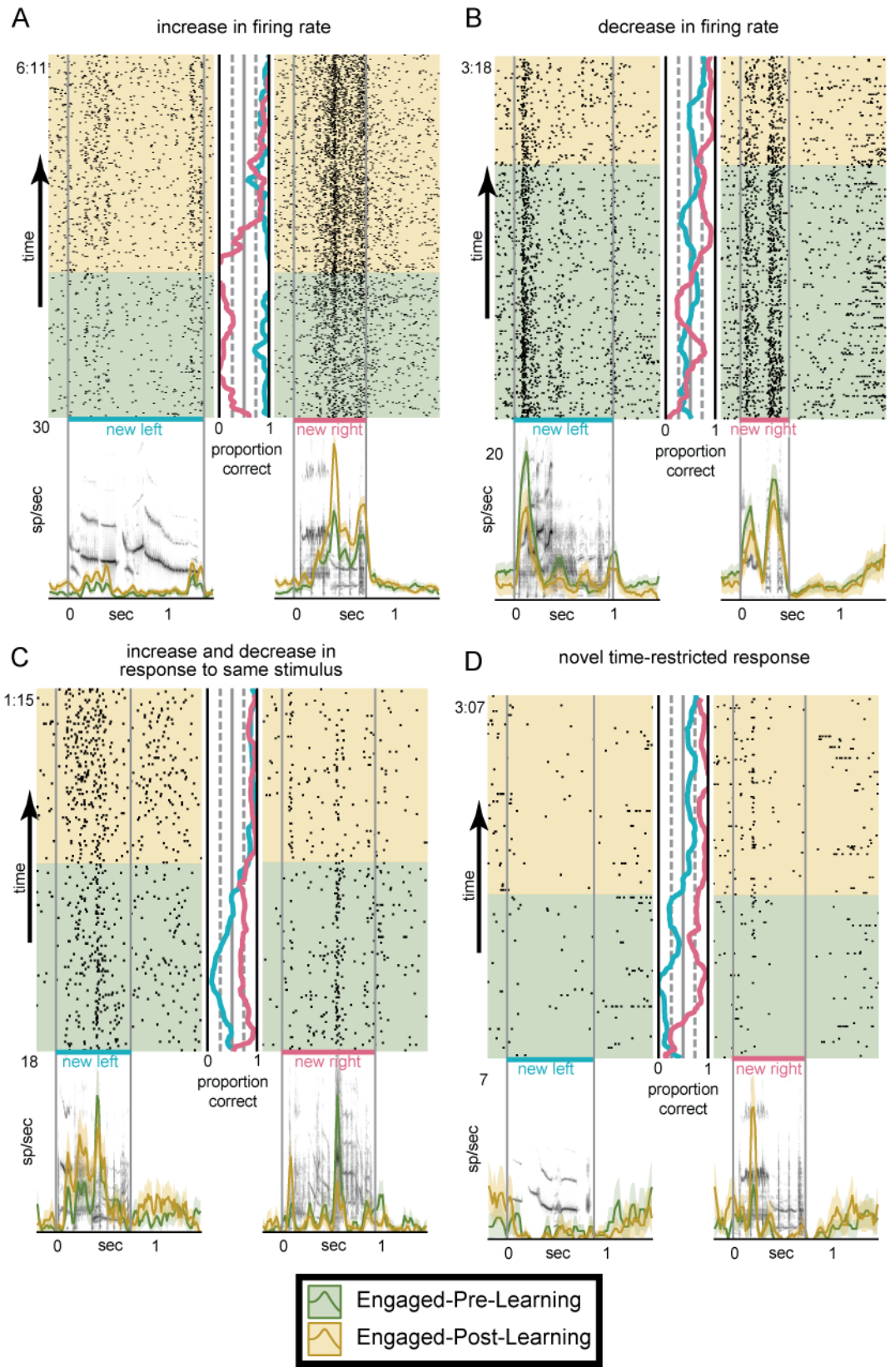

A
spikes/second
60
40
20
0
Δ after-before learning
15
0
−15
firing rate increases with learning
firing rate decreases with learning
B
spikes/second
20
10
0
Δ after-before learning
3
0
−3
before learning
after learning
over-trained
previously-trained
newly-trained
untrained
significant difference
C
spikes/second
30
20
10
0
Δ after-before learning
3
0
−3
individual motifs

bird was actively engaged in learning the proper classification of the newly-trained motifs in the Engaged-During-Learning condition. We observed neurons that both increased (Figure 7A) and decreased (Figure 7B) their firing rate to certain motifs across the Engaged-Pre-Learning to Engaged-Post-Learning boundary, and neurons that increased their firing rate to one part of a motif while decreasing it to another portion of the same motif (Figure 7C, right motif). Interestingly, we also observe neurons that gain a response to a motif for which there was previously none (Figure 7D).

Task-engagement dependent changes in firing rate develop over the course of learning

We noticed that, although there were changes in firing rate between the pre and post learning conditions, and that often responses appeared to change during the learning period itself, it did not appear that there was always a smooth transition from the response observed in the Non-Engaged-Pre-Learning condition via the change observed in the Engaged-During-Learning condition to the response observed in the Non-Engaged-Post-Learning condition. Instead, there were often discontinuities in the responses observed between the non-engaged and engaged task conditions that reminded us a great deal of the changes in firing rate due to task engagement we recently reported in recordings made from CMM neurons in birds performing similar, well-learned classification tasks (Knudsen and Gentner, 2013). Some of these firing rate differences between task-engaged and non-engaged recording conditions appeared

Figure 7: Neurons show temporally restricted changes in firing rate that correlate with learning novel object discriminations. Each panel depicts the response of a neuron as the bird learns to correctly classify the newly-trained left (left side of panel, teal) and the newly-trained right (right side of panel, magenta) motifs in the Engaged-Pre-Learning (green shading) and Engaged-Post-Learning (yellow shading) behavioral conditions. For each motif, the top of the panel plots the spike rasters for each trial, where each row is a single trial, and the presence of a dot on the row is the time of a recorded action potential. Earlier trials are toward the bottom, and the total recording time is plotted to the left of the topmost trial (reported in hours:minutes). Between the rasters for the two motifs is a record of the bird's behavior during recording and reports the proportion of trials that were answered correctly in the previous 20 trials. The solid grey line denotes chance, and the dashed grey lines are the 95% confidence interval around chance. Below the rasters are the peri-stimulus time histograms (PSTH) of the response of the neuron to both the Engaged-Pre-Learning trials (green, shading denotes SEM) and the Engaged-Post-Learning trials (yellow). Behind the PSTHs is the spectrogram of the respective motif. (A) a neuron that tended to increase its response during the learning period, (B) a neuron that tended to decrease its response, (C) a neuron that increased its response to the left motif, and showed an interesting pattern of increasing its response with learning toward the beginning of the motif, and decreasing its response later in the motif, (D) a neuron that had no regions of responsiveness before learning, but gained a strong peaked and time-restricted response after learning.


A
increase in firing rate
6:11
time
30
new left
new right
0
1
proportion
correct
sp/sec
0
sec
1
B
decrease in firing rate
3:18
time
20
new left
new right
0
1
proportion
correct
sp/sec
0
sec
1
C
increase and decrease in
response to same stimulus
1:15
time
18
new left
new right
0
1
proportion
correct
sp/sec
0
sec
1
D
novel time-restricted response
3:07
time
7
new left
new right
0
1
proportion
correct
sp/sec
0
sec
1
Engaged-Pre-Learning
Engaged-Post-Learning

to be present only after successful learning, while others that were already present appeared to diminish with learning.

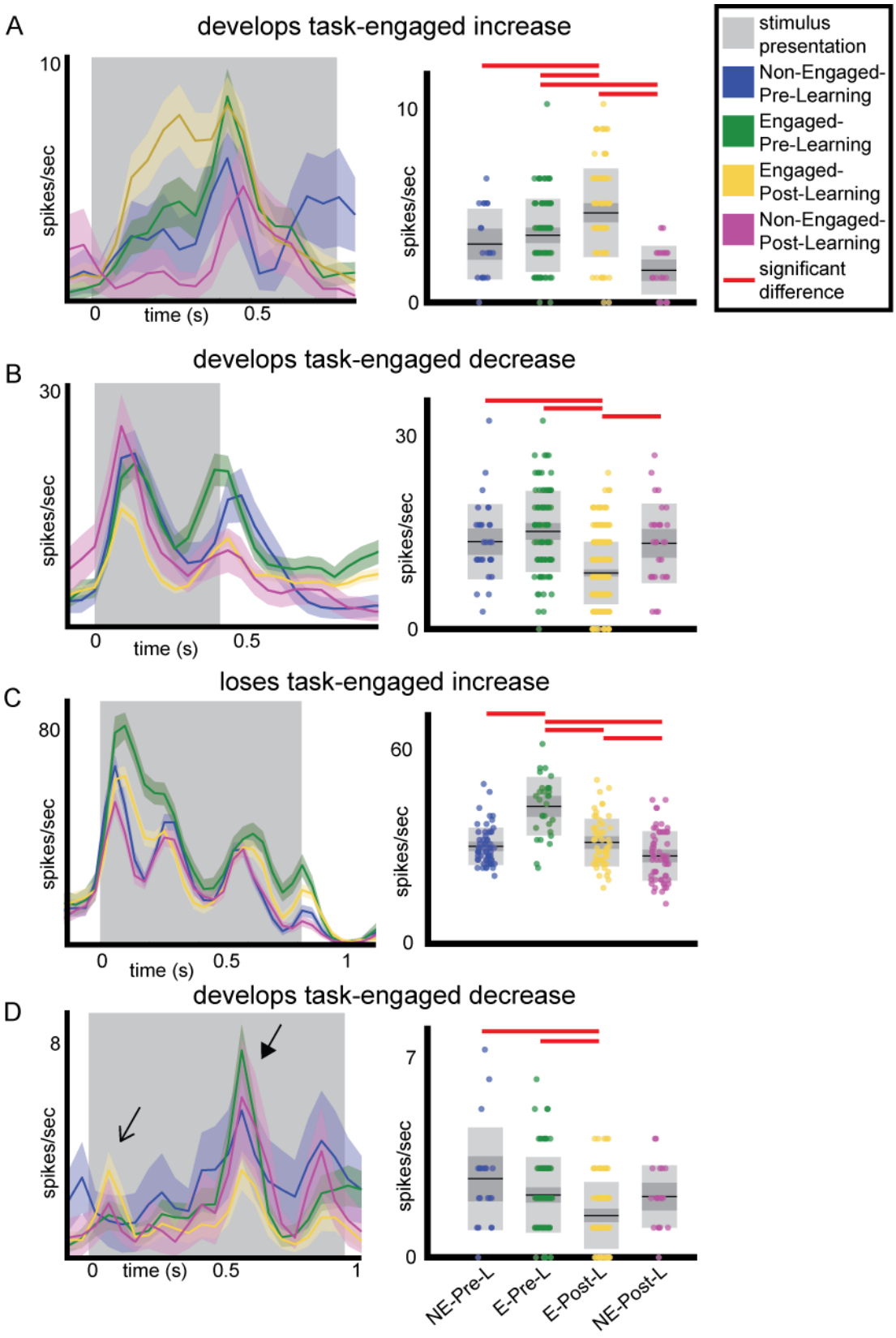
To determine whether neurons gained or lost differences due to task-engagement in their responses to a given motif over the course of learning, we compared the distribution of firing rates recorded from a neuron across the four behavioral conditions (Non-Engaged-Pre-Learning, Engaged-Pre-Learning, Engaged-Post-Learning, Non-Engaged-Post-Learning). We assessed these differences in each significantly driven motif in each learning instance (N = 205) by comparing the firing rate of the neuron across behavioral conditions with an ANOVA and then, if the ANOVA showed a significant main effect of behavioral condition, we compared significant differences between groups, looking for patterns consistent with a gain or loss of a task-dependent alteration in firing rate. To determine that a neuron gained a task-engagement-dependent effect, we looked for responses that were similar in the Non-Engaged-Pre-Learning, Engaged-Pre-Learning, and Non-Engaged-Post-Learning conditions and different in the Engaged-Post-Learning condition. Likewise, to determine that a neuron had lost a task-engagement-dependent effect, we looked for responses that were similar in the Non-Engaged-Pre-Learning, Engaged-Post-Learning, and Non-Engaged-Post-Learning conditions and different in the Engaged-Pre-Learning condition (see Methods: “Task-engagement dependent changes in firing rate” for the full logic).

This analysis revealed 11 instances of the gain of a task-engagement-dependent effect over the course of a learning instance. The development of firing rate

changes as a function of task engagement occurred as both increases in the response to the motif (e.g., Figure 8A), and decreases (e.g., Figure 8B,D). Interestingly, the distribution of the classes of motifs that developed these effects (newly-trained, 9; over-trained, 2) was significantly different than if one expected to see equal numbers of newly-trained and over-trained motifs developing such changes $X^2(1, N = 11) = 4.46$, $p = 0.035$. In addition to gains, we observed four instances of a loss of a task-engagement-dependent firing rate changes in firing rate (over-trained, 1; newly-trained, 3; e.g., Figure 8C).

Interestingly, we also see temporally restricted regions within a given motif response in which task engagement effects appear to be developing in different directions. In Figure 8D, we see that there is a region toward the beginning of the motif (marked by the open arrowhead) that elicits a similar response in both Non-Engaged conditions, and the Engaged-Pre-Learning condition that appears to show an increased response during task engagement after learning. In the neuron's response to the same motif (closed arrowhead) we again see similar response strength for both Non-Engaged conditions as well as the Engaged-Pre-Learning condition, which decreases after the bird has learned and is still performing the task. That is, the first portion of the motif (open arrowhead) does not elicit a response to the neuron until after the bird is successfully performing the classification task; the second portion (closed arrowhead) elicits a lower response after learning, and the neuron returns to its

Figure 8: Neurons develop task-engagement dependent effects over the course of learning. Each panel (A-D) depicts the response of a single neuron to a particular motif over the course of a learning instance. Left: mean (colored shading, SEM) time varying firing rate of the neuron in each of the four behavioral conditions (blue: Non-Engaged-Pre-Learning; green: Engaged-Pre-Learning; yellow: Engaged-Post-Learning; magenta: Non-Engaged-Post-Learning) in response to the motif (grey shading). Right: distributions of the firing rate for each trial, measured across the motif duration, and grouped by behavioral condition (light grey shading, SD; dark grey shading, SEM; horizontal black line, mean). Red lines show significant differences between groups (post-hoc Tukey's HSD after showing significant main effect of behavioral class by one-way ANOVA). Some neurons developed task-engagement dependent increases in their firing rate (A), some decreases (B), and others lost their task-engagement dependent effects (C). In (D), a task-engagement-dependent decrease forms with learning, but it appears that this may be due to the development of two different effects: a time-restricted task-engagement dependent increase (open arrow), and a larger task-engagement-dependent decrease (closed arrow).


A
develops task-engaged increase
10
spikes/sec
0
time (s)
0.5
10
spikes/sec
0
stimulus presentation
Non-Engaged-Pre-Learning
Engaged-Pre-Learning
Engaged-Post-Learning
Non-Engaged-Post-Learning
significant difference
B
develops task-engaged decrease
30
spikes/sec
0
time (s)
0.5
30
spikes/sec
0
C
loses task-engaged increase
80
spikes/sec
0
time (s)
0.5
1
60
spikes/sec
0
D
develops task-engaged decrease
8
spikes/sec
0
time (s)
0.5
1
7
spikes/sec
0
NE-Pre-L
E-Pre-L
E-Post-L
NE-Post-L

Non-Engaged-Pre-Learning response strength in both response regions when the bird is no longer performing the task. It is likely that averaging over the entirety of the motif is washing out many of these sub-motif level response modifications. Further analyses are underway to isolate these sub-motif response regions and characterize the development of task-engagement effects in these smaller time windows.

Neurons show adaptation in the pre-learning phase, and many show different adaptation rates to different motifs

Since the modal change in response was a decrease in firing rate over learning, we sought to understand whether this might be caused in part by stimulus adaptation, a phenomenon by which neurons in many sensory systems decrease their stimulus-evoked response over the course of repeated presentations of the same stimulus. We first focused on the Non-Engaged-Pre-Learning condition, to determine whether neurons decreased their firing rate in response to repeated presentations of the same motif. For each of the 80 neurons recorded in the Non-Engaged-Pre-Learning condition, we performed a linear regression to model firing rate as a function of the sequential number of motif presentation. We included motif identity as a factor to isolate the effects of repeated presentations of each motif separately and to test the interaction between motif identity and presentation number.

All 79 of 79 neurons (one neuron had only a single motif that elicited a significant response, precluding comparisons across motifs) showed a main effect of motif identity ((probability > $F_{2\text{-}19}$) < 0.001 for all neurons), indicating that neurons

responded with different numbers of spikes in response to different motifs. In addition, we observed a main effect of presentation number in 47 of 79 neurons ((probability>F_1) < 0.05), indicating that the number of previous presentations of a given motif significantly affected the firing rate of these neurons (e.g., Figure 9A). The median slope of the fit of presentation number to firing rate for these 47 neurons was -0.084 (sp/sec)/presentation (quartiles -0.151, -0.013). The distribution of these values is significantly less than zero (Wilcoxon signed-rank test, p = 0.002), indicating that, on average, this population of neurons decreases its firing rate in response to repeated presentations of the same motif. In addition to these main effects, we observed that 41 neurons showed an interaction between motif presentation number and motif identity ((probability > $F_{2\text{-}19}$) < 0.05), which indicates that neurons are changing their firing rates in response to repeated motif presentations differently for different motifs, and supports the notion that these neurons display stimulus-specific adaptation (e.g., Figure 9B).

Adaptation is altered by behavioral state in many neurons

To test whether stimulus-specific adaptation might be a mechanism under the control of top-down modulation due to differences in birds' behavioral states, we explored the possibility that the adaptation rate for a given motif might change when birds move from passively listening during the Pre-Learning-Non-Engaged condition to performing the learning task in the Engaged-During-Learning condition. In each learning instance, for each of the over-trained and newly-trained motifs that

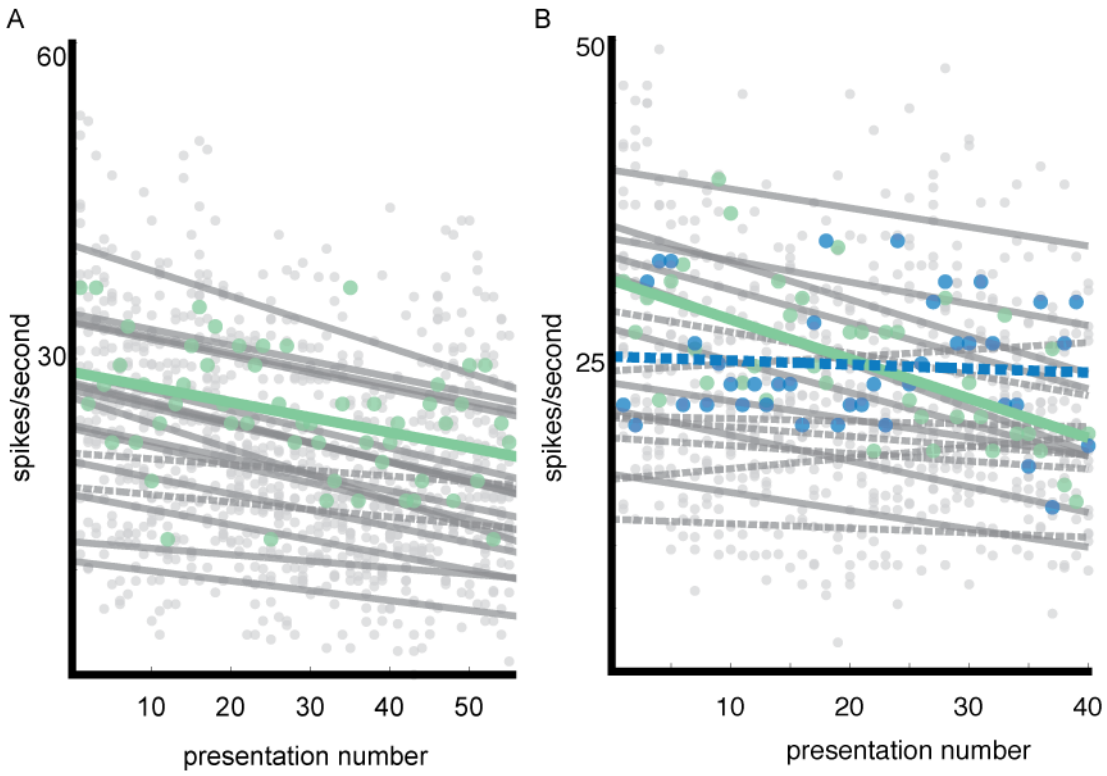


Figure 9: CMM neurons adapt to repeated presentations of a motif, and many neurons show evidence of stimulus-specific adaptation. (A) depicts linear fits of firing rate as a function of motif presentation number for 16 motifs that drove this neuron significantly above baseline in the Non-Engaged-Pre-Learning condition. Solid grey lines denote neurons that showed significant adaptation, while dotted grey lines show fits with a slope that is not different from zero. Grey dots are firing rates from individual trials to which the lines were fit. The green dots and line show a representative neuron that decreased its firing rate as a function of motif presentation number. Though many motifs showed evidence of adaptation, the adaptation rates were not different across motifs. (B) same as (A) for a different neuron which shows evidence that the response of the neuron adapts differently for different motifs. The motif represented by green shows significant adaptation, while the motif represented by blue does not show evidence of adaptation across repeated motif presentations, despite having a similar mean firing rate to the green motif.

showed a significant firing rate in each behavioral condition, we again performed a linear regression to model firing rate as a function of presentation number, now including the behavioral condition as a factor (Non-Engaged-Before-Learning, Engaged-During-Learning). In 55 of 170 motif-learning instance pairs, we observed a significant interaction between presentation number and behavioral condition ((probability > F_1) < 0.05), indicating that the slope of the best-fit line was significantly different in the Non-Engaged-Pre-Learning condition than it was in the Engaged-During-Learning condition. These effects were distributed such that 33 of 52 learning instances (two learning instances were excluded because none of the four over-trained or newly-trained motifs significantly drive the neuron in both behavioral conditions) had motifs that showed different adaptation rates between the Non-Engaged-Pre-Learning and the Engaged-During-Learning recording conditions. A neuron that does not adapt to a particular motif when not engaged in the learning task may show adaptation to this same motif while the bird is actively learning the motif classification, or vice versa, or may simply show different levels of adaptation between the two behavioral conditions. These results suggest that in many neurons, stimulus-specific adaptation may indeed be under the control of top-down mechanisms employed during the learning task as a way to modify the responses of neurons to specific motifs of interest.

DISCUSSION

To begin to elucidate the short-term correlates of the long-term experience-dependent plasticity known to occur in the avian auditory cortex, we recorded the acoustically driven extracellular spiking activity from 80 well-isolated neurons in the starling auditory cortical region CMM in response to auditory stimuli with which the birds had varying behavioral experience. For 44 of these neurons, we additionally recorded across the entire process of a bird learning to correctly classify two novel motifs, and for 5 of these neurons we were able to hold the neuron over multiple learning periods. This led to 54 learning instances in which a neuron's activity was recorded in response to a panel of motifs before, during, and after successful acquisition of the novel behavioral classification. Most neurons changed their responses to auditory stimuli over the course of the learning task; in only three learning instances did we fail to observe a change in firing rate to any motifs. We observed long-term population-level differences in response to stimuli with different training histories, and shorter-term changes in single neurons over the course of learning a novel set of motif classifications. These included an overall decrease in responsiveness to motifs—especially the newly-learned motifs—and the observation that many neurons gained and some lost the types of task-engagement-dependent changes in firing rate that have been previously shown to enhance the coding ability of neurons in CMM. Finally, we explored the possibility that these effects might be, at

least in part, mediated by stimulus-specific adaptation that may be under top-down behavioral control and serve as a mechanism for the short- and long-term changes observed in the neural coding in this system.

Responses in the Non-Engaged-Pre-Learning behavioral condition should reflect the long-term effects due to experience-dependent plasticity that lead to differences in the neurons' responses to motifs with different operant training histories. We found that the over-trained motifs with which birds had extensive experience tended to be represented by a diminished spiking response when compared to previously-trained motifs, with which the bird had much less training experience but was at some point capable of successfully classifying, or untrained motifs that the birds had never before experienced in an operant training context. When we looked for changes in individual neurons, we again saw that it tended to be the motifs that had more training that elicited fewer spikes from neurons than motifs with less training history.

This long-term difference in firing rate may reflect the long-term bias in coding that we have observed in this and other avian auditory system regions, but the direction of the effect is different from that observed in CMM in previous studies, where neurons were previously reported to display an increased response to learned versus novel stimuli. There are many reasons the direction of the effect here may be different from previous CMM studies, including differences in experimental conditions and sampling bias that can have a large effect on these sorts of population-level differences grouped across multiple recording sessions (single-wire, fully head-

restrained and anesthetized in previous studies; multi-electrode arrays mounted to the skull in awake birds, in the current study). However, perhaps a more parsimonious explanation is that once the birds in this study learned the meta-task, that is—once they knew that their task on a given day would include listening to a panel of motifs and then needing to learn to classify a subset of them—they may have begun to covertly attend to those motifs which they might be likely to hear in the upcoming learning session, and in doing so suppressed those for which they had a great deal of previous experience in classifying.

In addition to the long-term changes observed by recording neurons' responses in the Non-Engaged-Pre-Learning condition, we observed that neurons, on average, decreased their driven neural activity over the course of learning. Interestingly, they displayed the largest decreases for the newly-trained motifs, to the point of having significantly larger decreases for these neurons than for the over-trained motifs. Though the population of neurons tended to decrease driven firing rates across learning, there was a diversity of changes to the driven firing rate in response to single motifs across neurons. Most learning instances showed significant changes to single-motif responses that consisted of only decreases (27/54), but there were some learning instances in which neurons showed the opposite pattern: only significant increases (9/54). Perhaps most interesting were the learning instances in which neurons showed both increases and decreases in their responses to individual motifs after learning (15/54), since these responses cannot be described by simple changes in spiking

threshold, but require a more complicated neural explanation for their pattern of response changes across different stimuli.

The observation of stimulus-specific adaptation and its apparent modification by behavioral state provides a possible mechanistic basis for these effects. We observed that neurons on average adapted their responses to repeated presentations of the same motif, firing fewer and fewer spikes in response to each subsequent stimulus presentation. In addition, many neurons had significantly different decay rates across different motifs, suggesting that the responses to different motifs may be modulated by different sets of inputs impinging on each neuron. Studies performed in our lab, and currently in preparation for submission, support this notion. Single CMM neurons in anesthetized starlings can adapt to repeated presentations of auditory stimuli, and this adaptation is specific to features present in those stimuli; after adapting to repeated presentations of a particular stimulus, a neuron may still fire at the initial firing rate observed for a different stimulus to which the neuron had not been adapted (Kozlov & Gentner, in preparation). This suggests that different groups of afferent inputs may be differentially affected by adaptation, each providing the substrate for a unique portion of a neuron's receptive field, and each a possible target for learning-induced modification.

The implication of stimulus-specific adaptation in learning-induced changes in auditory neurons is not without precedent. A neighboring region in the avian auditory forebrain, NCM, shows strong adaptation to rapidly repeated presentations of brief stimuli in anesthetized zebrafinches, and this has been proposed as a mechanism of

sensory learning (Chew et al., 1995). Likewise, neurons in the A1 of cats show differential rates of adaptation to stimuli that occur with different presentation probabilities and act over timescales from hundreds to tens of thousands of milliseconds (Ulanovsky et al., 2004). The authors suggest that the stimulus-specific adaptation in these neurons appears to participate in encoding sensory memory on these timescales. Whether the stimulus-specific adaptation we observe underlies the long-term experience-dependent effects in CMM, or simply participates in the short-term decorrelation of the inputs to these neurons is unknown, but should not be ignored in future studies of the system.

In a recent study, we showed that engagement in an auditory task alters the motif-driven firing rates of most neurons in CMM compared to non-engagement. This study focused on birds that were performing a classification task with well-learned stimuli (analogous to the over-trained motifs in this study), and the effects due to task-engagement were such that neurons conveyed more information about the behavioral classes of motifs while the birds were engaged in a task than while they were not task-engaged (Knudsen and Gentner, 2013). That the changes with task engagement were specific to the learned behavioral categories suggests that they were established as a result of birds learning appropriate motif classifications during their experience in the operant training paradigm. The evidence we present in this report and the examples shown in Figure 8 appear to support this notion; we observed the development of such task-engagement effects over the course of learning. For the neurons that develop such a dependence on task-engagement, there is no particular modulation of the response to

the motif when the bird begins the learning task; in the Engaged-Pre-Learning condition, the neuron appears to be driven to the same level as it was in the Non-Engaged-Pre-Learning state. As the animal learns the proper association between motif, response, and reward, some modulatory action that was not present before the proper classification was learned comes into play either at the level of CMM or in upstream nuclei to increase or decrease the driven response to that particular motif. These increases and decreases then disappear when the animal is no longer engaged in the task, suggesting they are actively imposed by some a modulatory mechanism that is only present during behavior. This modulation may come as a function of trial-by-trial variations in dopaminergic or cholinergic input due to reward feedback or differences in object-based attention as the bird learns the proper contingencies of the classification task. The involvement of auditory attention is likely, as a set of recent experiments in the lab has shown strong attentional effects by carefully controlling starlings' auditory attention while recording in the caudal mesopallium. In these studies, neurons show time-restricted increases and decreases in driven firing rate that are only present when a bird's attention is properly directed to certain features in the stimulus and not others (Caporello and Gentner, in preparation). Whether or not the effects we observe here are maintained and reappear in subsequent task-engaged epochs, and whether or not these particular task-related changes lead to an enhanced representation of the motifs remains to be determined; the mechanism underlying these task-engagement and attentional effects remains unknown, and provides exciting directions for future study.

In previous studies of learning-induced modification of auditory representations, modifications tended to improve the system's ability to encode stimulus features that are relevant to an animal. How might the effects we observe here improve the representation of features that birds use to perform the classification task? Since the space of all auditory features that are possibly represented by the set of neurons in CMM is likely much larger than the space of features a given bird uses in any particular learning instance, we might expect that the majority of CMM neurons do not code for task-relevant features. In this case, a general decrease in the driven responses in CMM might be a way to reduce the overall noise of the system, while an increase in signal might be achieved when a small minority of neurons that have receptive fields that contain the features to which the bird is attending increase their driven firing rate. The observation that a certain minority of neurons, and some individual motif responses for a particular neuron, tend to increase their firing rate across learning agrees with this active reshaping of the representational landscape. In this case a general suppression of the less-informative features represented by most neurons accompanied by a selective retention or increase of the response to those features that are informative acts to renormalize the representation of useful features every time the bird needs to represent new meaningful stimulus associations. The general mechanism of this normalization may be the observed adaptation—cells respond more weakly to stimuli that are presented multiple times and don't need to be preferentially represented. Meanwhile, features that are useful for the classification task are rescued from this adaptation by neuromodulatory mechanisms such as

attention and reward prediction involved during behavior. Some or all of these behavioral modulations might then be maintained after the learning session, to be consolidated into the network representation and leading to the long-term experience-dependent effects that are observed in this region.

While the above description includes several assumptions and requires further experimentation to validate, it may be a helpful framework for interpreting the current results and designing future experiments. These experiments should work to focus animals' behavior to particular stimulus features of interest, ideally based on a model of a neuron's receptive field. Receptive field models allow a parametric mapping between stimulus and response, and will provide useful predictions about how a neuron's response should change as a bird learns novel stimulus classifications that can either support or invalidate the model described above. It may be necessary to control for the stimulus-specific adaptation we observe, or as suggested here, this adaptation may be an integral mechanism in the system's representation. Finally, a description of how the short-term behaviorally induced changes are consolidated and lead to long-term differences in the representation of different stimuli will require experiments that test the observed effects over longer timescales, and may require pharmacological manipulation aimed at blocking known mechanisms of plasticity or consolidation. There remain many challenges to understanding the experience-dependent reorganization of avian auditory cortex, and the results described in this report serve to focus future studies on those behavioral and stimulus dimensions that can explain the important sources of variance in this system.

REFERENCES

Bao S, Chan VT, Merzenich MM. Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature* 412: 79–83, 2001.

Catchpole CK, Slater PJB. *Bird Song: Biological Themes and Variations.* Cambridge, UK: Cambridge University Press, 1995.

Chew SJ, Mello C, Nottebohm F, Jarvis E, Vicario DS. Decrements in auditory responses to a repeated conspecific song are long-lasting and require two periods of protein synthesis in the songbird forebrain. *Proceedings of the National Academy of Sciences of the United States of America* 92: 3406–3410, 1995.

Dahmen JC, King AJ. Learning to hear: plasticity of auditory cortical processing. *Current opinion in neurobiology* 17: 456–64, 2007.

DiCarlo JJ, Lane JW, Hsiao SS, Johnson KO. Marking microelectrode penetrations with fluorescent dyes. *Journal of Neuroscience Methods* 64: 75–81, 1996.

Edeline JM, Weinberger NM. Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. *Behavioral neuroscience* 107: 82–103, 1993.

Eens M, Pinxten R, Verheyen RF. Temporal and Sequential Organisation of Song Bouts in the Starling. *Ardea* 77: 075–086, 1989.

Fritz J, Shamma S, Elhilali M, Klein D. Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature neuroscience* 6: 1216–23, 2003.

Fritz JB, Elhilali M, Shamma SA. Adaptive changes in cortical receptive fields induced by attention to complex sounds. *Journal of neurophysiology* 98: 2337–46, 2007.

Froemke RC, Merzenich MM, Schreiner CE. A synaptic memory trace for cortical receptive field plasticity. *Nature* 450: 425–9, 2007.

Gentner T, Hulse S. Perceptual mechanisms for individual vocal recognition in European starlings, Sturnus vulgaris. *Animal behaviour* 56: 579–594, 1998.

Gentner TQ, Margoliash D. Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669–74, 2003.

Gentner TQ. Temporal scales of auditory objects underlying birdsong vocal recognition. *The Journal of the Acoustical Society of America* 124: 1350–9, 2008.

Jackson N, Muthuswamy J. Artificial dural sealant that allows multiple penetrations of implantable brain probes. *Journal of neuroscience methods* 171: 147–52, 2008.

Jeanne JM, Thompson J V., Sharpee TO, Gentner TQ. Emergence of learned categorical representations within an auditory forebrain circuit. *The Journal of Neuroscience* 31: 2595–606, 2011.

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

Knudsen DP, Gentner TQ. Active recognition enhances the representation of behaviorally relevant information in single auditory forebrain neurons. *Journal of neurophysiology* 109: 1690–1703, 2013.

Lee C-C, Middlebrooks JC. Auditory cortex spatial sensitivity sharpens during task performance. *Nature neuroscience* 14: 108–14, 2011.

McLin DE, Miasnikov AA, Weinberger NM. Induction of behavioral associative memory by stimulation of the nucleus basalis. *Proceedings of the National Academy of Sciences of the United States of America* 99: 4002–7, 2002.

Metherate R, Weinberger NM. Cholinergic modulation of responses to single tones produces tone-specific receptive field alterations in cat auditory cortex. *Synapse (New York, N.Y.)* 6: 133–45, 1990.

Otazu GH, Tai L-H, Yang Y, Zador AM. Engaging in an auditory task suppresses responses in auditory cortex. *Nature neuroscience* 12: 646–54, 2009.

Recanzone GH, Schreiner CE, Merzenich MM. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. *Journal of Neuroscience* 13: 87, 1993.

Thompson J V, Gentner TQ. Song recognition learning and stimulus-specific weakening of neural responses in the avian auditory forebrain. *Journal of neurophysiology* 103: 1785–97, 2010.

Thompson J V, Jeanne JM, Gentner TQ. Local inhibition modulates learning-dependent song encoding in the songbird auditory cortex. *Journal of neurophysiology* 109: 721-33, 2013.

Ulanovsky N, Las L, Farkas D, Nelken I. Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience* 24: 10440–10453, 2004.

Weinberger NM. Associative representational plasticity in the auditory cortex: a synthesis of two disciplines. *Learning & memory (Cold Spring Harbor, N.Y.)* 14: 1–16, 2007.

GENERAL DISCUSSION

Songbirds have a complex and ever-changing relationship with the information that makes up their auditory world, making them a powerful model for understanding auditory cognition in animals and humans alike. In the wild, birds must be able to perceive and adapt to important auditory information, including encoding and recognizing important stimuli such as the song of their tutor, a rival, or a potential mate. We took advantage of these natural behaviors by bringing European starlings into the lab to examine the manner in which auditory stimuli are represented both perceptually and neurally, and how these representations change based on a bird's relationship to particular stimuli.

We showed in Chapter 2 that birds build distributed perceptual representations of stimuli; even when presented with very limited acoustic information from a previously learned stimulus, they are able to use this information to perform an action that leads to reward. The neural bases for these learned representations have been

studied before in anesthetized birds, but by 'waking up' our experimental subjects in Chapters 3 and 4, and putting them in a variety of experimental contexts that altered their relationship to auditory stimuli, we demonstrated that auditory representations are modified by a bird's behavioral state and by its recent experience with particular stimuli.

Neurons in the caudal mesopallium (CM) of starlings respond selectively to complex conspecific vocalizations. In Chapter 3, we trained birds to classify sets of these vocalizations, and saw that when they were engaged in this classification task, the neurons changed their selective responses in such a way as to encode more information about the relevant stimulus classes than when the birds were not performing the task. These are likely the effects of selective auditory attention, though we did not explicitly control for attention, and ours is only a coarse measure of neural encoding.

In Chapter 4, we attempted to see how such short-term changes as those observed in Chapter 3 might lead to the longer-term experience-dependent changes previously shown in CM of anesthetized birds. We trained birds to quickly recognize new auditory stimuli, and recorded from neurons before, during, and after these learning sessions. We again observed changes in the neural representation due to the animals' changing relationship to the stimuli; this time seeing a complex pattern of changes with learning, including both a general decrease in the response to stimuli for most neurons, and selective increases for others. We also witnessed the development of the types of task-dependent modulations observed in Chapter 3. How these

learning-induced changes relate to changes in the relative representation of different classes of stimuli remains to be determined, but we suggest that this might be explained by a combination of the suppression of uninformative features, and a selective retention and enhancement of behaviorally relevant information. It is possible that the changes we observe here are the basis for the long-term experience-dependent enhancements of behaviorally relevant stimuli previously described, but there is much to be done before this relationship is fully understood.

One of the most important factors for better understanding in this system is the knowledge of the relationship between a neuron's response properties and the specific acoustic features a bird is using to perform a particular task. The more we can match the variability in a bird's behavior to the variability in a neuron's receptive field, the more explanatory power we gain with each behavioral trial. As we better understand the response properties of these neurons by constructing the appropriate receptive field models, we will uncover the acoustic dimensions that provide the best description of neural response. Experiments can then be carefully constructed to tune birds' behavior to attending to some small set of these acoustic dimensions. In order to accomplish this, the desire for a complete understanding of the representation of natural stimuli must be balanced with the experimental expedience and analytical power of using artificial stimuli. Effort should be made to find a reduced stimulus that is parametric in nature, but which keeps the complexity inherent in the natural stimulus.

In addition to selecting the appropriate stimulus and behavior, future studies must select the appropriate targets for recording. CM is made up of many more

neurons than those from which we recorded. Any complete description of auditory representation, and the changes in this representation, will necessitate assessing activity in multiple neurons at once. It's very likely that the effects we observe here are a result of the entire auditory system modifying its response properties at multiple stages in a manner that may be hard to understand when recording from single neurons. The stimulus-specific adaptation results described in Chapter 4 suggest changes across independently modifiable inputs that might reflect plasticity happening in nuclei upstream of CMM such as NCM or CLM, or in Field L. Likewise, the neurons we record from in what we conceptualize as sensory structures somehow participate in the selection of an appropriate behavioral response. Whether they simply send a feedforward signal to downstream decision networks, or whether neurons in the auditory system in some way encode action selection will be an interesting question for future studies.

With these first studies of behavioral electrophysiology in the avian auditory forebrain we have extended the experimental range of a model system that is well suited to answering important questions about stimulus representation, motivation, attention, learning, and auditory cognition in general. We've made important technical advances in studying this system in both the control and harnessing of birds' behavior while recording from well-isolated neurons, in the design of experiments regarding both technical hardware and software issues, and in making the very important link between behavior and neural activity that make these kinds of data interpretable. As we make progress in explaining this system we'll be rewarded with a greater

understanding of how nervous systems use experience to tune themselves to the important features in their environments, allowing animals to use specialized sensory representations to successfully interact with the world.